

The translation of the original article

Glazunova O.A., Shavkunov K.S., Tutukina M.N., Panyukov V.V., Ozoline O.N. *Mathematical Biology and Bioinformatics*. 2016. V. 11. № 2. P. 394–405. doi: [10.17537/2016.11.394](https://doi.org/10.17537/2016.11.394)

===== TRANSLATIONS OF PUBLISHED ARTICLES =====

UDC: 577.218:579.253.43

Integration of foreign genetic material provokes local mutagenesis in the recipient genome

**Glazunova O.A.¹, Shavkunov K.S.¹, Tutukina M.N.¹, Panyukov V.V.²,
Ozoline O.N.*¹**

¹142290, *Institute of Cell Biophysics of Russian Academy of Sciences, Pushchino, Moscow Region, Russian Federation*

²142290, *Institute of Mathematical Problems of Biology – the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, Moscow Region, Russian Federation*

Abstract. Exchange of genetic material is a key factor of evolution in bacteria, though the mechanisms of integration of horizontally acquired genes into the regulatory networks of the new host remain practically uninvestigated. The present study is devoted to analyzing this adaptation and, in particular, involvement of the bacterial transcription machinery in this process. Therefore, two foreign genes in the genome of *Escherichia coli* K12 MG1655 were substituted with their copies taken from the genomes of presumed donor bacterium, and long-term growth of mutant and control cultures was initiated. Following 2000 (for gene *sfmA*) and 4000 (for gene *yhZ*) generations, the modified and native genomic regions were amplified and used for population sequencing. Spontaneous mutations were analyzed, and it was found that substitutions of G/C- to A/T-pairs occurred more frequently in the modified regions as compared to non-modified, whereas A/T → G/C substitutions were rarer. This implies that the host genome responds to integration of foreign genetic material with an adaptive reaction aimed to enrich the modified region with A/T-pairs. In the course of long-term evolution, it may result either in “silencing” of an adverse gene with the involvement of H-NS – a specific suppressor of foreign genes, or in the creation of a suitable promoter within a *promoter island* for adequate expression of useful gene.

Key words: *bacterial evolution, horizontal gene transfer, single nucleotide polymorphism, promoter islands.*

INTRODUCTION

A specific feature of evolution in bacteria resides in the ability of prokaryotes to acquire and assimilate genes from other organisms [1, 2]. Such exchange of genetic information is called horizontal transfer, in contrast to heritable “vertical” transfer of genetic information to offspring cells. Most frequently horizontal transfer occurs as a result of conjugation in combination with replication, or transduction; however, even simple uptake of the free DNA fragments (transformation) can result in their incorporation into the cell genome [3–5]. Random DNA fragments containing coding and non-coding nucleotide sequences, parts of genes or gene clusters can be subject to horizontal transfer. The length of transferred

*ozoline@rambler.ru

fragments may vary from several to several thousand base pairs [6]. Due to the continuous process of gene acquisition and loss, comparison of any pair of genomes from the same species often yields significant variations [7].

Genes acquired by chance can be distinguished from host genes due to peculiarities in their nucleotide composition, differences in codon usage frequency and phylogenetic analysis. Such genome-wide data assessment testified that a significant portion of functionally relevant genes were obtained by contemporary bacteria specifically by assimilation of alien genes [8, 9], and their acquisition in some cases promoted formation of novel species [10].

The mechanism of foreign gene assimilation in the new genome is less clear. Nevertheless, genomic areas harboring alien genes are known to be enriched with promoter-like sequences [11], and many of them include unusual regions called *promoter islands* [12–16]. Such sequences ranging from 300 to 1100 bp in length were discovered in a search for potential promoters in the *E. coli* genome [12], but drew special attention owing to their unusual functional properties. Using ChIP-on-chip data obtained in [17, 18], it was established that the *islands* efficiently interact with RNA polymerase, and their ability to form transcriptionally competent complexes was confirmed experimentally [12]. Still, RNA synthesis from *island* promoters is commonly limited by formation of short (9–14 nt) products ([13, 16], according to the results of [19]). In other words, *island*-associated transcription complexes perform abortive synthesis, saving substrate resources for the cell. Apparently, histone-like protein H-NS provides the highest contribution to this suppression, since according to [20] it covers up to 90 % of their total length [13], and deletion of *hns* activates the synthesis of RNAs from all the studied *islands* [14]. Being a specific inhibitor of horizontally transferred genes [21, 22], H-NS can thus utilize A/T-rich *island* sequences to suppress the synthesis of adverse products in the cell.

According to various estimations, the number of alien genes in *E. coli* makes from 10 to 30 % [11, 23–25], and the lower limit of this range corresponds to the number of *promoter islands* identified in the *E. coli* genome. Thus, 75 of 78 islands composed of promoters with σ^{70} -binding modules [12] and 370 of 434 islands with mixed σ -specificity [16] were located near genes obtained by *E. coli* in the result of horizontal transfer. Since these genes possess conventional transcription signals in the genomes of potential donors, it was presumed that the formation of promoter-like sites is an adaptive reaction of the genome to the introduction of foreign genetic material. On the one hand, the developed *island* can “switch off” the expression of unwanted genes by the ability to effectively bind H-NS repressor, while on the other hand it can also create a promoter for integration of the novel gene into the host’s regulatory networks.

