

SCIENTIFIC REPORTS



OPEN

Natural tuning of restriction endonuclease synthesis by cluster of rare arginine codons

Iwona Mruk¹, Tadeusz Kaczorowski^{1,2} & Agata Witczak¹

Restriction–modification (R–M) systems are highly widespread among bacteria and archaea, and they appear to play a pivotal role in modulating horizontal gene transfer, as well as in protecting the host organism against viruses and other invasive DNA particles. Type II R–M systems specify two independent enzymes: a restriction endonuclease (REase) and protective DNA methyltransferase (MTase). If the cell is to survive, the counteracting activities as toxin and antitoxin, must be finely balanced *in vivo*. The molecular basis of this regulatory process remains unclear and current searches for regulatory elements in R–M modules are focused mainly at the transcription step. In this report, we show new aspects of REase control that are linked to translation. We used the EcoVIII R–M system as a model. Both, the REase and MTase genes for this R–M system contain an unusually high number of rare arginine codons (AGA and AGG) when compared to the rest of the *E. coli* K-12 genome. Clusters of these codons near the N-terminus of the REase greatly affect the translational efficiency. Changing these to higher frequency codons for *E. coli* (CGC) improves the REase synthesis, making the R–M system more potent to defend its host against bacteriophages. However, this improved efficiency in synthesis reduces host fitness due to increased autorestriction. We hypothesize that expression of the endonuclease gene can be modulated depending on the host genetic context and we propose a novel post-transcriptional mode of R–M system regulation that alleviates the potential lethal action of the restriction enzyme.

Restriction–modification (R–M) systems are highly widespread among prokaryotes^{1–3}. The most well-studied Type II R–M systems are composed of two independent enzymes: a restriction endonuclease (REase) and a DNA methyltransferase (MTase). Both proteins recognize the same short specific DNA sequence, but differ in function. The REase cleaves at the recognition sequence while the MTase adds methyl group to modify such sites, to make them protected from the action of the cognate REase⁴. This way the genomic DNA is marked as the self, in contrast to any foreign DNA entering the cell that lacks such modification (non-self) and become easily the target for degradation by host REase. Since their discovery, R–M systems have been recognized as the potent tools to combat bacteriophages. Along with CRISPR systems, they form the active line of bacterial cell defence^{5–7}. By limiting the DNA flux into host cells, Type II R–M systems function as modulators of horizontal gene transfer and such may contribute to bacterial speciation^{8–11}. Due to the potent endonucleolytic activity of REase, the R–M systems might also be considered as potentially hazardous cargo for their hosts. If the expression of REase and MTase is not balanced, then the host death is likely outcome^{12–14}. The challenge then for R–M and CRISPR like systems is one of fine tuning the regulatory control over toxic genes in a manner that promotes high anti-phage immunity while minimizing self-destruction^{15–17}. For mobilizable R–M genes^{11,18}, tight control over REase is especially important at the stage of transfer to a new host cell. At this stage the genome has not been protected by MTase and is susceptible to degradation. R–M systems must have a time-dependent control that initially favours the expression and action of the MTase if the newly introduced gene are to be acquired by the new host^{19–22}. The details underlying these molecular processes are unclear, though it seems that genetic feedback loops play a central role. In this case of R–M gene expression regulation has been studied more extensively for fixed genes, at steady-state. In general, three modes²³ of gene control for Type II R–M systems are in use: regulatory MTase^{24–26}, antisense RNA^{27,28}, and a dedicated transcription factor called the C protein^{29–34}. Co-regulatory modes are very likely to be discovered.

¹Department of Microbiology, Faculty of Biology, University of Gdansk, Wita Stwosza 59, Gdansk, 80-308, Poland.

