Table of contents

II. List of abbreviations

Ap	ampicillin
BIME	bacterial interspersed mosaic element
Cm	chloramphenicol
cTc	chlortetracycline
DSB	double strand break
Kan	kanamycin
FLP	“flip”-recombinase of yeast 2μ plasmid
FRT	FLP-recombinase target
IRU	intergenic repeat unit
IS	insertion sequence
MDS	multi-deletion strain
MSS	Ma-Sandri-Sarkar
MLL	maximum likelihood method
OD	optical density
ORF	open reading frame
PAI	pathogenicity island
PCR	polymerase chain reaction
Rhs	rearrangement/recombination hot spot
Rp	rifampicin
RSA	repetitive sequence A
wt	wild type

III. Goals

1.Optimize the linear DNA fragment method for serial deletion construction

2.Construct an *E. coli* strain that lacks all mobile genetic elements

3.Develop a technique to measure the overall rate and composition of mutations in bacterial cells

4.Compare phenotypic traits of the multi-deletion and the parental strain

1. Introduction

1.1. *Escherichia coli,* and its genome

Escherichia coli is a facultative anaerobe, gram-negative, rod shaped bacterium. It colonizes the lower gut of animals, but due to its highly adaptive capabilities, can survive outside of its natural hosts, making it suitable for widespread dissemination (Neidhardt *et al.*, 1996). *E. coli* strain K-12 was originally isolated from the stool of a diphteria patient in 1922 (Lederberg, 1951). Since then, it became the most studied and best-known biological model-organism. It is also the most frequently used laboratory bacterium, which made the birth and the advance of molecular biology possible. Today, a vast repertoire of genetic techniques is available for this species, making it useful for almost any experiment involving DNA-cloning. In addition, it is widely used for biotechnological purposes, such as protein overproduction or small molecule biosynthesis.

E. coli is also known for being the causative agent of several diseases. Among these, the most common are the urinary tract infections, mostly caused by the uropathogenic strains. The enterotoxinogenic- (ETEC), enteropathogenic- (EPEC), enteroinvasive- (EIEC), enteroaggregative- (EAggEC), enterohemmorhagic- (EHEC) and enteroadherent- (EAEC) strains of *E. coli* cause enteric infections, with various clinical manifestations. Acute bacterial meningitis is a dangerous disease, caused at neonatal age mostly by *E. coli* K1. In a case of bowel-perforation, *E. coli* can be an agent of polymicrobial infections resulting in peritonitis or intra-abdominal abscesses. Ascending cholangitis and cholecystitis can also be caused by *E. coli*, especially in patients with biliary obstruction (Go *et al*., 2004).

It is not surprising, that the first sequencing project to target a living cell was the *Escherichia coli* K-12 MG1655 strain's genome-sequencing program, conducted by Frederick Blattner at the University of Wisconsin-Madison, USA. Due to its large size of 4,639,221 bp though, its completion was preceded by the sequencing of five smaller prokaryotic and one eukaryotic genome. 78.8% of the K-12 genome was reported to encode proteins, 0.8% coded stable RNAs and 0.7% consisted of non-coding repeats (BIME, IRU, BoxC, RSA and Rhs sequences). The remaining 11% is responsible for regulatory and other functions. A total of 4288 protein-coding genes were annotated, but more than one third of them had no attributed function. 1385 proteins have at least one paralogous pair, and 86 belong to distinct izozyme-families. 44 insertion sequences (ISs) and eight defective or cryptic prophages, integrated at various locations of the genome, were identified. The genome was found to have a 50.8%

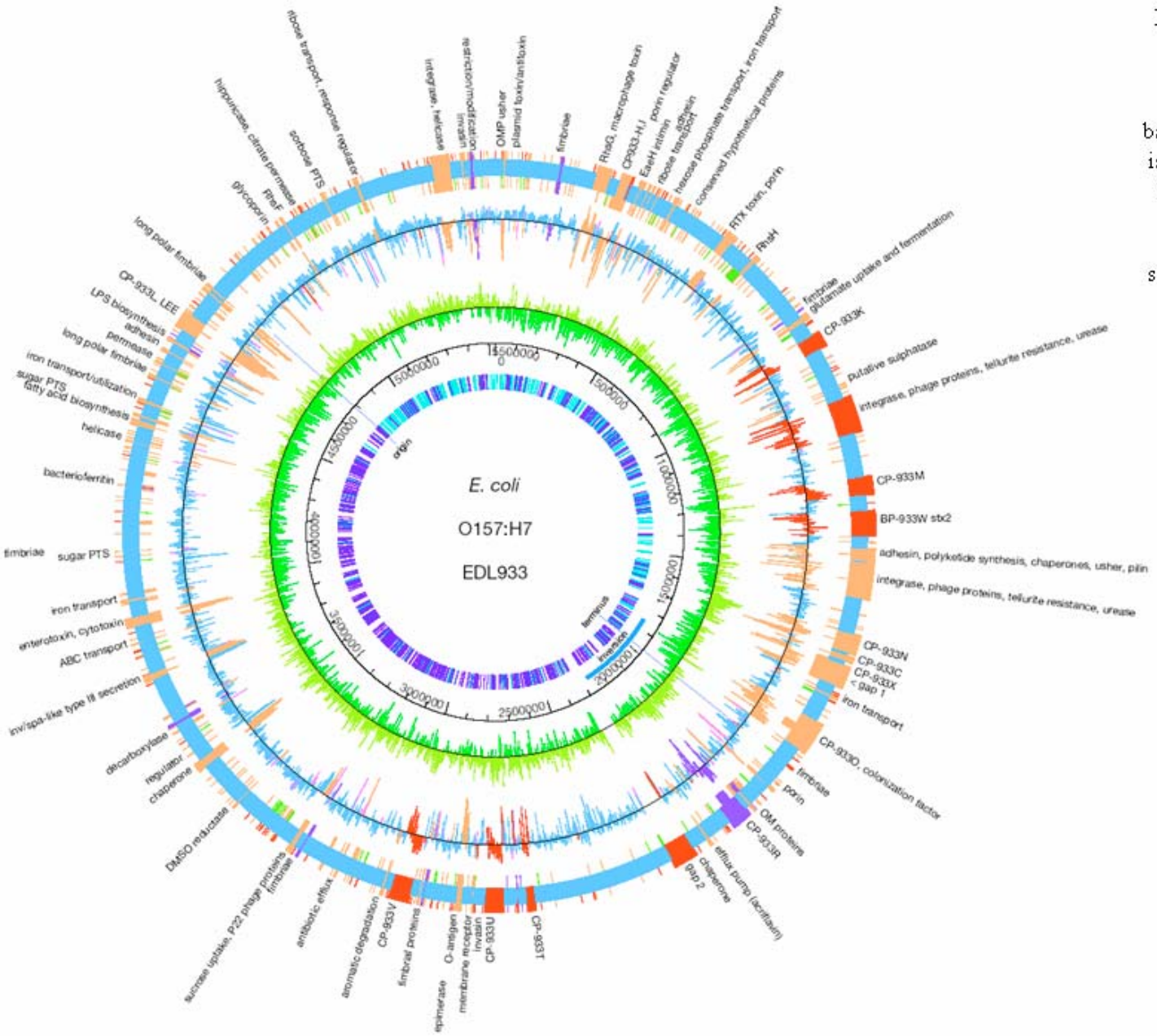


Figure 1.1. The genome map of *E. coli* O157:H7. On the outermost ring, the blue bar represents the conservative backbone. O-islands are red, K-islands green, while tan blocks show K- and O-islands in the same position. Purple bars represent hypervariable sequences. (Perna *et al.*, 2001.)

overall G+C content with a significant GC-skew favoring guanines in the leading strand (Blattner, 1997).

Less than 4 years later, the 5.5Mbp genomes of the EDL933 and the Sakai isolates of enterohaemorrhagic *Escherichia coli* (EHEC) O157:H7 were also sequenced (Perna *et al*., 2001; Hayashi *et al*., 2001, respectively). The K-12 and the O157:H7 strains evolved from a common ancestor, that existed 4.5 million years ago (Reid *et al.*, 2000). Comparison of the EHEC genome with that of the K-12 revealed several interesting facts: The strains' genomes were shown to harbor a highly conserved backbone gene set, with a sequence similarity >98% between the two strains. This "core" set of genes comprises 3541 ORFs and is 4.1Mbp long. The order of genes is highly colinear, except for a 422kbp long inversion spanning the replication terminus of the EDL933 isolate of the EHEC strain. 75% of the proteins differ in at least one amino acid, denoting the effect of parallel evolution. A few homologous pairs, like fimbriae and restriction-modification systems show only a low degree of similarity, down to a mere 34%. The conservative backbone sequence is interrupted by genomic islands specific for the K-12 or the O157:H7 strain, called K- or O-islands, respectively (**Figure 1.1**). The 234 K-islands comprise of 528 ORFs, while the 177 O-islands contain 1387. This implies, that O-islands are usually longer than their K-counterparts. Strain-specific islands can be formed either by losing genomic segments of a common ancestor in one of the lineages, or by the acquisition of foreign DNA (Eisen *et al*., 2001) by natural transformation, conjugation or phage transduction (Salyers and Whitt, 2001). The low codon-adaptation index (Sharp and Li, 1987), the unusual GC content or the presence of phage-related genes within such islands (Hayashi *et al*., 2001) mostly imply the latter mechanisms, which is usually referred to as horizontal- or lateral gene transfer (abbreviated HGT or LGT, respectively). Apart from the seemingly randomly located genomic islands, there are 106 examples of K-islands and O-islands present at the same location, relative to the conserved backbone, possibly marking specific integration sites of foreign DNA (Perna *et al*., 2001).

The sequencing of the uropathogenic *Escherichia coli* (UPEC) strain CFT073 further confirmed the existence of integration hotspots (Welch *et al.*, 2002). Moreover, 13 CFT073, 10 EDL933 and 11 Sakai strain-specific islands use tRNA genes as integration sites (Hayashi *et al*., 2001; Welch *et al.*, 2002), already known to be favored by bacteriophages (Cheetham, Mol Mi 1995; Hacker Mol Mi 1997). A further interesting feature is that the O157:H7 strain-specific islands carry tRNA genes for the codons abundant within the genes of the islands, but rare in the backbone-ORFs (Hayashi *et al*., 2001). Comparison of the non-redundant protein set of the three strains (**Figure 1.2**) reveals a surprisingly low percentage of shared

components (Welch *et al.*, 2002). Only 11% of the non-backbone genes of CFT073 were found in EDL933, indicating the different niches and virulence mechanisms of the two pathogenic strains (Welch *et al.*, 2002; Thomson and Parkhill, 2003).

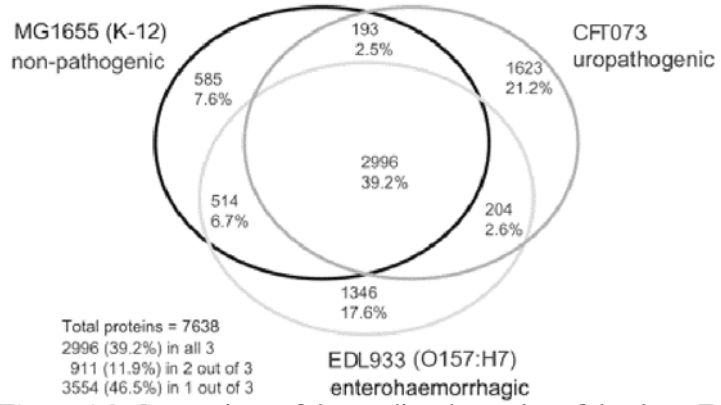


Figure 1.2. Comparison of the predicted proteins of the three *E. coli* strains to show the number of orthologs in each shared category and the numbers of strain-specific proteins. (Welch *et al.*,2002)

For example, the uropathogenic coli's genome is rich in genes coding for fimbrial adhesins, autotransporters, iron sequestration systems and phase-switch recombinases, but lack a type III secretion system or toxins typical of the enterohaemorrhagic strain. Even among different UPEC isolates, there's a peculiar difference concerning the localization, gene-content or gene-order of pathogenicity islands (Welch *et al.*, 2002).

Bacterial pathogenicity islands (PAI) were first described by Jörg Hacker *et al*. in 1990. They found that genes coding fimbriae and haemolysins in uropathogenic *E. coli* lie closely linked in certain segments of the genome. Deletion of these regions was found to occur both in vitro and in vivo, leading to the loss of the strain's virulence. These findings brought them to the hypothesis, that gain and loss of such PAIs could represent a general mechanism of bacterial virulence modulation. They defined PAIs as large regions of the genome, that carry virulence-associated genes, have a G+C content different than the rest of the chromosome, are frequently associated with tRNA genes, are often flanked by repeat-structures, contain integrase, transposase or other mobility genes and are often unstable (Hacker *et al*., 1997 and 2003). Later the term was widened to "genomic islands" (GI), as knowledge of similar DNA-segments coding for niche-specific function emerged concerning non-pathogenic bacteria. (A good example of the latter is the high pathogenity island, or HPI found in pathogenic Yersinia, commensal *E. coli* and non-pathogenic strains of Salmonella enteritica, too (Hacker *et al*., 2003).) A theory emerged, stating that bacterial genomes consist of a stable, conserved,

species-specific backbone, and a flexible, more rapidly changing combination of genes derived from a gene-pool available to many strains and species.

1.2. Minimal-genomes

Precise sequence data and sufficient gene-function knowledge was available to start a large-scale genome-engineering project in *E. coli* in 1999. Certain techniques for such purposes were also available, but their perfection was needed to conveniently carry out repetitive genomic deletions in the same bacterial cell. Prior to our project, only one work dealing with the systematic deletion of multiple genomic segments was published: Asai *et al*. (1999a) deleted all seven rRNA operons of *E. coli*, and studied the viability and growth characterisics of the mutants with heterologous rRNA genes carried on multicopy plasmids within the cell. They also reported the growth rate of *Escherichia coli* carrying a single copy of its original rrn operons to be 56% of the wild type (wt) (Asai *et al*. 1999b).

Determining the minimal gene set required to maintain a living cell is a question that's answer has long been sought by several microbial geneticist groups. Itaya *et al*. (1995) made random disruptions in a small set of *Bacillus subtilis* genes. From the fraction of dispensable loci, they calculated the *B. subtilis* genome to contain 300 to 500 indispensable genes. A systematic gene-inactivation project conducted by Kobayashi *et al*. (2003) later defined 271 open reading frames (ORFs) of *B. subtilis* to encode essential functions. This falls within the range of the 265 to 350 essential genes of *Mycoplasma genitalium*, estimated using transposon mutagenesis (Hutchison *et al*., 1999). A similar approach was used to assign indispensable functions to 478 *Haemophilus influenzae* and 620 *Escherichia coli* ORFs (Akerley *et al*., 2002, Gerdes *et al*., 2003, respectively). Knuth *et al*. (2004) cloned small genomic segments of *Salmonella enteritica serovar Typhimurium* into a conditionally replicative plasmid, and attempted to obtain genomic cointegrates by switching to non-permissive conditions. In cases of failure to achieve such cointegrates, the genomic segment was classified as essential. Taking the sensitivity of the method into consideration, the 257 essential genes identified this way probably represented 50% of the indispensable gene set of *Salmonella enteritica serovar Typhimurium*. It is important to note, that the works listed above gave only a list of the essential gene set of wt organisms. The minimal gene set required for an individual living cell is probably significantly larger. This difference is caused by the presence of synonymous genes, coding for an essential function. One-by-one, they are dispensable, but a cell lacking all of them is not viable (Goryshin *et al*., 2003). It is also important to define in such projects the conditions, under which the cells' viability is tested.

In this work, I present the deletion of 41 segments from the *Escherichia coli* genome. Our project was based on the hypothesis, that the genes present in the strain-specific genomic islands are dispensable under common culturing conditions. Our group's goal was to delete as many K-islands in the MG1655 strain as possible. We investigated how the accumulation of such deletions changes cellular physiology and applicability for biotechnological purposes.

1.3. Genome-modification techniques

Modification of the genetic material of bacterial cells originates from the earliest days of molecular biology, with the transformation of bacterial cells with plasmids carrying genes of interest. Soon thereafter, modification of the chromosome also became feasible. Genomes were randomly mutagenized, and mutants showing certain phenotypes were isolated for further characterization. A possible way for the identification of the mutated gene was by complementation with a genomic library, and searching for revertants to the original phenotype. Localization of the given gene relative to other known loci was possible by conjugation using Hfr donors or co-transduction with known chromosomal markers. As the extent of sequence information concerning certain bacterial species increased, so grew a need for a method capable of introducing modifications into the genome in a targeted manner. Early ways to generate deletions by P2 eduction (Sunshine and Kelly, 1971) or imprecise excision of Tn10 (Kleckner *et al*., 1977) were too random and poorly reproducible. Modern methods of genome manipulation usually involve the introduction of a suicide plasmid or a linear DNA molecule into the cell, which integrate into the chromosome by homologous recombination.

1.3.1. Suicide plasmid-deletion methods

Suicide plasmids are conditionally replicative circular DNA molecules. They can only be propagated in permissive hosts. Upon entering a non-permissive target cell or transfer to non-permissive conditions, they are either recombined into the chromosome, or lost from the cell. Recombination takes place between the genome and a short, homologous section of the plasmid, and is mediated by the cellular RecA protein (**Figure 1.3.A**). Selection for the recombinants is made possible by the antibiotic-resistance genes of these vectors. The next step of allelic exchange is the resolution of such cointegrates, which can result in either the wt or the plasmid-derived, mutant version of the gene remaining in the chromosome (**Figure 1.3.B**). Screening for the latter case can be done by phenotypic tests, Southern-hybridization, colony PCR or sequencing. The first such system routinely used (Gutterson and Koshland,

1983) exploited the fact, that plasmids harboring a ColE1 origin cannot replicate in PolA$^-$ hosts. Later, the application of the temperature-sensitive pSC101 replicon

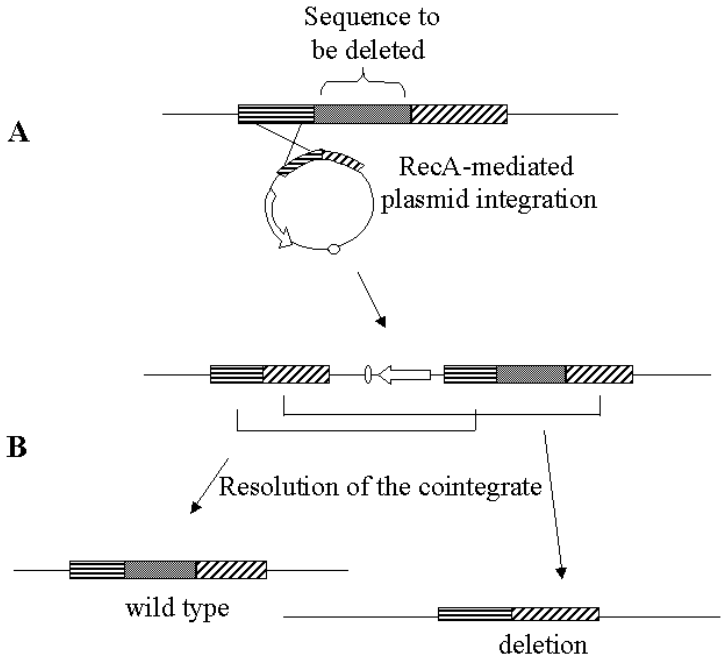


Figure 1.3. General scheme of the suicide plasmid-deletion method

made allelic exchange in wt hosts possible, too (Hamilton *et al*., 1989). Plasmids with an R6K γ-type origin are also useful for such purposes, since they require the π-protein, the product of the *pir* gene for replication (Miller and Mekalanos, 1988). The spontaneous resolution rate of plasmid-cointegrants is approximately $1.7*10^{-3}$/generation (Pósfai *et al*., 1999). Since this is inconveniently low, several procedures were developed to mediate- or to select for the result of this process. In the case of pSC101^{ts} replicons, growth of the cointegrate at permissive temperatures impairs chromosomal replication, thereby enriching the culture for cells with chromosomes lacking the integrated plasmid (Hamilton *et al*., 1989). A more effective way to select for the loss of the integrated plasmid is SacB-counterselection (Blomfield *et al*., 1991). Cells harboring a *sacB*-gene are not viable on 5% sucrose, due to the toxic effect of the encoded levansucrase enzyme. (Gay *et al.*, 1985) Some drawbacks of this method are false positive colonies caused by spontaneous mutations in the *sacB* gene, and the procedure's strain- medium- and temperature dependence (Blomfield, 1991; Link *et al*., 1997; Pósfai *et al*., 1999). To surpass these drawbacks, Pósfai *et al*., (1999) introduced the 18 bp-long target site of an ultra-rare cutting endonuclease, I-SceI into a suicide plasmid. I-Sce has no restriction sites in the *E. coli* chromosome, unless the given suicide plasmid is integrated

into it. Upon expression from a separate plasmid, I-SceI introduces a double-strand-break into the cointegrate, thereby stopping the cell's replication. This not only provides counterselection, but, by generating free DNA ends, also stimulates intermolecular recombination, increasing the efficiency of resolution by two to three orders of a magnitude (Pósfai *et al*., 1999). Single plasmids can be used to generate insertions, point mutations or deletions. If two suicide plasmids, both carrying FRT-sites or I-SceI sites are integrated into the genome, the flanked DNA segment can be rescued as a plasmid (Pósfai *et al*., 1997) or deleted (Pósfai *et al*., 1999) by expression of the FLP or I-SceI enzymes, respectively.

1.3.2. Linear DNA-based methods

The second large group of current genome-engineering tools is the recombination of a linear DNA fragment into a circular replicon. Jasin and Schimmel (1984) were the first to delete a gene this way from an *sbcB, recBC E. coli* target cell's genome. In wt *E. coli* cells, the RecA enzyme is responsible for the DNA bindig and strand exchange-, while the RecBCD (exonucleaseV) fulfils the helicase and strand-specific endonuclease functions (Lloyd and Low, 1996), which are all needed for the homologous gene-exchange seen in the suicide plasmid method. RecBCD also degrades linear DNA (Benzinger *et al*., 1975). Unfortunately *recBC*$^-$ strains grow extremely poorly, are defective in recombination and do not support the replication of many plasmids (Yu *et al*., 2000). The earliest solution of this problem came with the realization, that mutations in the *sbc* operon suppress the drawbacks of the *recBC* phenotype. Deleterious mutations in the *sbcA* gene derepress the *recE/recT* genes (Barbour *et al*., 1970; Kushner *et al*., 1971), thereby providing appropriate enzymes for recombination, even if the cellular *recA/recBCD* pathway is impaired. RecE is a 5'-3' exonuclease (Kushner *et al*., 1971; Joseph and Kolodner, 1983), while RecT is a ssDNA-binding protein that promotes annealing, strand transfer and strand invasion, in vitro (Clark *et al*., 1993; Hall *et al*., 1993; Hall and Kolodner, 1994; Noirot and Kolodner, 1998). Their genes are part of the Rac cryptic prophage in the *E. coli* K-12 strain's genome (Clark 1974; Willis *et al*., 1985). Later, *recD* derivatives of *E. coli* were found to be recombinase-proficient but exonuclease-deficient, and so were also used in gene disruption experiments (Russel *et al*., 1989). Zhang *et al.* (1998) made this technology applicable in a wide-variety of strains by constructing pBADETγ. This plasmid, in addition to encoding RecT and the inducible RecE, also expresses Gam, a λ phage protein that inhibits the linear DNA-degradative activity of the RecBCD complex (exonucleaseV) (Murphy, 1991). The λ phage orthologues of RecE and RecT, called Redα and Redβ, respectively, were shown to be two to three times more

effective at stimulating homologous recombination (Muyrers *et al.*, 1999). In this work, we routinely used the latter recombinases, carried on the inducible plasmid pBADαβγ. Others use *E. coli* strains with the three *red*αβγ genes implanted into the chromosome, under the control of the lac repressor (Murphy *et al*. 2000), or a temperature-sensitive λ cI-repressor (Yu *et al*., 2000). The linear DNA fragments used in such experiments are usually made by PCR from a template plasmid carrying a marker gene for the easy selection of the cointegrates. The primers make up the terminal homologies of the linear DNA fragment, which are the sites of the double-crossover upon integration into the genome. This method is capable of the insertion-, the deletion- or simply the interruption of certain sequences in the bacterial chromosome, in a BAC (bacterial artificial chromosome) or in a plasmid. Perhaps the most important advantage of using the *recE*/*T* or *redαβγ* genes for genome engineering instead of cellular *recA* is the need for much shorter homologies. A minimum homology length of 20bps is mandatory, but nearly maximal recombination efficiencies are reached at lengths between 50 and 100 bps, as long as circular molecules are targeted (Muyrers *et al*., 2000; Yu *et al*., 2000).

In this work, I present the perfection and routine use of a linear fragment-based deletion method, along with a suicide plasmid-based procedure, both developed in our laboratory.

1.4. Dynamics of the genome: measuring mutation rates

A major aim of our strain-improvement project was to increase the genomic stability of *E. coli.* To monitor this cellular trait, a trustworthy and reproducible method was needed that allows the statistically correct comparison of the mutation rate of different bacterial strains. The genomic mutation rate is a fundamental evolutionary parameter of any population. Bacterial cells maintain a balance between faithful replication and mutation-generating mechanisms. While most mutations are deleterious, genetic diversity on which selection can work is a necessity for long-term survival in a changing environment. Mutations in bacteria occur typically at a rate of $<10^{-6}$/gene/replication (Drake *et al*., 1998). Analysis of such rare events usually requires the use of a screen, preferably a positive selection system which filters out the mutants from a much larger population. The ideal mutation detection system selects for mutations which (i) can be of any type, (ii) are neutral in respect of cell growth, (iii) do not require a specific genetic background, (iv) have a simple phenotype-genotype relationship and (v) can be easily and reproducibly counted.

A great number of mutation detection systems have been used in *E. coli* for mutagenesis studies (LaRossa, 1996), however, the use of them is frequently limited by various factors.

Variants of the widely used assays utilizing the *lac* operon provide only a screen, limiting the quantitative analysis. Among the positive selection systems, the frequently used rifampicin test detects only base substitutions and short in frame insertions/deletions, affecting one to four amino acid residues of RpoB (Wehrli, 1983; Jin and Gross, 1988). The test based on the activation of the cryptic *bgl* operon detects mostly IS transpositions (Hall, 1998). Assays based on reversal of auxotrophy require a specific genetic background and detect only specific mutations. Analysis of mutants resistant to ciprofloxacin and other fluoroquinolones is complicated by multiple resistance mechanisms: mutations in the *gyrA* (Oram and Fisher, 1991), *marRAB* (Goldman, 1996), *nfxD* (Breines *et al*., 1997), and in the *hip*A or *hipQ* genes (Wolfson *et al*, 1998) can all lead to decreased sensitivity to fluoroquinolones. Similarly, selection using beta-lactam antibiotics can yield, beside others, *crp, cyaA, ompF* (Jaffe *et al*, 1983), *lyt* (Shimmin *et al*., 1984), *alaS* or *argS* (Vinella *et al*., 1992) mutant *E. coli* cells. Several selection systems detect mutations in genes carried on multicopy plasmids, further complicating the analysis. The above list of mutation rate measurement techniques is far from complete, but explain the difficulty of finding or developing an adequate system for such purposes. In the course of our work of constructing multi-deletion (MDS) *E. coli* cells and removing insertion elements (ISs), we developed an assay, based on the selection of *cycA* mutants. This system readily detects base substitutions, frameshifts, short and long deletions, as well as the insertion of IS elements, and proved to be particularly useful for the analysis and comparison of mutation rates and spectra.