To confirm such a masterful combination of biological functions, one needs to prove that the transfer of an “adverse” gene provokes evolutionary process aimed to enrich the gene itself and its genetic environment with A/T-pairs, which are necessary for effective binding of both the H-NS suppressor and RNA polymerase. The present study is devoted to verification of this hypothesis. Two genes, apparently originated in the *E. coli* genome from *Salmonella*, were selected as model objects. They are associated with promoter islands and exhibit virtually no transcription activity. Therefore, they can be deemed “unwanted” for *E. coli*. These genes were removed from the genome and replaced by the orthologous coding sequences from *Salmonella*, having normal promoters. In May 2015 we initiated chronic growth of the initial and mutant strains, with daily culture passaging. Following 2000 generations in case of one recombinant strain, and 4000 generations for the other one, the nucleotide sequences of the insertions were subjected to populational sequencing, the results of which are discussed in the present work.

MATERIALS AND METHODS

Genomes used in the study

The genome sequences and gene maps of *Escherichia coli* K12 MG1655 and *Salmonella enterica* Typhimurium 14028S were taken from NCBI GenBank (NC_000913.3 and NC_016856.1, respectively) and used for analysis.

Obtaining recombinant *E. coli* strains

DNA fragments for recombination were amplified using *S. enterica* Typhimurium 14028S DNA (kindly procured by K. Neuhaus, Technical University of Munich). The fragment associated with *sfmA* gene was obtained with primers F1 and R1, while primers F2 and R2 were used for the *ydhZ*-associated fragment (Table 1). Amplification program was the following: predenaturation at 95 °C – 3 min; 30 cycles of: 95 °C – 30 s, 55/54 °C – 30 c (for *sfmA* and *ydhZ*, respectively), 72 °C – 1.5 min; complete synthesis 72 °C – 2 min. Polymerase chain reaction (PCR) was run on Biometra T-1 Thermocycler (Biometra, Germany) according to the protocol of reagent manufacturers (Evrogen and Alpha Enzyme, Russia). Recombination was done following “gene doctoring” protocol proposed in [26]. Kanamycin resistance gene introduced in the course of recombination was not removed from the genome to preserve the possibility of purifying selection in case of culture contamination with external flora. The insertions were sequenced to confirm the absence of mutations.

Table 1. Primers used in the study

№	Name	Primer sequence*	Application
1	F1	5'-GCCCGAATTCGCGCCATGATTAACCAAAAAGA-3'	Synthesis of <i>sfmA</i> from the genome of <i>S. enterica</i>
2	R1	5'-GCCTAAGCTTATCATCTGCTATGTTCTCTA-3'	
3	F2	5'-TAGAGAATTCTGTTATATTCTGTTG-3'	Synthesis of <i>ydhZ</i> from the genome of <i>S. enterica</i>
4	R2	5'-CCCTAAGCTTCGCTAACGGC-3'	
5	F3	5'-CTGTCAGGTTTCGTTTCAG-3'	Synthesis of native and recombinant <i>sfmA</i> region
6	R3	5'-CATCATAGTTAATACGCCTTGTG-3'	
7	F4	5'-TGTGTGCTTGTGTCTCAGC-3'	Synthesis of recombinant <i>ydhZ</i> region
8	R4	5'-CAGTCATAGCCGAATAGCCT-3'	
9	F5	5'-TCGTTGAGTTCTACAGTCCA-3'	Synthesis of native <i>ydhZ</i> region
10	R5	5'-CTAATACGCTGCTGACTCTG-3'	
11	F6	5'-AAAAGAATTCAACTTGTTAGACAATAAAT TGTTATCAAAGATGACTTCTGTTATATTCTGTTG-3'	Amplification of <i>ydhZ</i> -containing fragment for homologous recombination (Fig. 1)
12	R6	5'-CTAACTCGAGAGTGTATGAGAAAAGATTA ACGCCGATAAACAATATCCTCCTTAGTCC-3'	
13	<i>ydhZ</i> _forw	5'-AAAAGAGGATGAACTGTACCG3'	Expression analysis
14	<i>ydhZ</i> _rev	5'-ACGCCTGCGATAACAATAT-3'	
15	<i>sfmA</i> _forw	5'-TGGCGGTACAATCACTT-3'	
16	<i>sfmA</i> _rev	5'-CAAGAATCTCAATCCCGAC-3'	

*Double underlining in primer sequences denotes restriction sites

Culture media and growth conditions

The conditions for long-term growth were selected using the *E. coli* expression database GenExpDB [27]. According to the available data, both genes are poorly transcribed under nearly all growth conditions, though *ydhZ* expression is even lower under anaerobic growth in minimal medium M9 [28]. Consequently, the cultures of unmodified *E. coli* cells (*E. coli-sfmA_control*) and cells carrying a substitution of the *sfmA* genomic locus for a *Salmonella* ortholog region (*E. coli-sfmA_rec*) were grown under conditions commonly used for *E. coli* culturing (15 ml of 0.5X Luria-Bertani medium in 50 ml flasks). In contrast, bacteria with a substitution of *ydhZ* gene associated with two islands (*E. coli-ydhZ_rec*), as well as a respective control culture (*E. coli-ydhZ_control*), were grown in 25 ml M9 medium supplemented with 0.5 % glucose in 25 ml flasks (moderately anaerobic conditions). The cells

of mutant cultures were grown in the presence of 50 mg/ml kanamycin. Bacterial growth rate was estimated by optical density using Jasco UVIDEK-4 spectrophotometer (Jasco Spectroscopic, Japan) at 600 nm.