²Laboratory of Extremophiles Biology, Department of Microbiology, Faculty of Biology, University of Gdansk, Wita Stwosza 59, Gdansk, 80-308, Poland. Correspondence and requests for materials should be addressed to I.M. (email: shamrock127@hotmail.com)

Thus far, studies of the regulatory elements involved in Type II R-M system gene expression have been focused at the stage of transcription, and to our knowledge, no report has yet revealed the mechanism(s) related to the post-transcriptional or protein synthesis level. In general, a primary notion has been that the regulatory mechanisms must be relatively host independent to facilitate the interspecies transfer of genes coding for R-M systems. Any reliance on the specific cellular host factor may limit the spread of genes. On the other hand, the successful installation of genes requires an optimal fitting to host gene expression machinery including the recognition of promoter sequences by host RNA polymerase and refactoring to host codon usage. Sequence analysis of numerous R-M systems revealed genes that are different from host genes in features such as GC content, dinucleotide frequencies and codon usage³⁵. This led us to focus on the codon usage for a few R-M systems isolated from *E. coli*. Of the systems evaluated, EcoVIII of *E. coli* E1585-68, turned out to be especially rich in rare codons for *E. coli*^{36–38}.

We show that an abundance of rare codons are clustered within the EcoVIII REase gene and may seriously impact its translation. We hypothesise that the expression of the endonuclease is modulated by host genetic context, and we propose a novel, post-transcriptional mode of R-M system regulation that may help alleviate the lethality of unbalanced restriction enzyme gene expression.

Material and Methods

Bacterial strains and plasmids. The *E. coli* K-12 strains used in this study are described below. MC1061 [*araD139* Δ (*ara, leu*)7697, Δ *lacX74*, *galU*, *galK*, *hsdR*, *strA*] was used in *lacZ* reporter assays³⁹. *E. coli* DH5 α and MM294 were used for all other purposes including the cloning steps. *E. coli* Rosetta and BL21(DE3) were employed in gene overexpression^{40,41}. MP060 and MP064 strains containing the yellow fluorescent protein (YFP) fused to the promoter of *sulA* (*P_{sulA}-yfp*) was employed to test the SOS response¹². The plasmids used are listed in Table 1. Plasmids constructed in this study were deposited in the Collection of Plasmids and Microorganisms, University of Gdansk, Poland.

T7 RNA polymerase dependent protein induction and pulse-chase assay. Single clones of BL21(DE3) [pLysS/pRARE] cells, which inducibly express T7 RNA polymerase and constitutively express T7 lysozyme (pLysS; Novagene) or genes encoding tRNAs for rare arginine codons AGA, AGG, and CGA, glycine codon GGA, isoleucine codon AUA, leucine codon CUA, and proline codon CCC (pRARE, Novagene), were grown overnight at 37 °C in M9 minimal medium⁴² supplemented with 0.2% glucose, ampicillin (100 μ g/ml) and chloramphenicol (34 μ g/ml). Cells were then diluted 100-fold in the same medium without glucose. When the cultures reached OD_{600nm} of 0.4, they were induced with 1 mM IPTG. Rifampicin (SIGMA) was used 30 min later at 200 μ g/ml to block transcription by *E. coli* RNA polymerase. For labeling the synthesized proteins, 300 μ l aliquots of cultures were collected at given times and incubated with 7 μ Ci/ml [³⁵S]methionine (1.45 Ci/mmol, Amersham Corp.) for 20 min at 37 °C. Cells were then pelleted, washed and dissolved in SDS-PAGE lysis buffer (10 mM Tris-HCl pH 8.0, 10% sucrose, 1 mM EDTA, 2.5% SDS, 0.001% bromophenol blue). Proteins were run on a 10% SDS-PAGE, and the gels were dried, blotted and exposed to X-ray film (KODAK).

Construction of *ecoVIII::lacZ* reporter and LacZ activity assay. The translational fusions of *ecoV-IIIIR* fragment carrying arginine codons fused to the *lacZ* reporter gene were generated as follows. The 280 bp of *ecoVIIIIR* gene fragment containing the natural promoter and ribosome binding site was amplified by PCR (primers: GAGCTCGAGTTAAAGCGTGGG and CAGGATCCCCACTTAATTTGAC introducing XhoI and BamHI site, respectively), cleaved with XhoI and BamHI and cloned into a pLex3B vector⁴³ that had been linearized with the same restriction enzymes. This yielded pLexRR (arginine codons WT- AGA AGA) in which *lacZ* is preceded by the *ecoVIIIIR* promoter and 29 codons in the same reading frame. All constructs generated were confirmed by sequencing. The LacZ assays were based on hydrolysis of *o*-nitrophenyl- β -D-thiogalactoside (ONPG, SIGMA) using exponentially growing cells in M9-glucose minimal medium at 37 °C as described elsewhere^{21,44}.