2. Materials and Methods

2.1. Strains and media

Plasmids were generally prepared from DH5α (Woodcock *et al*., 1989). Single deletions were made in *Escherichia coli* K-12 strain MG1655 (Blattner *et al*., 1997). In the experiments involving salicin and cycloserine, the minimal medium described by Hall (1998) was used, which will be referred to as MT medium. In all other cases, standard laboratory media (LB, M9/glucose minimal medium) and agar plates were used (Sambrook *et al.,* 1989, pp. A.1-A.3). Antibiotics were used in the following end-concentrations: ampicillin (Ap): 50 μg/ml, kanamycin (Kan): 25 μg/ml, chloramphenicol (Cm): 25 μg/ml and rifampicin (Rp): 100 μg/ml, cycloserine (Cyc): 0.04mM. Chlortetracycline (cTc) was suspended in LB to make a stock solution of 1 mg/ml, autoclaved for 20 minutes, then stored in the dark. It was used in final concentrations of 20 and 33 μg/ml. Salicin was used in an end-concentration of 0.5% (m/v). Casamino acids were added as a supplement to M9/glucose medium in an end-concentration of 0.5% (m/v). .

2.2. Plasmids

pSG76C or pSG76CS plasmids (accession numbers Y09893 and AF402780, respectively) were used as templates for generating linear fragments by PCR. The nucleotide sequences of pST98AS, and pSTKST are available under accession numbers AF170483 and AF406953, respectively. pSTAST was constructed by cleavage of pST98AS with EcoRI and PstI, blunting with Klenow fragment polymerase and ligation. pBADαβγ (Muyrers *et al*., 1999) was a kind gift of Francis Stewart. The nucleotide sequences of pST76-A and pKSUC1 are available under accession numbers Y09895 and AF402779, respectively. All plasmids termed "ST" have a temperature sensitive mutant form of the pSC101 replication origin, while plasmids termed "SG" have an R6K origin, thus require the Pir protein for replication.

2.3. Competent cells and transformation

"Traditional" competent cells were made by growing the cells in LB to an OD_{540} of 0.5-0.6. The cells were washed twice by centrifugation at 2200xg, and resuspended in 1.2 ml buffer containing 60 mM $CaCl_2$, 10 mM Tris-HCl, and 15 % glycerol. Afterwards, the cells were incubated on ice for 1 hour, then divided into 200 μl-aliquots. Transformation was carried out by the following sequential steps: addition of DNA to the aliquots, incubation on

ice for 1 h, two-minute heat shock at 42°C, addition of 1 ml of LB, shaking at 30 or 37 °C for 1 h, and spreading on LB plates containing the appropriate antibiotic.

Electro-competent cells were prepared by standard protocols. Briefly, cells were grown in a 100 ml culture to an OD_{540} of 0.5-0.6. Cells were then washed twice in ice-cold water, and once in ice-cold 10% glycerol. At the final step, the pellet was suspended in 200 µl of 10% glycerol, divided into aliquots of 40 µl, and stored at –80 °C. The cells were electroporated either with a linear DNA fragment or circular plasmid, at a voltage of 1800 V and a resistance of 150 Ω in a 0.1 cm electroporation cuvette using the Electroporator II from Invitrogen. Cells were diluted afterwards in 1 ml of LB, shaken at 30 °C or 37°C for 1-2 h, and finally plated on selective medium.

2.4. Constructing deletions using suicide plasmids

The method was described by Pósfai *et al.* in 1999. Briefly, using overlapping inner primers, an approximately 1 kbp genomic region containing the deletion was reproduced by a two-step PCR, and cloned into pST76-A (**Figure 2.1)**. The recombinant plasmid was electroporated into the deletion host, and plated on LB+Ap at 30 °C. A few colonies were restreaked on LB+Ap plates, and were incubated at 42 °C. Large colonies were screened by PCR using primer pairs **e** and **T7** or **d** and **Sce2** (depending on the homology used) for the

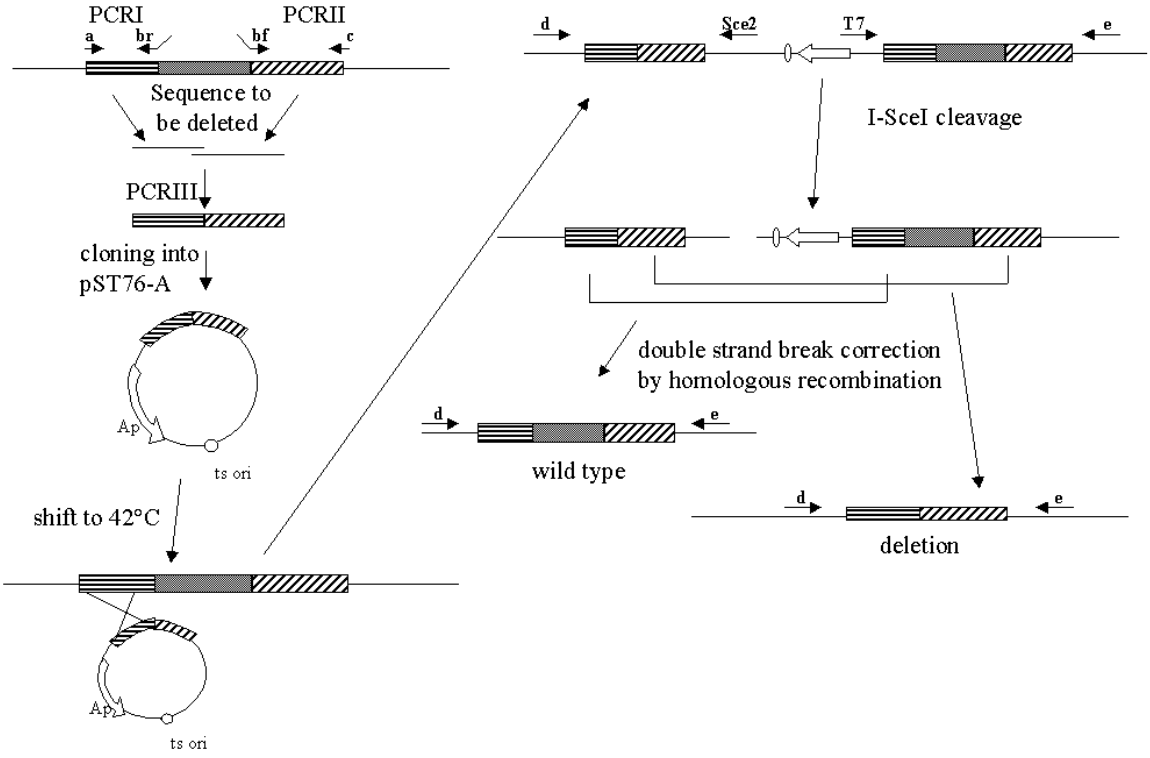


Figure 2.1. Deletion-construction using suicide plasmids

detection of recombination of the temperature-sensitive plasmid into the genome at the deletion-site. Positive colonies usually contained cells harboring unintegrated plasmids, too, so they were restreaked on LB+Ap plates, and screened again by PCR. The cells were then transformed by pSTKST, and expression of the I-SceI was induced by the addition of cTc. Cells were grown overnight at 30 ^{o}C, then 100 µl of the 10^5 dilution of the fully grown culture was spread on LB+Kan+cTc plates. The obtained colonies were screened by PCR with primers **d** and **e** for recombinants that lost the sequence that was to be deleted.

An alternative is to use the pSG76-A suicide plasmid. In this case, the Pir protein must be provided by a temperature-sensitive helper plasmid in the same cell. After shifting the incubation temperature to 42^{o}C, the Pir protein is diluted from the cells due to the loss of the helper plasmid. This usually takes slightly longer time than the loss of pST76-A, but the following steps of the deletion protocol are the same as above.

2.5. Constructing deletions using linear fragments

To generate primer **ab** (**Figure 2.2A**), 10 pmol of primer **a** was mixed with 10 pmol of primer **b** and PCR was performed in a total volume of 25 µl. Cycle parameters were: 15x(94^{o}C 40sec/57^{o}C or lower [depending on the extent of overlap between primers **a** and **b**] 40sec/72^{o}C 15sec). Next, 2 µl of this PCR product was mixed with 40 pmol of primers **a** and **c** (Fig.) each, 50 ng of pSG76-CS template and a second round of PCR was performed in a volume of 100 µl. Cycle parameters were: 28x(94^{o}C 40sec/57^{o}C 40sec/72^{o}C 80sec). The resulting, PCR-generated linear DNA-fragment was purified by Promega Wizard PCR purification kit, and suspended in 20 µl water. Elimination of the template plasmid (e.g., by *Dpn*I digestion) is not needed, since the template plasmid can only replicate in cells carrying the *pir* gene (for example, DH5απ cells). Electrocompetent MG/pBADαβγ cells were electroporated usually with 0.2-0.5 µg of linear DNA fragment, as indicated above, diluted afterwards in 1 ml of LB, shaken at 37°C for 2 h and finally plated on LB+Cm-plates. Colonies appearing after 24-36 hours, were screened for the correct recombination event by colony-PCR using primers **d** and **e** (**Figure 2.2E**). In order to facilitate the loss of the exogenous inserts from the genome, cells were then transformed either with I-SceI-expressing pKSUC1, or I-SceI production was induced in the cells from pSTKST, as described above. The resulting scarless deletions were verified by PCR using primers **d** and **e**.

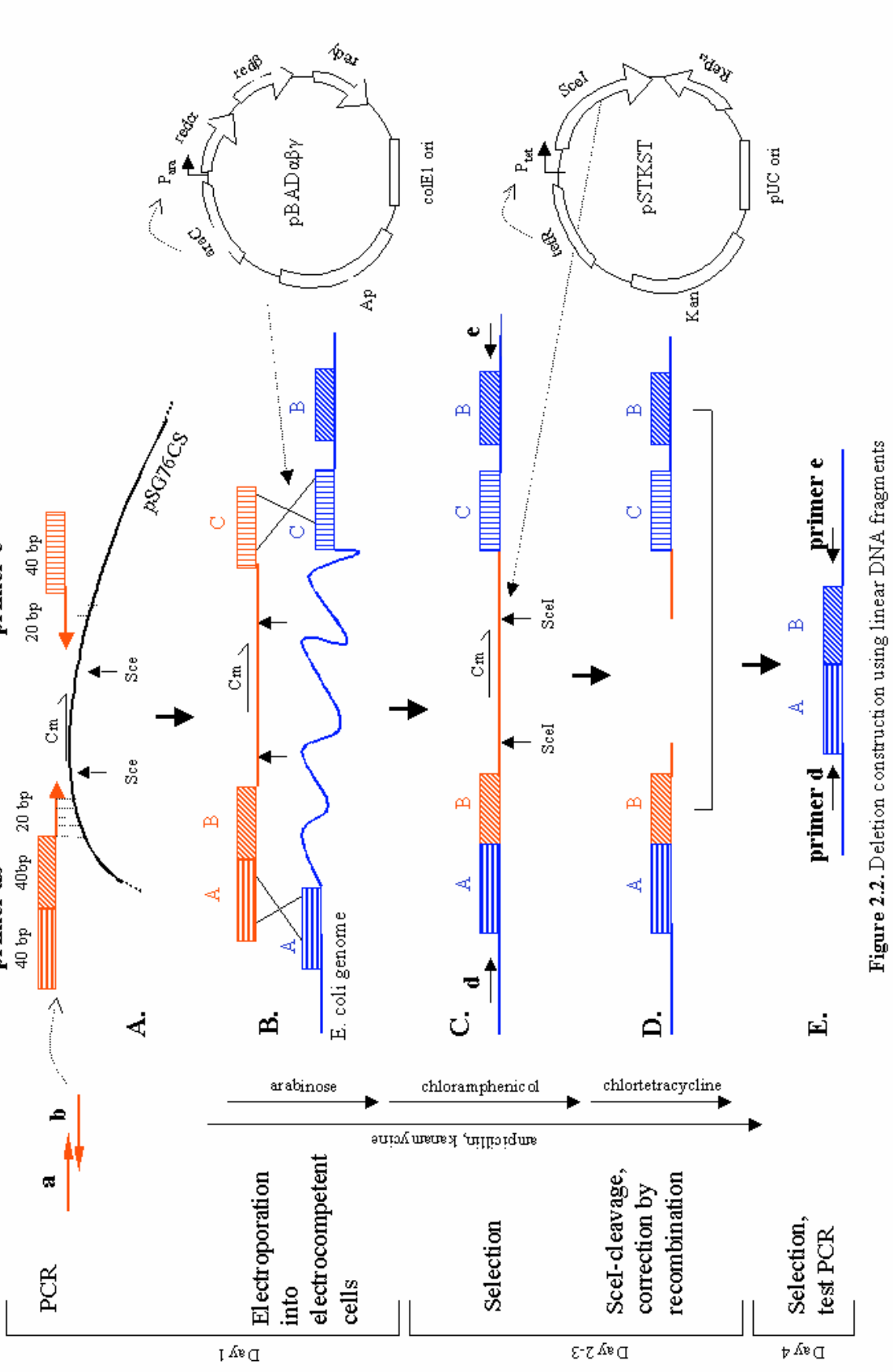


Figure 2.2. Deletion construction using linear DNA fragments

2.6. Localization of IS elements in the genome by inverse PCR

Inverse PCR has been described elsewhere (Ochman *et al*., 1988). Briefly, the appropriate bacterial genome was digested with several restriction endonucleases (one at a time), that do not cut into the type of IS being searched for. The restriction fragments were ligated into circles, and PCR was performed on them using primers pointing outward from the ends of the ISs (**Figure 2.3**). Primer pairs IS1A1 and IS1A2; UK1R and UK2R; and IS5ki1 and IS5ki2 were used to find IS1, IS2 and IS5, respectively. The products of the PCR

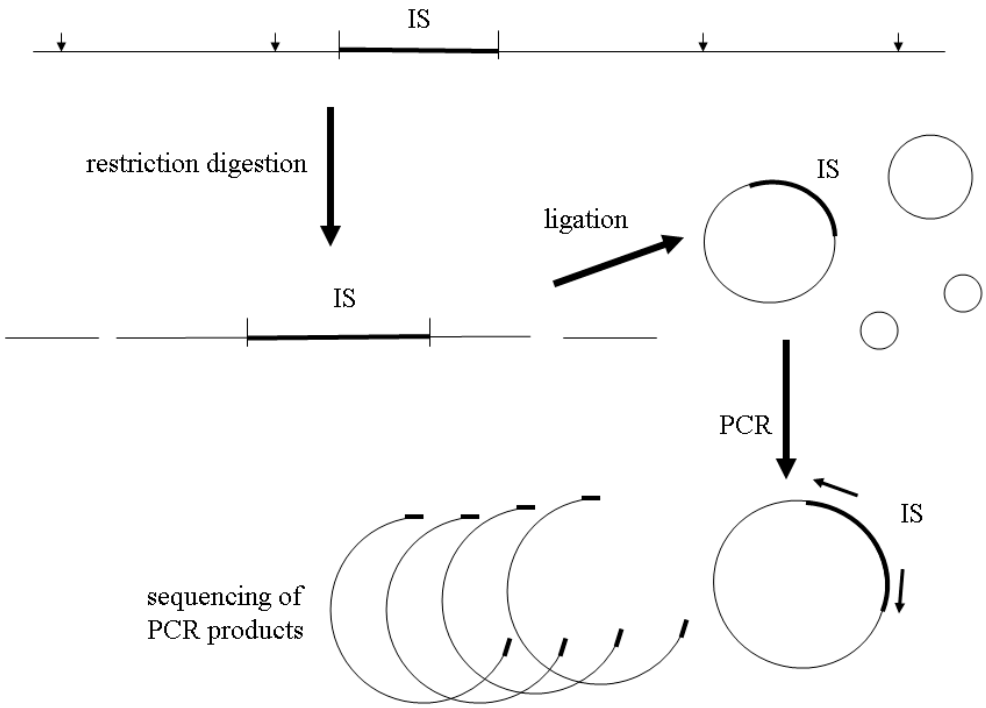


Figure 2.3. Search for extra (unsequenced) IS elements using inverse PCR

were examined by the electrophoresis of small aliquots on 1 % agarose gels. If a specific PCR product was visible, it was cleaned using the Promega Wizard® or the Viogene PCR-M™ PCR cleaning kit, according to the manufacturer's directions, and sequenced. The sequence data were compared to the *E. coli* genome using the BLAST program (Altschul *et al.,* 1990) to identify the exact position of the IS in the chromosome.

2.7. Growth rate measurements

1 ml of overnight grown, fresh bacterial suspension was added to 100 ml flasks of LB or M9. These flasks were shaken at 230 rpm at 30°C, 37°C or 42°C. The optical density (OD) of the suspension was measured every 30 minutes at 550 nm using a WPA Colourmax CO75000 colorimeter. The OD values were plotted against time to obtain a growth curve,

using the toolbox of Microsoft Excel. The doubling times of the strains were derived from the equations of the trendlines fitted to the curves.

2.8. Mutational rate measurements

2.8.1. Inactivation frequency of cycA resulting in cycloserine resistance

Briefly, 10^4 cells were inoculated into 20 tubes containing 1 ml glucose-MT each, and fully grown at 37°C in 36 hours (**Figure 2.4**). 50 µl from each tube was then spread on glucose-MT plates containing D-cycloserine, and incubated at 37°C for 30 hours. From the number of colonies counted the next day, the average number of mutations per 50 µl was approximated using the Ma-Sandri-Sarkar maximum-likelihood method (MSS-MLL)(Sarkar *et al.*, 1992). The obtained value, valid for 50 µl was extrapolated to the 1 ml volume of the tube using the following equation: $\mathbf{m}_{total}=\mathbf{m}_{fraction}(z-1)/(z*\ln(z))$

(eq. 41 given by Stewart *et al.*, in 1990).

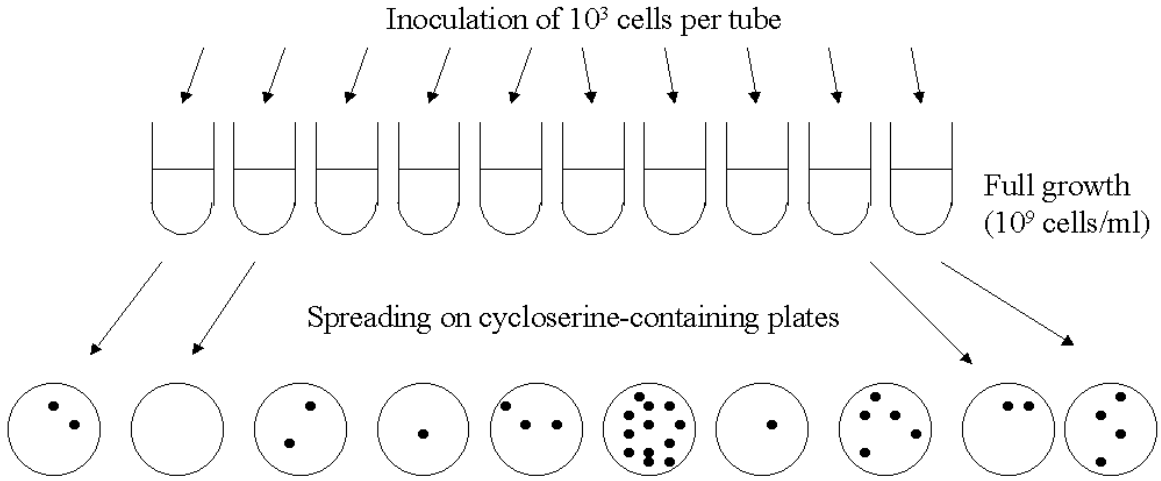


Figure 2.4. Experimental overview of mutation rate-meaurements using D-cycloserine-resistance

The distribution of the **m** values calculated this way can be simply transformed to a near-normal distribution by taking their natural logarithm. This allows us to statistically compare the ln(**m**) values of two strains using a two-tailed, unpaired t-test. To do this, the standard deviation of ln(**m**) was calculated the following way: $\sigma\approx 1.225\mathbf{m}^{-0.315}/C^{0.5}$, where C

is the number of parallel cultures (eq. 23 of Rosche and Foster, 2000). We only made statistical comparison of **m** values when the difference in total cell number was negligible (<3%, $p \geq 0.6$ with a two tailed, unpaired t-test). The total number of cells in a tube was calculated by spreading dilutions from 4 tubes onto non-selective plates. To obtain the mutational rate of *cycA* (mutations/cell/generation), the **m** value was divided by the total cell number.

To classify the *cycA* mutans, colony-PCRs were made using inward pointing *cycA*-specific primers (cycA1: 5'-ctgatgccggtaggttct, cycA2: 5'-gcgccatccagcatgata), and the lengths of the products were compared with those obtained from wt colonies using agarose-gel electrophoresis. Insertion-mutants were further sub-classified by PCRs using cycA1 and outward pointing primers specific for certain IS elements (IS1: 5'-tcgctgtcgttctca, 5'-aagccactggagcac; IS2: 5'-tcgcaggcataccatcaa, 5'-cagacgggttaacggca; IS5: 5'-gacagttcggcttcgtga, 5'-gctcgatgacttccacca; IS150: 5'-acgtgccgagatgatcct, 5'-cagacctatatgcctcgt) or by sequencing.

2.8.2. Frequency of mutations resulting in adaptation to salicin-minimal medium

The protocol was described by Hall in 1998. Briefly, 10^4 cells were inoculated into 20 tubes, each containing 1 ml of MT minimal medium. The cultures were shaken overnight at 37°C, then spread on 20 plates of salicin-MT. The plates were incubated at 37°C for 8-10 more days, and the new colonies (the adaptive mutants) were always marked, and counted. (The total colony numbers were plotted against time, yielding the adaptation curves of the compared strains.) The colony numbers observed on the plates were divided by the total plated cell number, and the arithmetic mean was calculated for each strain, every day. Total plated cell numbers were calculated by plating appropriate dilutions from 2-5 parallel cultures onto rich medium. For the convenience of dealing with small, positive numbers (between 0 and 3000), the observed mean colony number per $2.5 \cdot 10^9$ cells were plotted on our graphs. The values obtained for the different strains every day were compared using a two-tailed, unpaired, non-parametric Mann Whitney test. To classify the Bgl^+ mutans, colony-PCRs were made using inward pointing *bglR*-specific primers (BglR1: 5'-GTGGCGATGAGCTGGAT and BglR2: 5'-CCGACTTCACCAGTATTC), and the length of the products was compared with that obtained from a wt cell using agarose-gel electrophoresis. Insertion-mutants were further sub-classified by PCRs using outward pointing primers specific for the IS elements listed above.

2.9. Analysis of the IS-contamination of plasmid preparations using PCR

PCRs were performed in a total volume of 25 µl. Inward pointing primers specific for the different IS elements were used in end concentrations of 0.4 pmol/µl, each (IS1: 5'-tgagaacgacagcgac, 5'-ctccagtggcttctgtt; IS2: 5'-ttgatggtatgcctgcga, 5'-tgccgttaacccgtctg; IS5: 5'-tcagcagtaagcgccgt, 5'-agaccgtttcttcgccat). 1 µl of plasmid preparation was used as template. Cycle parameters were: 28x(94^{o}C 40sec/57^{o}C 40sec/72^{o}C 80sec).

3. Results

3.1. Optimization of the linear-fragment deletion method

3.1.1. Outlines of the linear-fragment deletion method

Four years ago, our group developed a genomic deletion method that leaves no extrachromosomal sequences behind, and can be repeated in the same cell in an unlimited number of times (Kolisnychenko *et al*. 2002). (For details, see **2.5**.) This method comprises two major steps: (1) recombination of a linear DNA fragment, carrying a selectable gene, into the bacterial chromosome via terminal sequences homologous to the regions flanking the deletion site, (2) introduction of a double strand break (DSB) in the inserted fragment to facilitate removal of the exogenous sequence by recombination between genomic segments doubled as a result of the insertion. The steps of the procedure are explained in detail below.

A 50-100 base pair sequence on both sides of the planned deletion must be known, for these will be the homology boxes involved in recombination. A PCR is made on the template plasmid pSG76CS, with primers that carry the homology boxes (**Figure 2.2.A**). Note, that on one end of the PCR fragment, two such boxes must be present. Since the synthesis of such long primers is inefficient, primer **ab** is made by filling up two overlapping oligonucleotides (**a** and **b**) in a PCR-like reaction. When the PCR fragment, made with primers **ab** and **c** is completed, the fragment contains the homology boxes **A, B** and **C**, a chloramphenicol resistance gene and two I-SceI restriction sites. The linear fragments are electroporated into cells containing the pBADαβγ plasmid (Muyrers *et al*., 1999). This plasmid carries the λ phage recombinase genes: *redα, β,* and *γ* under the control of the P_{ara} promoter. After arabinose induction, these recombinases protect the linear fragment from degradation, and enhance the recombination of homology boxes **A** and **C** with their respective pairs present in the genome. (Murphy, 1991; Muyrers *et al.*, 1999). The large segment in the genome is thus replaced by the short PCR-product via homologous recombination (**Figure 2.2.B, C**). Note, that due to the insertion process, box **B** becomes doubled. The recombinants are selected on chloramphenicol plates, and are checked routinely by PCR using primers **d** and **e**, which anneal to sequences of the genome neighbouring the deletion site.

In the second step, a DSB is made by the I-SceI meganuclease (**Figure 2.2.D**). This is an ultra rare-cutting endonuclease, with an 18 bp target sequence (Monteilhet *et al*, 1990). The wt *E. coli* MG1655 strain lacks this target site, so the genome is only cut within the insert. Bacterial cells carrying a DSB within their chromosome are not viable, unless the cell manages to correct this lack of genomic continuity. In bacterial cells, DSBs are repaired by

RecA-mediated homologous recombination, for which the most readily available substrates are the **B** boxes. The introduction of a DSB not only selects for colonies that have lost the insert, but also enhances its loss via RecA-mediated recombination by producing free DNA-ends (Pósfai *et al*., 1999). As a result of recombinational repair, all sequences between boxes **A** and **B** are lost from the genome, without any extra insertions remaining (**Figure 2.2.E**).

