Chapter 1
Genomics, Biological Features, and Biotechnological Applications of *Escherichia coli* B: "Is B for better?!"

Sung Ho Yoon, Haeyoung Jeong, Soon-Kyeong Kwon, and Jihyun F. Kim

Contents

1.1 Introduction . . . 2
1.2 Genomics of *E. coli* B Strains . . . 2
  1.2.1 Genomic Comparison of *E. coli* REL606 and MG1655 . . . 2
  1.2.2 Comparative Genomics of *E. coli* BL21(DE3) . . . 4
  1.2.3 Comparative Omics Analysis of *E. coli* B and K-12 . . . 6
1.3 Biological Features of ***E. coli*** B . . . 6
  1.3.1 Catabolism and Acetate Metabolism . . . 7
  1.3.2 Anabolism . . . 7
  1.3.3 Utilization of Carbon Sources . . . 8
  1.3.4 Cell Surface Features . . . 9
  1.3.5 Heat Shock Proteins . . . 9
  1.3.6 Cell Cycle and Growth . . . 10
  1.3.7 Secretion Capacity . . . 10
1.4 Usage of B Strains . . . 11
  1.4.1 As an Academic Lab Rat . . . 11
  1.4.2 As an Industrial Workhorse . . . 12
1.5 Future Prospects . . . 13
  References . . . 14

Abstract Strains of *Escherichia coli* B, especially BL21, have been widely used for overproducing recombinant proteins, ethanol, and other biomolecules. Almost all laboratory strains of *E. coli* are derivatives of non-pathogenic K-12 or B strains. While most genetic and metabolic studies have been performed with K-12 strains, little has been done on B strains. Recently, genome sequences of two *E. coli* strains of the B lineage, REL606 and BL21(DE3), have been determined, and results of multi-omics analyses were compared between B and K-12. As compared to K-12,

J.F. Kim (✉)
Industrial Biotechnology and Bioenergy Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), 111 Gwahangno, Yuseong, Daejeon 305-806, Republic of Korea;
Field of Functional Genomics, School of Science, University of Science and Technology (UST), 113 Gwahangno, Yuseong, Daejeon 305-333, Republic of Korea
e-mail: jfk@kribb.re.kr

S.Y. Lee (ed.), *Systems Biology and Biotechnology of Escherichia coli*,
DOI 10.1007/978-1-4020-9394-4_1, © Springer Science+Business Media B.V. 2009

B strains show a number of phenotypes such as faster growth in minimal media, lower acetate production, higher expression levels of recombinant proteins, and less degradation of such proteins during purification. In this review, we summarize the unique biological features of the B strains and overview their academic and industrial applications.

1.1 Introduction

Escherichia coli, a common inhabitant of the mammalian intestines, undoubtedly has been one of the best studied organisms and plays important roles in biological sciences, medicine, and industry. *E. coli* strain B was named by Delbrück and Luria in 1942 (Delbruck and Luria 1942), but the early history of B is less well established whereas the origin of *E. coli* K-12 is clear. In any case, derivatives of *E. coli* B have been serving not only as a research model for the study of phage sensitivity, restriction systems, and bacterial evolution in the laboratories, but as a major workhorse for protein expression in the biotechnological industry. However, genetic bases of the apparent superiority of B in many industrial setups have been restricted to a limited number of topics and the rest have been left largely undetermined.

The genomes of two derivative strains of B recently have been deciphered through an international collaboration among scientists in Korea (KRIBB; our group), the United States (Michigan State University and Brookhaven National Laboratory), and France (Genoscope). This has opened a new possibility of examining B through various omics technologies including DNA microarray for gene expression profiling and two dimensional gel electrophoresis followed by MALDI-TOF identification of proteins. In this chapter, we review the current understanding of the biological features of B strains in the context of genomic information and results of the multi-omics analyses together with their utility in scientific studies and biotechnical applications.

1.2 Genomics of *E. coli* B Strains

1.2.1 Genomic Comparison of E. coli *REL606 and MG1655*

The first complete genome sequence of a B strain was determined by an international consortium (Jeong et al. submitted). The strain of choice was REL606, an Ara$^-$ clone derived from chemical mutation of Bc251 (F$^-$ *mal*$^+$ λ^s) (Lederberg 1966). REL606 has been long used as a founder strain for long-term evolution experiments by Richard E. Lenski at Michigan State University (Cooper and Lenski 2000, Lenski et al. 1991). A Sanger chemistry-based, standard shotgun approach was exploited for the genome sequencing of REL606. KRIBB participated in the initial shotgun sequencing and final process for genome annotation, while Genoscope led genome sequencing to completion and automatic annotation based on MaGe (Vallenet 2006).

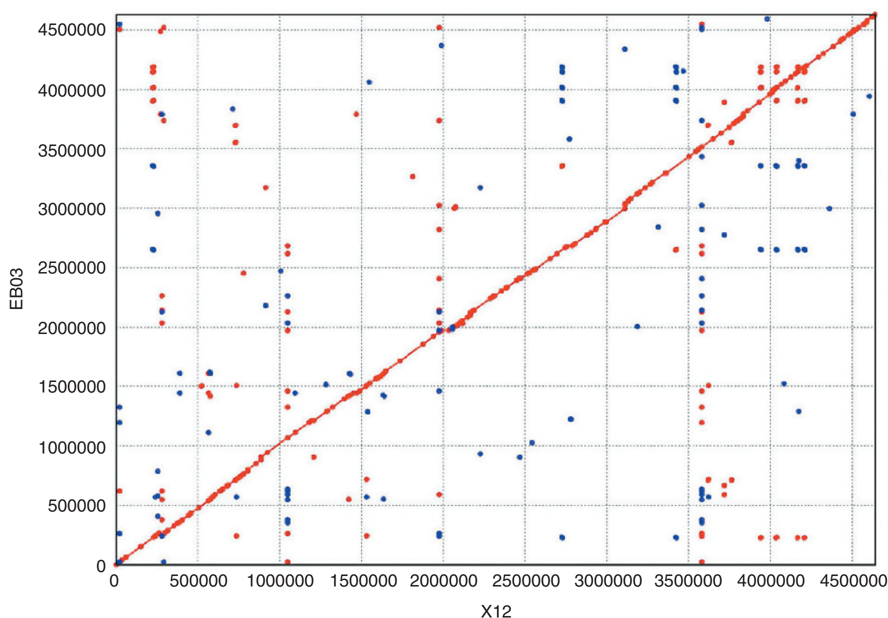

Fig. 1.1 Whole-genome alignment of *E. coli* K-12 MG1655 (*x*-axis) and B REL606 (*y*-axis). NUCMER script in MUMMER 3.0 was used for the generation of alignment with default parameters (http://mummer.sourcefourge.net/)

E. coli B REL606 has a single circular chromosome of 4,612,812 bp with no plasmid, which makes it the most compact genome among the completely sequenced strains of *E. coli*. Its chromosome size is most similar to that of K-12 MG1655 (4,639,675 bp), and as shown by the whole-genome alignment plot no chromosomal rearrangement was observed other than some insertions or deletions (Fig. 1.1). When MUMMER was used as an alignment generator, the total length of aligned regions between REL606 and MG1655 amounted to more than 96%. Average percent identity of the aligned regions is 97.5%, and it further increases up to 99.09% if it is length-weighted average.