Fluorescence assay. Cells were grown with shaking to exponential phase in LB medium⁴², gently pelleted, washed once with PBS buffer and then resuspended again in 500 μ l of PBS buffer. Half of the sample was used to monitor the optical density (600 nm) of bacteria and the other half was used to read the YFP intensity (emission at 515 nm with an excitation at 545 nm) in a 96-well plate reader (EnSpire Multimode; Perkin Elmer). Relative fluorescence was corrected by subtracting the level of fluorescence of non-YFP bacterial cells and dividing by the optical density.

Relative restriction activity assay. Restriction activity of *E. coli* cells carrying EcoVIII/EcoRI R-M system and its variants was measured using the plaque formation efficiency (EOP) of phage λ vir. There are 6 EcoVIII (HindIII) sites and 4 EcoRI sites in the λ vir genome. The EOP of λ vir was calculated as the ratio of plaques formed on *E. coli* MG1655 containing plasmid with no R-M system to those formed on the same strain containing a plasmid with the EcoVIII/EcoRI R-M system or their variants.

Analysis of fitness effect by mixed culture competition experiments. Two comparably sized colonies of *E. coli* MG1655 picked from fresh transformation on LB-agar plates (with appropriate selective antibiotics) were inoculated into 5 mL of M9-glucose media. One colony carried the pRR plasmid with WT R-M system and was selected on ampicillin. The second colony was selected on chloramphenicol, and contained its variant – the pFFcm plasmid wherein two rare arginine codons at the *ecoVIIIIR* gene were replaced by a higher frequency CGC codon. In addition, the *cat* gene was cloned to disrupt the *bla* gene of pFF to change the plasmid selection to chloramphenicol resistance. Control cultures were grown in parallel and contained cells with plasmids bearing restriction-negative and modification-positive variants (pRR0 vs. pFF0cm) or just empty vectors (pT7-3 vs. pT7-3 cm). These cultures were used to test whether the antibiotics-resistance genes alone give some advantage in

Name	Relevant features/genotype	Reference
pRARE	pACYC184 derivative carrying genes for tRNAs of the rare codons of <i>E. coli</i> : <i>argU</i> (AGG,AGA), <i>ileX</i> (AUA), <i>leuW</i> (CUA), <i>proL</i> (CCC), <i>glyT</i> (GGA), Cm ^R	40
pLysS	pACYC184 derivative carrying gene of T7 lysozyme, a natural inhibitor of T7 RNA polymerase, Cm ^R	Novagene
pKRP10	chloramphenicol resistance cassette, Cm ^R	99
pEC156	natural plasmid carrying EcoVIII R-M system	18,37,100
pT7-6/pT7-3	inducible ϕ 10 promoter of T7 phage, Ap ^R	101
pT7-EcoVIIIIM	as pT7-6, but gene for EcoVIII MTase cloned under, ϕ 10 promoter as transcriptional fusion with its own promoter	36
pT7-3cm	as pT7-3, but chloramphenicol resistance cassette introduced to break the bla gene, Cm ^R	This study
pRR	as pT7-3, EcoVIII R-M genes cloned under ϕ 10 promoter,	This study
pFF	as pRR, but two codons of EcoVIII REase gene substituted as followed R(AGG)16→R(CGC); R(AGG)17→R(CGC)	This study
pRF	as pRR, but one codon of EcoVIII REase gene substituted R(AGG)17→R(CGC)	This study
pFR	as pRR, but one codon of EcoVIII REase gene substituted R(AGG)16→R(CGC)	This study
pRR0	as pRR, but EcoVIII REase gene inactivated (R–M+)	This study
pFFcm	as pFF (R + M+), but chloramphenicol resistance cassette introduced to break the bla gene, Cm ^R	This study
pFF0cm	as pFFcm, but EcoVIII REase gene inactivated (R–M+), Cm ^R	This study
pLex-3B	vector for testing gene translational fusion to <i>lacZ</i> reporter, pBR322 ori, promoter-less <i>lacZ</i> gene, Ap ^R	43
pLexRR	as pLex3B, but WT fragment encompassing the natural gene promoter, rbs and first 29 codons of EcoVIII REase gene fused in frame to <i>lacZ</i> , Ap ^R	This study
pIM-RM	pACYC184 derivative carrying genes of WT EcoRI R–M system, R + M+; Cm ^R	28
pIM27	as pIM-RM, but EcoRI REase gene inactivated, R–M+; Cm ^R	28
pIMEcoRI-FF	as pIM-RM, but WT rare codons of REase AGG_CUA at position 9 and 10 replaced by frequent ones CGC_CUG; R + M+; Cm ^R	This study