To avoid the possible accumulation of unwanted rearrangements among occasional short homologous sequences at further loci in the genome due to the action of the λ-Red recombinases, the two-step process described above was not repeated in the same cell. Instead, single deletion intermediates were constructed individually in MG1655, and the genomic segments carrying the inserts were transduced by P1 phages (Miller, 1992) into the MDS, using the Cm-resistance gene within the inserts for selection. Thus, (starting from

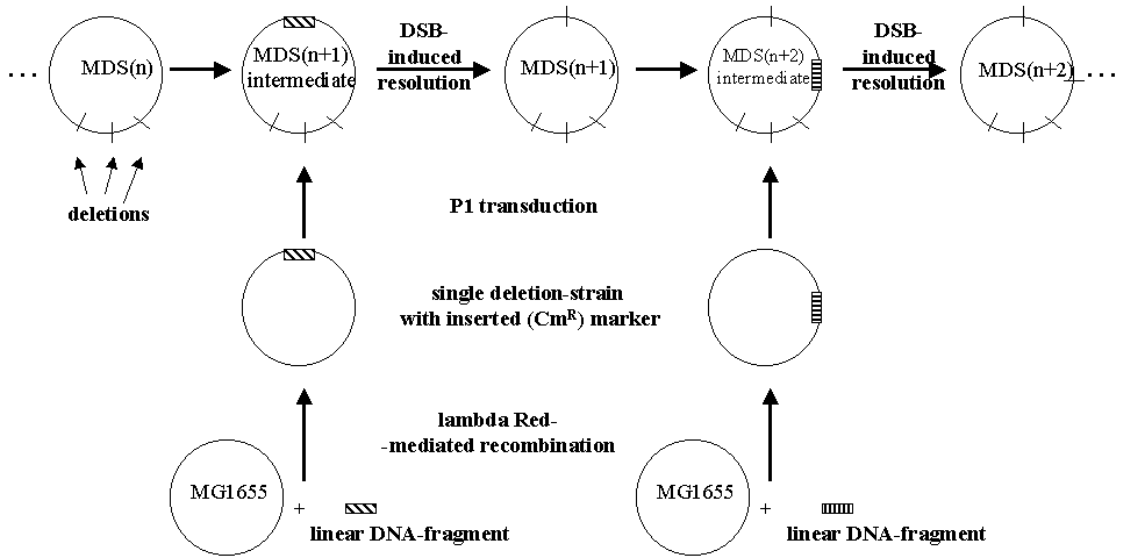


Figure 3.1. Protocol of the serial introduction of deletions into E. coli

MDS7), the second recombination step took place in the MDS (**Figure 3.1**). To verify the proceeding of this second step, colonies were replica plated onto LB/LB+Cm, and the Cm^S colonies were checked by PCR using primers **d** and **e**.

3.1.2. An alternative method for I-SceI cleavage

After P1 transduction, the MDS cells had to be made competent in order to be transformed with I-SceI-expressing pKSUC1. After the scarless deletion was completed, the

cells had to be cured from pKSUC1 prior to the next P1 transduction. To avoid repeating the time-consuming and tedious steps of transforming and curing the MDS cells after each deletion, my aim was to construct an inducible I-SceI meganuclease system. This was accomplished by constructing pSTAST, a derivative of pST98AS. Plasmid pST98AS carries an I-SceI meganuclease gene under the control of a P_{tet} promoter. The product of the tet repressor gene (*tetR,* also carried on the plasmid) represses the P_{tet} promoter, unless cTc is present. pST98AS also carries a restriction site for I-SceI, which had to be removed to avoid the plasmid from cutting itself upon induction (see Materials and Methods). The end product, pSTAST can be seen on **Figure 3.2**. Later on, the group used the Kan^R version of the plasmid (pSTKST), which enabled its use together with pBAD in the same cell.

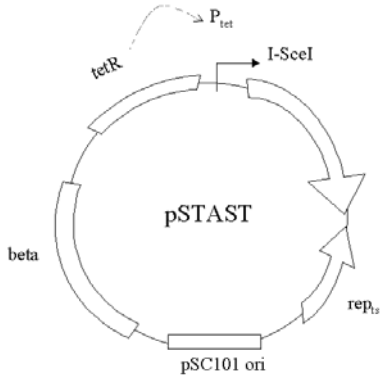


Figure 3.2. The pSTAST plasmid

3.1.3. Optimizing the induction of I-SceI by cTc from pSTAST

For full induction of I-SceI, we wanted to administer the maximum usable concentration of heat-inactivated cTc inducer. Heat-inactivated cTc is considered non-toxic, however, we found that higher concentrations of the compound had a negative effect on growth rate. **Figure 3.3** shows typical results of experiments measuring toxicity: 10 μg/ml of cTc has little, if any effect on growth rate, 25 and 50 μg/ml cTc slow it down, and 100 or 200 μg/ml cTc completely stops growth. Next, it was tested whether 20, 33 or 50 μg/ml cTc is more effective for induction, and for how long the inducer should be applied (ranging from 4 to 48 hours) for optimal effect (data not shown). It was found that the most effective and convenient protocol consisted of 24 hours of induction by 33 μg/ml cTc in liquid culture, followed by plating on agar plates supplemented with 20 μg/ml cTc.

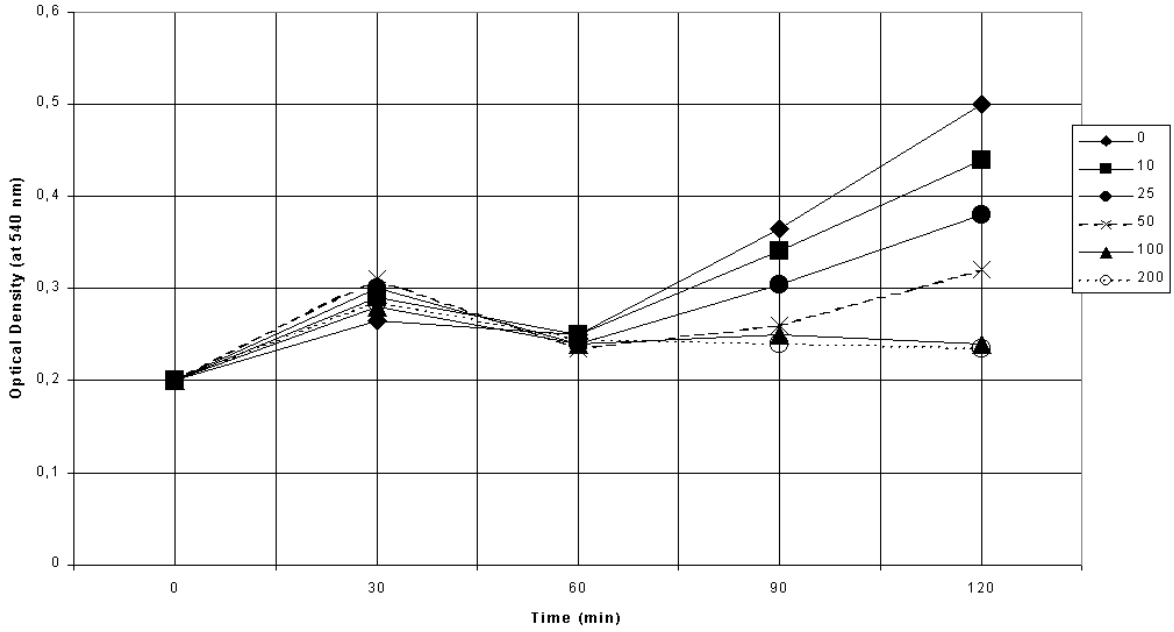


Figure 3.3. The effect of cTc on MG 1655 cell growth.
cTc was administered at 30 min. The numbers in legend show concentrations of cTc in micrograms/ml.

The efficiency of recombination had a great deviation, but was usually near 20-50%, or higher, when induced by the protocol described above. It could be raised even more, when M9+glycerol minimal medium was used instead of LB, but the induction was slower and more tedious this way, so we usually used LB, which gave satisfactory results (data not shown).

3.1.4. Inspecting the effect of the DSB on the mutation frequency

The protocol used for constructing the MDS calls for repeated introduction of DSBs into the genome. DSBs are known to activate the SOS response, resulting in the induction of error-prone DNA-polymerases. As a consequence, unwanted mutations might accumulate in the genome of the MDS cells. I carried out pilot experiments to measure the effect of the DSB on the mutation frequency by measuring the accumulation of rifampicin resistant cells, with and without the introduction of a DSB into the genome. Rifampicin is an ihibitor of the RNA polymerase β subunit. Cells resistant to it usually harbour a point mutation in their *rpoB* gene. cTc alone (without an inducible plasmid in the cell) was also tested for mutagenicity. The results of five independent experiments are shown in **Figure 3.4**. Although the values demonstrate an extensive variation (caused by factors discussed later), the major trends are visible: cTc alone did not seem to increase the frequency of mutation in the *rpoB* gene

compared to the control group, but when the plasmid-containing cells were treated with cTc

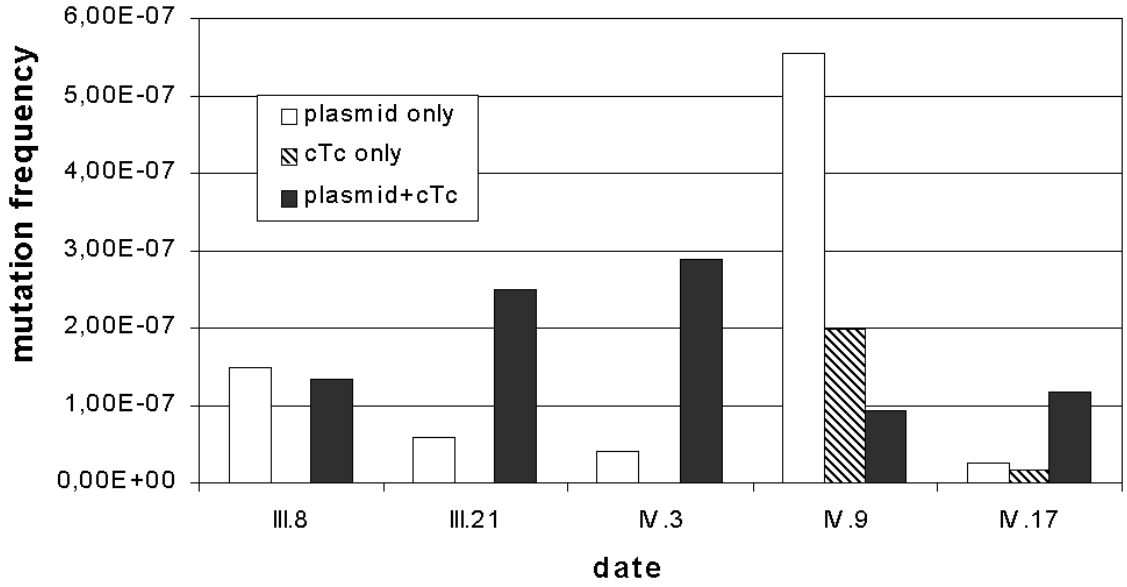


Figure 3.4. The mutation frequencies in five different experiments when cells carrying the marked plasmid were not treated ("plasmid only"), treated with cTc ("plasmid+cTc") or were treated with cTc, but they did not contain the plasmid carrying the inducable I-SceI ("cTc only")

(and thus caused a DSB in the genome), the mutational rate was increased. However, the increase was smaller than one order of magnitude, and the mutation rate was never higher than 10^{-6} mutants/cell/generation. The only exception was the experiment on IV.9, when we observed the opposite effect, probably as a result of a technical mistake.

3.2. Construction of a multi-deletion *Escherichia coli* cell

We hypothesised, that the genes present in the strain-specific genomic islands are dispensable under common culturing conditions. Our group's goal was to delete as many strain specific islands in the K-12 strain as possible, for two main reasons: (1) as a basic scientific research goal, we wanted to see how far we can go with the deletions without impairing the viability of the bacterium under laboratory conditions, (2) and we wished to develop a strain that is more useful for DNA-cloning experiments or other biotechnological purposes.

To achieve these goals, we started out by deleting the eleven largest, and six smaller K-islands present in *E. coli* MG1655. Proper and detailed planning by Guy Plunkett III always preceded each deletion. For annotated genes, the function of their product was usually sufficient to determine their deletability. In the case of unknown genes, however, a

comparative method was used by János Pósfai to make such conclusions. Namely, each predicted protein was searched using the reciprocal BLAST protocol against the twenty-four microbial proteomes available at the time of this work. Apart from phage integrase genes, non of the open reading frames (ORFs) had homologues in more than nine of the sequenced strains which minimized the possibility of attempting to delete highly conserved, essential genes. To avoid the deletion of important regulatory sequences, deletions were designed to leave as much of the intergenic regions behind as possible. In the cases of deletions MD1, MD2, MD4 and MD9, six to fifteen kbp regions adjacent to the K-islands, harbouring genes of unknown function were also eliminated. If two islands' proximity and the genes lying in between permitted, we included both in the same deletion. Deletions MD1 and MD5 comprise three K-islands each, while MD6 contains two. The primers **a**, **b** and **c** had to fulfil more or less the regular criteria for PCR oligonucleotide primers, concerning the number of nucleotide runs, hairpins and the illegitimate dimers formed with each other. This limited our possibilities of designing the joints for the deletions. Thus, in some cases, a few tens of nucleotides of the genes to be deleted were left in the genome. Care was taken not to create any fusion proteins this way. Overall, by cumulating the first twelve deletions (MD1 through MD12), thereby creating the MDS12 strain, seventeen K-islands were removed, reducing the genome by 376,180 nucleotides or 8.1% (Kolisnychenko *et al*., 2002). For exact endpoints and sizes of the deletions, see **Table 1.**

MD1 was deleted by György Pósfai using the method he published in 1997. It was based on recombining two suicide plasmids carrying FLP-recombinase target (FRT) sites into the genome, up- and downstream of the sequence to be removed. By inducing the FLP recombinase from a plasmid in the cell, all sequences were eliminated in between the FRT sites. This resulted in the first deletion, but also left an FRT site behind. All other deletions were made using the linear fragment method, unless otherwise stated.

Having deleted 8.1% of the genome, we tested certain phenotypic characteristics of the strain (see **3.3**.). Proceeding with the accumulation of deletions in the MDS, we aimed to remove all genetic elements that are potentially mobile. Eight cryptic prophages (CP4-6, Rac, CP-Eut, CP4-57, CP4-44, Qin, e14, and DLP12) and 24 insertion sequences (IS) were already removed in MDS12. According to the published sequence of *E. coli* MG1655 (Blattner *et al*., 1997), there were 20 more IS elements to be deleted. Thus, the next 19 segments to be removed (GP1-13, GP15, GP23-27) were chosen in a way to include all remaining ISs. Deletion of these segments, done by Vitaliy Kolisnychenko resulted in strain MDS31. The chronological order of the deletions made can be deduced from **Table 1.** All together, seven

Deletion	Endpoints	Size (bp)	Strain
MD1	263080, 324632	61553	
MD2	1398350, 1480279	81930	MDS2
MD3	2556711, 2563500	6790	MDS3
MD4	2754180, 2789270	35091	MDS4
MD5	2064327, 2078613	14287	MDS5
MD6	3451565, 3467490	15926	MDS6
MD7	2464565, 2474198	9634	MDS7
MD8	1625542, 1650785	25244	MDS8
MD9	4494243, 4547279	53037	MDS9
MD10	3108697, 3134392	25696	MDS10
MD11	1196360, 1222299	25940	MDS11
MD12	564278, 585331	21054	MDS12
GP1	15388, 20562	5175	MDS13
GP2	602688, 608572	5885	MDS14
GP3	2507651, 2515969	8309	MDS15
GP4	379334, 387870	8537	MDS16
GP5	389122, 399029	9908	MDS17
GP6	2993014, 2996890	3877	MDS18
GP7	3182797, 3189712	6916	MDS19
GP8	687083, 688267	1185	MDS21
GP9	1386912, 1396645	9734	MDS20
GP10	2099418, 2135738	36321	MDS24
GP11	2284421, 2288200	3780	MDS22
GP12	3359797, 3365277	5481	MDS23
GP13	3648921, 3651342	2422	MDS25
GP14	1128620, 1140210	11590	MDS37
GP15	1960590, 1977353	16764	MDS30
GP16	1995135, 2021700	26566	MDS39
GP17	4553059, 4594581	41523	MDS38
GP18	522062, 529349	7288	MDS33
GP19	728588, 738185	9598	MDS32
GP20	1525916, 1531650	5735	MDS34
GP21	3616622, 3623309	6688	MDS35
GP22	3759620, 3767869	8250	MDS36
GP23	1041254, 1049768	8515	MDS27
GP24	1085330, 1096545	11216	MDS31
GP25	2163173, 2175230	12058	MDS28
GP26	3578769, 3582673	3905	MDS26
GP27	3718263, 3719704	1442	MDS29

Table 1. Nomenclature of the first 39 deletions carried out in MG1655, and accumulated by P1 transduction. Endpoints denote the first and last deleted nucleotide. The order of the deletions accumulated in the multi-deletion series can be deduced from the MDS-number.

IS1s, seven IS2s, five IS3s, two IS4s, twelve IS5s, four IS30s, one IS 150, three IS186s, one IS600 and two IS911s were eliminated. A remnant of the P2 prophage was also removed in deletion GP25.

To continue our efforts of reducing and stabilizing the genome of *E. coli* even more, I deleted the Rearrangement hotspots (Rhs). Deletions GP18 through GP22 eliminated RhsD, C, E, B and A, respectively. Next, I deleted segment GP17. This removed the *EcoK*

restriction modification system (genes *hsdRMS*), the methylated cytosine restriction system (*mcrBCD*) and the methylated adenine recognition and restriction system (*mrr*). Finally, I deleted the genes responsible for flagellum synthesis (*flgA* through *flgN* in GP14 and *fliA* plus *fliC* through *fliT* in GP16). The *flh, che, mot, tap* and *tar* operons, which also play a role in flagellar synthesis, motility and chemotaxis, were removed in GP15.

Deletion	Change relative to theoretical joint sequence
MD2	nucleotides 1398350-1398351, 1480279 are missing at the theoretical joint
MD10	nucleotides 3134395-3134396 are missing downstream of joint
MD11	4-bp deletion 14-17 bp downstream of the joint (1222317-1222320) remade in connection with IS5 removal from oppA region
GP4	1-bp deletion downstream of the joint (387878)
GP6	perfect seq (possible 1-bp downstream deletion at 2996930)
GP14	one T (1140210) is missing at the joint
GP23	nucleotides 1049770, 1049776-1049784 missing
GP26	missing nucleotides downstream of joint (3582679,3582681,3582683-3582684)
GP27	possible 1-bp deletion 37 bp upstream of joint (3718226)

Table 2. Empirical joint sequences

At this point again, we decided to characterize the phenotype and genotype of our strain, MDS39. We sequenced the joints of the deletions, and in most of the cases, obtained the sequence identical to the predicted one. The few exceptions showing point mutations are listed in **Table 2.** Our American collaborators compared the genome of MDS39 and MG1655 on an Affymetrix *E. coli* DNA chip. As expected, the intensities of the signals obtained on the DNA-chip corresponding to the deleted segments were much lower in the MDS than in the wt. However, they obtained high signal peaks at multiple points well inside some areas that were removed from the genome. (An example is marked by an arrow in **Figure 3.5.**) The oligonucleotide probes that gave these signals corresponded to certain segments

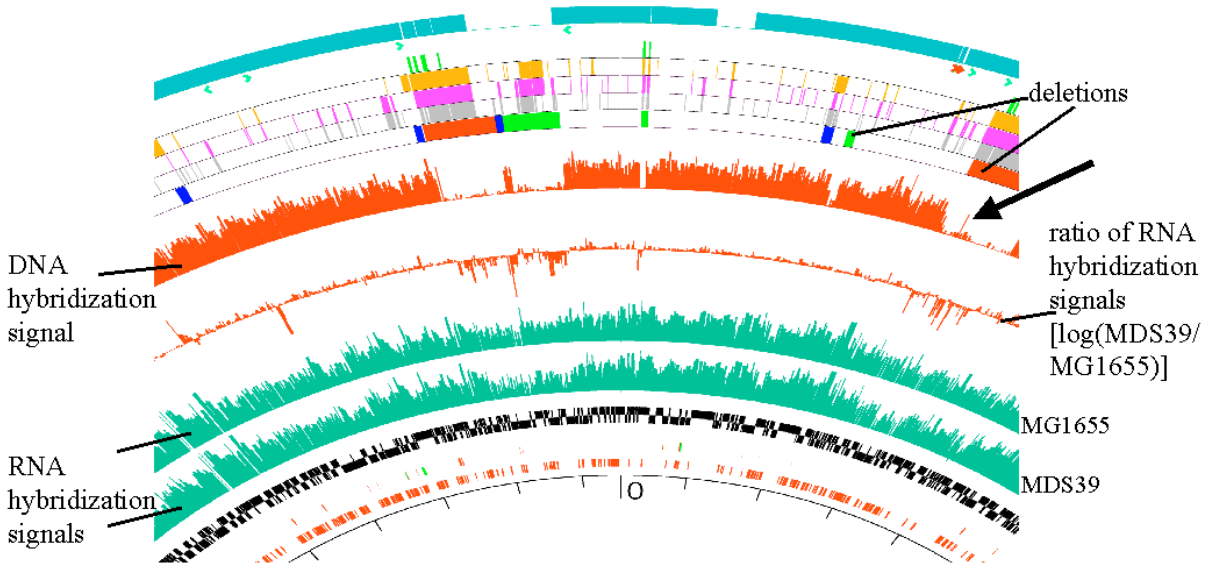


Figure 3.5. Segment of MDS39 genome marking deletions (red and green boxes), DNA and RNA hybridization signals

of IS1, IS2, IS5 and IS911. This indicated that these ISs are still present in the genome of MDS39, at unknown locations. Locating and deleting them was necessary to declare our strain free of mobile genetic elements.

My job was to define the exact endpoints of the extra IS elements, and delete them. For the localization of a known sequence in a genome, I tested two methods previously described. The first one (Wang *et al.,* 1991) uses one outward pointing primer, and assumes, that, under adequate conditions, the same primer will anneal to the opposite strand of the genome in an unspecific manner within a convenient distance, to yield a PCR fragment. Unfortunately, I obtained a large number of PCR products with this method, and found it hard -if possible- to select for the one that is produced by the mechanism described above (data not shown). The second method, inverse PCR, uses a primer pair pointing outward from the two ends of the IS element, and a set of restriction endonucleases that do not cut into the IS. PCRs were made after the digestion and ligation of the genome, as described in Materials and Methods. This procedure also gave several unspecific PCR fragments, but these were the same in the cases of different endonucleases, and were thus easy to recognize. If the ligated genomic circle carrying the IS element was short enough to yield a specific PCR fragment using the primer pair (**Figure 2.3**), the unspecific fragments disappeared, leaving only one band in the gel. Sequencing this fragment gave us the exact endpoints of the IS element in the genome.

We found three copies of IS1 in MDS39, that were not present in the sequenced genome of MG1655. They were integrated into *crl, ais* and *yeaJ*, flanked by the sequence-duplications listed in **Table 3**. One copy of IS2 and IS5 each were found, inserted into *yddA* and *oppA*, respectively, flanked by shorter duplications. The presence of the IS911 transposase (b4285) could not be confirmed by PCR, even though three different primer pairs were used (data not shown). DNA microarray experiments performed later on in the course of this project found the MDS free of IS911. Using primers specific for the interrupted genes, we checked whether the original cell, MG1655 carried these insertion sequences. Unexpectedly, three of them (those in *crl, yeaJ* and *oppA*) were already present,

Deletion	Endpoints	Duplication at IS insertion	Strain
YeaJ+IS1	1870052, 1871490	cggatgcac	MDS40A
IS2 (in YddA)	1576110, 1576111	taaac	MDS40B
tonA	167483, 173448		MDS41A
IS1 (in ais)	2363057, 2363058	cgaattccc	MDS41B
IS5 (in opp)	1298721, 1298722	ctaa	MDS41C
IS1 (in crl)	257907, 257908	tgaaggta	MDS41D
MD1 (widening)	262739, 324635		MDS41E=MDS41
endA	3088367, 3089074		MDS42

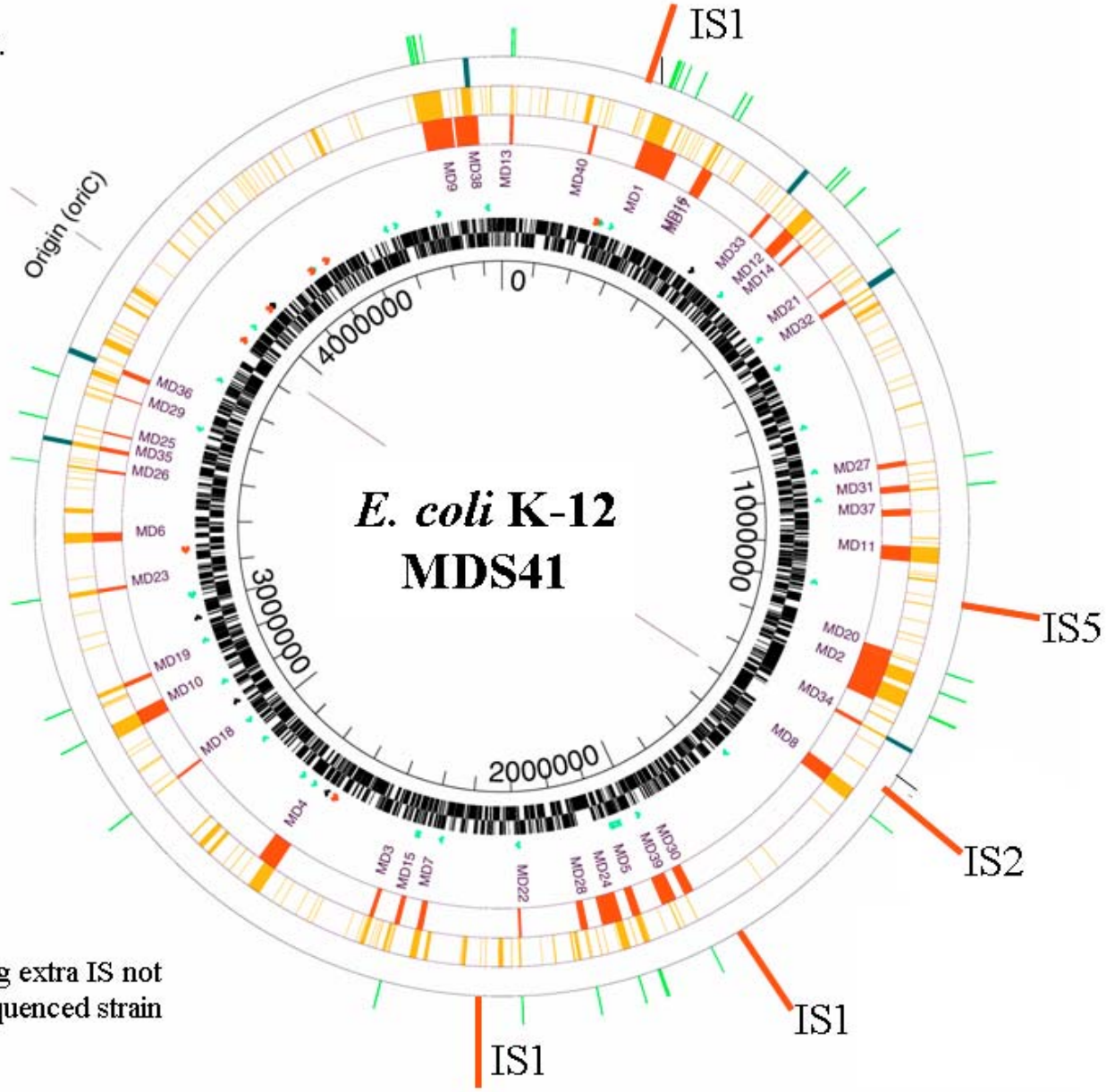
Table 3. Nomenclature of the additional eight deletions performed, together with the sequence duplications observed at the sites of extra IS integration. Endpoints mark the first and last nucleotide left unaffected by the deletion.

which means that the MG1655 used in our laboratory was different from the one sequenced, although both of them originated from the Blattner lab. By performing PCRs on our MDS-collection`s intermediate strains, we found that the other two IS insertion events occurred during the MDS construction. *YddA* became interrupted during the establishment of the 9th or 10th deletions and *ais* obtained an IS during construction of the 16th deletion.