Though overall genome organization is very similar between two strains, several prominent factors contribute to shaping peculiarities of each genome (Fig. 1.2, see below). First, highly divergent regions, occupying the equivalent positions on each genome, are readily identified by the broken lines appearing on the whole-genome alignment plot (Fig. 1.3). Most of them are related to genes involved in surface characteristics (e.g., LPS core oligosaccharide biosynthesis), which are probably ones under strong positive selection. Second, there are several horizontally transferred genomic segments that represent genome-specific regions. Fitness island encoding gene sets for the metabolism of aromatic hydrocarbon is a good example.

Lastly, distribution of mobile genetic elements such as prophages and insertion sequence (IS) elements are significantly different between the two strains. Specifically, IS seems to exert most dramatic effect to its host genome, since it can deacti-

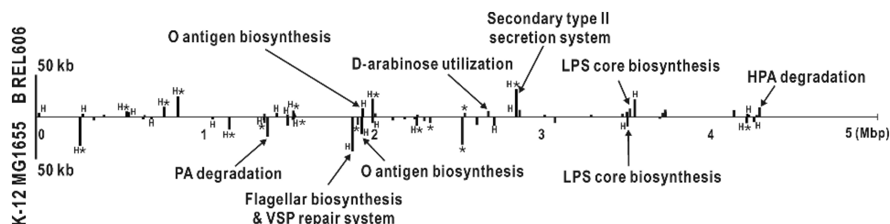

Fig. 1.2 Genomic regions unique to *E. coli* B REL606 or K-12 MG1655. The horizontal axis represents the map coordinates of backbone regions common to REL606 and MG1655; *vertical bars* denote locations and lengths of strain-specific regions larger than 2 kb. A H denotes the region overlapping predicted genomic islands (Yoon et al. 2005). Genomic regions containing prophage genes are marked by asterisks. Abbreviations: LPS, lipopolysaccharide; HPA, 3-hydroxyphenylacetic acid and 4-hydroxyphenylacetic acid; PA, phenylacetic acid; VSP, very short-patch

vate functional genes by insertion events, and can result in genomic rearrangement by homologous recombination as well. For example, REL606 has a 41-kb deletion ($\sim$0.8% of the entire chromosome), probably mediated by IS*1* (the most abundant IS in B), from *yecF* to *yedS*. The deleted segment, that includes *sdiA*, *amyA*, *rcsA*, *vsr*, *dcm*, and the *fli* cluster, produce relevant B-specific phenotypes that were previously known. Functions affected by IS transpositional inactivation include porin expression, restriction-modification, and Lon protease (saiSree et al. 2001, Schneider et al. 2002).

1.2.2 Comparative Genomics of E. coli *BL21(DE3)*

BL21(DE3) is a specifically engineered B descendant harboring the T7 RNA polymerase gene for high-level expression of recombinant proteins (Studier and Moffatt 1986). With other B-specific traits such as deficiency in proteases and amenability to high-density culture as mentioned above, BL21(DE3) has become the most widely used strain for biotechnological applications. Since BL21(DE3) and REL606 are very close to each other being diverged just dozens of years ago from their common ancestor, we applied a hybrid approach that combines hybridization-based genome resequencing by NimbleGen's CGS (Comparative Genome Sequencing) technology and 454 pyrosequencing by Roche GS 20 at about 10X coverage to assemble genome sequence of BL21(DE3) using the REL606 genome (Jeong et al. submitted). All potential mutagenic sites were confirmed by Sanger sequencing of PCR-amplified products, and genomic rearrangement that could be only identified by pairwise comparison between the reference sequence and *de novo*-assembled 454 contigs were also verified by the same method. Over 98% of the final BL21(DE3) sequence could be covered by 454 pyrosequencing contigs, and $\sim$166 kb were confirmed by conventional sequencing.

We confirmed 415 SNPs, 16 insertions, and 28 deletions in the genome of BL21(DE3) with respect to REL606. 79% of SNPs (326 out of 415) occur in a nar-