qRT-PCR

Total RNA fraction was extracted from 25 ml of continuously growing cultures using TRIzol reagent (Ambion, USA) and treated with DNase I (New England Biolabs, USA) according to the protocols of the manufacturers. The primers used for expression analysis are listed in table 1. cDNA was synthesized in a 25 μ l reaction mixture, containing 2 μ g RNA and 80 U RevertAid reverse transcriptase (Thermo Scientific, Lithuania) according to the manufacturer's protocol. Two microliters of this mixture were taken for PCR in a 25 μ l volume with 5 μ l of qPCRmix-HS SYBR (Evrogen, Russia) using DTLite 4 thermocycler (DNA Technology, Russia). The PCR program included predenaturation at 95 °C (3 min) and 35 cycles of: melting at 95 °C – 30 s, annealing at 54 °C – 40 s, synthesis at 72 °C – 30 s. SYBR Green I fluorescence was registered at the end of each cycle for 15 s. Samples without reverse transcriptase were used as a negative control. Antisense RNA to the *ysaA* gene was used as a reference amplicon [14]. Transcription activity of the recombinant genes was estimated in comparison to that of the assimilated *E. coli* orthologs by $2^{-\Delta\Delta C}$, where C – threshold cycle number registered by the machine on the dynamic curve of amplicon synthesis.

Library preparation and sequencing

Bacterial cultures after 2000 and 4000 generations were grown to stationary phase at 37 °C and harvested by centrifugation. DNA was isolated using Wizard Genomic DNA Purification Kit (Promega, USA) according to the manufacturer's protocol. The analyzed genomic regions were amplified using primer pairs F3-R3 (*E. coli-sfmA_rec* and *sfmA_control*), F4-R4 (*E. coli-ydhZ_rec*) and F5-R5 (*E. coli-ydhZ_control*). Approximately 100 ng genomic DNA was added to PCR mix (Evrogen, Russia). The reaction was as follows: melting at 95 °C – 3 min; 7 cycles: 95 °C – 30 s, 54 °C – 30 s, 72 °C – 5 min; 23 cycles: 95 °C – 30 s, 54 °C – 30 s, 72 °C – 3 min; complete synthesis at 72 °C – 2 min.

Products of 1170 and 975 bp amplified from the two control strains *E. coli-sfmA_control* and *E. coli-ydhZ_control*, respectively, were purified using QIAquick Gel Extraction Kit (QIAGEN, Germany), mixed in equal proportions (200 ng each) and fragmented with Ion Xpress Plus Fragment Library Kit (Life Technologies, USA) to a final length of 200–300 bp. Amplicons from the recombinant strains *E. coli-sfmA_rec* and *ydhZ_rec* had significantly different lengths (2683 and 907 bp, respectively). Therefore, they were fragmented with different incubation times prior to mixing. The obtained fragments were purified using Agencourt AMPure XP Reagent (Life Sciences, USA), ligated to adapters using Ion Fragment Library Kit and again purified with Agencourt AMPure XP Reagent. The quality of fragmentation was assessed on E-Gel SizeSelect Agarose Gels (Thermo Fisher Scientific, USA), and fragments of approximately 330 bp were collected. Following gel purification, the selected fragments were amplified on DTLite 4 as per the program: predenaturation 95 °C – 5 min; 7 cycles: 95 °C – 15 s, 58 °C – 15 s, 70 °C – 1 min. Quantitative normalization of libraries was done using Ion Library Equalizer Kit (Thermo Fisher Scientific, USA), giving ~100 pM concentration of DNA in the samples, as measured on Qubit® 3.0 Fluorometer (Invitrogen, USA).

Emulsion PCR and subsequent enrichment of samples containing DNA was carried out on Ion OneTouch 2 system (Life Technologies, USA). Samples were sequenced using Ion 314™ Chips v2 on Ion Torrent PGM (Life Technologies, USA) according to the manufacturer's protocol.

Data analysis

The derived reads varying in length from 25 to 280/350 nucleotides were quality filtered on Galaxy server (instrument Filter by Quality) [29] with a threshold level of ≥ 20 , corresponding to a 99 % probability that all nucleotides in each read are properly identified. The initial data format was preliminarily converted on Galaxy server using Fastq Groomer instrument.

The filtered reads were fragmented to 25 nt sequences for subsequent analysis. They were mapped on the respective genomic regions in the genomes of *E. coli* K12 MG1655 (control) or recombinant strains using Matcher software [30], which is freely available at <http://www.mathcell.ru/DnaRnaTools/Matcher.zip>. The remaining fragments were analyzed for the presence of substitutions (one substitution per single read) using MisMatcher software (available by request).