Table 1. Plasmids used in this study and their relevant features.

a mixed cell population. At time zero of the competition experiment, 1:1 mixed cultures of competing cells were inoculated. Every 15–18 hrs of incubation at 37 °C with shaking, the co-cultures were diluted 10⁶ into fresh minimal media without antibiotics. Samples of each mixed competition culture were immediately diluted and then spread quantitatively onto either LB-agar containing the appropriate selective antibiotic or onto LB agar without antibiotics. The colonies were counted and the ratio of colony-forming units (CFU) from the two competing cell populations was calculated using $T = (CFU_{cm}/CFU_{amp})$; their generation number was also determined. Data were normalized using the results from the vector control [$V = (CFU_{cm}/CFU_{amp})$]. The relative competitive fitness (W) was then calculated as $W = \log[T/V]$ for each tested generation time-point, as previously described²⁹.

Results

The rare arginine codon clusters of the EcoVIII R–M system are features of a horizontally transferred cassette. We analyzed the DNA sequence, for the *E. coli* E1585-68: restriction endonuclease (REase) and DNA methyltransferase (MTase). We found a significant difference in the overall GC content of the R-M system unit (less than 38%) compared to its natural carrier - plasmid pEC156 (49.3%) and to the *E. coli* genome (50.8%). We also modelled the codon usage pattern for the EcoVIII R–M system genes using the codon adaptation index (CAI)^{45,46}. Genes with low CAI indices reflect the optimal codon usage of their former host³⁵. The CAI values for REase and MTase are 0,151 and 0,185 respectively³⁶, and indicate that both belong to the group of *E. coli* genes that were most likely acquired by horizontal gene transfer. The results are consistent with data on codon usage deviation for other R–M systems isolated from *E. coli* (EcoRI and EcoRV) and from other members of the *Enterobacteriaceae* family (KpnI, SmaI, SinI, and PvuII)³⁵.

Our analysis of the *E. coli* codons in the EcoVIII R–M system revealed an overrepresentation of rare arginine (AGG and AGA) (>70%) and rare isoleucine (AUA) codons (Fig. 1A,C, Table S1). Out of the 20 arginine codons in the REase gene 18 are rare, and 13 of 15 are rare in the MTase gene. A high number of these rare codons cluster close to the C-terminus of REase, whereas the rare codons in the MTase remain in the central part of the gene (Fig. 1A). The rare isoleucine codons for MTase are closer to N-terminus, but distributed evenly across the REase. The arginine codons, however, form a cluster of two rare codons (tandem AGA) and one frequent codon starting at amino acid position 16 near the N-terminus of the REase. In case of MTase the AGG tandem is at position 167–168 (Fig. 1B). It has been shown that the presence of rare codon cluster close to the gene start may cause a translational complication⁴⁷.