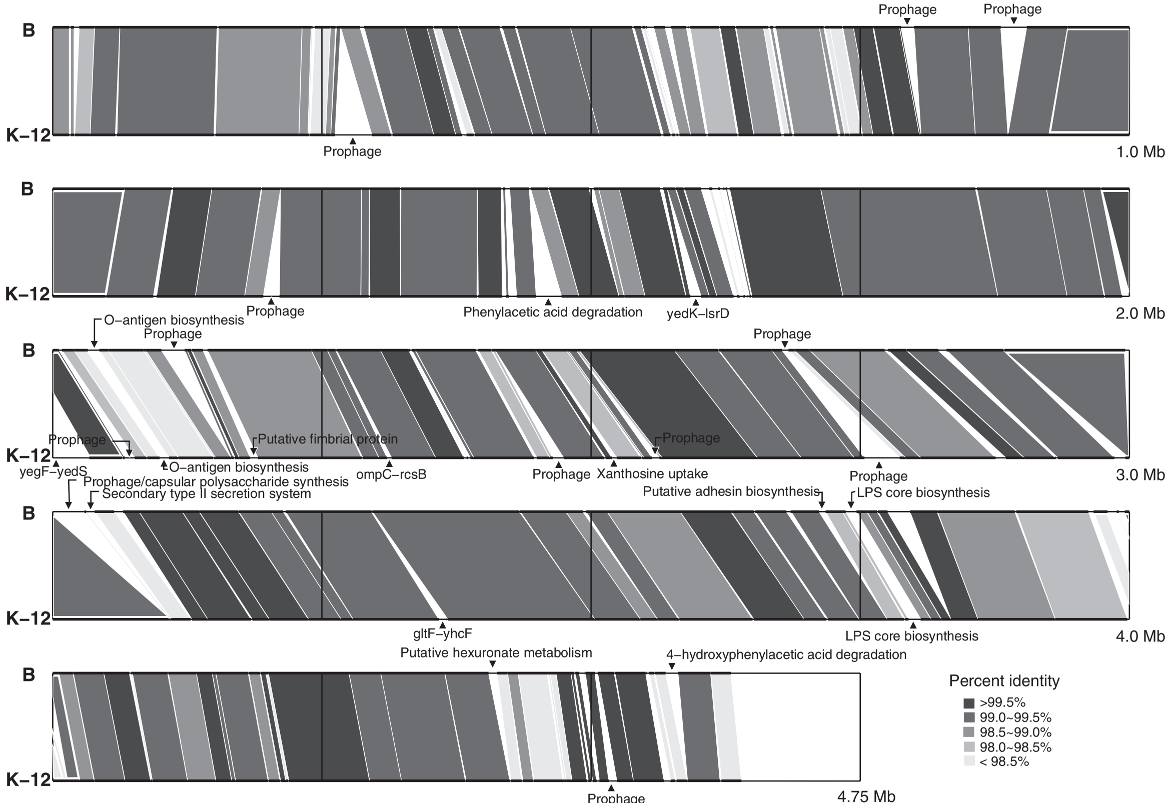

Fig. 1.3 A schematic diagram showing alignment of the blocks of homologous sequences (*thick lines*) between *E. coli* B REL606 (*top lines*) and K-12 MG1655 (*bottom lines*). Percent identity for each block is indicated as a grey parallelogram of a different shade. Texts denote genes or functions encoded by specific genomic segments

row region around 4.2 Mb-position, which occupies only 1.4% of the chromosome. Chemical mutagenesis applied to Bc251 to produce the progenitor to REL606 only cannot accounts for such a big differences. Close inspection revealed the highly divergent region was indeed transferred from W3110 by P1 transduction to produce Bc251. Another P1 transduction to create BL21 replaced again most of the W3110 DNA with B DNA, which resulted in K-12-derived sequence remaining only in REL606.

To obtain a complete genome sequence of BL21(DE3) based on 454 contigs, we recently constructed a fosmid library and produced paired-end sequences. Scaffolds have been made by a mixed assembly approach using fosmid end reads and verified 454 contigs, and gap closure is in progress (Jeong and Kim, unpublished data).

1.2.3 Comparative Omics Analysis of E. coli *B and K-12*

Recent advances in high-throughput omics technology are providing us with the possibility of deciphering an organism's genotype-to-phenotype relationships. Recently, we have carried out comparative and integrated analysis of the genome, transcriptome, proteome, and phenome data of B and K-12 strains that are closely related (Yoon et al. submitted). We also reconstructed an *in silico* metabolic network of B that accommodates the multidimensional omics data. From the study, we identified many important differences in cellular metabolism and physiology between B and K-12.

Lack of flagellar biosynthetic genes and low expression of motility-related genes make B non-motile. This is an important property of B when used as a cell factory because flagella biosynthesis is energy-intensive and is not necessary under an industrial setup of constant agitation and generous supply of nutrients (Posfai et al. 2006, Yu et al. 2002). Differences in the composition of the LPS core and expression of outer membrane proteins may influence the permeability and integrity of the cell envelope, which presumably result in alterations to screening barriers that control import and export of materials such as antibiotics, nutrients, and proteins. Importantly, the existence of the second T2S system and enhanced capability for protein release qualify B strains as the first choice for extracellular production of recombinant proteins. Information on naturally exported proteins can be useful in developing a strategy of excretory protein production, as exemplified by the use of OmpF fusion approach for the extracellular production of human proteins (Jeong and Lee 2002). B strains exhibited up-regulation of many amino acid biosynthetic genes, and showed lower expression of proteases. These characteristics are desirable for the enhanced production of recombinant proteins.

1.3 Biological Features of *E. coli* B

Elucidation of biological features of a strain is always the starting point for its biotechnological applications. B and its derivatives have been widely used for the

production of recombinant proteins and biomolecules. B strains have also served as research models for studies of phage sensitivity, restriction systems, mutagenic assays, and bacterial evolution (Cooper and Lenski 2000, Herrera et al. 2002, Swartz 1996). Although the genome sequences of B and K-12 are highly similar based on comparison of IS elements (Schneider et al. 2002), B often shows phenotypes distinct from those of K-12 (Swartz 1996).

1.3.1 Catabolism and Acetate Metabolism

B strains have been widely used for overproduction of recombinant proteins because they offer faster cell growth in minimal media and lower production of acetate than K-12 derivatives. Such differences between B and K-12 strains can be attributed to their genetic backgrounds. However, from the analysis of genome sequences of two B strains, REL606 and BL21(DE3) (Jeong et al. submitted), we found that there was no genetic difference in genes involved in glycolysis, TCA cycle, pentose phosphate pathway, glyoxylate pathway, gluconeogenesis, and acetate production among genomes of B and K-12. Thus, it is possible that the metabolic genes are under different regulation of in these two groups of *E. coli*. Transcriptome analysis of BL21 and JM109 (B derivative) during batch fermentation with high initial glucose concentration (Phue et al. 2007) demonstrated that genes involved in glyoxylate shunt, TCA cycle, fatty acid, gluconeogenesis and anaplerotic pathways were expressed differently between the two strains, while no apparent differences were detected for those in glycolysis and pentose phosphate pathway.

Acetate accumulation is one of the major problems encountered during high cell density cultivation of *E. coli*, because it inhibits cell growth and production of foreign proteins (Eiteman and Altman. 2006). Generally, B strains accumulate less acetate than K-12 strains during high cell density cultivation with glucose as a carbon source. A common explanation for low acetate accumulation by B is the active glyoxylate shunt which is the main pathway for acetate utilization due to the high expression of acetate operon (*aceBAK*). Analyses of DNA microarray and Northern blot demonstrated that BL21 showed high activity in glyoxylate shunt, TCA cycle, gluconeogenesis pathway, conversion of acetate to acetyl CoA, and fatty acid degradation irrespective of glucose concentration in culture media (Phue et al. 2005). In case of JM109, a K-12 derivative, the trend was similar at low glucose culture conditions, while it was lowered at high glucose conditions. Phue et al. (2005) suggested that insensitivity of BL21 to glucose concentration can be attributed to absence of a regulatory mechanism and possibility of altered activity of FruR, a transcriptional regulator of the control of carbon and energy metabolism, in BL21. Much effort should be made to fully understand the metabolism of glucose and acetate of the B and K-12.