RESULTS

Selection of genetic loci for recombination and search for a potential donor bacterium

To avoid improper selection of a model organism, we substituted alien genes located near *promoter islands* in the *E. coli* genome for their orthologs from the genome of a potential donor. To pick out model genetic loci we analyzed all alien genes located near to the 370 *mixed promoter islands* [16]. BLAST (NCBI) was used for a search for orthologs. In 68 cases the genes adjacent to the *islands* appeared to be unique, with no analogues available in other genomes. Orthologs of only one out of two neighboring genes were found for 120 islands.

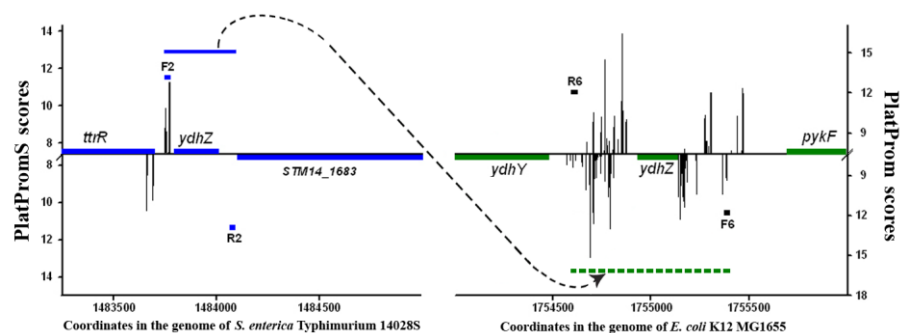


Fig. 1. Genetic background of gene *ydhZ* in the genome of *S. enterica* Typhimurium 14028S (left) and *E. coli* K12 MG1655 (right). Colored rectangles mark gene disposition in both genomes. Vertical bars show potential transcription start points predicted by algorithms optimized for promoter search in the corresponding genomes. F2, R2, F6 and R6 – primers used for recombination (Table 1).

Transfer of such gene, even together with its regulatory region, does not allow complete preservation of the genetic context. However, we picked gene *ydhZ* from this group, which encodes an enzyme of iron assimilation, since it has two adjacently located *island* sequences (Fig. 1). In 107 cases, the *islands* together with their surrounding environment were found in several or many bacterial genomes. They can possibly include the original donors, but their identification required complex phylogenetic analysis, which has not been conducted, yet. Thirteen island-associated genes were identified, along with *E. coli*, in the genomes of a certain genus, but having alternate location on the chromosome, they were also unlikely applicable for a model experiment. Three islands were localized between convergent genes. Their biological role is obscure, thus, such regions were also excluded from consideration. The remaining 59 islands were subjected to a more detailed analysis.

The possible donors included *Salmonella enterica*, *Salmonella bongori*, *Citrobacter koseri*, *Citrobacter rodentium*, *Klebsiella oxytoca*, *Enterobacter cloacae*, *Dickeya dadantii*, *Cronobacter sakazakii*, *Yersinia enterocolitica*, *Edwardsiella tarda* and *Morganella morganii*. In every considered case, a distribution profile of transcription start points predicted

in silico in the genome of the potential donor was obtained. Target genomic DNA of *Salmonella* was scanned with PlatPromS algorithm adjusted to the context of its promoters [31], and the results were compared with the distribution profile of transcription start points in *E. coli* predicted by PlatProm. In other cases, due to the lack of training promoter compilations, the genomes of potential donors and *E. coli* were scanned with a unified algorithm PlatPromU, with threshold levels established to ensure equal levels of significance, as described in [32]. It appeared that in 12 genomes of potential donors the target genes also possess islands, which may be indicative of their horizontal transfer from other genomes. The model object was picked from the remaining 47 candidates according to the following considerations: donor accessibility, similar size of intergenic locus, availability of information on the functional significance of neighboring genes and the dependence of their expression from growth conditions. As a result, we chose a region located between an *E. coli*-born gene *fold* (encodes an enzyme with dehydrogenase and cyclohydrolase activities) and “alien” *sfmA* (protein involved in adhesion), as a second model *promoter island*. Its genetic background is demonstrated in Figure 7 in [13].

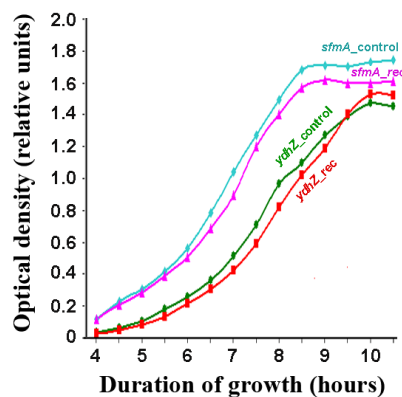


Fig. 2. Growth dynamics of the initial (*E. coli-sfmA_control* and *E. coli-ydhZ_control*) and mutant (*E. coli-sfmA_rec* and *E. coli-ydhZ_rec*) cultures straight after recombination. Later in the course of continuous growth the mutual arrangement of the curves showed no significant alteration. A delay in the growth rate of the mutant strain with a substituted *ydhZ* region and its respective control from the two other cultures is accounted for less suitable culturing conditions – insufficient oxygen and low-nutrient M9 medium instead of LB.