Effects of rare arginine codon cluster on EcoVIII enzymes synthesis. To test the effects of rare arginine codons on EcoVIII REase and MTase synthesis in *E. coli* cells, we made transcriptional fusions and used ϕ 10 promoter of T7 phage. In this system, the T7 RNA polymerase gene is under the control of an IPTG inducible *lacUV5* promoter that is located on *E. coli* BL21(DE3) chromosome⁴¹. In addition, we used the pLysS plasmid to provide a T7 lysozyme, which is a natural inhibitor of T7 RNA polymerase⁴⁸, to decrease the basal expression of the toxic REase gene. EcoVIII MTase and REase synthesis was monitored using a pulse-chase assay carried out in minimal media in the presence of [³⁵S]-methionine after IPTG induction and rifampicin addition to block *E. coli* RNA polymerase, but not T7 RNA polymerase (Fig. 2).

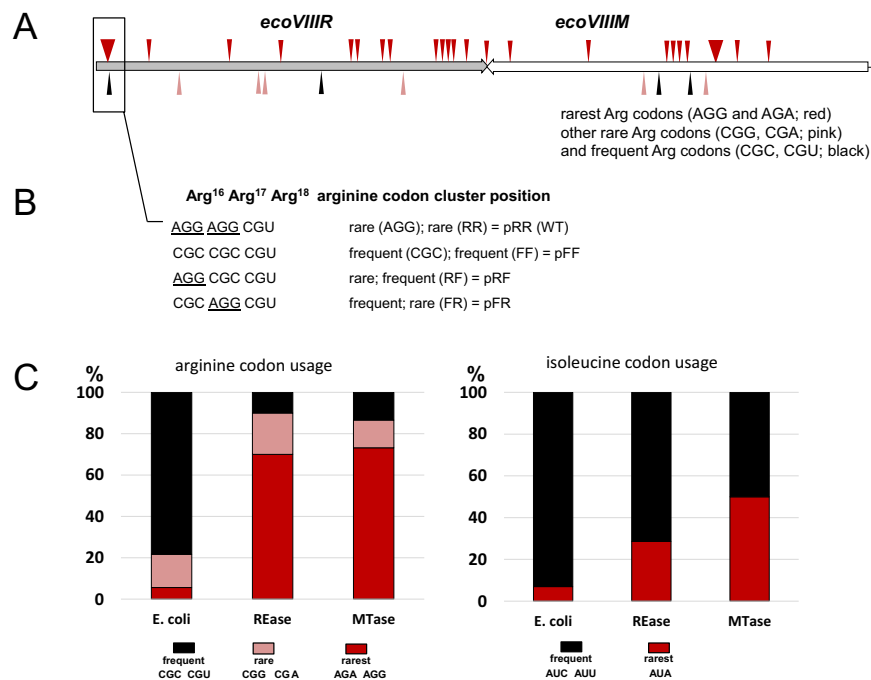


Figure 1. Representation of arginine codons in the EcoVIII REase and MTase genes. **(A)** Distribution of arginine (Arg) codons (in scale). The positions of rarest AGG and AGA Arg codons are indicated above gene arrows in red, whereas other rare CGG and CGA Arg codons are annotated below the genes, marked in pink. The most common Arg codons for *E. coli* are also marked below the genes, in black. Two Arg clusters are marked as thick triangles, others represent the single occurrences. **(B)** The position of cluster of rare Arg codons within the WT REase gene is denoted as RR in pRR plasmid (rare codon at 16th; rare codon at 17th). Codon variants were mutagenized to generate REase variants encoded as FR (frequent codon at 16th; rare codon at 17th); RF (rare codon at 16th; frequent codon at 17th) or FF (both frequent codons). At all gene variants the REase amino acid sequence remains unchanged. **(C)** Codon usage for arginine and leucine for *E. coli* genes compared to EcoVIII REase and MTase genes represented here as the percentage content of particular codon (frequent/rare/rarest) in entire pool of codons for analyzed amino-acid (Supplementary Table S1). *E. coli* values are based on 4332 coding sequences (1372057 codons) for *E. coli* K12 deposited at Codon Usage Database (www.kazusa.or.jp/codon/).