1.3.2 Anabolism

Sequence differences in amino acids biosynthetic pathways have been found in genes for L-arginine and branched-chain amino acids biosyntheses. In K-12, for-

mation of the enzymes involved in arginine biosynthetic pathway is under feedback inhibition by arginine, while those enzyme levels are not affected by arginine in B. The different regulation mode is due to the differences in a single amino acid in the arginine repressor protein (ArgR), changing from the proline residue in K-12 to leucine in B (Tian et al. 1994). K-12 contains two genes encoding ornithine carbamoyltransferase in arginine biosynthesis, *argI* and *argF*, while B, *E. coli* W, and other species in *Enterobacteriaceae* have only *argI* or its equivalents (Legrain et al. 1976).

B is insensitive to extracellular valine while K-12 cannot grow in the presence of valine (Yoon et al. submitted). The first reaction in the biosynthesis of branched-chain amino acids (leucine, isoleucine and valine) is catalysed by three isozymes, acetohydroxy acid synthase (AHAS) I, II, and III encoded by *ilvBN*, *ilvGM*, and *ilvIH* respectively. It is known that valine exerts feedback inhibition on isozymes I and III (Umbarger 1996). Thus, exogenous valine can inhibit cell growth because the valine makes the cell unable to synthesize leucine and isoleucine, which is known as valine toxicity. It is thus essential that the functional isozyme II is expressed for cell growth in the presence of exogeneous valine. K-12 has a frameshift mutation in the *ilvG* gene, but B has an intact *ilvG* mediating valine resistance.

1.3.3 Utilization of Carbon Sources

Ability to utilize a variety of substrates is quite different between B and K-12, which can be attributed largely to genetic discrepancy in nutrient uptake systems. Most enteric bacteria cannot grow on D-arabinose which is uncommon in the natural environments. As enzymes for L-fucose utilization can degrade D-arabinose to dihydroxyacetone phosphate and glycoaldehyde, regulatory mutations of the L-fucose pathway in K-12 led to growth on D-arabinose (LeBlanc and Mortlock 1971). In contrast to K-12, B strains cannot grow on L-fucose because of the lack of L-fuculose-1-phosphate aldolase (Boulter et al. 1974). Interestingly, B strains can degrade D-arabinose without mutation. This is due to the possession of gene cluster for converting D-arabinose to D-xylulose 5-phosphate, which appears to have acquired through horizontal gene transfer (Elsinghorst and Mortlock 1994).

Eliminating environmental pollutants such as aromatic compounds by microorganisms is a competitive alternative to the commonly used chemical processes (Pieper and Reineke 2000). Aromatic compounds are highly abundant in soil and water, and *Pseudomonas* strains and other soil bacteria can catabolize a wide range of aromatic compounds. Unexpectedly, some *E. coli* strains and other enteric bacteria are reported to be able to degrade aromatic amino acids (Diaz et al. 2001). *E. coli* B and C can grow on 3- and 4-hydroxyphenylacetic acid (HPA) but not on phenylacetic acid (PA), while K-12 grow on PA but not on 3-HPA and 4-HPA (Diaz et al. 2001). Recent studies that compared the genomes of B and K-12 (Yoon et al. submitted) demonstrated that each has a different gene cluster for the catabolism of aromatic compounds – the *paa* cluster for the catabolism of phenylacetic acid in K-12 and the *hpa* cluster for the degradation of 3- and 4-HPA in B.

1.3.4 Cell Surface Features

Cell envelope is the principal stress-bearing and shape-maintaining element in *E. coli*, and its integrity is of critical importance to cell viability. B strains have been widely used for mutagenic assays and toxicological studies because they show higher membrane permeability than does K-12 (Herrera et al. 2002). Structural studies on the LPS core oligosaccharides have revealed that K-12 is devoid of the O antigen while B lacks the O antigen plus the distal part of the polysaccharide core of the outer membrane (Jansson et al. 1981). Sequence comparison has revealed that the outer membrane structure of B is quite different from that of K-12 (Jeong et al. submitted). IS elements were found to be inserted at the gene clusters for O antigen biosynthesis: at *wbbL* for K-12 and between *manC* and *wbbD* for B strain. In the B genome, the core part of LPS was further disrupted by the insertion of IS*1* at *waaT* encoding the UDP-galatose:(glucosyl) LPS α1,2-galactosyltransferase.

Importantly, flagellar biosynthesis genes are missing in B (Jeong et al. submitted). A 38-kb region of K-12 from *yecF* to *yedS*, containing *fliYZACDSTEFGHIJKLMNOPQR* genes was deleted in the genome of B. Thus, B cannot form the flagella and thus is non-motile.

Porin proteins control the permeability of polar solutes across the outer membrane and play important roles in the nutrient uptake process (Nikaido 1996). In K-12, though the total amount OmpC and OmpF is constant, their relative proportion changes depending on the culture medium osmolality, which is controlled by the EnvZ-OmpR regulatory system. By contrast, B strains express only OmpF in large quantity (Pugsley and Rosenbusch 1983). This is attributed to the fact that IS insertion in the B genome results in the deletion of the first 114 bp of *ompC* and the upstream region containing *micF* which posttranscriptionally prevents the production of OmpF (Schneider et al. 2002). Noxious agents such as antibiotics and bile acids diffuse far better through OmpF because OmpF produces a larger channel than OmpC (Nikaido 2003). Thus, *ompF* mutants became highly resistant to β-lactam compounds (Harder et al. 1981). In the phenotype microarray test, we discovered that B displayed sensitivity to various stress conditions of osmolarity, pH stress, and antibiotics much higher than K-12 (Yoon et al. submitted).

1.3.5 Heat Shock Proteins

Heat shock proteins (HSPs) including molecular chaperones and proteases make sure cellular proteins being in the right shape and in the right place at the right time (Gross 1996). Thus, they are required both during stress and normal growth conditions. Among the ATP-dependent proteases, B strains are naturally deficient in the major protease Lon which degrades abnormally folded proteins. This is due to the insertion of IS*186* in the promoter region of *lon* (saiSree et al. 2001). Additionally, the BL21 cells lack the OmpT outer membrane protease. Besides its major role in protein quality control, Lon is involved in many biological processes such as cell differentiation, pathogenicity, motility, stringent response to amino acid starvation,

and regulation of the toxin-antitoxin module (Tsilibaris et al. 2006). Lon mutants are viable, but display sensitivity to ultraviolet light and overproduce capsular polysaccharide, which are the result of the elevated levels of regulatory proteins (SulA and RcsA) that are normally degraded by Lon.