According to NCBI database, *ydhZ* exhibited high similarity (87 %) only with sequences in *Salmonella* genomes, while homology with *sfmA* (83–86 % coverage, 66–68 % identity) was found in *Shigella* bacteria closely related to *E. coli*, in 3 genomes of *Enterobacter cloacae* and in all genomes of *Salmonella enterica*, which is evolutionarily more distant from *E. coli*. Thus, *Salmonella* could have possibly been a donor of both genes for *E. coli*. It appeared that orthologs of *sfmA* (*fimA*) and *ydhZ* in all *Salmonella* bacteria possess usual promoter regions (as on left panel of Fig. 1), while in various *E. coli* strain the same genes harbor promoter islands, which could hence be products of adaptive evolution. Therefore, these two genes were used for recombination. Their replacement in the *E. coli* genome with *S. enterica* orthologs caused hardly any effect on the growth of the bacterial cultures (Fig. 2).

Expression of recombinant genes is altered during long-term culturing

The level of expression of the transferred genes was estimated during the first seven months of growth (Fig. 3). Right after obtaining the mutant clones, RNA samples were isolated from their descendants grown to stationary phase. Subsequently, RNA was isolated from the growing cultures containing heterogeneous cell populations, which accumulated spontaneous mutations both in the experimental and control samples. The transcriptional activity of the recombinant strains was compared to gene expression in simultaneously growing control samples.

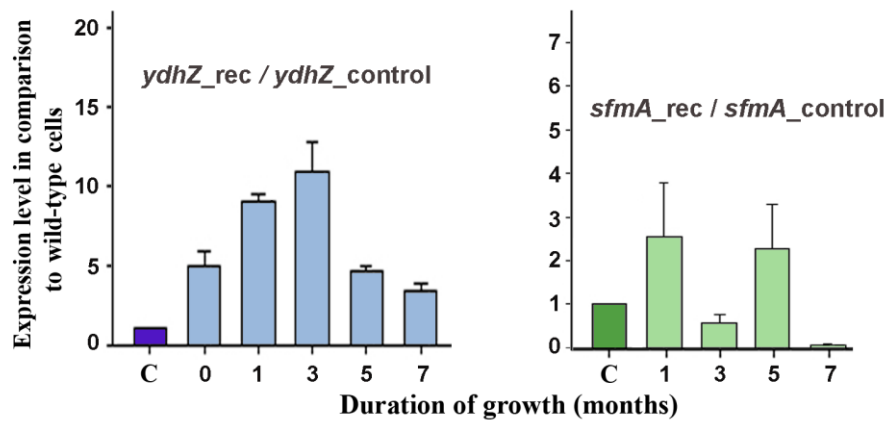


Fig. 3. Relative changes in the expression efficiency of the recombinant genes during long-term culturing. The amount of products registered in the mutant cultures was normalized for the amount of products registered in the control samples (schematically shown by bars marked as C). The measurements obtained at the zero growth point for mutant *E. coli_sfmA_rec* yielded high dispersion, and the corresponding results are not shown.

The transcriptional activity of the transferred *ydhZ* gene was higher than that of its ortholog deleted from the *E. coli* genome (bars 0 and C on the left panel of Fig. 3). This implies that the *ydhZ* promoter in the genome of *Salmonella* is well recognized by the transcription apparatus of *E. coli*. It could not be excluded that RNAs synthesized from the *Salmonella* gene are poor substrates for *E. coli* RNAases, thus explaining their higher level in the recombinant cells. However, the intracellular amount of alien RNAs demonstrated a significant steady increase during a three-month period of growth. This is due to spontaneous mutations randomly occurring in different cells of both bacterial populations, but in the recombinant culture they increased the expression of a foreign gene by an order of magnitude, as compared to the population of cells with the already assimilated ortholog. Further growth was accompanied by a decrease in the expression intensity, which may indicate successful suppression of synthesis of the unwanted gene by the cellular systems. The results obtained for *sfmA* were less straightforward (right panel in Fig. 3), though a similar trend can be observed. In any case, both recombinations provoked an adaptive response in the recipient genome.

The types of mutations within the recombinant regions differ from spontaneous substitutions in the substituted sequences

Libraries of amplicons derived from the recombinant strains *E. coli_ydhZ_rec* (primers F4 and R4 in Table 1) and *E. coli_sfmA_rec* (primers F3 and R3) were jointly sequenced using Ion Torrent PGM with Ion 314™ Chip v2. As a result, we obtained 406119 reads, whereas 560699 reads resulted from combined sequencing of the control regions.

Table 2. Amount of perfectly mapped reads and reads with 1–3 errors for each culture type

Strain	Amplicon length (bp)	Number of generations	Number of perfectly mapped reads	Number of 25 nucleotide reads with substitutions			
				one	two	three	%
<i>ydhZ_control</i>	975	4000	73787	559	136	132	1.1
<i>ydhZ_mutant</i>	907		36136	286	72	79	1.2
<i>sfmA_control</i>	1170	2000	272198	1983	385	363	0.99
<i>sfmA_mutant</i>	2683		146309	883	292	300	1.0

Following quality filtering at a threshold level of $Q \geq 20$ (99 % significance for all nucleotides in each read) and fragmentation to 25 nt sequences, we got 174963 and 376222 reads for two mutant and two control regions, respectively. Most of them were perfectly mapped to the analyzed areas, but approximately one percent of reads contained from one to three substitutions (Table 2). Fragments containing one substitution were used for further analysis.