The induction of *ecoVIIIIM* expression led to a prominent and exclusive MTase synthesis (36 kDa) (Fig. 2A). In contrast, EcoVIII REase synthesis was not observed under the same conditions (Fig. 2B). As a control, we used a plasmid, wherein the *ecoVIIIIR* precedes the *bla* gene in the same orientation (Fig. 2C). Only β -lactamase (precursor and mature form) synthesis was detected, regardless of the presence or absence the preceding EcoVIII REase gene (Fig. 2B). We concluded that the observed translational defect may be due to the presence of the two consecutive rare arginine codons in the REase gene (AGG¹⁶AGG¹⁷), close to the translation initiation site (within 25 first codons). The presence of the AGG/AGA codons in this location may confer a severe effect on initiation of protein synthesis^{49–52}. In addition, no tandem of rare codons was detected within the 25 first codons of *ecoVIIIIM* gene, but there are single rare isoleucine codons (AUA) present at positions 11, 22, 25. It seems these codons do not cause any problem for MTase overproduction in this case (Fig. 2A). The rare arginine tandems have no severe effect when are present beyond the first 25 triplets⁵³. We also found rare arginine codons located in the central part of the EcoVIII MTase gene (AGA¹²²AGA¹²³), but still observed that efficient MTase synthesis was achieved (Figs 1 and 2A).

The presence of two consecutive rare arginine codons in *ecoVIIIIR* affects translation efficiency. To test the effects of the tandem rare arginine codons on *ecoVIIIIR* translation efficiency, we used a quantitative *lacZ* reporter assay. We generated translational fusions of WT *ecoVIIIIR* with the promoter-less *lacZ* gene in the pLex-3B vector. The *lacZ* reporter gene in such construct was preceded by the natural promoter region, ribosome binding site and 29 N-terminal *ecoVIIIIR* codons in the same reading frame (pLex-RR). The reporter assay shows the high LacZ activity was achieved only when the plasmid pRARE carrying the genes for tRNAs for the rare codons including tRNA^{AGG/AGA} was present (Fig. 3).

To further investigate the effect of clustered arginine codons, we replaced the rare codons with higher frequency variants and measured whether REase synthesis is overcome after induction of *ecoVIIIIR* gene expression from the T7 promoter. To do this, we constructed variants of the REase gene by site-directed mutagenesis, wherein each rare codon cluster (AGG, pRR) is replaced by a tandem of frequent codons (CGC; pFF/pRF/pFR; Fig. 1B). Next we performed overexpression experiments as shown before (Fig. 2). We compared the total protein