ISs in *oppA* and *crl* genes were removed using the suicide-plasmid method, resulting in the recontruction of the intact genes. The *yeaJ* gene could neither be corrected by the suicide-plasmid, nor the linear fragment procedure, but we managed to delete the whole gene, including the IS1 disrupting it. These deletions, like all prior ones, were made in MG1655.

The relevant genomic segments of their intermediates were transduced into the MDS. The IS

Figure 3.6. The simplified *E. coli* genome.



Legend:

- deletion
- K-island
- IS element
- rhs sequence
- genes
- deletion removing extra IS not present in the sequenced strain

elements disrupting *yddA* and *ais* were not present in MG1655, so we eliminated them simply by co-transducing the correct genomic fragment with the closely lying cointegrates of GP20 and GP11, respectively, used as selection markers. The corrected genes were checked by colony-PCR, followed by sequencing. Three more genomic segments were also removed by György Pósfai to make our strain more useful as a laboratory host, using the suicide-plasmid method: *tonA* (encoding a T1 and Φ80 phage receptor), *endA* (coding for a periplasmic DNAse) and the FRT site left over from deletion MD1. This way, a total of 42 segments of the sequenced *E. coli* MG1655 strain were removed. **Figure 3.6** depicts these deletions (except *endA*), relative to the K-islands. Also visible are the IS elements present and not present in MG1655 at the time of its sequencing (green and red marks on the outermost ring, respectively).

3.3. Characterization of the multi-deletion strain

Compared to MG1655, the genome of MDS42 was reduced by 662,606 bps (14,28%), and it was free of mobile genetic elements. This was confirmed by the DNA-microarray studies conducted in the Blattner-laboratory. Our aim now was to compare the MDS's phenotype with that of the wt. We were also hoping to be able to prove, that the new strain can not only conveniently be used for cloning experiments, but that it is even advantageous to use due to the absence of the mobile genetic elements.

3.3.1. Transformability

In parallel to our deletion work, I made chemical transformation and electroporation tests with the arising MDSs. We usually compared the transformability of the MDS with that of the wt strain, MG1655. **Figure 3.7.** shows the chemical transformation efficiencies of the two strains, using a small plasmid, pBR322. The transformability of MDS41 is two orders of a magnitude lower, compared to MG1655, with a p value smaller than 0.01.

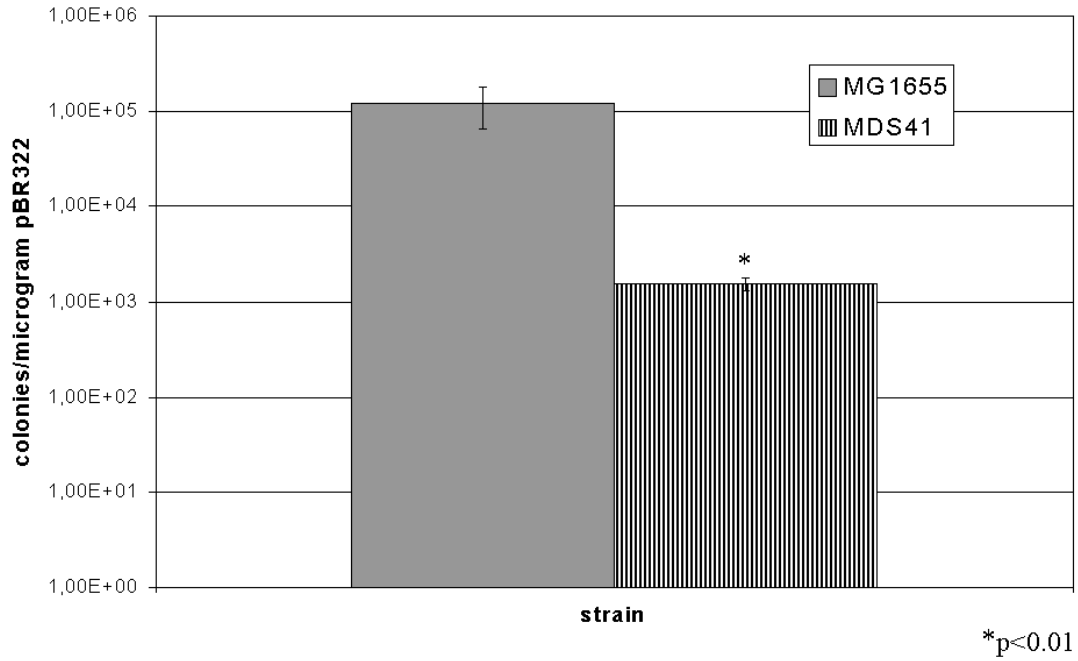


Figure 3.7. Chemical tranformation efficiencies of MG1655 and MDS41. Mean values of five measurements. Note logarithmic scale of y axis.

Figure 3.8 compares the electroporation efficiencies of three strains using pBR322. MDS39

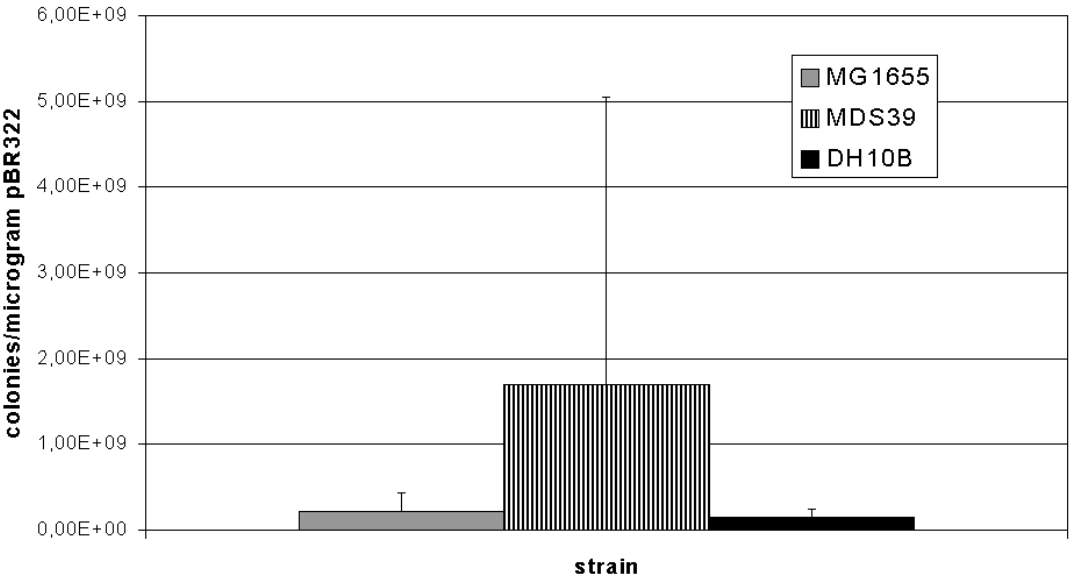


Figure 3.8. Electroporation efficiencies of strains MG1655, MDS39 and DH10B, using plasmid pBR322. Shown values are means of four to seven measurements.

seemed to be eight to ten times better transformable this way, than either MG1655 or DH10B, but due to the great variation of the values obtained, this difference wasn't significant.

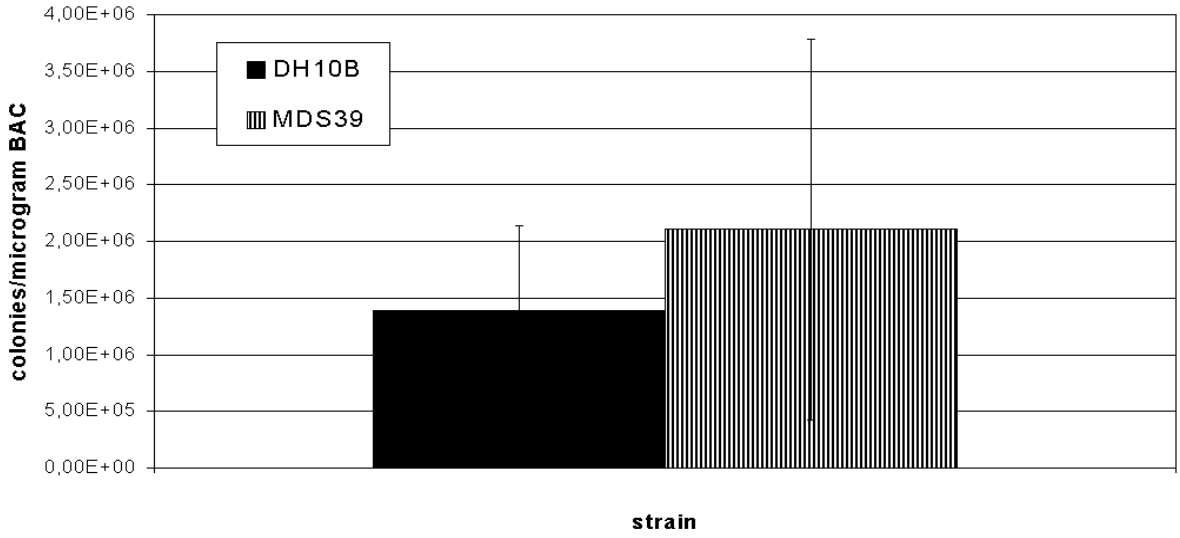


Figure 3.9. Electroporation efficiencies of strains MDS39 and DH10B, using a 100kbp large BAC. Shown values are means of six to nine measurements.

Finally, **figure 3.9** shows the electroporation efficiencies of MDS39 and DH10B, using a 100 kbp long bacterial artificial chromosome.

3.3.2. Growth rate

Growth rates of the sequentially constructed MDSs were regularly measured by our group. **Figure 3.10** shows an example for the growth curves of MG1655 and MDS41 in M9+glucose minimal medium, at 37°C. The log phase of the multi-deletion and the wt are similar. However, it was observed, that in the case of using overnight starter cultures, the MDSs need a significantly longer lag phase to start division (data not shown). Doubling times, calculated using the data obtained during exponential growth in LB and M9 at 37°C are shown on **figure 3.11.** In parallel with the increasing number of deletions, a trend of a slight but not significant elongation of doubling times can be observed in LB. The doubling times measured in M9 do not show any significant change with the increasing number of deletions.

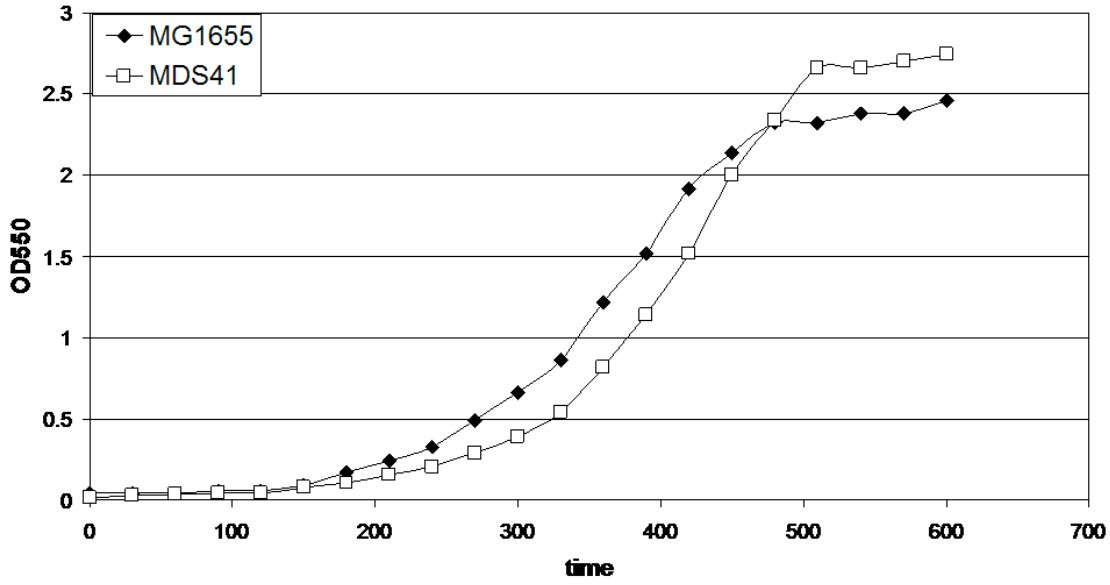


Figure 3.10. Growth curves of *E. coli* strains measured in M9+glucose minimal medium (data of Kinga Umenhoffer).

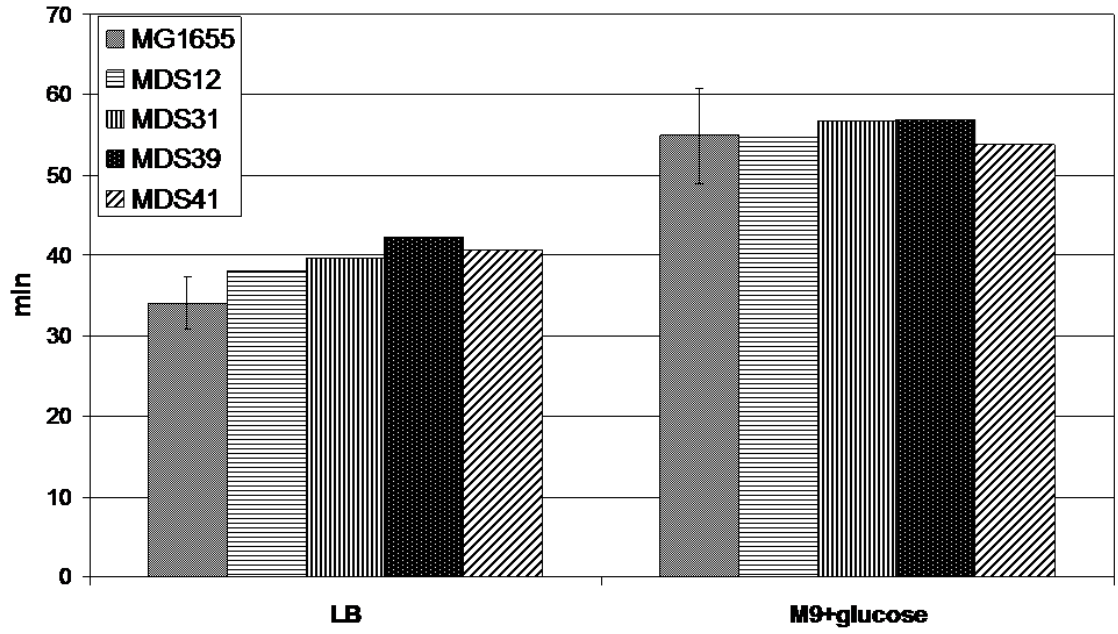


Figure 3.11. Doubling times of *E. coli* strains harboring an increasing number of deletions, measured in LB or M9+glucose liquid media

Our collaborator, David Frisch also recorded the growth curves of several single-deletion and multi-deletion strains, but he used a 96-well microplate shaker connected to an automatic absorption meter. This instrument allowed him to study many strains under strictly the same

conditions. His results are shown on **Figure 3.12.** The log phases of the MDSs are similar to that of the wt, MG1655. The calculated doubling times are 27.6 and 30.1 min for MG1655 and MDS42, respectively. The lag phases of the MDSs were usually longer than that of MG1655. An additional piece of data that he obtained was that all MDSs having 31 or more deletions tend to exhibit a decrease in their OD after reaching the plateau phase, indicating a decrease in cell-count.

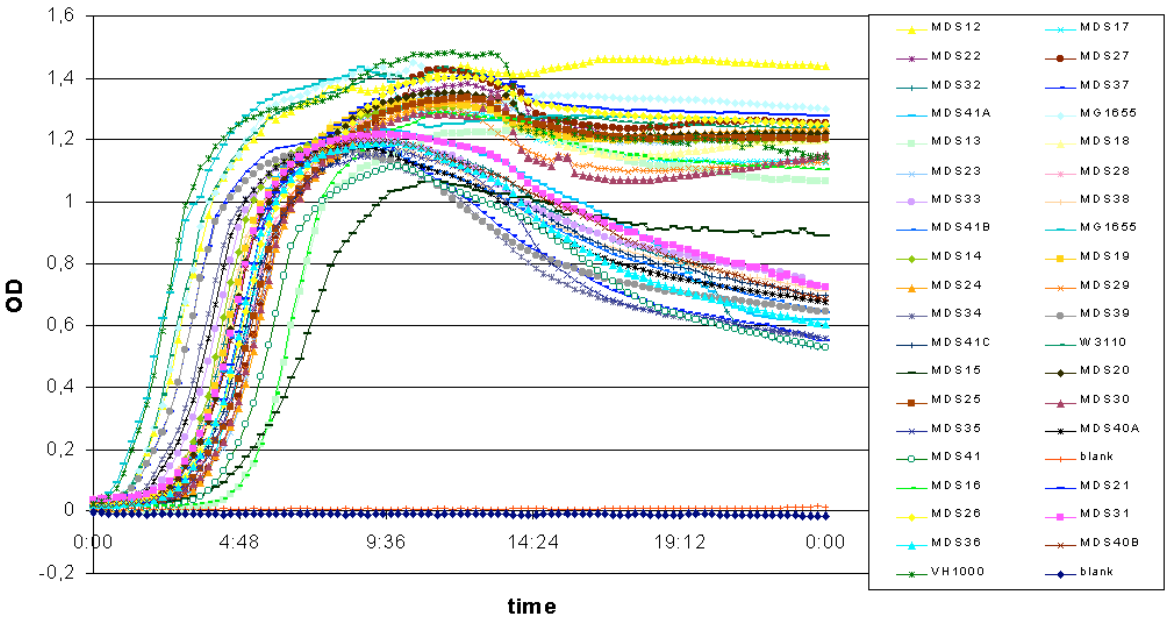


Figure 3.12. Growth curves of *E. coli* strains harboring an increasing number of deletions, measured in LB (data of David Frisch)

3.3.3. Mutation rate measurements

We hypothesized that removal of the mobile genetic elements stabilizes the genome. To measure the frequency of IS hopping, and to quantitate the share of these events in the entire

mutation spectrum, we had to develop adequate mutation detection systems. In the first two attempts, we measured the inactivation frequency of a gene (a methyltransferase or toxic endonuclease) present on a plasmid. The procedures involved growth of the plasmid in the strain to be tested, preparation of plasmids, and electroporation of the plasmids in an indicator strain, usually DH5α (data not shown). These techniques were inadequate, for reasons discussed later. The following two methods involve analysis of genes carried on the chromosome.

3.3.3.1. Measuring the adaptation to salicin-minimal medium

Salicin is a beta-glycoside sugar. The functions needed for its uptake and hydrolysis are encoded by the cryptic *bgl* operon in *E. coli* (Schnetz *et al*., 1987). This operon is kept in an inactive state by silencer elements (Schnetz, 1995), but can be activated by the disruption of these silencers. This is most commonly caused by IS elements (usually IS1 or IS5) inserting into *bglR* (Schnetz and Rak, 1988), but can also be a result of the deletion of the sequence upstream of the CAP binding site (Schnetz and Rak, 1992, Schnetz, 1995), or two base substitutions within the CAP binding site (Schnetz and Rak, 1992). Rarely, mutations in *gyrA, gyrB* (DiNardo *et al*., 1982), *hns* (Defez and De Felice, 1981) or in *bglJ* can also activate the *bgl* operon in trans (Giel *et al*., 1996). Other cryptic elements that can be activated to turn the phenotype of *E. coli* to Bgl^+ are the *cel* (Kricker and Hall, 1984) and *asc* operons (Hall and Xu, 1992) and the *arbT* transporter gene (Kricker and Hall, 1987). Despite all the possibilities listed, 98% of the Bgl^+ mutants obtained during growth arise from the insertion of an IS element into the *BglR* region (Schnetz, 1987). In the absence of growth, though, only 79% of the Bgl^+ mutants are the results of insertions. The rest is mostly caused by mutations in *hns* (Hall, 1998).

The ratio of insertions among the adaptive mutations in this system is thus high enough to effectively demonstrate the presence or absence of insertion sequence-transposition in the

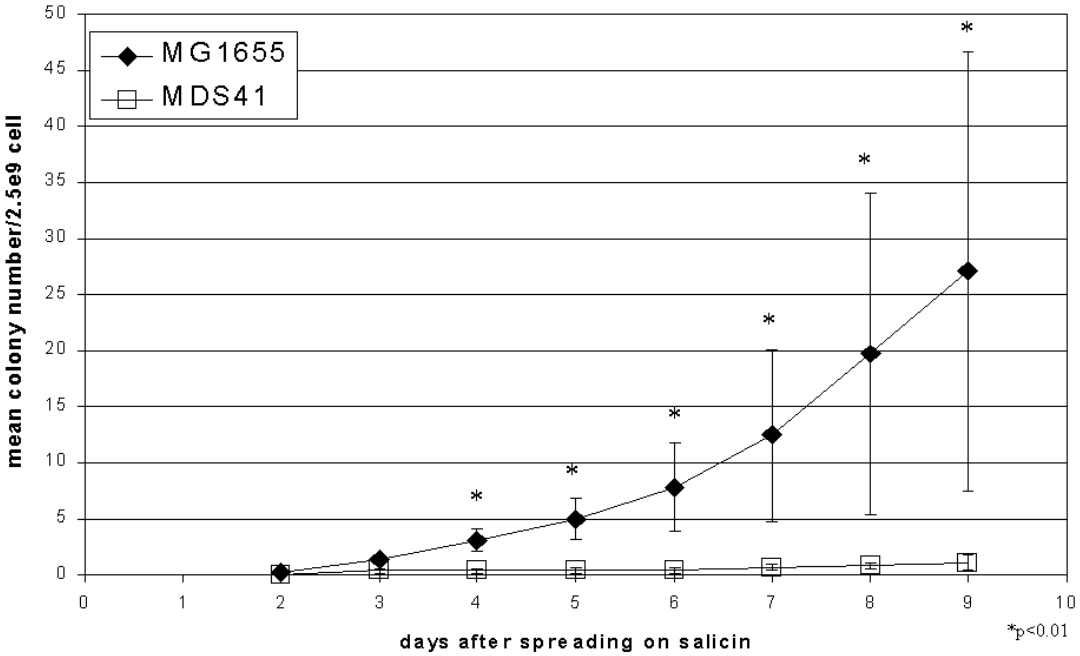


Figure 3.13. Adaptation of strains MG1655 and MDS41 to salicin-minimal medium. Data is an average of three experiments, all involving 2x20 plates

strains tested. These experiments involve growing parallel cultures in glucose-minimal medium, followed by plating on salicin-minimal medium. The colonies that appear two days after are the mutants that arose during growth in the culture. The mutants arising later (the adaptive mutants) are then counted every day for a period of ~10 days. By plotting the average number of colonies of the parallel plates against time, we can obtain the adaptation curves of the strains. The adaptation curves of MG1655 and MDS41 are compared on **Figure 3.13**. As expected, the adaptation of MDS41 to salicin-minimal medium is much slower than that of MG1655 (10-20%). The composition of mutants was screened in both strains by PCR-amplifying the *bglR* region, and running the products on agarose gels. Fragments longer than those obtained from wt colonies were termed insertion mutants, while fragments with equal lengths were termed as "other" type of mutants. Deletion mutants of the *bglR* region were not observed using this method. Using primers specific for different IS elements, the type of IS disrupting the *bglR* region could also be identified. The mutant compositions of the two strains are shown on **Figure 3.14**. About 92% of the mutations in the *bglR* region of MG1655

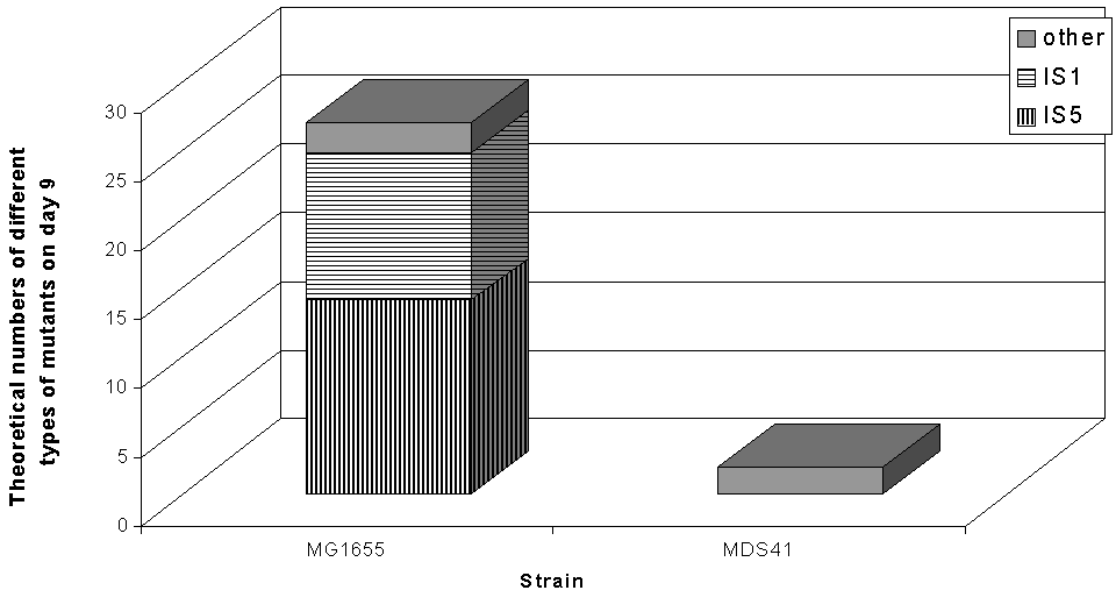


Figure 3.14. Composition of Bgl$^+$ mutants on day 9. Data based on 98 and 15 PCRs made on MG1655 and MDS41 colonies, respectively.

were caused by IS element hopping, while no insertion mutants could be detected in *bglR* of IS-free MDS41.