As HSPs are up-regulated by the heat-shock sigma factor σ^{32} encoded by *rpoH* when cells are exposed to stress condition such as temperature upshift and production of recombinant proteins, they can be used as a stress probe for monitoring cellular stress (Cha et al. 1999, Vostiar et al. 2004). When cellular stress levels of BL21 and K-12 strains (JM105, HB101, and TOP10) were measured by fusing promoters of heat-shock genes (*rpoH*, *dnaK* or *clpB*) to the reporter gene (*gfp*), BL21 exhibited the lowest cellular stress level and expressed the highest foreign protein (Seo et al. 2003). Possibly, lower cellular stress level of B strain is one of the reasons for high capacity in foreign protein production.

1.3.6 Cell Cycle and Growth

Bacterial cell growth is closely coordinated with DNA replication and chromosome segregation (Haeusser and Levin 2008). The cell cycle of slowly growing bacteria can be divided into three time periods: (i) period B, cell division to the initiation of chromosome replication, (ii) period C, chromosome replication, and (iii) period D, termination of replication to cell division. From the cytometry data that measured periods C and D of *E. coli*, the D period in B/r is much shorter than in K-12 strains (Michelsen et al. 2003).

Normally, B strains grow faster than K-12 in minimal media. The widely used K-12 strains, MG1655 and W3110, grow slowly in a pyrimidine-free medium than in a medium containing uracil. In the *rph-pyrE* operon involved in *de novo* pyrimidine biosynthesis of these strains, *rph* is frame-shifted to produce truncated RNase PH, and the premature translation stop leads to decreased expression of *pyrE* encoding orotate phosphoribosyltransferase (Jensen 1993). However, the *rph* gene is intact in the B strains (Yoon et al. submitted).

1.3.7 Secretion Capacity

Bacterial extracellular proteins perform important biological processes such as assembly of flagella and fimbriae, nutrient acquisition, cell-to-cell communication, and pathogenesis. In Gram-negative bacteria, excretory proteins are much less than in Gram-positive species because they should cross the two membranes of the cell envelope. Laboratory *E. coli* strains normally does not secrete extracellular proteins because the genes encoding type II secretion (T2S) pathway operon (*gsp*) are silenced by H-NS (Francetic et al. 2000). B strains released more proteins according to analyses of the extracellular proteomes of B and K-12 during flask culture (Yoon et al. submitted) and high cell density cultivation (Xia et al. 2008). This could be at

least partly due to an additional gene cluster for T2S in the B strains (Yoon et al. submitted). Phylogenetic analysis revealed that the T2S system commonly found in REL606 and MG1655 were clustered into the clade of *E. coli* and other genera in the *Enterobacteriaceae* family, whereas the sequence of the additional T2S system in REL606 was grouped into the branches of *E. coli* strains having multiple T2S systems and families other than *Enterobacteriaceae*. This implies that the two T2S systems of the B strains have evolved independently or more specifically the latter might have been introduced.

1.4 Usage of B Strains

1.4.1 As an Academic Lab Rat

Since 1940s, K-12 and B strains have been widely used as a laboratory strain and have had significant impact on biological sciences, medicine, and industry (Daegelen et al. submitted, Lederberg 2004). While K-12 strains have been mainly used for developing recombinant DNA techniques, B strains have been the subject of physiological studies (Swartz 1996). B strains also served as hosts for the historical studies of T1-T7 bacteriophages (Delbruck 1946), which led to the construction of BL21(DE3) (Studier and Moffatt 1986). Due to the rapid growth in minimal media and enhanced membrane permeability, B strains were favored by physiologists.

A radiation-resistant mutant, *E. coli* B/r, was isolated after UV-irradiation (Witkin 1946), and has been used for determining cell cycle-related parameters (Helmstetter 1968, Michelsen et al. 2003). Chemical composition measurements of B/r was measured (Neidhardt and Umbarger 1996), which is essential to estimate growth requirements such as energy distribution, reducing power (Neijssel et al. 1996) and metabolic fluxes in a genome-scale metabolic model (Feist et al. 2007).

Due to the increased membrane permeability, B strains have been used widely for mutagenic assays and toxicological studies. Mutants of B strain WP2 (e.g. WP2 *uvrA* and WP2 *uvrA*/pKM101) have been used as a tester strain in mutagenic assays (Gatehouse et al. 1994) and officially included in the OECD guideline for bacterial reverse mutation test (OECD guideline for testing of chemicals: bacterial reverse mutation test, 1998). These mutants are sensitive to oxidizing mutagens, cross-linking agents and hydrazines (Blanco et al. 1998, Herrera et al. 1993, Wilcox et al. 1990). The higher permeability can make B a primary choice for functional studies by flow cytometry and fluorescence microscopy. When B and K-12 strains were stained with several fluorochromes, B strain showed higher uptake of fluorescent dyes and higher fluorescent intensity (Herrera et al. 2002).

Evolution experiments with microorganisms are of a great interest because they allow one to investigate genetic and phenotypic evolution in action under the controlled environment (Elena and Lenski 2003, Philippe et al. 2007). For decades, B has served as a research model for long-term bacterial evolution. Twelve popu-

lations derived from a common ancestor have propagated by daily serial transfer in a glucose-limited minimal medium for more than 40,000 generations (Cooper et al. 2003). In the experiment, all the populations have adapted to the growth environment via beneficial mutations. Critical issues in evolution are being addressed using laboratory populations of bacteria, e.g. the dynamics of evolutionary adaptation, the genetic bases of adaptation, interactions between different genotypes in a population, and between interacting microbial species (Elena and Lenski 2003). Application of evolutionary principle to strain development and process optimization, which is so called evolutionary engineering, is becoming an important strategy in the field of metabolic engineering (Sauer 2001). Recently, the efficiency of natural selection using long-term evolution experiment is successfully exploited to improve industrial strains (de Crecy et al. 2007, Fong et al. 2005).