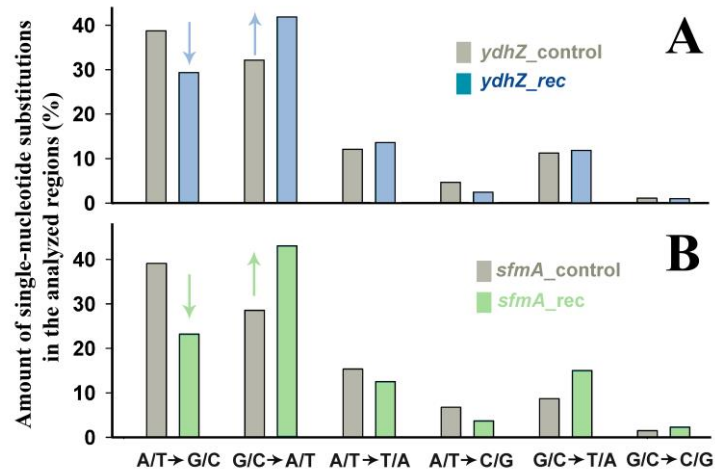


Fig. 4. Normalized per the length of amplicons the number of spontaneous mutations detected in the recombinant regions of mutant genomes and the respective areas of *E. coli* MG 1655 chromosomes. Mutant and parental cells were grown in parallel and harvested after 2000 (*ydhZ_rec*, *ydhz_control*) or 4000 (*sfmA_rec*, *sfmA_control*) generations. To assess intra-population variability, the amplicons were obtained from sampled mixtures of cell cultures, rather than from single clones. The data are shown as percentage of the total number of spontaneous mutations registered for each of the analyzed regions in reads with single substitutions.

Figure 4 features the content of various substitutions identified using MisMatcher. Identical mutations registered in the same position of different reads were considered as a single mutation. Substitutions were mapped on both strands of the studied regions, and those resulting in the origination of the same complementary pair were also accounted as one mutation. Different mutation types in the same position were considered independently.

As expected, in all cultures purine-purine and pyrimidine-pyrimidine substitutions in each DNA strand occur with the highest frequency. Nevertheless, the number of spontaneous transformations of G/C-pairs to A/T-pairs was higher in both mutant cultures, than that in the control cells, while A/T→G/C mutations were contrariwise less frequent. This discrepancy from the control cultures represents another evidence for the adaptive response of bacteria on the transfer of genetic material, conforming to the assumption that made substitutions provoked an increase in the frequency of spontaneous A/T-pair formation near the modified regions in the genome.

DISCUSSION

The study summarizes preliminary results of a long-term evolutionary experiment initiated a year and a half ago and aimed at elucidation of the mechanisms of alien gene integration in the genome. For this purpose, two previously assimilated genes and closely located *promoter islands* in the genome of *E. coli* were replaced by their respective orthologs from *Salmonella enterica* Typhimurium 14028S transcribed from normal promoters (see Fig. 1, as well as Fig. 7 in [13]). Since both genes demonstrate low expression in *E. coli*, we selected growth conditions that do not require their transcription activity. The adaptive response of cells on the introduction of alien genetic material was manifested through a slight difference in the growth curves (Fig. 2), alteration of transcription activity (Fig. 3) and a specific type of spontaneous mutations in the areas of recombination (Fig. 4). All these

changes were observed only at the populational level, because none of the mutations were stabilized in all cells of the community.

In a similar, but more global long-term experiment running without any evolutionary pressure in the laboratory of R. Lenski (University of Michigan, USA) for about 29 years, 12 simultaneously growing *E. coli* cultures accumulated a mere 627 fixed substitutions over 40000 generations [33, 34]. Hundreds of thousands point substitutions emerged in each population, but most of them appeared to be neutral and were thus not fixed in the genomes. If the only source of mutations is inaccuracy of DNA polymerase, working with a precision rate 10^{-10} – 10^{-11} errors per base pair per generation, the probability of occasional encountering of a single fixed mutation in the analyzed recombinant regions of ~1000–2000 bp over 2000–4000 generations is extremely low ($\sim 10^{-4}$ – 10^{-5}). However, the observed particular pattern of mutagenesis allows expecting a higher rate of adaptive evolution.

Substitutions G/C→A/T can originate as a result of deamination of methylated cytidines, and bacterial cells possess appropriate enzymes for this process. Moreover, the mutant substituted for *sfmA* turned out to feature a higher rate of G/C→T/A transversions. This may be due to guanine oxidation to 8-oxoguanine, which forms a noncanonical complementary pair with adenine. At the next round of replication adenine serves as a template for incorporation of thymine. This is indirectly confirmed by the absence of this effect in the anaerobically growing cell culture substituted for *ydhZ*.

Considering the indicated observations, the obtained data confirmed the hypothesis of adaptive reaction aimed at enrichment of a modified region with A/T-pairs in response to the introduction of foreign genetic material. At a long run of evolution it can result in either “silencing” of an unwanted gene with the involvement of specific alien gene suppressor H-NS, or creation of a suitable promoter within a *promoter island* for adequate expression of useful gene.

The expression analysis and creation of Matcher and MisMatcher were carried out with the financial support from the Russian Science Foundation (grant 14-14-00985, <http://rscf.ru/node/8>, ONO (PI), OAG, MNT, KSS). The search for model genes and obtainment of mutants were carried out with the financial support of the Russian Foundation for Basic Research (grant 16-04-01570, ONO (PI), OAG).

