Vezető kutató: Fehér Tamás (Principal investigator) OTKA ID: 72719 Típus: PD (Type) 2007/1

Final Report

Mechanisms of mutation in *Escherichia coli*: analysis, engineering and applications

Summary

Spontaneous mutations arising in prokaryotes are the basis of several problematic phenomena, e.g. antibiotic resistance of pathogens, reduced production yield of industrial producer strains, or sequencing errors due to alteration of cloned sequences. We have previously narrowed the mutational repertoire of *Escherichia coli* by removal of all mobile DNA elements (insertion sequences, prophages, recombination hot spots) in a large scale genome-reduction effort (Pósfai *et al.*, 2006). In this work, we demonstrate that the lack of transposable elements aids in the cloning and expression of a heterologous protein, which is otherwise rapidly inactivated by insertion sequences (ISes) disrupting its gene with a high frequency (Umenhoffer *et al.*, 2010). To prevent recontamination of the IS-free bacterial genome by transposable elements during generalized phage transduction, we have removed the resident *IS1* element from the genome of phage P1vir, commonly used for strain engineering (Fehér *et al.*, 2011a). We further decreased the mutation-generating potential of *E. coli* by removal of the genes of three error-prone DNA polymerases, thereby significantly reducing the rate of point mutations. In addition, we have developed a method to measure the selective advantage conferred by a single IS element residing in the bacterial genome, compared to an IS-free host. This system uses competition experiments of fluorescently labeled IS-containing and IS-free cells, and was developed to systematically map the environments that activate IS element transposition. We used this technique to "replay evolution", and demonstrate that the historic invasion and expansion of IS-elements in enterobacterial genomes was most probably fueled by the selfish (replicative) nature of the elements, and not by the selective advantage they conferred to their hosts (Fehér *et al.*, 2011b).

Introduction

Mutations are the permanent and heritable changes of the genetic material of living organisms. Mutations comprise single nucleotide alterations in the nucleotide sequence (point mutations), loss of longer stretches of DNA (deletions), integration of novel DNA sequences (insertions) or virtually any other type of rearrangement (inversions, duplications, etc.). Importance of the analysis of mutations in prokaryotes is well demonstrated by the fact that pathogens often evade therapeutic antibiotics with the help of mutations emerging within the gene encoding the drug target (Abigail and Whitt, 1994). Besides investigating mutation-generating functions, their modulation is also of interest: increasing the mutation rate, combined with proper selection can lead to the generation of useful phenotypes. Decreasing the rate of mutations is perhaps even more important to conserve such useful phenotypes under conditions where they pose a significant burden to the cell: e.g. during overproduction of proteins or certain metabolites for industrial purposes, which results in growth retardation. Mutations reducing production yield accelerate growth, thereby allowing the mutants to overgrow the population within the fermentor, and eventually lead to the cease of production (Madigan et al., 1997). Repeated sterilization and re-inoculation of fermentors puts a significant financial load onto the

biotech industry, which could be circumvented by the use of strains with reduced capacities of mutating their genomes.

Effect of the elimination of transposable elements

Mobile genetic elements (insertion sequences, prophages, transposons, integrons, recombination hot spots, etc.) are autonomous segments of genome that encode the functions necessary for facilitating their own transposition to new loci. Depending on the genetic and physiological context, their contribution to gene inactivation ranges from 3.9% (Halliday and Glickman, 1991) to 98% (Hall, 1998). In addition to insertional inactivation, mobile elements are also capable of activating gene expression and enhancing deletions, inversions or duplications by providing substrates for homologous recombination (Zeyl and Bell, 1996). In a previous work, we have reported the elimination of all transposable elements from the genome of *E. coli*. The resulting reduced genome strain, MDS42 displayed a genomic mutation rate 25 % lower than the maternal MG1655 cell (Pósfai *et al.*, 2006). Although this reduction in mutation rate may seem negligible, numerous examples can be found when IS elements predominate the mutations inactivating a cloned sequence (Blumenthal et al., 1985; Fernandez et al., 1986; Muller et al., 1989; Rawat et al., 2009). We chose one such case to investigate the background of IS-predominance and to test the efficiency of cloning in the IS-free MDS42 strain.