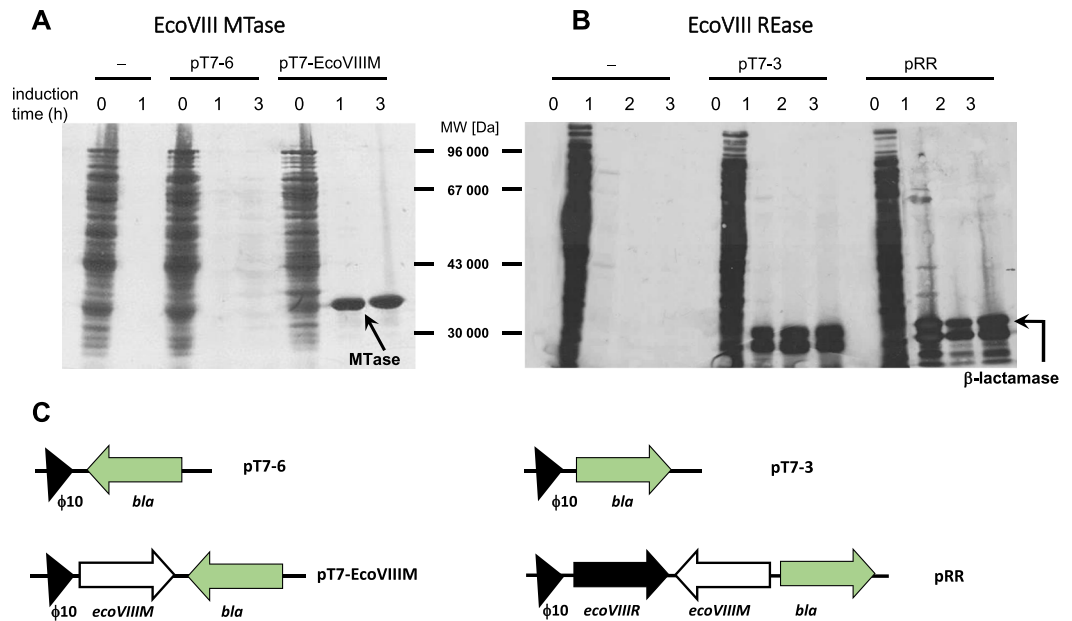


Figure 2. Effect of rare arginine codons on EcoVIII enzyme synthesis. *E. coli* BL21(DE3) pLysS cells containing a compatible plasmid from pT7 series carrying EcoVIII MTase gene (pT7-EcoVIIIIM) (A) or EcoVIII REase gene (pRR) (B) both under T7 promoter, as well as their parallel controls, were grown until log phase, induced with 1 mM IPTG and labeled with [³⁵S]-methionine at the post-induction time (one, two and three hours) as described under “Material and Methods”. Time point 0 represents a sample taken before induction. 200 µg/ml rifampicin was added to cultures 30 min after IPTG induction to block the host RNA polymerase and allow for exclusive expression from the T7 promoter. Time-point samples were taken and their proteins were resolved on 10% SDS-PAGE and visualized by autoradiography. Molecular weight markers are shown. Arrows indicate the labeled products: EcoVIII MTase (36 kDa) (A) and the precursor and mature form of β-lactamase (31.5 and 28.9 kDa) (B). A schematic representation of plasmid constructs used is shown in (C).

content from cells with and without IPTG induction to determine if R.EcoVIII (36,7 kDa) was generated for each of the variants. We observed, that replacing a single rare arginine codon from the tandem is sufficient to unleash REase overproduction (compare pRR versus pFF; pRF and pFR induction positive lanes in Fig. 4A left panel).

Moreover, if *E. coli* cells with pRR plasmid carrying WT EcoVIII R–M system were supplemented with tRNA^{Arg} from the pRARE plasmid then inducible REase overproduction occurs indicating that the translational suppression is relieved (Fig. 4A right panel). To confirm these observations, we also performed a pulse-chase assay for cells carrying plasmids with the REase gene, having the rare arginine codons substituted. We used the same conditions as in the experiment shown in Fig. 2B. The experiments revealed the exclusive presence of a comparably intense band for the EcoVIII REase for each variant when at least one rare arginine codon was replaced (Fig. 4B). We also tested whether the REase mRNA stability is affected by the presence of genes for rare codons carried by pRARE plasmid. We used the approach, where the cultures of *E. coli* MG1655 carrying pRR plasmid with and without pRARE, were treated with the rifampicin, which inhibits the initiation of transcription. The results clearly show that no significant change in mRNA level for REase gene has been found. The determined half-life of REase mRNA was about 2 min. (Supplementary Fig. S2).

Overall, these results indicate that the inhibition of R.EcoVIII synthesis under overexpression conditions is primarily caused by the presence of tandem rare arginine codons at the N-terminus of the *ecoVIIIIR* gene (positions 16 and 17).

Substitution of the arginine tandem increases REase restriction activity and induces the cell filamentation.