1.4.2 As an Industrial Workhorse

A primary goal of the bioprocess development is the cost-effective production of desired products such as therapeutic and industrial proteins on a large scale. B and K-12 are preferred production hosts because of fast growth, facility in genetic modification and cultivation, and high yields for many recombinant proteins. As mentioned above, B and its derivatives have salient features desirable for high cell density culture such as low acetate production even when grown on excess glucose, faster growth in minimal media, protease deficiency, and simple cell surfaces that enhance permeability. Thus, they have been widely used for the overproduction of recombinant proteins, ethanol, and other biomolecules on a large scale (Choi et al. 2006).

The most popular strains, BL21 and its derivatives (Studier and Moffatt 1986), are derived from B, thus are naturally deficient in the major protease Lon. Additionally, their chromosomes are deleted from the gene for the outer membrane protease OmpT. The absence of these proteases can lead to higher expression levels of recombinant proteins and less degradation of such proteins during purification. A derivative of BL21, BL21(DE3), was constructed to harbor a recombinant phage λ carrying the T7 RNA polymerase gene under the control of the *lacUV5* promoter (Studier and Moffatt 1986). Addition of isopropyl β-D-thiogalactopyranoside (IPTG) into growth media is required to express the T7 RNA polymerase, which then transcribes target genes located in a plasmid under the control of the T7 promoter. Due to its high selectivity and activity, BL21(DE3) is extremely popular for mass-production of recombinant proteins which are toxic to the host cells (Choi et al. 2006). Various versions of the T7 RNA polymerase-based expression system have been developed to use cheap and nontoxic inducers instead of IPTG or to minimize basal expression of the cloned gene (Sorensen and Mortensen 2005).

Membrane proteins (MPs) account for more than 50% of all drug targets and are of major pharmaceutical and biotechnological interests. Generally, a large amount of MPs is required for their functional and structural studies. However, in many

cases, over-expression of MPs is lethal to host cells. To overcome this difficulty, mutant hosts, C41(DE3) and C43(DE3) were derived from BL21(DE3) over-producing some membrane proteins (Miroux and Walker 1996). These two mutant hosts, especially C43(DE3), showed reduced toxicity of expressed MPs and are widely used for a variety of MPs. Although the genetic mutation(s) responsible for the changes have not yet been identified, comparative genome sequence analyses of C41(DE3), C43(DE3), and their parental BL21(DE3) revealed several interesting genetic changes (Kwon and Kim, unpublished data). Comparative analysis revealed that there are six SNPs in C41(DE3) and seven in C43(DE3) as compared to BL21(DE3). Interestingly and perplexingly, only two of them overlaps between the two strains. Also, there are two IS-mediated deletions that have been observed – one in both C41(DE3) and C43(DE3) and the other only in C43(DE3). It is reported that C41(DE3) and C43(DE3) are also superior to BL21(DE3) in the production of some cytoplasmic proteins and in the stability of their cloning plasmid (Dumon-Seignovert et al. 2004). However, it is hard to predict an expression host and system working best for a target protein, and so, screening process is required to some extent.

1.5 Future Prospects

Until now, most genetic and metabolic studies of *E. coli* have been performed with K-12 or its derivatives. Also, a variety of omics analyses (Choi et al. 2003, Franchini and Egli 2006, Han and Lee 2006, Ishii et al. 2007, Nandakumar et al. 2006, Yoon et al. 2003) and *in silico* metabolic modeling of K-12 (Covert et al. 2004, Feist et al. 2007) have been accelerated by the availability of the complete genome sequences of the MG1655 and W3110 strains (Blattner et al. 1997, Hayashi et al. 2006). In contrast to K-12, little studies have been performed for B strains. This can be attributed to the fact that K-12 strains are the best fit in the current recombinant DNA techniques and a wealth of safety information makes them preferred recombinant organisms by the biosafety communities (Swartz 1996). Additionally, many B derivatives as industrial hosts have been developed in private companies, which can make some difficulties in academic and public research. However, various features of B strains are beneficial for the overexpression of foreign proteins and studies of *E. coli* physiology. B strain is now in its early stages of global omics studies and systems biology (Xia et al. 2008, Yoon et al. submitted). With the availability of genome sequences of B strains (Jeong et al. submitted), the omics information on the cellular metabolism and physiology should be pivotal in better understanding the underlying biological networks and is invaluable for designing strains having customized genomes as well as establishing rational fermentation strategies.

Acknowledgments This work was supported by the 21C Frontier Microbial Genomics and Applications Center Program, Ministry of Science and Technology, Republic of Korea.

References

Blanco, M., A. Urios, and A. Martinez. 1998. New *Escherichia coli* WP2 tester strains highly sensitive to reversion by oxidative mutagens. Mutat. Res. **413**:95–101.

Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277**:1453–74.

Boulter, J., B. Gielow, M. McFarland, and N. Lee. 1974. Metabolism of D-arabinose by *Escherichia coli* B/r. J. Bacteriol. **117**:920–3.

Cha, H. J., R. Srivastava, V. N. Vakharia, G. Rao, and W. E. Bentley. 1999. Green fluorescent protein as a noninvasive stress probe in resting *Escherichia coli* cells. Appl. Environ. Microbiol. **65**:409–14.

Choi, J. H., K. C. Keum, and S. Y. Lee. 2006. Production of recombinant proteins by high cell density culture of *Escherichia coli*. Chem. Eng. Sci. **61**:876–85.

Choi, J. H., S. J. Lee, and S. Y. Lee. 2003. Enhanced production of insulin-like growth factor I fusion protein in *Escherichia coli* by coexpression of the down-regulated genes identified by transcriptome profiling. Appl. Environ. Microbiol. **69**:4737–42.

Cooper, T. F., D. E. Rozen, and R. E. Lenski. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **100**:1072–7.

Cooper, V. S., and R. E. Lenski. 2000. The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature **407**:736–9.

Covert, M. W., E. M. Knight, J. L. Reed, M. J. Herrgard, and B. O. Palsson. 2004. Integrating high-throughput and computational data elucidates bacterial networks. Nature **429**:92–6.

Daegelen, P., F. W. Studier, R. E. Lenski, S. Cure, and J. F. Kim. Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). submitted.

de Crecy, E., D. Metzgar, C. Allen, M. Penicaud, B. Lyons, C. J. Hansen, and V. de Crecy-Lagard. 2007. Development of a novel continuous culture device for experimental evolution of bacterial populations. Appl. Microbiol. Biotechnol. **77**:489–96.

Delbruck, M. 1946. Bacterial viruses or bacteriophages. Biol. Rev. **21**:30–40.