REFERENCES

- Ochman H., Davalos L.M. The nature and dynamics of bacterial genomes. *Science*. 2006. V. 311. P. 1730–1733. doi: [10.1126/science.1119966](https://doi.org/10.1126/science.1119966).
- Cohan F.M., Koeppel A.F. The origins of ecological diversity in prokaryotes. *Curr Biol*. 2008. V. 18. P. R1024–R1034. doi: [10.1016/j.cub.2008.09.014](https://doi.org/10.1016/j.cub.2008.09.014).
- Wybouw N., Pauchet Y., Heckel D.G., Van Leeuwen T. Horizontal gene transfer contributes to the evolution of *Arthropod Herbivory*. *Genome Biol. Evol.* 2016. V. 8. P. 1785–1801. doi: [10.1093/gbe/evw119](https://doi.org/10.1093/gbe/evw119).
- Levin B.R. Frequency-dependent selection in bacterial populations. *Philos Trans R Soc Lond B Biol Sci*. 1988. V. 319. P. 459–472. doi: [10.1098/rstb.1988.0059](https://doi.org/10.1098/rstb.1988.0059).
- Domingues S., Harms K., Fricke W.F., Johnsen P.J., da Silva G.J., Nielsen K.M. Natural transformation facilitates transfer of transposons, integrons and gene cassettes between bacterial species. *PLoS Pathog*. 2012. V. 8. Article № e1002837. doi: [10.1371/journal.ppat.1002837](https://doi.org/10.1371/journal.ppat.1002837).
- Ochman H., Lawrence J.G., Groisman E.A. Lateral gene transfer and the nature of bacterial innovation. *Nature*. 2000. V. 405. P. 299–304. doi: [10.1038/35012500](https://doi.org/10.1038/35012500).
- Studier F.W., Daegelen P., Lenski R.E., Maslov S., Kim J.F. Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. *J. Mol. Biol.* 2009. V. 394. P. 653–680. doi: [10.1016/j.jmb.2009.09.021](https://doi.org/10.1016/j.jmb.2009.09.021).

8. Nakamura Y., Itoh T., Matsuda H., Gojobori T. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat. Genet.* 2004. V. 36. P. 760–766. doi: [10.1038/ng1381](https://doi.org/10.1038/ng1381).
9. McDaniel L.D., Young E., Delaney J., Ruhnau F., Ritchie K.B., Paul J.H. High frequency of horizontal gene transfer in the oceans. *Science.* 2010. V. 330. P. 50. doi: [10.1126/science.1192243](https://doi.org/10.1126/science.1192243).
10. Gogarten J.P., Doolittle W.F., Lawrence J.G. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 2002. V. 19. P. 2226–2238. doi: [10.1093/oxfordjournals.molbev.a004046](https://doi.org/10.1093/oxfordjournals.molbev.a004046).
11. Huang Q., Cheng X., Cheung M.K., Kiselev S.S., Ozoline O.N., Kwan H.S. High density transcriptional initiation signals underline genomic islands in bacteria. *PLoS ONE.* 2012. V. 7. Article № e33759. doi: [10.1371/journal.pone.0033759](https://doi.org/10.1371/journal.pone.0033759).
12. Shavkunov K.S., Masulis I.S., Tutukina M.N., Deev A.A., Ozoline O.N. Gains and unexpected lessons in genome-scale promoter mapping. *Nucl. Acids Res.* 2009. V. 37. P. 4919–4931. doi: [10.1093/nar/gkp490](https://doi.org/10.1093/nar/gkp490).
13. Panyukov V.V., Ozoline O.N. Promoters of *Escherichia coli* versus promoter islands: function and structure comparison. *PLoS ONE.* 2013. V. 8. Article № e62601. doi: [10.1371/journal.pone.0062601](https://doi.org/10.1371/journal.pone.0062601).
14. Purtov Y.A., Glazunova O.A., Antipov S.S., Pokusaeva V.O., Fesenko E.E., Preobrazhenskaya E.V., Shavkunov K.S., Tutukina M.N., Lukyanov V.I., Ozoline O.N. Promoter Islands as a platform for interaction with nucleoid proteins and transcription factors. *J. Bioinform. Comput. Biol.* 2014. V. 12. P. 1441006. doi: [10.1142/S0219720014410066](https://doi.org/10.1142/S0219720014410066).
15. Glazunova O.A., Kiselev S.S., Shavkunov K.S., Bykov A.A., Panyukov V.V., Ozoline O.N. Promoter islands in the genome of *E. coli*: comparative analysis against AT-rich sequences. *Math. Biol. Bioinform.* 2014. V. 9. No. 2. P. 563–574. doi: [10.17537/2014.9.563](https://doi.org/10.17537/2014.9.563).
16. Panyukov V.V., Kiselev S.S., Shavkunov K.S., Masulis I.S., Ozoline O.N. Mixed promoter islands as genomic regions with specific structural and functional properties. *Math. Biol. Bioinform.* 2013. V. 8. No. 2. P. 432–448. doi: [10.17537/2013.8.432](https://doi.org/10.17537/2013.8.432).
17. Reppas N.B., Wade J.T., Church G.M., Struhl K. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Mol. Cell.* 2006. V. 24. P. 747–757. doi: [10.1016/j.molcel.2006.10.030](https://doi.org/10.1016/j.molcel.2006.10.030).
18. Herring C.D., Raffaele M., Allen T.E., Kanin E.I., Landick R., Ansari A.Z., Palsson B.O. Immobilization of *Escherichia coli* RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *J. Bacteriol.* 2005. V. 178. P. 6166–6174. doi: [10.1128/JB.187.17.6166-6174.2005](https://doi.org/10.1128/JB.187.17.6166-6174.2005).
19. Dornenburg J.E., DeVita A.M., Palumbo M.J., Wade J.T. Widespread antisense transcription in *Escherichia coli*. *mBio.* 2010. V. 1. Article № e00024-10. doi: [10.1128/JB.187.17.6166-6174.2005](https://doi.org/10.1128/JB.187.17.6166-6174.2005).
20. Kahramanoglou C., Seshasayee A.S., Prieto A.I., Ibberson D., Schmidt S., Zimmermann J., Benes V., Fraser G.M., Luscombe N.M. Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res.* 2011. V. 39. P. 2073–91. doi: [10.1093/nar/gkq934](https://doi.org/10.1093/nar/gkq934).
21. Lucchini S., Rowley G., Goldberg M.D., Hurd D., Harrison M., Hinton J.C. H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog.* 2006. V. 2. Article № e81. doi: [10.1371/journal.ppat.0020081](https://doi.org/10.1371/journal.ppat.0020081).
22. Dorman C.J. H-NS, the genome sentinel. *Nat. Rev. Microbiol.* 2007. V. 5. P. 157–161. doi: [10.1038/nrmicro1598](https://doi.org/10.1038/nrmicro1598).
23. Langille M.G., Brinkman F.S. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics.* 2009. V. 25. P. 664–665. doi: [10.1093/bioinformatics/btp030](https://doi.org/10.1093/bioinformatics/btp030).