CTXVP60 is a gene developed for vaccination purposes by fusing the genes of Cholera Toxin (CTX) and Rabbit Haemorrhagic Virus Viral Protein 60 (VP60). Propagation of the pCTXVP60 plasmid in any conventional *E. coli* host resulted in severe growth retardation, marked by small colony size on agar plates. In a short period of time, however, an increasing number of large, rapidly growing colonies appeared on these plates, the 92% of which harbored CTXVP60 genes interrupted by an IS element. In liquid medium, these rapidly growing mutants quickly took over the population, as explained above. Transcription of the gene was necessary to elicit the toxic, growth-retarding effect, as verified by an inducible construct. Its overexpression caused elongation of cells, and appearance of multiple nucleoid structures, as visualized using confocal laser scanning microscopy. We hypothesized, that the large number of rare Arg codons within the CTXVP60 gene was responsible for the burden posed on the cells. A codon optimized version lacking rare Arg codons (CTXVP60opt) relieved the cells from the burden, however, increasing the number of rare Arg codons (CTXVP60deopt) had a similar effect, too. Introducing a frameshift mutation in the 5' sequence of the gene did not relive the growth retardation, indicating that it is not the encoded CTXVP60 protein *per se*, that is responsible for the toxic effect. Finally, a closer analysis of the sequence revealed that in a different reading frame, a 238 amino acid long leucin-rich peptide (ORF238) is encoded by the genetic construct. Isolated expression of ORF238 elicited the same growth retardation and cell morphological changes as the full length CTXVP60. Its toxic effect is presumably a result of its hydrophobic nature, possibly causing membrane perturbations. Both CTXVP60opt and CTXVP60deopt eliminated ORF238, explaining the loss of toxicity.

Importantly, when using an IS-free host cell (MDS42) for propagation of pCTXVP60, colony morphology was much more uniform, and the frequency of appearance of rapidly growing, mutant colonies was reduced approximately 50-fold, despite the fact that the toxic effect was still observable. pCTXVP60 plasmids of correct sequence were readily isolated from the IS-free host. Reintroducing a single copy of *IS1* into MDS42 dramatically elevated the inactivation-frequency of the CTXVP60 gene, indicated by the higher frequency of clones losing their growth retardation (**Figure 1**). This example demonstrates, that reducing the evolutionary potential of a

bacterial cell by eliminating a mechanism of mutation can aid in cloning and overexpression of a foreign gene.

To examine, whether this case is an exception, or an example of a more general phenomenon, we examined the raw sequences of shotgun libraries. These libraries were constructed in *E. coli* hosts by cloning multitudes of genomic segments originating from various species for the aim of whole genome sequencing. By performing nucleotide level BLAST searches on the raw sequence reads using stringent parameters, we found 109 cases of IS elements originating from *E. coli* interrupting a cloned genomic sequence. This parameter is approximately two orders of a magnitude higher than one would expect by chance, indicating that such IS-mediated inactivation of ectopic genes inhibiting the growth of the *E. coli* host cannot be regarded as irrelevant and rare events, but might happen in many cases, leading to the selection of altered clones. This further underlines the importance of minimizing the evolutionary potential of cells used in molecular biology and biotechnology.

This work was published in Microbial Cell Factories (Umenhoffer *et al.*, 2010). Kinga Umenhoffer made a major contribution to the experiments belonging to this topic.