Delbruck, M., and S. E. Luria. 1942. Interference between bacterial viruses. I. Interference between two bacterial viruses acting upon the same host, and the mechanism of virus growth. Arch. Biochem. **1**:111–41.

Diaz, E., A. Ferrandez, M. A. Prieto, and J. L. Garcia. 2001. Biodegradation of aromatic compounds by *Escherichia coli*. Microbiol. Mol. Biol. Rev. **65**:523–69.

Dumon-Seignovert, L., G. Cariot, and L. Vuillard. 2004. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). Protein Expr. Purif. **37**:203–6.

Eiteman, M. A., and E. Altman. 2006. Overcoming acetate in *Escherichia coli* recombinant protein fermentations. Trends Biotechnol. **24**:530–6.

Elena, S. F., and R. E. Lenski. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat. Rev. Genet. **4**:457–69.

Elsinghorst, E. A., and R. P. Mortlock. 1994. Molecular cloning of the *Escherichia coli* B L-fucose-D-arabinose gene cluster. J. Bacteriol. **176**:7223–32.

Feist, A. M., C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. O. Palsson. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. **3**:121.

Fong, S. S., A. P. Burgard, C. D. Herring, E. M. Knight, F. R. Blattner, C. D. Maranas, and B. O. Palsson. 2005. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol. Bioeng. **91**:643–8.

Francetic, O., D. Belin, C. Badaut, and A. P. Pugsley. 2000. Expression of the endogenous type II secretion pathway in *Escherichia coli* leads to chitinase secretion. Embo J. **19**: 6697–703.

Franchini, A. G., and T. Egli. 2006. Global gene expression in *Escherichia coli* K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. Microbiology **152**:2111–27.

Gatehouse, D., S. Haworth, T. Cebula, E. Gocke, L. Kier, T. Matsushima, C. Melcion, T. Nohmi, T. Ohta, S. Venitt, et al. 1994. Recommendations for the performance of bacterial mutation assays. Mutat. Res. **312**:217–33.

Gross, C. A. 1996. Function and regulation of the heat shock proteins. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, D. Schneider, and H. E. Umbarger (eds.), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, 2nd ed. ASM Press, Washington, DC, pp. 1382–99.

Haeusser, D. P., and P. A. Levin. 2008. The great divide: coordinating cell cycle events during bacterial growth and division. Curr. Opin. Microbiol. **11**:94–9.

Han, M. J., and S. Y. Lee. 2006. The *Escherichia coli* proteome: past, present, and future prospects. Microbiol. Mol. Biol. Rev. **70**:362–439.

Harder, K. J., H. Nikaido, and M. Matsuhashi. 1981. Mutants of *Escherichia coli* that are resistant to certain beta-lactam compounds lack the *ompF* porin. Antimicrob. Agents Chemother. **20**:549–52.

Hayashi, K., N. Morooka, Y. Yamamoto, K. Fujita, K. Isono, S. Choi, E. Ohtsubo, T. Baba, B. L. Wanner, H. Mori, and T. Horiuchi. 2006. Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Mol. Syst. Biol. **2**:2006 0007.

Helmstetter, C. E. 1968. DNA synthesis during the division cycle of rapidly growing *Escherichia coli* B/r. J. Mol. Biol. **31**:507–18.

Herrera, G., A. Martinez, M. Blanco, and J. E. O'Connor. 2002. Assessment of *Escherichia coli* B with enhanced permeability to fluorochromes for flow cytometric assays of bacterial cell function. Cytometry **49**:62–9.

Herrera, G., A. Urios, V. Aleixandre, and M. Blanco. 1993. Mutability by polycyclic hydrocarbons is improved in derivatives of *Escherichia coli* WP2 *uvrA* with increased permeability. Mutat. Res. **301**:1–5.

Ishii, N., K. Nakahigashi, T. Baba, M. Robert, T. Soga, A. Kanai, T. Hirasawa, M. Naba, K. Hirai, A. Hoque, P. Y. Ho, Y. Kakazu, K. Sugawara, S. Igarashi, S. Harada, T. Masuda, N. Sugiyama, T. Togashi, M. Hasegawa, Y. Takai, K. Yugi, K. Arakawa, N. Iwata, Y. Toya, Y. Nakayama, T. Nishioka, K. Shimizu, H. Mori, and M. Tomita. 2007. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. Science **316**:593–7.

Jansson, P. E., A. A. Lindberg, B. Lindberg, and R. Wollin. 1981. Structural studies on the hexose region of the core in lipopolysaccharides from Enterobacteriaceae. Eur. J. Biochem. **115**:571–7.

Jensen, K. F. 1993. The *Escherichia coli* K-12 "wild types" W3110 and MG1655 have an *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels. J. Bacteriol. **175**:3401–7.

Jeong, H., V. Barbe, D. Vallenet, S.-H. Choi, C. H. Lee, S.-W. Lee, B. Vacherie, S. H. Yoon, D.-S. Yu, L. Cattolico, C.-G. Hur, H.-S. Park, B. Ségurens, M. Blot, D. Schneider, F. W. Studier, S. C. Kim, T. K. Oh, R. E. Lenski, P. Daegelen, and J. F. Kim. Genome sequencing and comparative analysis of *Escherichia coli* B REL606 and BL21(DE3). submitted.

Jeong, K. J., and S. Y. Lee. 2002. Excretion of human beta-endorphin into culture medium by using outer membrane protein F as a fusion partner in recombinant *Escherichia coli*. Appl. Environ. Microbiol. **68**:4979–85.

LeBlanc, D. J., and R. P. Mortlock. 1971. Metabolism of D-arabinose: a new pathway in *Escherichia coli*. J. Bacteriol. **106**:90–6.

Lederberg, J. 2004. *E. coli* K-12. Microbiol. Today **31**:116.

Lederberg, S. 1966. Genetics of host-controlled restriction and modification of deoxyribonucleic acid in *Escherichia coli*. J. Bacteriol. **91**:1029–36.

Legrain, C., V. Stalon, and N. Glansdorff. 1976. *Escherichia coli* ornithine carbamolytransferase isoenzymes: evolutionary significance and the isolation of λ*argF* and λ*argI* transducing bacteriophages. J. Bacteriol. **128**:35–8.

Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am. Nat. **138**.

Michelsen, O., M. J. Teixeira de Mattos, P. R. Jensen, and F. G. Hansen. 2003. Precise determinations of C and D periods by flow cytometry in *Escherichia coli* K-12 and B/r. Microbiology **149**:1001–10.

Miroux, B., and J. E. Walker. 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J. Mol. Biol. **260**:289–98.