3.3.3.4. Measuring the frequency of mutations causing cycloserine-resistance

The next system to be tested selected for cells in which the *cycA* (also known as *dagA*) gene was inactivated. This gene codes for a transporter, which is responsible for the uptake of the antibiotic D-cycloserine. Mutations in *cycA* confer the cells resistance to D-cycloserine, an inhibitor of cell wall synthesis (Wargel *et al*., 1970, Wargel *et al*., 1971). The CycA protein also imports D-alanine, D-serine and glycine into the bacterial cells, so the strains were grown on minimal medium during the experiment to avoid negative selection against the *cycA* mutants. The method (Materials and Methods) involves growing aliquots of parallel cultures and plating on glucose-MT-cycloserine/glucose-MT plates. The mutant number and total cell number, respectively, were calculated from the colony numbers counted on these plates, and the ratio of the obtained values yielded the mutation frequency of *cycA*. Despite the same total cell number, there was usually a great variation in the mutation frequencies calculated for twenty parallel-grown cultures. The reason for this great fluctuation in mutant numbers,

tube-by-tube, is the "jackpot-effect": the occurrence of a mutation in the early phase of culturing, as opposed to late-arising mutations, results in a much larger number of mutants in the end. Estimating the actual mutation frequency by averaging the obtained mutation frequencies is incorrect, because the distribution of the number of mutants in parallel cultures (also known as the Luria-Delbrück distribution) is asymmetric. The only value theoretically constant in parallel bacterial cultures is the number of mutations per tube (**m**). This value could not be directly measured, but could be estimated by the Ma-Sandri-Sarkar (Sarkar *et al*., 1992) Maximum Likelihood method (MSS-MLL). This method, using the Lea-Culson generating function, generated many mutant-number distributions that differed from each other only in the **m** parameter. Using our spreadsheet algorithm, made for this purpose, these distributions were mathematically compared to the observed mutant number distribution. The **m** value of the theoretical distribution that fitted the obtained data the best, was a good estimation of the actual number of mutations per tube. This value, divided by the total cell number gave us the mutation rate: the chance of every bacterial cell, during its lifetime, to develop a mutation in *cycA*, which is biologically relevant information. Another advantage of **m**, is that its logarithm has a nearly normal distribution. This means that the **m** values of two cultures having the same total cell number could be compared with a standard t-test (Rosche and Foster, 2000).

Figure 3.15 shows the result of a typical experiment. The **m** values for MG1655 and MDS41 were 69.3 and 54.6 mutations per tube, respectively, and the difference was significant ($p<0.001$). This experiment was repeated five more times. To compare the results

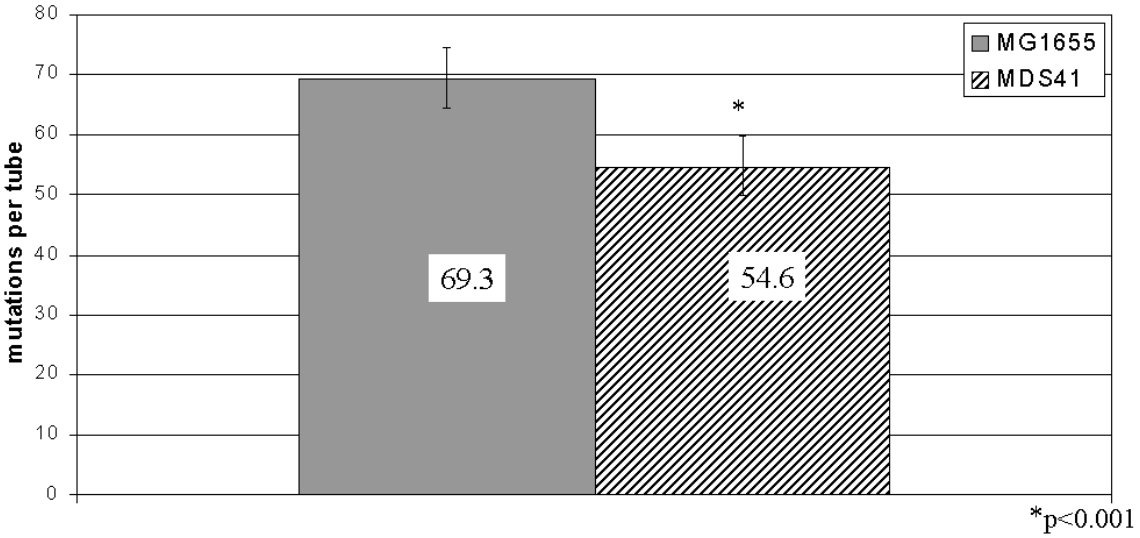


Figure 3.15. Mutations per tube resulting in cycloserine resistance in two *E. coli* strains. Data based on 2x20 parallel cultures.

of different experiments, mutation rates were calculated. The mutation rate of MDS41

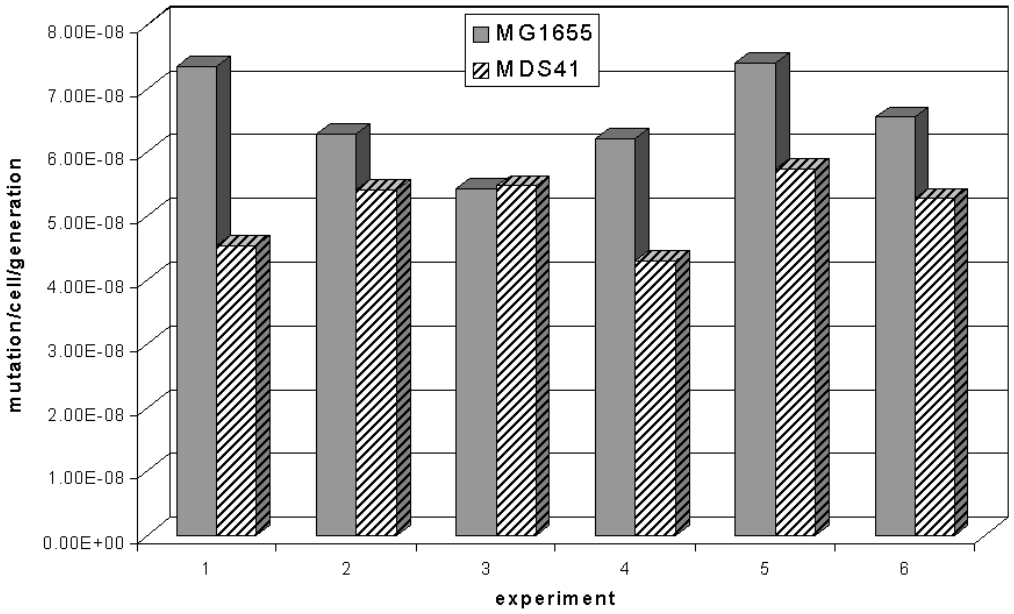


Figure 3.16. Six experiments comparing the mutation rates of the *cycA* gene within MG1655 and MDS41. All values were calculated using 20-20 parallel cultures.

ranged between 60 % and 100 % of the value for MG1655. The ratio of the means calculated for the two strains was 78.3% **(Figure 3.16)**.

The composition of mutants was screened by PCR- amplification of *cycA* of cycloserine-resistant colonies, and agarose-gel electrophoresis of the products. If the length of the PCR fragment obtained from the mutants was the same as that of the wt, we defined it as a point mutant. If it was shorter, or if no product was obtained, we classified it as a deletion mutant. Longer PCR product indicated an insertion mutant. The latter group was classified further into sub-classes, depending on which IS was inserted. The IS elements interrupting *cycA* were identified by PCR using primers specific for the IS elements, and a primer specific for *cycA*. If no PCR product could be obtained with any of our primers, we sequenced the amplified gene. **Figure 3.17** shows the mutant composition detected in MG1655 and MDS41 strains, based on 236 PCRs. Insertions, which account for

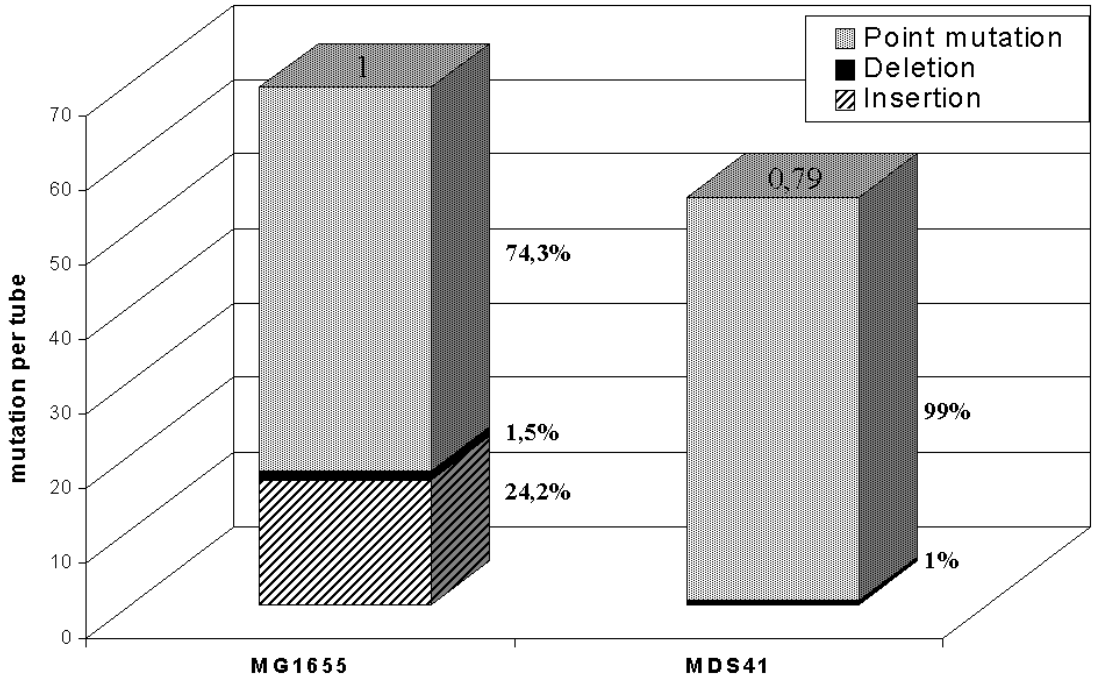


Figure 3.17. Composition and relative quantity of *cycA* mutants in strains MG1655 and MDS41, based on 128 and 108 tested colonies, respectively.

approximately 24.2 % of the mutations in MG1655, are totally missing in MDS41. The majority of the insertions are caused by IS150 and IS1 (**Figure 3.18**, data in part supplied by Botond Cseh).

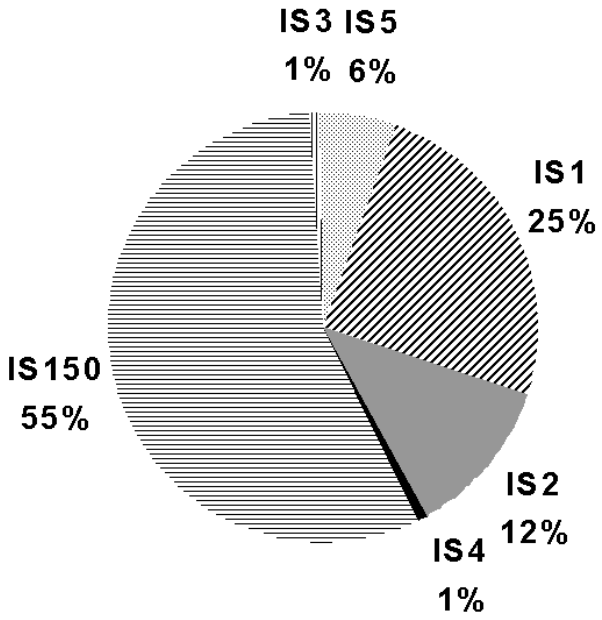


Figure 3.18. Composition of *cycA* insertion-mutants in strain MG1655, based on 118 independent colonies.

3.3.4. Analysis of plasmid preparations

Small scale preparations of plasmid pUC19 were made from *E. coli* strains MG1655, DH10B and MDS42. To test the possible presence of IS elements within these preparations,

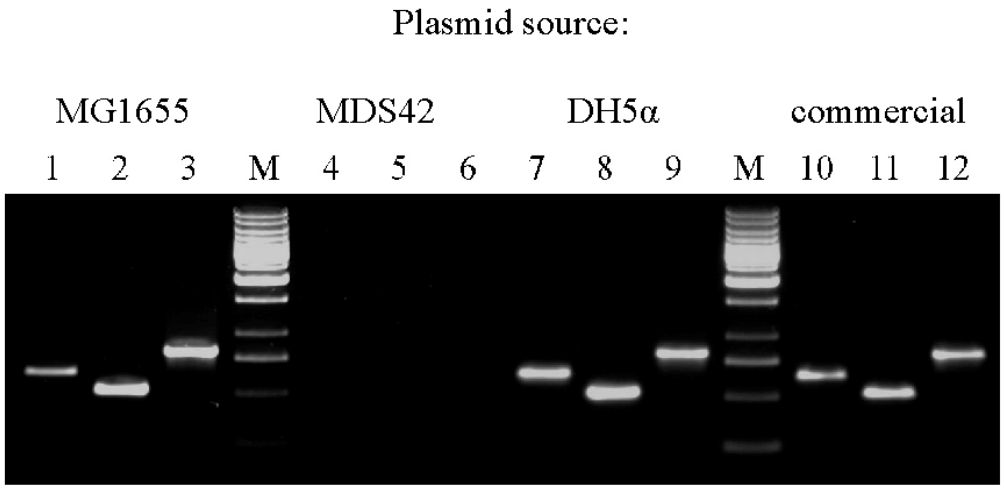


Figure 3.19. Agarose gel electrophoresis of PCR products made on four different pUC19 plasmid preparations. Specificity of primers used: lanes 1,4,7,10: IS1; lanes 2,5,8,11: IS2; lanes 3,6,9,12: IS5. M: molecular weight marker.

PCR reactions were made using them as template with inward pointing primers specific for three different insertion sequences (IS1, IS2 and IS5). A commercially available pUC19 stock solution was also tested this way. All plasmid preparations demonstrated IS contamination, except for the one originating from MDS42 (**Figure 3.19**).

3.4. Characterization of the cycA-method for mutation rate measurements

Prior to the evaluation of the results obtained using the cycA method, testing this procedure to see whether it fulfils the main requirements of an ideal mutation detection system was necessary.

3.4.1. Mapping the cycloserine-resistance phenotype

In order to confirm the genotype-phenotype relation of the cycloserine-resistant mutants, they were complemented by the wt *cycA* gene using plasmid pBADcycA. This is an arabinose-inducible expression vector, constructed by Botond Cseh. Twelve randomly

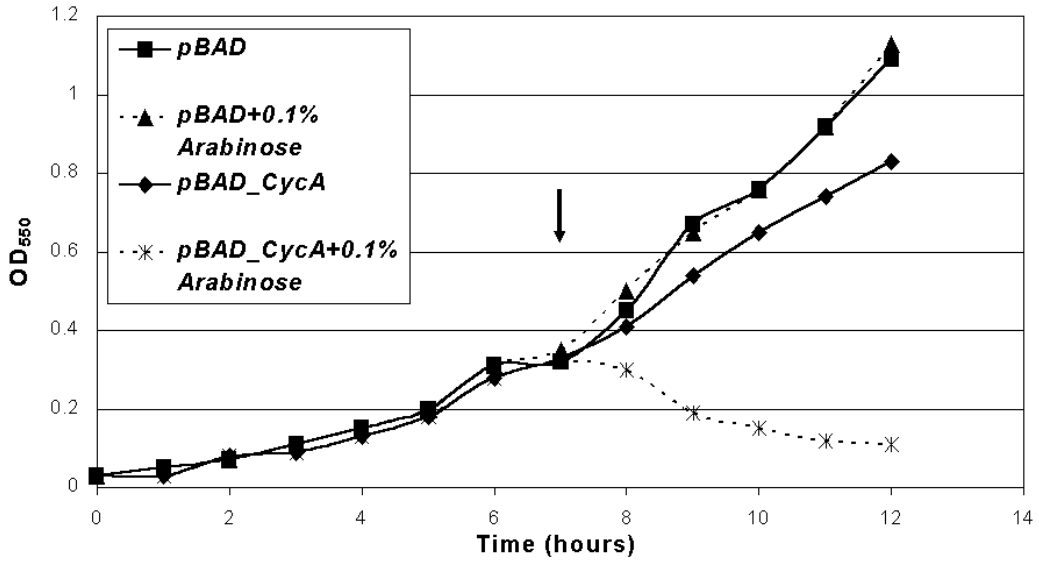


Figure 3.20. Effect of wt *cycA* expression on the growth of D-cycloserine resistant mutant MG1655 cells in liquid culture. Plasmids carried by the cells are indicated. Cultures were supplemented with 0.04 mM D-cycloserine, and 0.1% arabinose inducer was added at OD=0.3 (indicated by a vertical arrow), where relevant. All 12 randomly chosen mutants yielded the same results. Only one typical experiment is shown.

chosen D-cycloserine resistant mutants of MG1655 were transformed either by pBADcycA or the insert-less vector pBAD. Transformants were then spread on glycerol/minimal plates with or without D-cycloserine and/or arabinose inducer. Growth was observed in all cases except when the pBADcycA-harboring mutants were spread on plates containing arabinose and

cycloserine (data of B. Cseh, not shown). These results indicated that expression of the wt gene sensitized the resistant mutants to the antibiotic. In another experiment, another set of twelve randomly chosen D-cycloserine resistant mutants harbouring pBADcycA were separately grown in liquid cultures, and arabinose was added in early logarithmic phase. Expression of wt *cycA* resulted in growth inhibition in the presence of D-cycloserine (**Figure 3.20**). When arabinose was added to mutant cells carrying pBAD, no effect on growth was observed.

We also sequenced twenty colonies from MG1655 and MDS41, defined as point mutants. We were able to detect four transitions, five transversions, four (-1) and two (+1) frameshifts, one small (8-bp) deletion, one 3-bp and one 2-bp duplication among the mutants. Two independent cells carried the same transversion, while one carried multiple mutations within *cycA*. All base substitutions resulted in an amino-acid change (**Figure 3.21**).

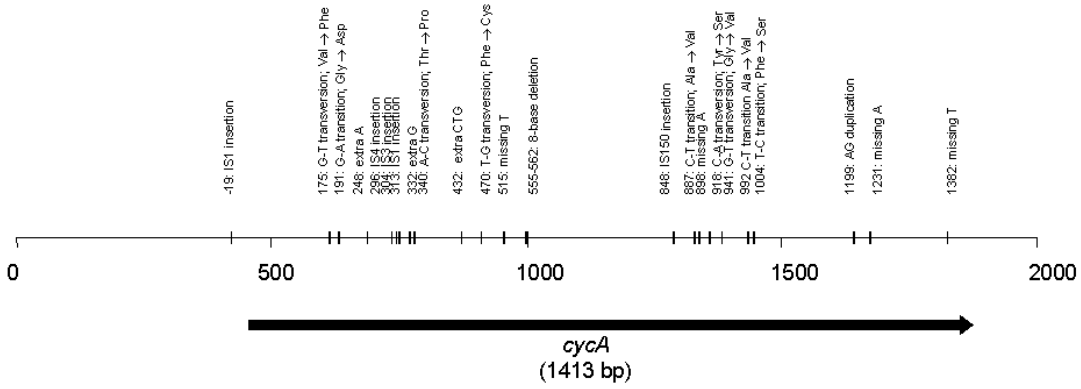


Figure 3.21. The location and identity of single mutations in *cycA* of D-cycloserine resistant cells. A sample of twenty randomly chosen *cycA* mutants harboring small mutations and five mutants harboring IS insertions were analysed. Mutation 780:A→C was observed in two independent mutants. Position 1 corresponds to the first nucleotide of the start codon of *cycA*.

3.4.2. The effect of cycloserine-resistance on growth under non-selective conditions

We hypothesized that in the absence of the transport of substrate amino acids (e.g., in minimal medium), mutations of *cycA* are neutral in respect to growth rate. To test this hypothesis, starter cultures of wt MG1655 or MDS42 and their randomly chosen D-cycloserine resistant mutants were mixed at a ratio of 50:50, and grown without antibiotic

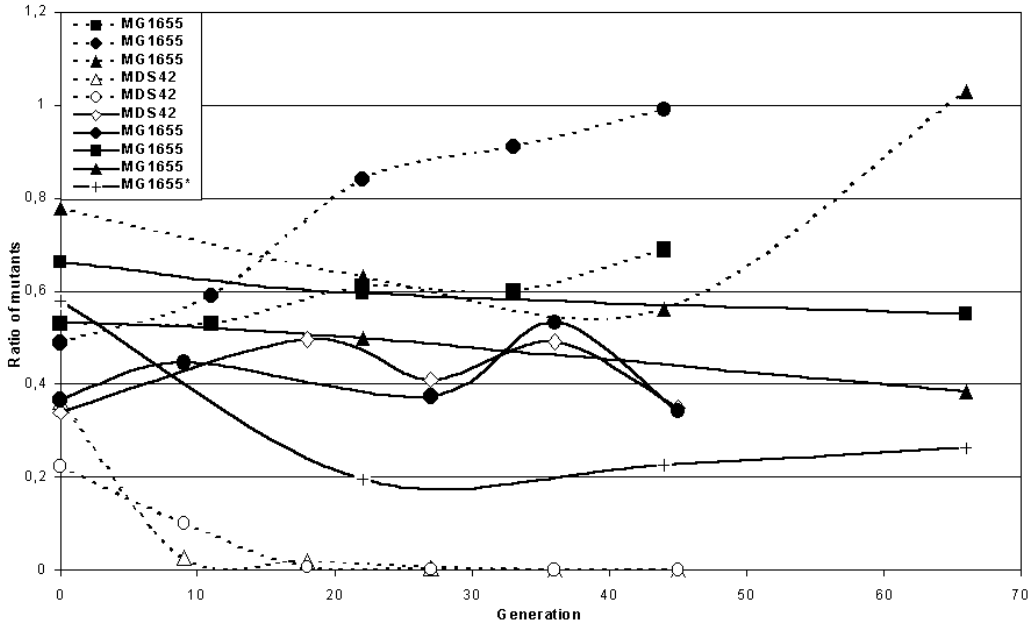


Figure 3.22. Competition of wt cells vs D-cycloserine resistant mutants (solid lines) and wt cells vs rifampicin resistant mutants (dashed lines) in liquid glucose-MT cultures. Each line represents an independent experiment involving a different, randomly picked mutant. All are point mutants, except one carrying an IS150 (*).

selection for 45-65 generations in minimal medium by repeated inoculum transfer from cultures reaching saturation into fresh medium. For comparison, mixed cultures of wt MG1655 or MDS42 and their randomly chosen rifampicin resistant (*rpoB*) mutants were grown in an identical way. After every ~10-20 doublings, the ratio of the antibiotic-resistant cells within the culture was determined by plating on glucose-MT plates with or without the specific antibiotic. 5-5 competition experiments were done using independent, randomly chosen *cycA* or *rpoB* mutants (**Figure 3.22**). Competition experiments can show wide fluctuations, but the observed tendency was clear: the ratio of D-cycloserine resistant cells remained largely constant in the mixed cultures, indicating that the *cycA* mutations did not cause growth defect or growth advantage in minimal medium. In contrast, rifampicin resistant mutants displayed variable results in similar competition experiments. While some of them were gradually overgrown by wt cells in the mixed culture, others outcompeted the wt cells.

4. Discussion

4.1. Optimization of the linear-fragment deletion method

4.1.1. *Construction and application of a plasmid expressing I-SceI*

Our group has developed a novel method for the construction of markerless deletions in the genome of *E. coli*. This procedure involves the replacement of a genomic segment by a PCR-generated linear DNA molecule. The cointegrates made this way are resolved after the I-SceI cleavage of the inserts, followed by a second round of homologous recombination. My primary task was to make the method more suitable for the accumulation of a large number of genomic deletions in a cell. This involved the construction of a plasmid that carries an inducible I-SceI meganuclease, to avoid the need of serial transformation of - and curing from - a plasmid that constitutively expresses the enzyme. I reached this goal by constructing pSTAST. Currently the Kan-resistant version pSTKST is used in the deletion work, allowing the parallel application of suicide plasmids carrying the Ap resistance marker. By measuring the growth retardation of *E. coli* MG1655 caused by different concentrations of heat-inactivated cTc, and testing the efficiency of different sublethal cTc concentrations for different periods of time, the optimal conditions could be selected for I-SceI induction. These refinements are now routinely used in the deletion protocols.

4.1.2. Inspecting the effect of the DSB on the mutation frequency

The DSB in the genome, created by the I-SceI meganuclease, presumably induces the SOS repair system of the cells (Krasin and Hutchinson, 1981; Pollard *et al.*, 1981). Activation of this system usually leads to an increased mutational rate (Walker *et al*,. 1996). A significant increase could mean that by applying the DSB-mediated method, the genome could become littered with unwanted point mutations. Pilot experiments were made to measure the effect of the DSB on the mutation frequency. Although the mutation frequencies measured in five different experiments had a quite wide distribution due to the jackpot effect, a trend showing that a DSB in the genome increases the frequency of point mutations by an order of a magnitude could be observed. Nevertheless, the absolute point mutation frequency was never higher than 1 mutant in 1’000’000 cells. This means, the method poses no practical disadvantage in terms of mutant contamination, even in the case of repeated use.

4.2. Construction of a multi-deletion *Escherichia coli* cell

Apart from a few exceptions (deletions MD1, *tonA*, *endA*, and the reconstruction of the *oppA* and *crl* genes), all of the deletions were made using the linear fragment-method. The deletions were routinely constructed in MG1655 harboring pBADαβγ (MG/BAD). The arabinose-induced MG/BAD could be stored at –70°C for several months in 40 µl aliquots without losing the ability to recombine a linear fragment into the genome (data not shown). This fact, in addition to the lack of cloning steps and the uniform PCR protocols made it possible to conveniently construct many deletion intermediates in parallel. The marked genomic segments – representing the deletion intermediates - were then incorporated in the multi-deletion strain by P1 transduction, followed by the optimized, DSB-mediated finishing of the scarless deletion process. This streamlined deletion process was a necessary prerequisite of the large-scale genome remodeling.

Only a minority of the deletions were performed using the more time-consuming suicide plasmid method. We preferred it in cases where the deletion endpoints were strictly defined (e.g., to delete an IS element to reconstruct an interrupted gene), limiting the choices for optimal primer design. Such limitation is more pronounced in the case of the long primers used by the linear fragment method.