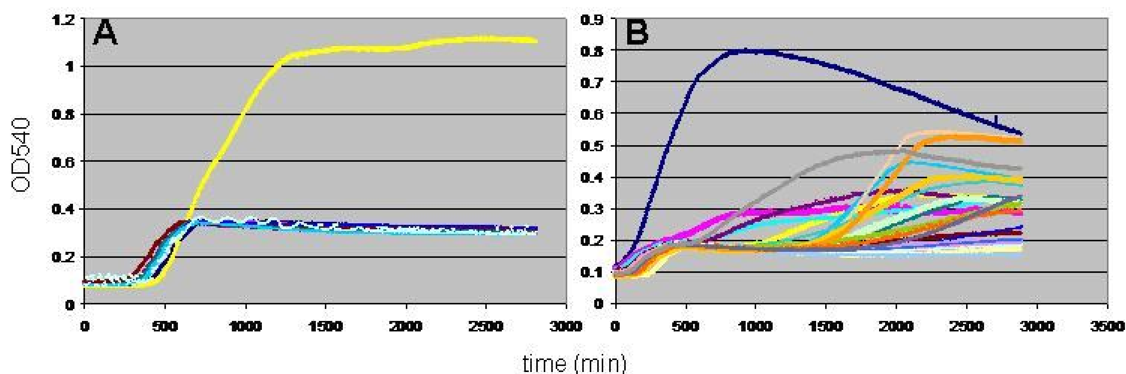

Figure 1. Growth curves of cultures harboring the pCTXVP60 plasmid. (A) 10 parallel cultures of MDS42; (B) 24 parallel cultures of MDS42*Yea*:*IS1*

Reducing the frequency of point mutations

To further reduce the mutational repertoire of MDS42, we set out to delete the genes of error prone DNA polymerases. Such polymerases, encoded by the *polB*, *dinB* and *umuDC* genes of *E. coli*, are induced during the SOS response, the heat-shock response, the stringent response and general stress response (Foster, 2005). They are responsible for cross-lesion synthesis upon severe DNA damage, and for the generation of diversity that selection can act on to evade the effects of stressors threatening the population. Due to their continuous presence in the cell, and their low-fidelity nature, we hypothesized that the elimination of their genes would lead to further reduction of the spontaneous mutation rate. Our hypothesis proved correct: removing either one of the three genes significantly lowered the rate of point mutations, measured using the method developed earlier in our laboratory (Fehér *et al.*, 2006)(**Figure 2**). In addition, combining the *polB* and *dinB* deletions had an additive effect. Interestingly, combining *umuDC* with other deletions failed to decrease the mutation rate further, possibly indicating an interaction between these polymerases.

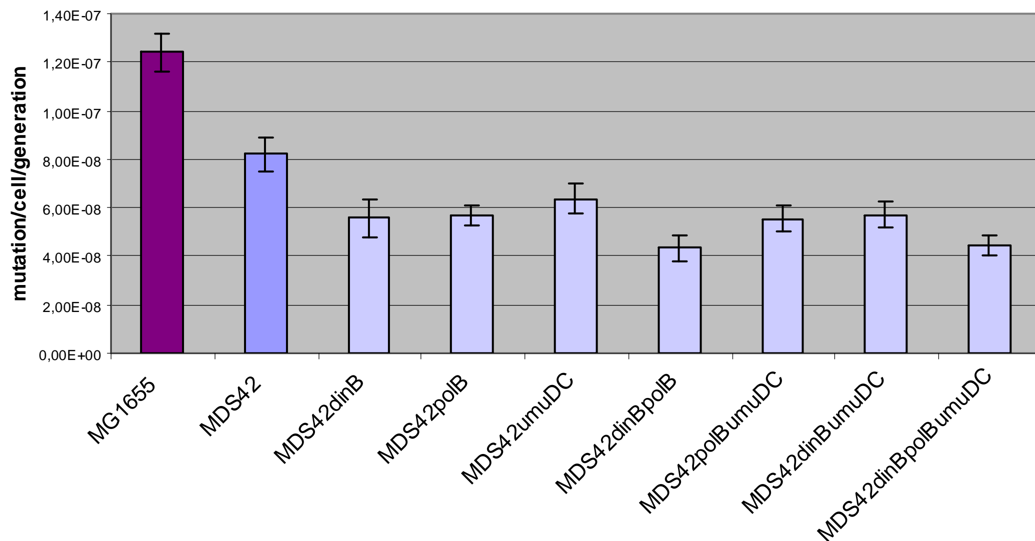

Figure 2. Mutation rates of various strains inferred from the frequency of resistance to D-cycloserine