Nandakumar, M. P., A. Cheung, and M. R. Marten. 2006. Proteomic analysis of extracellular proteins from *Escherichia coli* W3110. J. Proteome Res. **5**:1155–61.

Neidhardt, F. C., and H. E. Umbarger. 1996. Chemical composition of *Escherichia coli*. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, D. Schneider, and H. E. Umbarger (eds.), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, 2nd ed. ASM Press, Washington, DC, pp. 13–16.

Neijssel, O. M., M. J. Teixeira de Mattos, and D. W. Tempest. 1996. Growth yield and energy distribution. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, D. Schneider, and H. E. Umbarger (eds.), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, 2nd ed. ASM Press, Washington, DC, pp. 1683–92.

Nikaido, H. 2003. Molecular basis of bacterial outer membrane permeability revisited. Microbiol. Mol. Biol. Rev. **67**:593–656.

Nikaido, H. 1996. Outer membrane. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, D. Schneider, and H. E. Umbarger (eds.), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, 2nd ed. ASM Press, Washington, DC, pp. 29–47.

Philippe, N., E. Crozat, R. E. Lenski, and D. Schneider. 2007. Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. Bioessays **29**:846–60.

Phue, J. N., B. Kedem, P. Jaluria, and J. Shiloach. 2007. Evaluating microarrays using a semiparametric approach: application to the central carbon metabolism of *Escherichia coli* BL21 and JM109. Genomics **89**:300–5.

Phue, J. N., S. B. Noronha, R. Hattacharyya, A. J. Wolfe, and J. Shiloach. 2005. Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and Northern blot analyses. Biotechnol. Bioeng. **90**:805–20.

Pieper, D. H., and W. Reineke. 2000. Engineering bacteria for bioremediation. Curr. Opin. Biotechnol. **11**:262–70.

Posfai, G., G. Plunkett, 3rd, T. Feher, D. Frisch, G. M. Keil, K. Umenhoffer, V. Kolisnychenko, B. Stahl, S. S. Sharma, M. de Arruda, V. Burland, S. W. Harcum, and F. R. Blattner. 2006. Emergent properties of reduced-genome *Escherichia coli*. Science **312**:1044–6.

Pugsley, A. P., and J. P. Rosenbusch. 1983. OmpF porin synthesis in *Escherichia coli* strains B and K-12 carrying heterologous *ompB* and/or *ompF* loci. FEMS Microbiol. Lett. **16**:143–47.

saiSree, L., M. Reddy, and J. Gowrishankar. 2001. IS186 insertion at a hot spot in the *lon* promoter as a basis for Lon protease deficiency of *Escherichia coli* B: identification of a consensus target sequence for IS186 transposition. J. Bacteriol. **183**:6943–6.

Sauer, U. 2001. Evolutionary engineering of industrially important microbial phenotypes. Adv. Biochem. Eng. Biotechnol. **73**:129–69.

Schneider, D., E. Duperchy, J. Depeyrot, E. Coursange, R. Lenski, and M. Blot. 2002. Genomic comparisons among *Escherichia coli* strains B, K-12, and O157:H7 using IS elements as molecular markers. BMC Microbiol. **2**:18.

Seo, J. H., D. G. Kang, and H. J. Cha. 2003. Comparison of cellular stress levels and green-fluorescent-protein expression in several *Escherichia coli* strains. Biotechnol. Appl. Biochem. **37**:103–7.

Sorensen, H. P., and K. K. Mortensen. 2005. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J. Biotechnol. **115**:113–28.

Studier, F. W., and B. A. Moffatt. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189**:113–30.

Swartz, J. R. 1996. *Escherichia coli* recombinant DNA technology. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Shaechter, and H. E. Umbarger (eds.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, DC, pp. 1693–1712.

Tian, G., D. Lim, J. D. Oppenheim, and W. K. Maas. 1994. Explanation for different types of regulation of arginine biosynthesis in *Escherichia coli* B and *Escherichia coli* K12 caused by a difference between their arginine repressors. J. Mol. Biol. **235**:221–30.

Tsilibaris, V., G. Maenhaut-Michel, and L. Van Melderen. 2006. Biological roles of the Lon ATP-dependent protease. Res. Microbiol. **157**:701–13.

Umbarger, H. E. 1996. Biosynthesis of the branched-chain amino acids. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, D. Schneider, and H. E. Umbarger (eds.), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, 2nd ed. ASM Press, Washington, DC, pp. 442–57.

Vallenet, D., L. Labarre, Z. Rouy, V. Barbe, S. Bocs, S. Cruveiller, A. Lajus, G. Pascal, C. Scarpelli, and C. Medigue. 2006. MaGe: a microbial genome annotation system supported by synteny results. Nucl. Acids Res. **34**:53–65.

Vostiar, I., J. Tkac, and C. F. Mandenius. 2004. Off-line monitoring of bacterial stress response during recombinant protein production using an optical biosensor. J. Biotechnol. **111**:191–201.

Wilcox, P., A. Naidoo, D. J. Wedd, and D. G. Gatehouse. 1990. Comparison of *Salmonella typhimurium* TA102 with *Escherichia coli* WP2 tester strains. Mutagenesis **5**:285–91.

Witkin, E. M. 1946. Inherited differences in sensitivity to radiation in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **32**:59–68.

Xia, X. X., M. J. Han, S. Y. Lee, and J. S. Yoo. 2008. Comparison of the extracellular proteomes of *Escherichia coli* B and K-12 strains during high cell density cultivation. Proteomics **8**:2089–103.

Yoon, S. H., M. J. Han, H. Jeong, C. H. Lee, X. X. Xia, J. H. Shim, S. Y. Lee, T. K. Oh, and J. F. Kim. Comparative multi-omics systems analysis of closely related *Escherichia coli* strains. submitted.

Yoon, S. H., M. J. Han, S. Y. Lee, K. J. Jeong, and J. S. Yoo. 2003. Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. Biotechnol. Bioeng. **81**:753–67.

Yoon, S. H., C. G. Hur, H. Y. Kang, Y. H. Kim, T. K. Oh, and J. F. Kim. 2005. A computational approach for identifying pathogenicity islands in prokaryotic genomes. BMC Bioinformatics **6**:184.

Yu, B. J., B. H. Sung, M. D. Koob, C. H. Lee, J. H. Lee, W. S. Lee, M. S. Kim, and S. C. Kim. 2002. Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. Nat. Biotechnol. **20**:1018–23.