Completion of 42 deletions resulted in a 14.28% reduction of the genome. Surprisingly, all planned deletions could be accumulated in the same cell. This shows the feasibility of large-scale genome remodeling, based on rational design. Similar engineering certainly would not have been possible using a Cre/LoxP or FLP/FRT system for the second recombination step, since the accumulation of LoxP or FRT sites would have led to their unwanted recombination at the serial Cre or FLP enzyme inductions, respectively. Albeit all construction-steps could have theoretically been carried out in the same cell using our protocol, as a safety measure, the first steps of the deletions were constructed separately. This not only prevented the unwanted recombinational effect of the λ-Red recombinases, but also gave us the advantage of possessing every deletion separately (see below).

One of our major goals was to remove all mobile genetic elements with the intent of stabilizing the genome. Localization and deletion of these elements was not a trivial task. Removing defective prophages is a work without a sharp end point: for example, we deleted a gene (ogrK) coding for a transcription activator for the late genes of the P2 phage, surrounded by many genes of unknown function in GP25. These genes could both be the last remnants of a lysogenic phage, inserted into the genome a long time ago, or could be just members of a group of orthologous genes present in both phages and bacteria (Blattner, 1997). Similarly,

MD7 and MD9 were deleted as ordinary K-islands. However, based on their similarity to O157:H7 Sakai strain cryptic prophages, Hayashi *et al*. (2001) defined them as K-12 prophage- like elements KpLE1 and KpLE2.

The deletion of IS elements is more well defined, but more difficult, too. A total of 31 deletions had to be made to remove all 44 IS elements that were found in the genome sequence in 1997 by Blattner *et al*. The data obtained later on from DNA-microarray experiments indicated the presence of even more IS elements. Their localization and their deletion was time-consuming. Nevertheless, it was quite informative to prove the transposition event of two insertion sequences during our work, and even more illuminating to find that our MG1655 strain already had three IS elements not present in the sequenced MG1655. Naas *et al.* reported in 1994, that during storage in stab culture at room temperature, the frequency of IS transposition can be so high, that over 30 years, the original strain can eventually become lost by yielding many new daughter cell lines. It's important to see, that despite the origin of our strain being the Blattner lab, and the storage at -70°C whenever possible, the IS distribution is different, which means their transposition could not be prevented. These data emphasize the need for an IS-free bacterial strain.

Rearrangement hotspots (Rhs) were also deleted. The 6 copies of Rhs account for 0.8 % of the MG1655 genome. These sequences are immobile, but their 3' ends, called the H-rpt are homologous to ISAS2, an active IS element in *Aeromonas salmonicida* (Gustafson *et al*., 1994, Bachellier *et al*., 1996). Since they're present in multiple copies, they are potential substrates of *RecA*-dependent homologous recombination (Lin *et al*., 1984). The FRT site, remaining after the deletion of MD1 was also removed later on to prevent its recombination with other FRT sites possibly introduced in future versions of our MDS.

Synthesis of the flagella can cost 2 % of the biosynthetic energy expenditure of cells, and its operation can account for a further 0.1 % of their total energy consumption under growth conditions (Macnab, 1996). To presumably reduce the energy needs and the doubling time of our strain, we deleted a total of seven operons involved in flagellum synthesis and chemotaxis.

A few additional mutations, characteristic of typical cloning hosts were introduced into our strain by deleting complete genes. For example, the *EcoK* restriction and modification system was deleted for the ability of cloning DNA originating from *hsdM*$^-$ strains, or *EcoK*$^-$ organisms. The genes of the Methylated cytosine restriction system (*mcrABC*) were removed in order to avoid the restriction of CG-methylated DNA, characteristic of higher eukaryotic cells. The genes coding for the receptor of phages T1, T5, ϕ80 (*tonA)*, and toxins colicinM

and microcinJ25 (Braun *et al.*, 2002) were also eliminated. MDS42 is lacking the periplasmic DNAse *endA* gene as well. A *recA*$^-$ and a *lacM15* derivative of this strain were also made for cloning repetitive sequences and for the α-complementation of certain vectors, respectively. The majority of the deletions were performed using the linear fragment method.

4.3. Characterization of the multi-deletion strain

4.3.1. Transformability

The chemical transformability of MDS41 was found to be about two orders of a magnitude weaker than that of MG1655. The replacement of calcium with other ions in the transformation buffer raised the chemical competence of MDS42 to a level equal to or better than the parental strain (Sun Chang Kim, personal communication). A possible explanation for the negative effect of Ca^{++} could be the aggregation of cells by the divalent cations. In our laboratory, the accelerated sedimentation of the MDS was commonly observed under certain circumstances (data not shown). Furhter investigation is needed to clear this issue and to identify the genes of possible cell-surface structures whose deletion explains this phenomenon.

The electroporation efficiency of MDS39 was better than that of DH10B in the majority of the experiments using both a small plasmid, pBR322, or a 100 kbp BAC. We couldn't prove the statistical significance of the difference due to the large variation of the data, but the consistency of the experimental results convinced us that MDS39 is just as useful for transformation, if not better, than DH10B. Our American collaborators tested the electroporation efficiency of MDS42 using standard laboratory procedures, and found it to be an order of a magnitude lower than that of DH10B. However, when using the special protocol of Scarab Genomics, they obtained ten times higher electroporation efficiencies for the MDS with small plasmids, and two times higher with the BAC (Pósfai *et al*., 2006). In all measurements using MDS39 or higher-numbered MDS, we obtained better electroporation efficiencies than the original MG1655, regardless of the type and molecular weight of the DNA used. It is possible, that the absence of the flagellar apparatus and other cell-surface proteins in the MDS cause the slight increase in electroporation efficiency by freeing the access to the pores of the cell wall. The lack of the periplasmic endonuclease, *endA* could also enhance the electrocompetence of MDS42. Deletion of *rfc*, coding for the O-antigen polymerase is not expected to increase transformation efficiency in itself, since the O-antigen isn't synthesized in K-12 due to the *rfb-50* mutation, caused by an IS5-integration (Liu and Reeves, 1994).

Absolute transformation efficiency of the strain (colonies/μg DNA) is the most relevant information to the user. To avoid drawing false conclusions on transformability due to differences in the number of cells surviving the transformation, the ratio of survivor cells was also measured after transformation events. No significant difference was found among the strains (data not shown).

4.3.2. Growth rates

Reduction of the genome size by 14.28% (662,606 bp) did not significantly decrease doubling time, showing that DNA replication is not the rate-limiting step of cell division. This stands in good agreement with the previous assumptions that the interinitiation time defines the doubling time of rapidly dividing cells (Helmstetter, 1996). The elimination of the flagellar system, together with other known protein-coding genes did not accelerate cell division, either. The slight and insignificant increase in the doubling time of the MDSs compared to MG1655 observed in rich medium (LB) could be caused by the loss of unknown genes coding for proteins responsible for the uptake and/or direct utilization of certain biomolecules. The lack of any measurable difference in the growth rates in glucose-minimal medium demonstrates that the deletions left all the genes necessary for optimal biomolecule-synthesis under such conditions unchanged. The longer lag phase observed in the case of the MDSs could indicate a slower activation of the available gene pool. This might be due to the deletion of a regulatory gene present on a genomic island. A similar case was reported by Morschhäuser *et al.* (1994), showing that genes present in the pathogenicity islands of uropathogenic *E. coli* 536 positively control the expression of genes in the "core" genome.

The decrease in the OD_{550} of the cell culture after reaching the plateau phase could have a more specific explanation. Since this phenomenon was only seen in strains containing 31 or more deletions, it is likely to be caused by the 31st deletion, GP24. Microarray experiments analysing the transcriptome of MDS41 detected the upregulation of *rsp*, a gene lying close to deletion MD8. This could be caused by a normally distant activator sequence getting in the vicinity of *rsp* as a result of the deletion. Rsp is known to inhibit the production of the σ_s subunit of RNA polymerase (Huisman and Kolter, 1994) needed for cell entrance into stationary phase. The lack or decreased amount of σ_s leaves the cell unable to adapt to food limitations, thereby killing it, hence the decrease in OD_{550}. This does not explain, why deletion GP24, and not MD8 activated this phenomenon, but the deletion of one or more

further activators of σ_s cannot be ruled out. Future characterization of single deletion mutants and their appropriate combinations is needed to clear this issue.

4.3.3. Mutation rate measurements

4.3.3.1. Inactivation of a methyltransferase gene

To detect a possible change in mutation frequency caused by the deletions, we first attempted to measure the inactivation frequency of a methyltransferase gene carried on a plasmid (data not shown), but the wide variation of the data did not allow us to draw any reliable conclusions. There can be several sources of error in such an experiment. Two preparations of electrocompetent cells can have a two to ten-fold variation in electroporation efficiency, even when belonging to the same batch of the same strain (Sambrook *et al*., 1989, p.1.75). Also, cells of the indicator strain (DH5α) could occaisonally have a mutation in their mcrBC genes, giving false positive colonies. Most importantly, there is anecdotal evidence that DNA fragments acquire mutations during transformation with a frequency much higher than during cellular replication. Although others (Umenhoffer Kinga, personal communication, Kim and Wood, 1997) detected no increase in the *genomic* mutation rate of *E. coli* due to electroporation, a mutagenic effect of electroporation on incoming DNA cannot be excluded. In summary, we concluded that due to intrinsic sources of error, complex protocols involving plasmid preparation and electroporation are not suitable for measuring small changes in the mutation rate.

4.3.3.2. Inactivation frequency of a toxic endonuclease

In our second method, we measured the inactivation frequency of a toxic, heat sensitive endonuclease encoded on a plasmid (data not shown). In contrast to the previous experiment, only one transformation had to be made with the plasmid and no known mutations of the tester strain could give a false positive colony. The problem of the plasmids acquiring mutations during transformation still existed, though, and probably explained the wide variation of the measured mutational frequencies.

Accordingly, we focused on the investigation of genes carried on the genome in the following.

4.3.3.3. Measuring the adaptation to salicin-minimal medium

We measured the activation frequency of the cryptic bgl operon by plating the strains on salicin-minimal medium. We carried out this experiment for the effective demonstration of

the presence or absence of IS elements within the two strains. In 80-98% of the bgl^+ mutants, the activation of the cryptic operon is caused by an IS element interrupting the bglR region (Schnetz, 1987; Hall 1998), so we expected a dramatic difference between the wt and IS-free strains' colony number on salicin-minimal medium. This was obtained indeed, with a high degree of significance. The adaptation of MDS41 to salicin minimal medium was much slower than that of MG1655 (**figure 3.13)**. The composition of mutants was screened by PCR, and no insertion mutants were detected in MDS41. In contrast, 92% of the bgl^+ MG1655 colonies on day 9 carried an IS element in their bglR segment, which implies that the large difference in the adaptation rates of the two strains was caused by the deletion of IS elements.

4.3.3.4. Measuring the frequency of mutations causing cycloserine-resistance

In contrast to the previous method, which is a biological assay for detecting the presence of IS-elements, we performed this experiment to measure the overall mutation rates of the strains. The other important difference between the two systems is that the bgl test detected mutations in the absence of growth, while the cycA test measured the mutation rate during cell division.

The inactivation of the *cycA* gene can be easily detected by plating on D-cycloserine. To convert mutant numbers to the number of mutations per tube (**m**), or mutation rate, we used the MSS-MLL, considered to be the golden standard method for fluctuation analysis (Roche and Foster, 2000). The ratio of the average mutation rates calculated for MDS41/MG1655 was 78.3% (**figure 3.16)**. Inspection of the composition of mutants revealed that 24.2% of the mutation in the *cycA* gene is caused by IS-transposition. This allowed us to draw the conclusion that the decreased mutation rate of *cycA* in MDS41 is caused by the lack of insertion sequences.

4.3.4. Analysis of plasmid preparations

The IS contamination of pUC19 preparations originating from different hosts was tested by PCR. All tested positive, except for the one prepared from MDS42 (**Figure 3.19**). This indicates, that the danger of the introduction of an exogenous IS element into the cloning host is most probably present at every bacterial transformation, unless DNA originating from an IS-free MDS strain is used.

4.4. Characterization of the cycA-method for mutation rate measurements

During our genome engineering project, we developed a procedure of mutation rate measurement based on the detection of D-cycloserine resistant cells (Fehér *et al.*, 2006). We used this method to provide evidence for the increased genomic stability of MDS41, and to attribute this trait to the lack of IS elements. However, due to the novelty of the protocol, further tests were needed to prove that it fulfils the criteria of ideal techniques measuring mutation rate.

4.4.1. Mapping the cycloserine-resistant phenotype

A direct relation between phenotype and genotype is necessary for simple interpretation of raw data. D-cycloserine has multiple targets in the bacterial cell: D-alanyl-D-alanine ligases A and B, and alanine racemase (Lambert and Neuhaus**,** 1972; Zawadzke *et al*., 1991; Noda *et al*., 2004; Bugg and Walsh, 1992). This implies, that their simultaneous modification has a negligible probability. Our hypothesis was that practically all cells demonstrating resistance to D-cycloserine are the results of mutations in *cycA*, encoding the transporter responsible for its uptake. To prove this, complementation experiments were made by controlled overexpression of the wt *cycA* gene in D-cycloserine resistant mutants. Twelve independent mutants were complemented on solid medium, and further twelve in liquid medium (**Figure 3.20**). In all cases, overexpression of wt *cycA* led to inhibition of growth in the presence of D-cycloserine, proving that modification of *cycA* was responsible for resistance.

All types of mutations were detected within the *cycA* gene. With the use of colony-PCR, the rapid classification of point-, insertion- and deletion mutants was possible. Of the ten different species of IS elements present in MG1655, our non-exhaustive screen of *cycA* insertion mutants identified six (IS1, IS2, IS3, IS4, IS5, IS150). Further investigation of point mutants by sequencing revealed the presence of base substitutions (transitions, transversions), +1 and -1 frameshifts, small deletions and short duplications, also. It is noteworthy that loss-of-function-mutations of *cycA* were found along nearly the entire length of the gene. This means that the 1413-bp gene provides a large target for selectable mutations, making it easier to obtain sufficiently large number of mutants for quantitation.

4.4.2. The effect of D-cycloserine-resistance on growth under non-selective conditions

Fluctuation analyses assume that mutants arising prior to selection divide with the same rate as their wt counterparts. Therefore, it was essential to prove, that D-cycloserine

resistance poses no growth advantage or disadvantage compared to wt under conditions lacking the physiological substrates for the CycA transporter. Our results indicated (**Figure 3.22**) that mutations in *cycA* are neutral in respect of the growth rate in minimal medium.

In contrast, rifampicin-resistant mutants demonstrated improved, unchanged or impaired growth characteristics compared to their wt counterparts. This is not surprising, taking into account that mutations needed for resistance affect *rpoB*, coding for the RNA polymerase beta subunit, and might influence the processivity of the enzyme. Indeed, rifampicin resistant mutants of *E. faecium* were reported to have a fitness cost (Enne *et al*., 2004). Moreover, it was shown that rifampicin resistant *E. coli* mutants can have pleiotropic phenotypes (Jin and Gross, 1989). This variability introduces an unknown parameter into mutation rate-measurements using rifampicin-resistance. Our system of D-cycloserine-selection completely circumvents this source of error.

In summary, the technique for measuring mutation rates using D-cycloserine selection is based on a direct relation between phenotype and genotype. It detects practically all types of mutations, no specific genetic background is needed, and the mutants demonstrate no growth advantage or disadvantage in minimal medium. Conversion of mutant numbers to mutation rate by fluctuation analysis is not simple, but the reproducibility of the results make up for this slight drawback to yield an ideal method of genomic mutation rate measurement.

4.5. Possible uses and advantages of the multi-deletion strain

Practical benefits of the reduced genome strain could be numerous. Its high electroporation efficiency makes it suitable for routine cloning purposes. Unlike DH10B, a common cloning host, MDS41 is culturable in minimal medium. Due to the IS-free status, the strain is unique among the commonly used laboratory hosts, and offers a solution to certain cloning problems.

Earlier reports (Binns, 1993, Hill *et al*., 2000, Kovarik *et al*., 2001, Astua-Monge *et al*., 2002) pointed out that a large, and rapidly increasing number of eukaryotic sequences in the public database inadvertently contain bacterial IS elements. These apparently hopped from *E. coli* into eukaryotic DNA during a brief period while it was cloned. The most dramatic example for such an unwanted transposition can be found under GenBank accession number AY319289. This contig of the mouse genome is interrupted by an 11 kbp segment of the *E. coli* genome, flanked by two IS10 elements. These insertion sequences, left behind in strain DH10B from previous genome engineering work, apparently lied close enough to function as

a complex transposon, and interrupted the sequence of interest by hopping. Cloning DNA in MDS41 for the intent of sequencing would eliminate such sources of error.

The decrease in mutation rate could also be useful in the case of protein or small molecule overproduction for industrial purposes. Overexpression of heterologous proteins has been shown to increase the overall mutation rate of bacterial cells (Lukacsovich and Venetianer, 1991), probably through the induction of the SOS response (Sweasy *et al*., 1990, Walker *et al*., 1996, Aris *et al*., 1998). This explains the phenomenon well known in the industry, that mutations are accumulated in overproducing strains' genomes at a faster rate (Zucca *et al*., 1979, Schurr *et al*., 1994, Barrett *et al*., 2004). If overexpression of a gene poses a burden to the cell, thereby slowing down its growth, mutants that lose this trait quickly overgrow the rest of the population in the fermentor. Nakamura and Inouye (1981) found, that *E. coli* cells expressing the plasmid-borne outer membrane lipoprotein gene (*lpp*) of *Serratia marcescens* lose the overproducing trait within thirty genererations of growth. Upon inspection of six mutant plasmid preparations, they found a mobile element (IS1 or IS5) disrupting the 5' untranslated region of the gene in three cases. A further example for this phenomenon was the instability of plasmids carrying the tyrosine operon in *tyrR* strains of *E. coli*. The lack of the repressor resulted in tyrosine overproduction, but 60 out of 82 transformants lost or decreased overexpression by acquiring deletions or IS1 insertions within the plasmid (Rood *et al.,* 1980). Our group has also shown that expressing a highly toxic fusion protein in MG1655, IS transposition accounts for 39% of the mutation rate (data of Kinga Umenhoffer, not shown), in contrast to the basal ratio of 25%. The absolute number of IS-insertions in *cycA* increased four-fold. This indicates a direct induction of IS transposition, possibly as a result of stress caused by protein overproduction. The use of the IS-free MDS completely eliminated this component of the mutational spectrum (Pósfai *et al*., 2006), emphasizing its possible advantage under such conditions.

Plasmids prepared from common laboratory *E. coli* strains or originating from commercial sources are contaminated with IS sequences, as shown by IS-specific PCRs (**Figure 3.19.**). Our group has evidence that the IS elements are present in these preparations in the form of genomic contamination, as minicircles, and as plasmid-co-integrants (data not shown). Keeping the MDSs IS-free during manipulations involving plasmids, only preparations originating from IS-free MDS host should be used.

The IS-free strain could prove to be a unique tool to study mobile DNA. The deliberate re-implantation of ISs into the genome, in a single or arbitrary copy number, would allow

studies on the mechanism, kinetics and regulation of IS-transposition without a noisy genetic background produced by native ISs.

Although it is non-pathogenic, *E. coli* K-12 has been described as a "pathogen waiting to happen" **(**http://www.news.wisc.edu/4708.html**).** Indeed, coli-sepsis is a hazard not to be underestimated during laboratory work with micro injuries on the skin **(**József Schlammadinger, personal communication**)**. Moreover, *E. coli* expressing capsular antigen type K-12 has been described to be among the five most common *E. coli* strains to be cultured from patients of acute pyelonephritis (Kajiser *et al*., 1977). Unfortunately, information is not available regarding the other virulence genes that the K-12 antigen-expressing *E. coli* cells causing the infection harboured. Under conditions of urethral obstruction, a commonly used K-12 strain (HB101) was shown to cause bacteraemia and renal inflammation in mice

Deleted gene(s)	Function	Reference
wbbABCDEFIJKL, wzxC, wzb	lipopolysaccharide biosynthesis	Reeves *et al*., 1996
nfnB	nitrofurantoin resistance	Sastry and Jayaraman, 1984
mviN	*Salmonella typhimurium* virulence gene homologue	Blattner *et al*., 1997
YgaF, YeaJ	overexpressed in murine septicaemic infection	Khan and Isaacson 2002
eaeH	homologous to the attachment and effacement protein	Blattner *et al*., 1997
flu	outer membrane fluffing protein	Caffrey and Owen, 1989
borD	lambda phage Bor protein homolog	Minagawa *et al*., 2003
ycgH1, ycgH2	adhesin-analogue autotransporter couple	Zhai and Saier, 2002
fimABCDEFGHI	type 1 fimbriae	Orndorff and Falkow, 1985
fecABCDEFGHIR	citrate dependent ABC iron transporter	Pressler *et al*., 1988
hokACD, sokD	small toxic peptide-encoding genes	Gerdes *et al*., 1990

Table 4. Putative virulence genes deleted in MDS42

(Johnson *et al*., 1993). These data underline the importance of deleting the potential virulence factors of *E. coli* MG1655. Removing the genes of the flagellary apparatus, in addition to those listed in **Table 4**, presumably reduced the risk of infection during laboratory work and other applications even further. By introducing mutations resulting in auxotrophy, further steps of safety could be added.

Single deletion-bearing MG1655 strains were saved as glycerol stocks. These cell lines carry the possibility of constructing arbitrary combinations of the available deletions using P1

transduction. This procedure could prove to be useful in the search for the genes causing observed changes in phenotype, or to provide certain combinations of deletions on order.

As described above, the IS elements interrupting the *YddA* and *ais* genes were eliminated by P1-transduction of the correct genomic segment from single-deletion derivates of MG1655. This was made possible by the chloramphenicol resistance gene inserted in the near vicinity of the interrupted genes, which could be used as a selective marker during transduction. Later on in the course of our work, the genome of the Phi80 phage was discovered to be inserted in the chromosome at its regular attachment site (between 1308567 and 1308767). Its correction would have indicated planning and executing one more deletion. Fortunately, the cleanup of the *opp* gene was made in MG1655, and the P1-transduction of the cointegrate eliminated the prophage, as confirmed by the genome hybridization tests. These examples reflect a further use of our deletion intermediates: correction of unwanted genomic changes in any close relative of *E. coli* K-12 by a simple and easily reproducible P1 transduction. Taking into account, that 0.1 Mbp can be transduced in a P1 capsid about 46 evenly spaced markers are needed to cover the whole 4.639 Mbp genome of *E. coli* K-12. At the time of the writing of this thesis, the number of deletions constructed allow the nearly total coverage of the genome using phage P1.

Cronan (2003) developed a new method for in vivo cosmid-packaging, which is much faster and less tedious than the in vitro process. Compared to previous in vivo packaging protocols, his method can yield a high (10^{10}/ml) level of packaged cosmid DNA. However, recombination between the helper phage and the defective prophages found in the genome of the packaging host can lead to the production of infective phage particles, too. Avoiding this recombination is crucial to guarantee the purity of the packaged cosmid. For this reason, he recommends our MDS, lacking all cryptic prophages, for use as an in vivo cosmid-packaging host.

Poteete and coworkers (2004) transformed linearized λ DNA carrying fragments of the *lac* operon into Red-expressing *E. coli* cells. They found, that the insertion of the DNA fragment into the chromosome was generated in 10-60% of the cases by homologous recombination not via the *lac* operon, but via cryptic prophages present in the genome. Upon transformation of an MDS12-derivative, lacking cryptic prophages, with such DNA, the frequency of illegitimate insertions decreased by three orders of a magnitude. This indicates, that the unwanted recombination of λ vectors into the genome of *E. coli* is a realistic scourge, but could be circumvented by using our MDS.

Most advantages of the IS-free MDS listed above need experimental verification. Perhaps further modifications will be needed to fulfill these expectations. Nevertheless, promising data is already availabale from our collaborators concerning the cloning of multiple palindromic DNA sequences. The production of certain proteins is also significantly more effective in the MDS than in conventional hosts. Exploring the background of these traits offers exciting experiments for the future.

4.6. Parallel genome-modification projects

Yu *et al.* (2002) performed a 6.7% reduction of the *E. coli* MG1655 genome. Initially, they made a random Tn-mutant pool, using modified Tn5 transposons, carrying either Kan or Cm resistance markers. They mapped each mutation by sequencing. Next, they combined selected mutants using P1-transduction, and excised the flanked chromosomal segment by Cre-mediated recombination of the loxP sites within the transposons. They performed the combinatorial accumulation of such deletions, and managed to delete 313 kbp this way. Not all deletions could be combined: the "mutually exclusive" deletions are thought to harbor genes coding for similar, essential functions. These genes were not identified. Only the growth rate of the deletion strain was measured in their report, and it was not significantly different than that of the wt. The major drawback of their deletion method is its expensiveness: in order to reach a saturated Tn5 mutant pool, where all non-essential genes are interrupted, an extremely large number of transposon-mutants needs to be sequenced. Without such a saturated mutant-pool, not all deletions can precisely be engineered. Furthermore, the use of the Cre/loxP system results in exogenous sequences remaining at the deletion sites.

Goryshin *et al*. (2003) developed a modified Tn5 transposon-based deletion-system. After random integration into the genome, this modified composite transposon can, upon induction, perform a second, intramolecular transposition. This event can lead to local inversions or genomic deletions of random size. By repeating the deletion cycle 20 times, Goryshin *et al*. constructed several different multi-deletion strains, that lacked up to 262 kbps of the MG1655 strain's genome. Using a variation of this technique, they managed to rescue the genomic segments neighboring 15 different Tn-insertions on a conditionally-replicative plasmid, and were thus able to investigate the essentiality of the genes carried upon them. The advantages of their deletion method are the following: (A) No previous knowledge is needed concerning the sequence or dispensability of the targeted bacterial genes. (B) The Tn5 transposase was shown to be active in all tested bacterial species, and (C) the deleted

segments can be saved as plasmids. These make the technique highly adequate for the screening of different bacterial genomes for essential genes. In contrary, the randomness of the integration site and the deletion size make it much less useful for the precise, planned construction of a minimal genome.

Westers and coworkers (2003) reported the deletion of 320 kbp of the *Bacillus subtilis* genome, thereby reducing it by 7.7%. They deleted six genomic segments, using a suicide-plasmid-based method. An extensive comparison of the multiple-deletion- and the parental-strain concluded, that the growth-, viability-, carbon-metabolism-, protein secretion-, competence- and sporulation characteristics were unaffected by the deletions. They found the multiple-deletion strain to have a reduced motility in 0.25% agar, but an increased motility in 0.5% agar, compared to the wt strain. A major aim of their work is to engineer an improved cell for industrial scale enzyme-, antibiotic- or insecticide production. The secretion rate of a heterologous protein (AmyQ), though, was not changed by the deletions, which brought them to the conclusion that no large energy-resources were redirected towards product formation or secretion.