In our following experiments, we concentrated on the triple mutant (MDS42*polBdinBumuDC*), which had a spontaneous mutation rate of one third of the wild type MG1655. To test whether this ratio changes under conditions of stress, we challenged the cells with mitomycinC, or introduced the genes of GFP or ORF238 on a plasmid and induced their expression (**Figure 3**). The expression of ORF238 had the most dramatic effect: it elevated the mutation rate several fold in every cell type (BL21, MG1655, MDS42 and MDS42RecA), except in the triple mutant, where it elicited a mere 50% increase. MitomycinC, a DNA damaging agent had a similar effect, but BL21 and MDS42RecA failed to grow in its presence. The expression of GFP increased the mutation rate only in MG1655. Strikingly, neither of the stressors led to the increase of the mutation rate of the triple mutant strain by more than 50%. We could therefore conclude that elimination of the error prone DNA polymerases reduces the spontaneous mutation rate, and their loss allows the cell to dampen the mutation rate elevation induced by certain stressors, e.g. overexpression of heterologous proteins. In this part of the project, the experiments were carried out by Bálint Csörgő, and the manuscript being prepared on this topic will contribute to his PhD thesis.

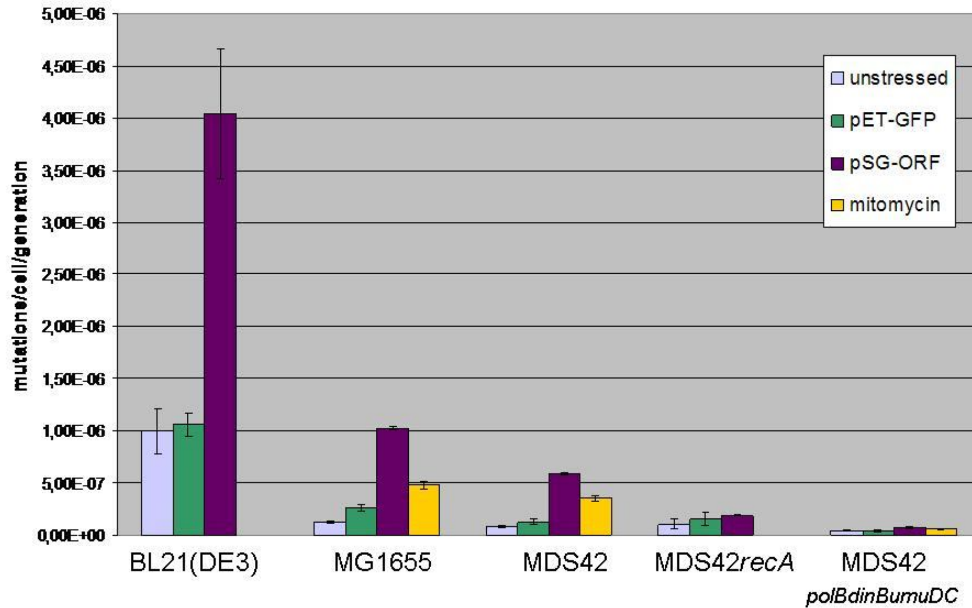

Figure 3. The effect of protein overexpression or mitomycinC on the mutation rate of various *E. coli* strains

Constructing an IS-free P1 phage for contamination-free genome engineering

We have shown in the experiments above, that re-introduction of an IS element into the IS-free MDS42 strain can dramatically accelerate the inactivation rate of a toxic cloned gene. It is thus important to avoid the reentry of transposable elements in the course of genome engineering carried out in MDS42 and its derivates. Phage P1 is one of the most widely used tools for general transduction in enterobacteria. The two P1 strains sequenced to date both carry single copies of *IS1* and *IS5* in their DNA (Lobocka *et al.*, 2004). This raised the possibility, that the use of P1 might result in the inadvertent transfer of IS sequences into the target genome. This would be undesirable and we decided to avoid it by constructing an IS free P1. As starting point we decided to use P1vir, the hypervirulent, non-lysogenizing version to pursue this goal because it also eliminates the possibility of another type of genome contamination, the production of P1 lysogens in the target. This introduced an additional complication, since phage genome engineering is most easily done in the lysogenic form. Bacteriophage recombineering using electroporated DNA (BRED), offered a possible solution by allowing the modification of phages in the lytic stage (Marinelli *et al.*, 2008). This method, described recently for mycobacteriophage engineering, employs bacterial overexpression of cloned phage recombinases to assemble phage genomes from purified phage DNA, and synthetic DNA fragments of choice.