24. Lawrence J.G., Ochman H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA*. 1998. V. 95. P. 9413–9417. doi: [10.1073/pnas.95.16.9413](https://doi.org/10.1073/pnas.95.16.9413).
25. Price M.N., Dehal P.S., Arkin A.P. Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. *Genome Biol.* 2008. V. 9. Article № R4. doi: [10.1186/gb-2008-9-1-r4](https://doi.org/10.1186/gb-2008-9-1-r4).
26. Lee D.J., Bingle L.E., Heurlier K., Pallen M.J., Penn C.W., Busby S.J., Hobman J.L. Gene doctoring: a method for recombineering in laboratory and pathogenic *Escherichia coli* strains. *BMC Microbiol.* 2009. V. 9. P. 252. doi: [10.1186/1471-2180-9-252](https://doi.org/10.1186/1471-2180-9-252).
27. *E. coli Gene Expression Database* (GenExpDB). URL: <http://genexpdb.ou.edu/main/> (accessed 26.11.2016).
28. Sangurdekar D.P., Srienc F., Khodursky A.B. A classification based framework for quantitative description of large-scale microarray data. *Genome Biol.* 2006. V. 7. Article № R32. doi: [10.1186/gb-2006-7-4-r32](https://doi.org/10.1186/gb-2006-7-4-r32).
29. Afgan E., Baker D., van den Beek M., Blankenberg D., Bouvier D., Cech M., Chilton J., Clements D., Coraor N., Eberhard C., Gruning B., Guerler A., Hillman-Jackson J., Von Kuster G., Rasche E., Soranzo N., Turaga N., Taylor J., Nekrutenko A., Goecks J. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucl. Acids Research.* 2016. V. 44. P. W3–W10. doi: [10.1093/nar/gkw343](https://doi.org/10.1093/nar/gkw343).
30. *Matcher*. URL: <http://www.mathcell.ru/DnaRnaTools/Matcher.zip> (accessed 26.11.2016).
31. Ozoline O.N., Deev A.A. Predicting antisense RNAs in the genomes of *Escherichia coli* and *Salmonella typhimurium* using promoter-search algorithm PlatProm. *J. Bioinform. Comput. Biol.* 2006. V. 4. P. 443–454. doi: [10.1142/S0219720006001916](https://doi.org/10.1142/S0219720006001916).
32. Kiselev S.S., Ozoline O.N. Structure-specific modules as indicators of promoter DNA in bacterial genomes. *Math. Biol. Bioinform.* 2011. V. 6. P. t1–t13. doi: [10.17537/2011.6.t1](https://doi.org/10.17537/2011.6.t1).
33. Barrick J.E., Yu D.S., Yoon S.H., Jeong H., Oh T.K., Schneider D., Lenski R.E., Kim J.F. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*. 2009. V. 461. P. 1243–1247. doi: [10.1038/nature08480](https://doi.org/10.1038/nature08480).
34. Barrick J.E., Lenski R.E. Genome-wide mutational diversity in an evolving population of *Escherichia coli*. *Cold Spring Harb. Symp. Quant. Biol.* 2009. V. 74. P. 119–29. doi: [10.1101/sqb.2009.74.018](https://doi.org/10.1101/sqb.2009.74.018).

Received 22 February 2017.

Published 06 March 2017.