Hashimoto *et al.* (2005) deleted 16 long genomic segments of *E. coli* MG1655, using linear PCR fragments. For the planning of the deletion end-points, they used the data concerning gene-function available in the literature, completed with their previous experimental results of separately engineering 75 medium-sized deletions. By transducing the long-sized deletions into one cell, a 29.7% genomic size reduction was accomplished. They observed the marked increase of the generation time of growing cells, more or less in parallel with the increasing number of deletions. They also described the multi-deletion cells to be longer and wider than the parental strain, but with a protein content reduced by 23-25%. Also, instead of having a single nucleoid midcell, or two nucleoids at ¼ and ¾ cell lengths, the multi-deletion cells had four or more, randomly distributed nucleoids in the cellular periphery. Their work provided important evidence, that careful planning of the deletions is needed to avoid having a detrimental effect on cellular physiology.

Recently, *Corynebacterium glutamicum* also became the object of genome engineering. Similarly to our work, Suzuki *et al.* (2005) defined the strain specific islands of the *C. glutamicum* R genome by comparison to strain ATCC 13032. Using suicide plasmids and the Cre/LoxP recombination system, they deleted eleven genomic islands, ranging in size from 9.8 to 55.6 kbp. They did not combine these deletions, though, probably because the presence of multiple (>2) loxP sites in the genome does not allow the predictable generation of deletions.

As the first version of this thesis was in press, Glass *et al.* (2006) repeated the Tn-mutagenesis of *M. genitalium* in search of the essential gene set. This time, care was taken to avoid the complementation of lethal mutants by wt cells in mixed colonies. As a result, a significantly higher number of protein coding genes (382) was found to be essential, than earlier (Hutchison *et al.,* 1999). Adding the 43 stable RNA genes, a complete set of genes coding for an ABC transporter, a lipoprotein and a glycerophosphoryl diester phosphodiesterase, they defined the minimal gene set of a bacterium to comprise at least 430 genes. Suprisingly, 110 of these are of unknown function, underlining the lack of knowledge concerning possibly basic pathways or structures required for life.

4.7. The future

The bacterial genome-engineering field will certainly see further genome reductions and targeted modifications. Our engineering work will be followed by a more detailed analysis of the arising MDSs. These include metabolic flux analysis of the bacterial cells, with the help of gas chromatography-linked mass spectrometry (GC-MS) or high-pressure liquid chromatography-linked mass spectrometry (HPLC-MS). Such an investigation, together with appropriate biochemical analysis software could give us a quantitative picture of the metabolic traffic within the cell. By comparing the 2D-protein gels obtained from wt and multi-deletion strains, the up-or down-regulation of certain proteins could also be catalogued. This would take us one step further than the microarray analysis of mRNA levels. Seeking the cause of the impairment of chemical competence, the increased aggregation and accelerated sedimentation of the cells are also interesting projects of the near future. As far as the deletions are concerned, perhaps it is not a distant goal to construct a "core" *E. coli* genome, harboring nothing but the conserved backbone sequence within its genome. That strain, being a common denominator of all present *E. coli*, could provide further experimental information about microbial genome evolution.

V. Summary

Escherichia coli K-12 and *E. coli* O157:H7 were among the first genomes that were completely sequenced. Comparison of the data revealed, that approximately 70-80% of their genomes are highly conserved. This backbone sequence is interrupted by many strain-specific islands. Based on these data, and other findings, the view emerged, that bacterial genomes consist of a stable, conserved, species-specific backbone, and a flexible, more rapidly changing combination of genes derived from a gene-pool available to many strains and species.

Since the genes of genomic islands are found in some, but not other strains, and their *in vivo* spontaneous loss was also described before, we hypothesized, that these genes are dispensable under common culturing conditions. Our group's goal was to delete as many strain specific islands in *E. coli* K-12 as possible, for two main reasons. First, as a basic scientific research initiative, we wanted to see how far one can go with the deletions without impairing the viability of the bacterium under laboratory conditions. Second, to develop a strain that is more useful for DNA-cloning experiments or other biotechnological purposes.

Numerous techniques have been developed for the targeted manipulation of bacterial chromosomes. These can be classified into two large groups, depending on whether suicide plasmids or linear DNA-fragments are the targeting molecules carrying the sequences homologous to the genome. Our deletion method is based on a λ-Red type recombination of a PCR-generated linear DNA fragment into the genome. Generation of such a cointegrate is followed by induction of a specific double stranded break within the insert, stimulating an intramolecular recombination event. The completed deletion process leaves no extrachromosomal sequences behind, and can be repeated in the same cell in an unlimited number of times.

This method was optimized for the rapid construction of serial deletions within the same cell. Plasmids responsible for the controlled overexpression of I-SceI, a highly specific endonuclease were engineered, allowing the precisely timed induction of the double strand break in the chromosome, needed for the second recombination step. A simple but highly effective protocol was developed for I-SceI induction. The potential mutagenicity of the procedure was investigated by measuring the mutation rate of *rpoB*. Although the mutation rate of this gene increased by an order of a magnitude, it never exceeded 10^{-6} mutation/cell/generation, and was thus considered insignificant.

42 deletions have been accumulated in the genome of *E. coli* strain MG1655. Among the deleted sequences were all the mobile genetic elements (44 IS elements, 9 prophages, 6 recombination hot spots), some damaged operons, and many genes of unknown function. Several putative virulence genes were removed, presumably increasing laboratory safety. These include the *wzxC, wzb, mviN, nfnB, YgaH, YeaJ, eaeH, borD, flu, ycgH1, ycgH2,* and *sokD* genes along with the *fla-, fim-, fec-, hok-* and *wbb* operons. Several genes, like *hsdRMS, mcrBCD, mrr, tonA*, or *endA* were deleted with the specific aim of strain improvement for DNA-cloning purposes.

The DNA-microarray tests made with the multi-deletion strain (MDS) showed, that although all known IS's were removed, further IS elements, absent at the time of sequencing, were present in the genome. Three IS1's, one IS2, and one IS5 were located using inverse PCR, underlining the mobility of these elements even in a time scale of three to four years. After deletion of these extra sequences, the MDS became free of mobile genetic elements, as proven by DNA micro-hybridizaton.

Phenotypic analysis was made to compare the MDS to the parental line. An insignificant elongation of doubling times in rich medium was detected along with the increasing number of deletions. No such effect was observed during growth in minimal medium. It is possible that genes responsible for the uptake and/or direct utilization of certain biomolecules were deleted, resulting in the slower growth in rich medium.

Compared to MG1655 or the commonly used cloning host DH10B, the MDS showed up to ten fold elevated electroporation efficiencies in the case of small plasmids. A slight advantage was observed even when 100 kbp-large BACs were electroporated. Due to the large variation of the data, the significance of these differences could not be proven. A possible explanation for the increase is that the partial lack of fimbriae facilitates the access of extracellular DNA to the cell surface. On the other hand, the deletions seemed to impair the chemical transformability of the cell, perhaps due to a tendency to coagulate in the presence of calcium ions.

Plasmids prepared from common laboratory *E. coli* strains or originating from commercial sources are contaminated with IS-elements, as shown by IS-specific PCRs. These pose the constant threat of unwanted inactivation of certain genes by integration, as well as the generation of deletions or inversions in the genome. The use of DNA originating from a multi-deletion host completely circumvents this problem.

Several protocols measuring the mutation rate were tested. Those detecting changes arising on a plasmid demonstrated poor reproducibility, and were inadequate for strain

comparison. A novel technique of mutation rate measurement was developed, based on the detection of D-cycloserine-resistant mutants. Mutations inactivating the CycA transporter render the cells resistant to this antibiotic. Mutant detection, combined with the Ma-Sandri-Sarkar maximum likelihood method of fluctuation analysis was used for the calculation of mutation rates. In this system, mutant selection is based on a direct relation between phenotype and genotype. It detects practically all types of mutations, no specific genetic background is needed, and the mutants demonstrate no growth advantage or disadvantage in minimal medium. These characteristics, together with its reproducibility make this method useful for statistical comparison of genomic mutation rates.

Mutation rates of the multi-deletion and wild type (wt) strains were compared. The overall inactivation rate of the *cycA* gene in the MDS was found to be 20-25% lower. PCR analysis of the mutants showed, that the inactivation of the *cycA* gene in the wt strain is caused by IS element-insertions in 24.2% of the cases. The difference between the mutation rate of the two strains can thus be explained by the deletion of IS elements.

For the uptake and hydrolysis of salicin, a beta-glycoside sugar, the activation of the cryptic *bgl* operon is needed. This is most commonly caused by IS elements inserting into its regulatory region. To perform a biological assay demonstrating the presence or absence of IS-elements, the activation of the *bgl* operon was measured by plating wt and MDS strains on salicin-minimal medium. The MDS demonstrated a 80-90 % lower overall activation frequency. Using PCR analysis, *bgl* was shown to be activated in wt cells in 92 % of the cases by IS element-insertion into the regulatory region of the operon. No such insertions were detected in the MDS. The difference in the activation frequency of the two strains can thus be explained by the deletion of IS elements.

The optimized linear fragment deletion method was proven to be feasible for the serial deletion of bacterial genomic islands. After careful planning, 14.28 % (662,606 bp) of the genome of *E. coli* was deleted without impairing the viability of the bacterial cell.

This work resulted in a strain that is advantageous for experiments involving the cloning of DNA fragments. Elevated electroporation efficiency and lower mutation rate are beneficial features. Earlier reports pointed out that a large number of eukaryotic sequences in the public database contain bacterial IS elements as contamination. Cloning DNA in the MDS for the intent of sequencing eliminates such sources of error. The decrease in mutation rate is also useful in the case of protein or small molecule overproduction for industrial purposes.

The IS-free strain is a unique tool to study the mechanism, kinetics and regulation of IS-transposition without a noisy genetic background produced by native ISs. The lack of

defective prophages in the genome could make it an ideal host for in vivo cosmid packaging. Single deletion-bearing strains are available, carrying the possibility of constructing arbitrary combinations of the available deletions using P1 transduction. These could be useful for mutant studies, or could serve as platforms for future genome engineering projects.

Results in brief:

- *E. coli* K-12 genome reduced by 14.28% (702 genes deleted in 42 segments)
- all major mobile elements removed (11 phages or phage remnants, 49 ISs, 6 Rhs)
- characterization of the multi-deletion strain:
  - -growth rate practically unchanged (LB / glucose minimal media)
  - -decreased chemical competence
  - -unchanged or increased electroporation efficiency
  - -plasmid preparations free of IS-contamination
  - -lowered mutation rate during growth and in stationary phase due to the absence of IS-generated mutations
- proof for five IS element-transpositions since sequencing
- development of a method capable of reproducible measurement and comparison of bacterial mutation rates and spectra

VI. Összefoglalás

Az elmúlt évtized egyik nagy molekuláris biológiai áttörése bizonyos fajok genetikai állományának teljes feltérképezése volt. Az *Escherichia coli* K12-es és O157:H7-es törzsei az első baktériumok közt szerepeltek, melyek genomjának meghatározták a bázissorendjét. Szekvenciájuk összehasonlítása arra az érdekes megfigyelésre vezetett, hogy genomjuk hozzávetőleg 70-80%-a nagy mértékben konzervált, szinte azonos. Ezt a „gerinc" szekvenciát számos helyen specifikus, csak az adott törzsben megtalálható szigetek szakítják meg. Más *E. coli* törzsek és baktériumfajok vizsgálata is hasonló struktúráltságot tárt fel. Így alakulhatott ki az az elmélet, mely szerint a prokarióta genomok két részből állnak: egy konzervált gerincből, és egy dinamikusan változó, niche-specifikus funkciókat ellátó génkészletből. Ez utóbbiak baktériumfajok közötti cseréje, az ún. laterális géntranszfer evolúciós skálán igen gyakori jelenség. Az uropathogén *E. coli* törzsben ilyen genomi szigetek spontán elvesztését is megfigyelték. Erre, és e szigetek törzsspecifitására alapoztuk munkahipotézisünket, mely szerint e szigetek nélkülözhetőek szokványos tápfolyadékokban való növesztés során. Csoportunk célja a lehető legtöbb törzs-specifikus sziget kiejtése volt az *E. coli* K12-es törzs genomjából, két alapvető célból: (i) alapkutatási szándékkal meg kívántuk állapítani, mennyi deléció végezhető el a törzsön anélkül, hogy életképessége – laboratóriumi körülmények között – súlyosan károsodna; (ii) alkalmazott kutatási célzatunk pedig egy olyan törzs kifejlesztése volt, amely kifejezetten alkalmas DNS-klónozásra, fehérjék vagy kismolekulák termelésére, illetve egyéb, igény szerinti biotechnológiai felhasználásra.

Napjainkban számos, a bakteriális kromoszóma módosítását lehetővé tévő technika áll rendelkezésre. Ezek működése többnyire homológ rekombináción alapul. Két nagy csoportra osztjuk őket aszerint, hogy cirkuláris vagy lineáris DNS szakasz rekombinációja megy végbe a kromoszómával. Munkánk során egy PCR-el generált lineáris DNS szakaszt rekombináltattunk a bakteriális genomba, a λ fág Red-rekombinázainak indukciója segítségével. Az így kapott kointegráns szakaszban egy extrém ritkán hasító restrikciós endonukleázzal (I-SceI) kettős szálú DNS-törést hoztunk létre, amelyet intramolekuláris rekombináció követett. E kétlépcsős folyamat eredményeként olyan deléciót alakítottunk ki, amelyhez nem társult extrakromoszómális DNS visszamaradása, és ezért tetszés szerinti számban ismételhető, ugyan abban a sejtben.

Az első feladat a delécios módszer optimalizálása volt, nagyszámú deléció kumulatív módon történő gyors és hatékony konstrukciója céljából. Az I-SceI precízen kontrollált expresszióját lehetővé tevő plazmidokat szerkesztettünk, melyekkel a kettősszálú DNS törés

és a következményes második rekombinációs lépés megfelelően időzíthető. Kifejlesztésre került egy egyszerű, de hatékony indukciós protokoll. Az eljárás potenciális mutagenitásának becslésére megvizsgáltam annak hatását az *rpoB* gén mutációs frekvenciájára. Megfigyeléseim szerint módszerünk egy nagyságrenddel megemeli az adott gén mutációs rátáját, ám az mindig 10^{-6} mutáció /sejt /generáció érték alatt marad, ezért ezt nem tekintjük szignifikáns hátránynak.

Az *E. coli* MG1655-ös törzs genomjában 42 deléciót halmoztunk fel. Ezzel kiejtettünk a sejtből minden mobilis genetikai elemet (44 IS elemet, 9 profágot, 6 rekombinációs forró pontot), több kriptikus operont és számos ismeretlen funkciójú gént. Eltávolítottunk több feltételezett virulencia-faktor gént is, mellyel a törzs esetleges pathogenitása is csökkenhetett. Ilyen a *wzxC, wzb, mviN, nfnB, YgaH, YeaJ, eaeH, borD, flu, ycgH1, ycgH2,* és a *sokD* gén, valamint a *fla, fim, fec, hok,* és *wbb* operon. Egyebeket, úgymint a *hsdRMS, mcrBCD, mrr, tonA*, és *endA* géneket azzal a kifejezett céllal ejtettük ki, hogy törzsünk hatékonyabb DNS klónozó gazdává váljék.

Noha minden ismert IS szekvenciát eltávolítottunk a genomból, a multideléciós törzs (MDS) genomján végzett DNS-mikrohibridizációs kísérletek további IS elemek jelenlétére világítottak rá. Ezek nyilvánvalóan nem voltak jelen a törzs szekvenálása idején, ami jól demonstrálja ezen elemek mozgékonyságát. Lokalizációjukat inverz PCR-el végeztük. Az így talált három IS1, egy IS2 és egy IS5 szekvencia eltávolításával az MDS mindennemű mozgékony genetikai elemtől mentessé vált, melyet DNS-csippel végzett vizsgálatok is megerősítettek.

Az MDS fenotípusát több aspektusból is összehasonlítottuk a vad típuséval. A deléciók számának növekedésévek párhuzamosan a törzs gazdag folyékony kultúrában mért duplázódási idejének nem szignifikáns mértékű megnyúlását észleltük. Minimál tápfolyadékban nem tapasztaltunk ilyen tendenciát. Ezek ismeretében különböző biomolekulák felvételéért illetve lebontásáért felelős gének delécióját feltételezzük, amelyek hiánya gazdag táptalajon lassabb növekedéshez vezet.

Az MDS sejtjeibe tízszer nagyobb hatékonysággal elektroporálhatók kis méretű plazmidok, mint mint az MG1655-ös vagy a közkedvelt klónozó gazda, a DH10B törzsbe. Még 100 kbp nagyságú bakteriális arteficiális kromoszómák elektroporációja esetén is kimutatható kis mértékű előny a multideléciós törzs javára. Az adatok szórása miatt a növekedés szignifikanciáját nem tudtam bizonyítani. Ha a különbségek valósak, azok oka feltehetően a sejtfelszíni fimbriák mennyiségének csökkenése, amely az extracelluláris DNS számára könnyebb hozzáférést biztosít a külső membránhoz. Ezzel szemben a deléciók

csökkentették a sejt kémiai transzformációjának hatékonyságát. Ez utóbbi feltehetően a Ca^{2+} ionok jelenlétében fennálló fokozott aggregációs hajlam következménye, amely jelenségre egyelőre nem találtunk magyarázatot.

PCR reakciókkal kimutattuk, hogy DNS klónozásra használt, szokványos *E. coli* törzsekből származó plazmid preparátumok IS-elemekkel szennyezettek. Ezek nemkívánatos mutációkat okozhatnak különböző génekbe való beékelődésükkel, de deléciókhoz, inverziókhoz és duplikációkhoz is vezethetnek. E folytonosan fennálló vezély csak úgy hárítható el, ha MDS-ből preparált DNS-t használunk molekuláris biológiai munkákhoz.

Számos, a mutációs ráta mérését célzó eljárást vizsgáltam. Azok a módszerek, amelyek plazmid-hordozta gének inaktivációs frekvenciáját mérték, gyenge reprodukálhatóságuk miatt alkalmatlannak bizonyultak bakteriális törzsek összehasonlítására. Ezért egy új eljárást dolgoztam ki, amely a kromoszómális *cycA* gén inaktivációjának gyakoriságát méri. E mutánsok D-cikloszerin rezisztenciájuk révén könnyen detektálhatóak. Párhuzamos kultúrákban kapott mutáns számokat a Ma-Sandri-Sarkar féle fluktuációs analízissel feldolgozva kiszámíthatjuk az adott törzs mutációs rátáját. Ez a mérési elrendezés direkt genotípus-fenotípus összefüggésem alapul, és gyakorlatilag minden mutáció-típust (inszerciót, deléciót, báziscserét, kereteltolódást) detektálni képes. További jellemzője, hogy nem igényel speciális genetikai hátteret, és a mutánsok minimál-tápfolyadékban növesztve nem mutatnak növekedési hátrányt vad típusú társaikkal szemben. E módszerrel kapott eredmények jól reprodukálhatóak és statisztikai módszerekkel könnyen elemezhetők, ezért a D-cikloszerines szelekciót ideálisnak találtuk bakteriális törzsek mutációs rátáinak összehasonlítására. PCR reakciókkal lehetőség van a D-cikloszerin rezisztens mutánsok *cycA* génjeinek amplifikációjára, és hosszuk alapján inszerciós-, deléciós- és pontmutánsként való osztályozására.

A multideléciós és az anyai MG1655-ös *E. coli* törzsek összehasonlítása azt az eredményt adta, hogy az előbbi mutációs rátája 20-25%-al alacsonyabb az utóbbihoz képest. Az anyai törzs mutánsai között a PCR-el kimutatott inszerciós mutánsok aránya 24,2% volt. Ahogyan vártuk, az MDS-ben nem találtunk inszerciós mutánst. A két törzs között fennálló mutációs rátabeli különbséget tehát az IS elemek deléciójával magyarázhatjuk.

A szalicin egy β-glikozidos cukor, amely felvételéhez és lebontásához *E. coli*ban a kriptikus *bgl* operon aktivációja szükséges. Ezt leggyakrabban az operon szabályozó régiójába beékelődő IS-elemek okozzák. Ezért a szalicin hasznosítás az IS elemek jelenlétét vagy hiányát demonstráló biológiai próbaként is felhasználható. Szalicin minimál táptaljra való kikenés után az MDS 80-90%-al alacsonyabb aktivációt mutatott, mint az anyai törzs.

PCR-el igazoltam, hogy a szalicinen növekedét mutató anyai sejtek 92%-ában IS inszerció történt a *bgl* operon szabályozó régiójába. Az MDS-ben nem találtam inszerciót e régióban. A szalicinen való növekedési képesség nagymértékű különbsége, amely a két törzs között fennáll, tehát az IS elemek delécijára vezethető vissza.

Az optimalizált, lineáris DNS fragmenseket használó deléciós módszerünkről bizonyítást nyert, hogy alkalmas bakteriális genomi szigetek sorozatos delécijára. Megfelelő tervezés után az *E. coli* genom 14,28%-át (662606 bp-t) ejtettük ki anélkül, hogy a sejt laboratóriumi életképességét számottevően rontottuk volna. Munkák eredménye egy olyan törzs lett, amely DNS-klónozási feladatokra is alkalmas. A magas elektroporációs hatékonyság és a csökkent mutációs ráta révén rutin molekuláris biológiai munkák mellett kifejezetten előnyös lehet ipari célú fehérje vagy kismolekula túltermeltetésre is. Az adatbázisokban napjainkra egyre nagyobb mennyiségben rendelkezésre álló eukarióta genomszekvenciákban számos bakteriális IS elemet találtak. Ezek az IS elemek feltehetően a bakteriális klónozás fázisában ugrottak a adott DNS szakaszra. Ha a szekvenálandó genomi fragmenst MDS-ben klónozzuk, az efféle szekvenálási műtermékek keletkezése elkerülhető.

Az IS mentes törzs egyedülálló lehetőséget biztosít a transzpozíció mechanizmusának, kinetikájának és szabályozásának vizsgálatára bevitt IS elemeken, natív mobilis elemek által produkált „zajos" háttér nélkül. A defektív profágok genomból való kiejtése révén az MDS igéretes lehet in vivo kozmid pakoló gazdaként való felhasználásra is. A rendelkezésre álló egyedi delécikkal P1 transzdukció révén tetszőleges kombinációjú multideléciós törzsek állíthatók elő. Ezek hasznosak lehetnek különböző mutánsok vizsgálatára, illetve további genommanipulációs vizsgálatoknak szolgáltathatnak kiindulási alapot a jövőben.

Eredmények pontokba foglalva:

- 14,28%-os genomcsökkentés *E. coli* K-12-ben (42 deléció, 702 gén)
- minden mobilis genetikai elem eltávolítása (11 profág, 49 IS, 6 Rhs)
- a multideléciós törzs jellemzése:
  - -gyakorlatilag változatlan növekedési ráta
  - -csökkent kémiai kompetencia
  - -megtartott vagy növelt elektroporációs hatékonyság
  - -IS –mentes plazmid preparátum
  - -az IS –transzpozíció hiánya révén csökkent mutációs ráta
- a szekvenálás óta öt IS elem spontán transzpozíciójának igazolása a törzsben
- egy hasznos eljárás kifejlesztése bakteriális mutációs ráták mérésére és összehasonlítására

V. Acknowledgements

This project was made possible by the careful planning, financing and supervision of György Pósfai. The technical help and expertise of Ágnes Szalkanovics, Vitaliy Kolisnychenko, Kinga Umenhoffer, Zsuzsa Győrfy, Ildikó Karcagi and Botond Cseh was essential. Information and materials provided by Frederick R. Blattner, Guy Plunkett III., Christopher D. Herring, David Frisch and János Pósfai were also of extreme value. Antal Kiss, Csaba Vízler, Trevor Twose and Sun Chang Kim also contributed to this project, for which I am grateful.

VII. References

Akerley BJ, Rubin EJ, Novick VL, Amaya K, Judson N, Mekalanos JJ. 2002. A genome-scale analysis for identification of genes required for growth or survival of Haemophilus influenzae. Proc Natl Acad Sci U S A. 99:966-971.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403-410.

Aris A, Corchero JL, Benito A, Carbonell X, Viaplana E, Villaverde A. 1998. The expression of recombinant genes from bacteriophage lambda strong promoters triggers the SOS response in Escherichia coli. Biotechnol Bioeng. 60:551-559.

Arnqvist, A, Olsen, A., Pfeifer, J, Russell, DG, Normark, S. 1992. The Crl protein activates cryptic genes for curli formation and fibronectin binding in Escherichia coli HB101. Mol. Microbiol. 6:2443-2452

Asai T, Condon C, Voulgaris J, Zaporojets D, Shen B, Al-Omar M, Squires C, Squires CL. 1999a. Construction and initial characterization of Escherichia coli strains with few or no intact chromosomal rRNA operons. J Bacteriol. 181:3803-3809

Asai T, Zaporojets D, Squires C, Squires CL. 1999b. An Escherichia coli strain with all chromosomal rRNA operons inactivated: complete exchange of rRNA genes between bacteria. Proc Natl Acad Sci U S A. 96:1971-1976.

Astua-Monge, G., Lyznik, A., Jones, V., Mackenzie, S. A. & Vallejos, C. E. 2002. Evidence for a prokaryotic insertion-sequence contamination in eukaryotic sequences registered in different databases. Theor. Appl. Genet. 104:48-53

Bachellier S, Gilson E, Hofnung M, Hill CW. 1996. Repeated Sequences. In *Escherichia coli and Salmonella,* Neidhardt, FC, Curtiss, RIII, Ingraham, JL, Lin, ECC, Low, KB, Magasanik, B, Reznikoff, WS, Riley, M, Schaechter, M and Umbarger, HE (eds.), ASM Press, Washington DC, USA, Vol. 2, pp. 2012-2019.

Barbour SD, Nagaishi H, Templin A, Clark AJ. 1970. Biochemical and genetic studies of recombination proficiency in Escherichia coli. II. Rec+ revertants caused by indirect suppression of rec- mutations. Proc Natl Acad Sci U S A. 67:128-135

Barrett E, Stanton C, Zelder O, Fitzgerald G, Ross RP. 2004. Heterologous expression of lactose- and galactose-utilizing pathways from lactic
acid bacteria in Corynebacterium glutamicum for production of lysine in whey. Appl Environ Microbiol. 70:2861-2866.

Benzinger R, Enquist LW, Skalka A. 1975. Transfection of Escherichia coli spheroplasts. V. Activity of recBC nuclease in rec+ and rec minus spheroplasts measured with different forms of bacteriophage DNA. J Virol. 15:861-871.

Binns, M. 1993. Contamination of DNA database sequence entries with *Escherichia coli* insertion sequences. Nucleic Acids Res. 21:779

Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of Escherichia coli K-12. Science. 277:1453-1474.

Blomfield IC, Vaughn V, Rest RF, Eisenstein BI. 1991. Allelic exchange in Escherichia coli using the Bacillus subtilis sacB gene and a temperature-sensitive pSC101 replicon. Mol Microbiol. 5:1447-1457.

Braun M, Killmann H, Maier E, Benz R, Braun V. 2002. Diffusion through channel derivatives of the Escherichia coli FhuA transport protein. Eur J Biochem. 269:4948-4959.