Using a collection of PCR primers specific for all ten IS-types of *E. coli*, we screened the P1vir genome for the presence of IS elements. Only *IS1*-specific primers gave a PCR product, and we found a copy of this element (designated $IS1_{\mathrm{P1vir}}$) in the same position as in the sequenced P1 genomes. To test the ability of $IS1_{\mathrm{P1vir}}$ for transposition, we cloned it in a plasmid, introduced it into the IS-free MDS42, and measured the frequency of arising bgl^{+} mutants on salicin minimal medium. By sequencing the *bgl* gene of a mutant, we confirmed the transposition of $IS1_{\mathrm{P1vir}}$ into

the *bgl* operon, and determined the frequency of transposition to be similar to other *IS1* elements, despite the single nucleotide polymorphisms of $IS1_{P1vir}$.

We then removed $IS1_{P1vir}$ from the genome of P1vir by applying BRED for the first time in a coliphage. This was also the first demonstration of using λ Red recombinases for this purpose. The resulting phage, P1virΔIS was negative for all IS elements targeted by our PCR test. In a rapid phenotypic characterization, we found that P1virΔIS is indiscernible from P1vir as far as plaque morphology, phage titer, or capacity for generalized transduction is concerned. By performing single-step phage growth experiments, we demonstrated that the latency time, rise period and burst size of P1virΔIS is also highly similar to that of P1vir. We also compared the relative fitness of the two phages in head-to-head competition experiments by growing their mixture on *E. coli* cells. We found no significant changes in their ratio over 40 generations, indicating no fitness effect of our engineering process.

Overall, we constructed an IS-free P1 phage to avoid recontamination of MDS42 during generalized transduction. More importantly for the general public, we demonstrated the use of the λ Red recombinases for BRED, which could aid the rapid engineering of hypervirulent phages targeting Gram- bacteria in phage therapy. The manuscript describing this work has been returned from Microbial Biotechnology for minor revision (Fehér *et al.*, 2011a). (The journal will receive its first impact factor in the summer of 2011.)

Studying the selective advantage conferred by IS elements

The fact that MDS42 lacks all types of IS elements from its genome allows us to reintroduce any IS element of choice to study it in an isolated fashion. We have constructed MDS42*Yea*:*IS1* for this purpose, which contains a single copy of *IS1* integrated in the genome. Exposing MDS42*Yea*:*IS1* to various environments and measuring the rate of IS transposition (e.g. with the *bgl* test mentioned above) could result in a list of conditions that induce *IS1* transposition. Since such tests are tedious and time consuming, we decided to precede them with a medium-throughput screening technique. This technique, developed in our lab, is based on the head-to-head competition of IS-free and IS-containing *E. coli* cells, growing in 1:1 mixtures. For rapid and automatic detection of competition outcome, both competing cell lines were labeled with inducible fluorescent markers to make MDS42*Yea*:*IS1*YFP (expressing yellow fluorescent protein) and MDS42CFP (expressing cyan fluorescent protein) (**Figure 4**), and competition was carried out using a large number of parallel cultures, grown in 96-well plates. Control competition in LB medium at 37°C demonstrated no selective advantage of the *IS1*-carrying strain. We hypothesized, that if under any condition, MDS42*Yea*:*IS1*YFP gains a selective advantage over MDS42CFP, it indicates two possibilities: i) IS-generated mutations are selected by that specific condition, ii) that specific condition induces IS transposition. The latter cause can then be verified by measuring the transposition rate, as mentioned above.