Breines DM, Ouabdesselam S, Ng EY, Tankovic J, Shah S, Soussy CJ, Hooper DC. 1997. Quinolone resistance locus nfxD of Escherichia coli is a mutant allele of the parE gene encoding a subunit of topoisomerase IV. Antimicrob Agents Chemother. 41:175-179.

Bugg TD, Walsh CT. 1992. Intracellular steps of bacterial cell wall peptidoglycan biosynthesis: enzymology, antibiotics, and antibiotic-resistance. Nat Prod Rep. 9:199-215.

Caffrey P, Owen P. 1989. Purification and N-terminal sequence of the alpha subunit of antigen 43, a unique protein complex associated with the outer membrane of Escherichia coli. J Bacteriol. 171:3634-3640.

Cheetham BF, Katz ME. 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. Mol Microbiol. 18:201-208.

Clark AJ. 1974. Progress toward a metabolic interpretation of genetic recombination of
Escherichia coli and bacteriophage lambda. Genetics. 78:259-271.

Clark AJ, Sharma V, Brenowitz S, Chu CC, Sandler S, Satin L, Templin A, Berger I, Cohen A. 1993. Genetic and molecular analyses of the C-terminal region of the recE gene from the Rac prophage of Escherichia coli K-12 reveal the recT gene. J Bacteriol. 175:7673-7682.

Curtiss R 3rd, Charamella LJ, Berg CM, Harris PE. 1965. Kinetic and genetic analyses of D-cycloserine inhibition and resistance in Escherichia coli. J Bacteriol. 90:1238-1250.

Cronan JE. 2003. Cosmid-based system for transient expression and absolute off-to-on transcriptional control of Escherichia coli genes. J Bacteriol. 185:6522-6529.

Defez R, De Felice M. 1981. Cryptic operon for beta-glucoside metabolism in Escherichia coli K12: genetic evidence for a regulatory protein. Genetics. 97:11-25.

DiNardo S, Voelkel KA, Sternglanz R, Reynolds AE, Wright A. 1984. Escherichia coli DNA topoisomerase I mutants have compensatory mutations at or near DNA gyrase genes. Cold Spring Harb Symp Quant Biol. 47 Pt 2:779-84.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. Genetics. 148:1667-1686.

Eisen JA. 2001. Gastrogenomics. Nature. 409:463

Enne, VI, Delsol AA, Roe JM, Bennett PM. 2004. Rifampicin resistance and its fitness cost in Enterococcus faecium. J Antimicrob Chemother. 53:203-207.

Fehér T, Cseh B, Umenhoffer K, Karcagi I, Posfai G. 2006. Characterization of cycA mutants of Escherichia coli. An assay for measuring in vivo mutation rates. Mutat Res. 595:184-190

Gay P, Le Coq D, Steinmetz M, Berkelman T, Kado CI. 1985. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. J Bacteriol. 164:918-921.

Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, Daugherty MD, Somera AL, Kyrpides NC, Anderson I, Gelfand MS, Bhattacharya A, Kapatral V, D'Souza M, Baev MV, Grechkin Y, Mseeh F, Fonstein MY, Overbeek R, Barabasi AL, Oltvai ZN, Osterman AL. 2003. Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. J Bacteriol. 185:5673-5684.

Gerdes K, Poulsen LK, Thisted T, Nielsen AK, Martinussen J, Andreasen PH. 1990. The hok killer gene family in gram-negative bacteria. New Biol. 2:946-956.

Giel M, Desnoyer M, Lopilato J. 1996. A mutation in a new gene, bglJ, activates the bgl operon in Escherichia coli K-12. Genetics. 143:627-35.

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CAIII, Smith HO, Venter JC. 2006. Essential genes of a minimal bacterium**.** PNAS 103:425-430.

Go CHU, Cunha BA. 2004. Escherichia coli infections. www.eMedicine.com

Goldman JD, White DG, Levy SB. 1996. Multiple antibiotic resistance (mar) locus protects Escherichia coli from rapid cell killing by fluoroquinolones. Antimicrob Agents Chemother. 40:1266-1269.

Goryshin IY, Naumann TA, Apodaca J, Reznikoff WS. 2003. Chromosomal deletion formation system based on Tn5 double transposition: use for making minimal genomes and essential gene analysis. Genome Res. 13:644-653.

Guerinot, ML, Morisseau, BA, Klapatch, T. 1990. Electroporation of Bradyrhizobium japonicum. Mol Gen Genet 221:287-290

Gustafson CE, Chu S, Trust TJ.. 1994. Mutagenesis of the paracrystalline surface protein array of Aeromonas salmonicida by endogenous insertion elements. J Mol Biol. 237:452-463.

Gutterson NI, Koshland DE Jr. 1983. Replacement and amplification of bacterial genes with sequences altered in vitro. Proc Natl Acad Sci U S A. 80:4894-4898

Hacker J, Bender L, Ott M, Wingender J, Lund B, Marre R, Goebel W. 1990. Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal Escherichia coli isolates. Microb Pathog. 8:213-225.

Hacker J, Blum-Oehler G, Muhldorfer I, Tschape H. 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol Microbiol. 23:1089-1097.

Hacker J, Blum-Oehler G, Hochhut B, Dobrindt U. 2003. The molecular basis of infectious diseases: pathogenicity islands and other mobile genetic elements. A review. Acta Microbiol Immunol Hung. 50:321-330.

Hall BG, Xu L. 1992. Nucleotide sequence, function, activation, and evolution of the cryptic asc operon of Escherichia coli K12. Mol Biol Evol. 9:688-706

Hall SD, Kane MF, Kolodner RD. 1993. Identification and characterization of the Escherichia coli RecT protein, a protein encoded by the recE region that promotes renaturation of homologous single-stranded DNA. J Bacteriol. 175:277-287. Erratum in: J Bacteriol 175:1211.

Hall SD, Kolodner RD. 1994. Homologous pairing and strand exchange promoted by the Escherichia coli RecT protein. Proc Natl Acad Sci U S A. 91:3205-3209.

Hall BG. 1998. Activation of the bgl operon by adaptive mutation. Mol Biol Evol. 15:1-5.

Hamilton CM, Aldea M, Washburn BK, Babitzke P, Kushner SR. 1989. New method for generating deletions and gene replacements in Escherichia coli. J Bacteriol. 171:4617-4622.

Hashimoto M, Ichimura T, Mizoguchi H, Tanaka K, Fujimitsu K, Keyamura K, Ote T, Yamakawa T, Yamazaki Y, Mori H, Katayama T, Kato J. 2005 Cell size and nucleoid organization of engineered Escherichia coli cells with a reduced genome. Mol Microbiol. 55:137-149.

Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T, Sasakawa C, Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H. 2001. Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. 8:11-22.

Helmstetter, CE. 1996. Timing of Synthetic Activities in the Cell Cycle. In *Escherichia coli and Salmonella,* Neidhardt, FC, Curtiss, RIII, Ingraham, JL, Lin, ECC, Low, KB, Magasanik, B, Reznikoff, WS, Riley, M, Schaechter, M and Umbarger, HE (eds.), ASM Press, Washington DC, USA, Vol. 2, pp. 1627-1639.

Hill, F., Gemund, C., Benes, V., Ansorge, W. & Gibson, T. J. 2000. An estimate of large-scale sequencing accuracy. EMBO Rep. 1:29-31

Huisman GW, Kolter R. 1994. Sensing starvation: a homoserine lactone--dependent signaling pathway in Escherichia coli. Science. 265:537-539.

Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC. 1999. Global transposon mutagenesis and a minimal Mycoplasma genome. Science. 286:2165-2169.

Itaya M. 1995. An estimation of minimal genome size required for life. FEBS Lett. 362:257-260.

Jaffe A, Chabbert YA, Derlot E. 1983. Selection and characterization of beta-lactam-resistant Escherichia coli K-12 mutants. Antimicrob Agents Chemother. 23:622-625.

Jasin M, Schimmel P. 1984. Deletion of an essential gene in Escherichia coli by site-specific recombination with linear DNA fragments. J Bacteriol. 159:783-786.

Jin DJ, Gross CA. 1988. Mapping and sequencing of mutations in the Escherichia coli rpoB gene that lead to rifampicin resistance. J Mol Biol. 202:45-58.

Jin DJ, Gross CA. 1989. Characterization of the pleiotropic phenotypes of rifampin-resistant rpoB mutants of Escherichia coli. J Bacteriol. 171:5229-5231.

Johnson DE, Russell RG, Lockatell CV, Zulty JC, Warren JW. 1993. Urethral obstruction of 6 hours or less causes bacteriuria, bacteremia, and pyelonephritis in mice challenged with "nonuropathogenic" Escherichia coli. Infect Immun. 61:3422-3428.

Joseph JW, Kolodner R. 1983. Exonuclease VIII of Escherichia coli. II. Mechanism of action. J Biol Chem. 258:10418-10424.

Kaijser B, Hanson LA, Jodal U, Lidin-Janson G, Robbins JB. 1977. Frequency of E. coli K antigens in urinary-tract infections in children. Lancet. 1:663-666.

Khan MA, Isaacson RE. 2002. Identification of Escherichia coli genes that are specifically expressed in a murine model of septicemic infection. Infect Immun. 70:3404-3412.

Kim CS, Wood TK. 1997. Creating auxotrophic mutants in Methylophilus methylotrophus AS1 by combining electroporation and chemical mutagenesis. Appl Microbiol Biotechnol. 48:105-108.

Kleckner N, Roth J, Botstein D. 1977. Genetic engineering in vivo using translocatable drug-resistance elements. New methods in bacterial genetics. J Mol Biol. 116:125-159.

Knuth K, Niesalla H, Hueck CJ, Fuchs TM. 2004. Large-scale identification of essential Salmonella genes by trapping lethal insertions. Mol Microbiol. 51:1729-1744.

Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Debarbouille M, Dervyn E, Deuerling E, Devine K, Devine SK, Dreesen O, Errington J, Fillinger S, Foster SJ, Fujita Y, Galizzi A, Gardan R, Eschevins C, Fukushima T, Haga K, Harwood CR, Hecker M, Hosoya D, Hullo MF, Kakeshita H, Karamata D, Kasahara Y, Kawamura F, Koga K, Koski P, Kuwana R, Imamura D, Ishimaru M, Ishikawa S, Ishio I, Le Coq D, Masson A, Mauel C, Meima R, Mellado RP, Moir A, Moriya S, Nagakawa E, Nanamiya H, Nakai S, Nygaard P, Ogura M, Ohanan T, O'Reilly M, O'Rourke M, Pragai Z, Pooley HM, Rapoport G, Rawlins JP, Rivas LA, Rivolta C, Sadaie A, Sadaie Y, Sarvas M, Sato T, Saxild HH, Scanlan E, Schumann W, Seegers JF, Sekiguchi J, Sekowska A, Seror SJ, Simon M, Stragier P, Studer R, Takamatsu H, Tanaka T, Takeuchi M, Thomaides HB, Vagner V, van Dijl JM, Watabe K, Wipat A, Yamamoto H, Yamamoto M, Yamamoto Y, Yamane K, Yata K, Yoshida K, Yoshikawa H, Zuber U, Ogasawara N. 2003. Essential Bacillus subtilis genes. Proc Natl Acad Sci U S A. 100:4678-4683.

Kolisnychenko V, Plunkett G 3rd, Herring CD, Fehér T, Pósfai J, Blattner FR, Pósfai G. 2002. Engineering a reduced Escherichia coli genome. Genome Res. 12:640-647.

Kovarik, A., Matzke, M. A., Matzke, A. J. & Koulakova, B. 2001. Transposition of IS*10* from the host *Escherichia coli* genome to a plasmid may lead to cloning artefacts. Mol. Genet. Genomics 266:216-222.

Krasin F, Hutchinson F. 1981. Repair of DNA double-strand breaks in Escherichia coli cells requires synthesis of proteins that can be induced by UV light. Proc Natl Acad Sci U S A. 78:3450-3453.

Kricker M, Hall BG. 1984. Directed evolution of cellobiose utilization in Escherichia coli K12. Mol Biol Evol. 1:171-182.

Kricker M, Hall BG. 1987. Biochemical genetics of the cryptic gene system for cellobiose utilization in Escherichia coli K12. Genetics. 115:419-429.

Kushner SR, Nagaishi H, Templin A, Clark AJ. 1971. Genetic recombination in Escherichia coli: the role of exonuclease I. Proc Natl Acad Sci U S A. 68:824-827.

Lambert MP, Neuhaus FC. 1972. Mechanism of D-cycloserine action: alanine racemase from Escherichia coli W. J Bacteriol. 110:978-987.

LaRossa RA. 1996. Mutant selections linking physiology, inhibitors, and genotypes. In: *Escherichia coli and Salmonella,* Neidhardt, FC, Curtiss, RIII, Ingraham, JL, Lin, ECC, Low, KB, Magasanik, B, Reznikoff, WS, Riley, M, Schaechter, M and Umbarger, HE (eds.), ASM Press, Washington DC, USA, Vol. 2., pp. 2527-2587.

Lederberg J. 1951. Genetic studies with bacteria. In *Genetics in the 20th century*. Dunn LC (ed), Macmillan, New York, USA, pp. 263-289.

Lewin, B. 1997. *Genes VI*. Oxford University Press; Oxford, England.

Lin RJ, Capage M, Hill CW. 1984. A repetitive DNA sequence, rhs, responsible for duplications within the Escherichia coli K-12 chromosome. J Mol Biol. 177:1-18.

Link AJ, Phillips D, and Church GM. 1997. Methods for generating precise deletions and insertions in the genome of wild-type Escherichia coli: application to open reading frame characterization. J Bacteriol. 179:6228-6237

Liu D, Reeves PR. 1994. Escherichia coli K12 regains its O antigen. Microbiology. 140:49-57.

Lloyd RG and Low KB. 1996. Homologous Recombination. In *Escherichia coli and Salmonella,* Neidhardt, FC, Curtiss, RIII, Ingraham, JL, Lin, ECC, Low, KB, Magasanik, B, Reznikoff, WS, Riley, M, Schaechter, M and Umbarger, HE (eds.), ASM Press, Washington DC, USA, Vol. 2, p. 2239.

Lukacsovich T, Venetianer P. 1991. Increased mutation rate in bacteria overexpressing a cloned foreign gene. Mutat Res. 264:25-28.

Macnab RM. 1996. Flagella and Motility. In *Escherichia coli and Salmonella,* Neidhardt, FC, Curtiss, RIII, Ingraham, JL, Lin, ECC, Low, KB, Magasanik, B, Reznikoff, WS, Riley, M, Schaechter, M and Umbarger, HE (eds.), ASM Press, Washington DC, USA, Vol. 1, p. 132.

Miller VL, Mekalanos JJ. 1988. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in Vibrio cholerae requires toxR. J Bacteriol. 170:2575-2583.

Miller, JH. 1992. *A short course in bacterial genetics. A laboratory manual and handbook for* Escherichia coli *and related bacteria.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. pp. 271-273.

Minagawa S, Ogasawara H, Kato A, Yamamoto K, Eguchi Y, Oshima T, Mori H, Ishihama A, Utsumi R. 2003. Identification and molecular characterization of the Mg2+ stimulon of Escherichia coli. J Bacteriol. 185:3696-3702.

Monteilhet, C, Perrin, A, Thierry, A, Colleaux, L and Dujon, B. 1990. Purification and characterization of the in vitro activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron. Nucleic Acids Res. 18:1407-1413.

Morschhäuser J, Vetter V, Emody L, Hacker J. 1994. Adhesin regulatory genes within large, unstable DNA regions of pathogenic Escherichia coli: cross-talk between different adhesin gene clusters. Mol Microbiol. 11:555-566.

Murphy KC. 1991. Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of Escherichia coli RecBCD enzyme. J Bacteriol. 173:5808-5821.

Muyrers, JP, Zhang, Y, Testa, G and Stewart, AF. 1999. Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res. 27:1555-1557.

Naas T, Blot M, Fitch WM, Arber W. 1994. Insertion sequence-related genetic variation in resting Escherichia coli K-12. Genetics. 136:721-730.

Nakamura K, Inouye M. 1981. Inactivation of the Serratia marcescens gene for the lipoprotein in Escherichia coli by insertion sequences, IS1 and IS5; sequence analysis of junction points. Mol Gen Genet. 1981;183(1):107-14.

Neidhardt, FC. 1996. The Enteric Bacterial Cell and the Age of Bacteria. In *Escherichia coli* and *Salmonella,* Neidhardt, FC, Curtiss, RIII, Ingraham, JL, Lin, ECC, Low, KB, Magasanik, B, Reznikoff, WS, Riley, M, Schaechter, M and Umbarger, HE (eds.) ASM Press, Washington DC, USA, Vol. 1, pp. 1-3

Noda M, Kawahara Y, Ichikawa A, Matoba Y, Matsuo H, Lee DG, Kumagai T, Sugiyama M. 2004. Self-protection mechanism in D-cycloserine-producing Streptomyces lavendulae. Gene cloning, characterization, and kinetics of its alanine racemase and D-alanyl-D-alanine ligase, which are target enzymes of D-cycloserine. J Biol Chem. 279:46143-46152.

Noirot P, Kolodner RD. 1998. DNA strand invasion promoted by Escherichia coli RecT protein. J Biol Chem. 273:12274-12280.

Ochman H, Gerber AS, Hartl DL. 1988. Genetic applications of an inverse polymerase chain reaction. Genetics. 120:621-623

Oram M, Fisher LM. 1991. 4-Quinolone resistance mutations in the DNA gyrase of Escherichia coli clinical isolates identified by using the polymerase chain reaction. Antimicrob Agents Chemother. 35:387-389.

Orndorff PE, Falkow S. 1985. Nucleotide sequence of pilA, the gene encoding the structural component of type 1 pili in Escherichia coli. J Bacteriol. 162:454-457.

Perna NT, Plunkett G 3rd, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Posfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dimalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC, Welch RA, Blattner FR. 2001. Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. Nature. 409:529-533.

Pollard EC, Fluke DJ, Kazanis D. 1981. Induced radioresistance: an aspect of induced repair. Mol Gen Genet. 184:421-429.

Pósfai G, Kolisnychenko V, Bereczki Z, Blattner FR. 1999. Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome. Nucleic Acids Res. 27:4409-4415.

Pósfai G, Koob MD, Kirkpatrick HA, Blattner FR. 1997. Versatile insertion plasmids for targeted genome manipulations in bacteria: isolation, deletion, and rescue of the pathogenicity island LEE of the Escherichia coli O157:H7 genome. J Bacteriol. 179:4426-4428.

Pósfai G, Plunkett G 3rd, Fehér T, Frisch D, Keil G, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, Harcum SW, Blattner FR. 2006. A reduced genome *Escherichia coli* lacking all transposable elements improves cloning. Submitted to Science

Poteete AR, Fenton AC, Nadkarni A. 2004. Chromosomal duplications and cointegrates generated by the bacteriophage lambda Red system in Escherichia coli K-12. BMC Mol Biol. 5:22.

Pressler U, Staudenmaier H, Zimmermann L, Braun V. 1988. Genetics of the iron dicitrate transport system of Escherichia coli. J Bacteriol. 170:2716-2724.

Reeves PR, Hobbs M, Valvano MA, Skurnik M, Whitfield C, Coplin D, Kido N, Klena J, Maskell D, Raetz CR, Rick PD. 1996. Bacterial polysaccharide synthesis and gene nomenclature. Trends Microbiol. 4:495-503.

Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS. 2000. Parallel evolution of virulence in pathogenic Escherichia coli. Nature. 406:64-67.

Rood JI, Sneddon MK, Morrison JF. 1980. Instability in tyrR strains of plasmids carrying the tyrosine operon: isolation and characterization of plasmid derivatives with insertions or deletions. J Bacteriol. 144:552-559.

Rosche WA, Foster PL. 2000. Determining mutation rates in bacterial populations. Methods 20:4-17

Russell CB, Thaler DS, Dahlquist FW. 1989. Chromosomal transformation of Escherichia coli recD strains with linearized plasmids. J Bacteriol. 171:2609-2613.

Salyers AA, Whitt DD. 2001. Bacterial Genetics III: Horizontal Gene Transfer in Bacteria. In *Microbiology: Diversity, Disease and the Environment*, Fitzgerald Science Press, Bethesda, Maryland, USA pp. 133-154.

Sambrook, J, Fritsch, EF and Maniatis, T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA,

Sarkar S, Ma WT, Sandri GH. 1992. On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. Genetica. 85:173-179.

Sastry SS, Jayaraman R. 1984. Nitrofurantoin-resistant mutants of Escherichia coli: isolation and mapping. Mol Gen Genet. 196:379-380.

Schnetz K, Toloczyki C, Rak B. 1987. Beta-glucoside (bgl) operon of Escherichia coli K-12: Nucleotide sequence, genetic organization, and possible evolutionary relationship to regulatory components of two Bacillus subtilis genes. J Bacteriol. 169:2579-90.

Schnetz K, Rak B. 1988. Regulation of the bgl operon of Escherichia coli by transcriptional antitermination. EMBO J. 7:3271-7.

Schnetz K, Rak B. 1992. IS5: a mobile enhancer of transcription in Escherichia coli. Proc Natl Acad Sci U S A. 89:1244-8.

Schnetz K. 1995. Silencing of Escherichia coli bgl promoter by flanking sequence elements. EMBO J. 14:2545-50.

Schurr MJ, Martin DW, Mudd MH, Deretic V. 1994. Gene cluster controlling conversion to alginate-overproducing phenotype in Pseudomonas aeruginosa: functional analysis in a heterologous host and role in the instability of mucoidy. J Bacteriol. 176:3375-3382.

Sharp PM, Li WH. 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15:1281-1295

Shimmin LC, Vanderwel D, Harkness RE, Currie BR, Galloway CA, Ishiguro EE. 1984. Temperature-sensitive beta-lactam-tolerant mutants of Escherichia coli. J Gen Microbiol. 130:1315-1323.

Stewart FM, Gordon DM, Levin BR. 1990. Fluctuation analysis: the probability distribution of the number of mutants under different conditions. Genetics. 124:175-185.

Sunshine MG, Kelly B. 1971. Extent of host deletions associated with bacteriophage P2-mediated eduction. J Bacteriol. 108:695-704

Suzuki N, Okayama S, Nonaka H, Tsuge Y, Inui M, Yukawa H. 2005. Large-scale engineering of the Corynebacterium glutamicum genome. Appl Environ Microbiol. 71:3369-3372.

Sweasy JB, Witkin EM, Sinha N, Roegner-Maniscalco V. 1990. RecA protein of Escherichia coli has a third essential role in SOS mutator activity. J Bacteriol. 172:3030-3036.

Thomson NR, Parkhill J. 2003. The building blocks of pathogenicity. Trends Microbiol. 11:66-67.

Vinella D, D'Ari R, Jaffe A, Bouloc P. 1992. Penicillin binding protein 2 is dispensable in Escherichia coli when ppGpp synthesis is induced. EMBO J. 1992 11:1493-1501. Erratum in: EMBO J 1992 11:5137.

Wang WP, Myers RL, Chiu IM. 1991. Single primer-mediated polymerase chain reaction: application in cloning of two different 5'-untranslated sequences of acidic fibroblast growth factor mRNA. DNA Cell Biol. 10:771-777.

Wargel RJ, Shadur CA, Neuhaus FC. 1970. Mechanism of D-cycloserine action: transport systems for D-alanine, D-cycloserine, L-alanine, and glycine. J Bacteriol. 103:778-788.

Wargel RJ, Hadur CA, Neuhaus FC. 1971. Mechanism of D-cycloserine action: transport mutants for D-alanine, D-cycloserine, and glycine. J Bacteriol. 105:1028-1035.

Walker, GC. 1996. The SOS response of Escherichia coli. In *Escherichia coli and Salmonella,* Neidhardt, FC, Curtiss, RIII, Ingraham, JL, Lin, ECC, Low, KB, Magasanik, B, Reznikoff, WS, Riley, M, Schaechter, M and Umbarger, HE (eds.), ASM Press, Washington DC, USA, Vol. 1, p. 1409.

Wehrli W. 1983. Rifampin: mechanisms of action and resistance. Rev Infect Dis. 5 Suppl 3:S407-11.

Welch RA, Burland V, Plunkett G 3rd, Redford P, Roesch P, Rasko D, Buckles EL, Liou SR, Boutin A, Hackett J, Stroud D, Mayhew GF, Rose DJ, Zhou S, Schwartz DC, Perna NT, Mobley HL, Donnenberg MS, Blattner FR. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. Proc Natl Acad Sci U S A. 99:17020-17024.

Westers H, Dorenbos R, van Dijl JM, Kabel J, Flanagan T, Devine KM, Jude F, Seror SJ, Beekman AC, Darmon E, Eschevins C, de Jong A, Bron S, Kuipers OP, Albertini AM, Antelmann H, Hecker M, Zamboni N, Sauer U, Bruand C, Ehrlich DS, Alonso JC, Salas M, Quax WJ. 2003. Genome engineering reveals large dispensable regions in Bacillus subtilis. Mol Biol Evol. 20:2076-2090.

Willis DK, Satin LH, Clark AJ. 1985. Mutation-dependent suppression of recB21 recC22 by a region cloned from the Rac prophage of Escherichia coli K-12. J Bacteriol. 162:1166-1172.

Wolfson JS, Hooper DC, McHugh GL, Bozza MA, Swartz MN. 1990. Mutants of Escherichia coli K-12 exhibiting reduced killing by both quinolone and beta-lactam antimicrobial agents. Antimicrob Agents Chemother. 34:1938-1943.

Woodcock, DM, Crowther, PJ, Doherty, J, Jefferson, S, DeCruz, E, Noyer-Weidner, M, Smith, SS, Michael, MZ and Graham, MW. 1989. Quantitative evaluation of Escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucleic Acids Res. 17: 3469-3478.

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. 2000. An efficient recombination system for chromosome engineering in Escherichia coli. Proc Natl Acad Sci U S A. 97:5978-5983.

Yu BJ, Sung BH, Koob MD, Lee CH, Lee JH, Lee WS, Kim MS, Kim SC. 2002. Minimization of the Escherichia coli genome using a Tn5-targeted Cre/loxP excision system. Nat Biotechnol. 20:1018-1023

Zawadzke LE, Bugg TD, Walsh CT. 1991. Existence of two D-alanine:D-alanine ligases in Escherichia coli: cloning and sequencing of the ddlA gene and purification and characterization of the DdlA and DdlB enzymes. Biochemistry. 30:1673-1682

Zhai Y, Saier MH Jr. 2002. The beta-barrel finder (BBF) program, allowing identification of outer membrane beta-barrel proteins encoded within prokaryotic genomes. Protein Sci. 11:2196-2207.

Zhang Y, Buchholz F, Muyrers JP, Stewart AF. 1998. A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet. 20:123-128.

Zucca J, Balassa G, Sousa JC, Silva MT. 1979. Genetic instability of sporulation-associated characters in a Bacillus subtilis mutant: relationship between sporulation, segregation and the synthesis of extracellular enzymes (kinetic studies). Gen Microbiol. 112:283-296