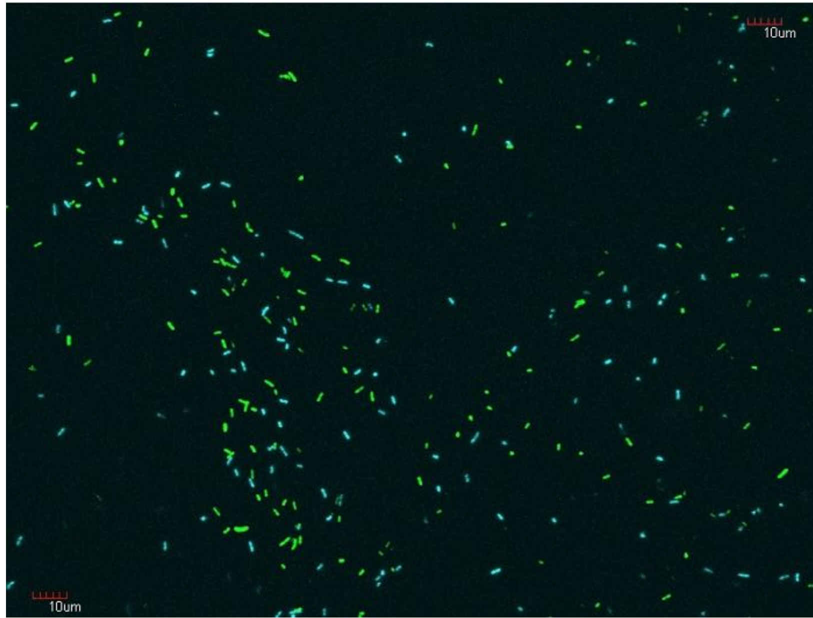

Figure 4. Fluorescent microscopic image of a mixture of *E. coli* cells expressing the YFP or CFP protein, respectively.

Cataloguing the transposition-inducing environments started with sublethal concentrations of various antibiotics. We found that Streptomycin, at 3.1 μg/ml significantly increases the selective advantage of MDS42*Yea*:*IS1*YFP over MDS42CFP. Generation of the list of conditions that induce *IS1* transposition is still ongoing. However, the technique itself allowed us to study another type of question, one that is sometimes referred to as “the holy grail” of evolutionary biology. Namely, was the invasion of *IS1* elements into enterobacterial genomes a result of the selective advantage that they provided to their hosts, or simply a result of their “selfish”, replicative nature? Our system permitted us to replay this evolutionary scenario. As an initial experiment, we carried out a positive control test with our bacterial competition system in salicin-minimal medium. This medium can only be broken down by the activation of the *bgl* operon, which is known to highly favor insertion-mutants over any other type of mutation (Hall, 1998). MDS42*Yea*:*IS1*YFP displayed a high, 29:1 selective advantage over MDS42CFP, as seen from results of parallel competition experiments (**Figure 5**). This is an example of IS elements

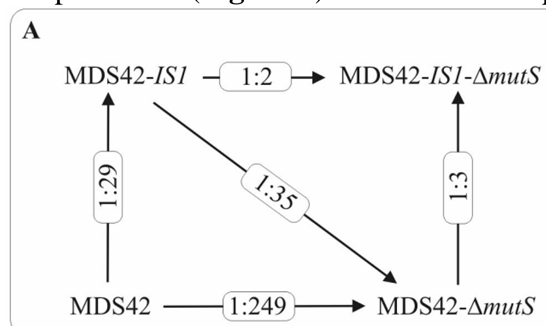

Figure 5. Overview of competition experiment results

fixing in a population as a result of the selective advantage that they confer to their hosts. However, drawing a general conclusion from this one example would be a superficial approach. Instead, we tested whether a *ΔmutS* mutator allele, which generates point mutants at a 100-fold

elevated rate, confers advantage to its host in the same environment, over a non-mutator, IS-free counterpart. The result was even more dramatic (**Figure 5**): MDS42*ΔmutS*CFP defeated MDS42YFP in 249 cases out of 250. Furthermore, MDS42*ΔmutS*CFP outgrew MDS42*IS1*YFP with a ratio of 45 to 1, despite the fact that salicin-minimal medium is an environment that highly favors IS-generated mutants. This brought us to our first conclusion: since mutator cells are likely to be present in any large population, they are likely to be the first ones to adapt to environmental changes, even if insertion-alleles are favored over point mutants. In other words, the invasion and spread of *IS1* in enterobacterial genomes was more of a stochastic event (e.g. an outcome of evolutionary drifts in population bottlenecks), and a consequence of their replicative nature, rather than a result of the selective advantage they provided to their hosts.

In addition, we observed that the selective advantage conferred by a *ΔmutS* mutator allele is lower, if an IS element is already fixed in the population (**Figure 5**; 2:1 vs. 249:1), and this is true the other way around, as well. We refer to this phenomenon, where the presence of one mutator allele inhibits the spread of another within a population, as the "evolutionary conflict theory". According to our hypothesis, it is caused by the saturation of the advantageous effect of increasing the mutation rate.

This work was done in collaboration with the Evolutionary Systems Biology Group of Csaba Pál. We summarized these results in a manuscript which is currently under review at Molecular Biology and Evolution (Fehér *et al.*, 2011b).

References

Abigail, S., and Whitt, D. (1994) *Bacterial Pathogenesis: A molecular approach.* : American Society for Microbiology, Washington, DC, USA

Blumenthal, R.M., Gregory, S.A., and Cooperider, J.S. (1985) Cloning of a Restriction-Modification System from Proteus vulgaris and its Use in Analyzing a Methylase-Sensitive Phenotype in *Escherichia coli*. *Journal of Bacteriology* **164**: 501-509.

Fehér, T., Karcagi, I., Blattner, F.R., and Pósfai, G. (2011a) Bacteriophage Recombineering in the Lytic State Using the Lambda Red Recombinases. *Microbial Biotechnology* **Under review**.

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

Fehér, T., Bogos, B., Méhi, O., Fekete, G., Csörgő, B., Kovács, K. et al. (2011b) Evidence for competition between bacterial agents of evolvability. *Molecular Biology and Evolution* **Under review**.

Fernandez, C., Larhammar, D., Servenius, B., Rask, L., and Peterson, P.A. (1986) Spontaneous Insertions Into Cosmid Vector - A Warning. *Gene* **42**: 215-219.

Foster, P.L. (2005) Stress responses and genetic variation in bacteria. *Mutat Res* **569**: 3-11.

Hall, B.G. (1998) Activation of the bgl operon by adaptive mutation. *Molecular Biology and Evolution* **15**: 1-5.

Halliday, J.A., and Glickman, B.W. (1991) Mechanisms of spontaneous mutation in DNA repair-proficient Escherichia coli. *Mutat Res* **250**: 55-71.

Lobocka, M.B., Rose, D.J., Plunkett, G., 3rd, Rusin, M., Samojedny, A., Lehnherr, H. et al. (2004) Genome of bacteriophage P1. *J Bacteriol* **186**: 7032-7068.

Madigan, M., Martinko, J., and Parker, J. (1997) *Brock biology of microorganisms*: Prentice Hall International, London, UK

Marinelli, L.J., Piuri, M., Swigonova, Z., Balachandran, A., Oldfield, L.M., van Kessel, J.C., and Hatfull, G.F. (2008) BRED: a simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. *PLoS One* **3**: e3957.

Muller, J., Reinert, H., and Malke, H. (1989) Streptokinase mutations relieving Escherichia coli K-12 (prlA4) of detriments caused by the wild-type skc gene. *J Bacteriol* **171**: 2202-2208.

Pósfai, G., Plunkett, G., 3rd, Fehér, T., Frisch, D., Keil, G.M., Umenhoffer, K. et al. (2006) Emergent properties of reduced-genome Escherichia coli. *Science* **312**: 1044-1046.

Rawat, P., Kumar, S., Pental, D., and Burma, P.K. (2009) Inactivation of a transgene due to transposition of insertion sequence (IS136) of Agrobacterium tumefaciens. *J Biosci* **34**: 199-202.

Umenhoffer, K., Feher, T., Baliko, G., Ayaydin, F., Posfai, J., Blattner, F.R., and Posfai, G. (2010) Reduced evolvability of Escherichia coli MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microb Cell Fact* **9**: 38.

Zeyl, C., and Bell, G. (1996) Symbiotic DNA in eukaryotic genomes. *Trends in Ecology & Evolution* **11**: 10-15.