To further understand the role of the rare arginine cluster in EcoVIII R–M expression, we examined how their presence or absence affects restriction activity under steady-state levels, when gene expression signals come from a constitutive promoter and other natural genetic elements. We used λvir bacteriophage to compare plaque formation for the WT EcoVIII R–M system and the three variants with codon substitutions in the tandem arginine codons (Fig. 1B). The rare arginine codons (AGG_AGG; pRR, WT), when substituted with CGC_AGG (pFR); AGG_CGC (pRF) or CGC_CGC (pFF) conferred comparable and almost seven-fold higher restriction than the wild-type (Table 2). We also tested the activity of REase variants *in vitro* using crude cell extracts and substrate DNA to calculate the enzyme activity in units per amount of total cellular protein content. The results were similar to the *in vivo* observations (not shown) and taken together clearly indicate the biological significance of the rare arginine clusters present at the 5' end of *ecoVIIIIR* gene. We also supplemented the WT EcoVIII R–M system with pRARE carrying the genes for tRNA of rare codons (arginine, leucine, isoleucine, glycine, proline; Table 1) and observed significantly higher levels of restriction (38-fold) (Table 2). In this case, we concluded, although the rare arginine codon cluster seems to be important for the initiation of translation,

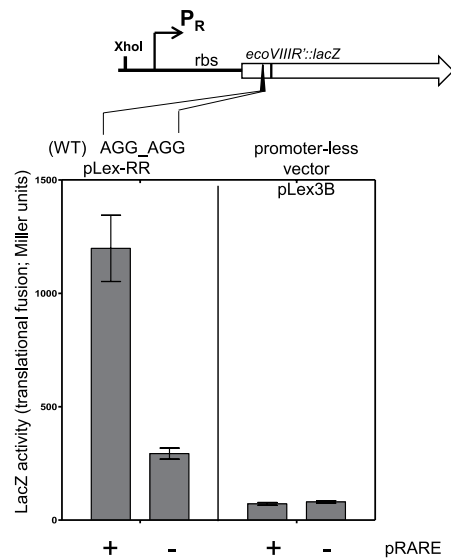


Figure 3. Cluster of rare arginine codons in REase gene blocks its translation. This effect is suppressed, when genes coding for rare tRNAs for *E. coli* are introduced into the cell by the pRARE plasmid. The expression of REase gene is measured by translational fusion to *lacZ* reporter assay presented on the scheme. The control plasmid (pLex3B) was used in parallel. Standard deviation from at least three measurements is indicated.

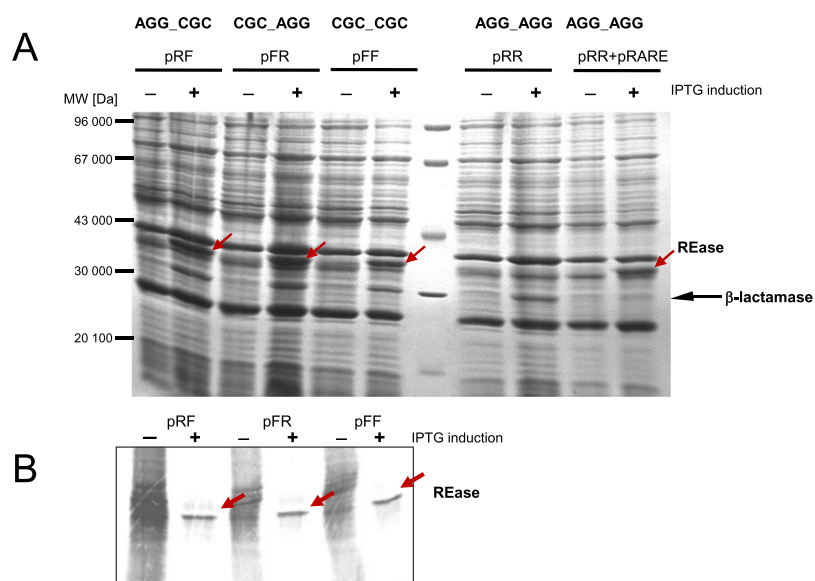


Figure 4. Synthesis of WT EcoVIII REase and its variants with rare codons substituted by high frequency codons for *E. coli*, as indicated on Fig. 1B. (A) *E. coli* BL21(DE3) cells containing a plasmid with REase variant were induced with 1 mM IPTG or not induced (control). Then, the samples were taken and proteins resolved on 10% SDS-PAGE and visualized by Coomassie staining. Molecular weight markers are shown. The arrows indicate the band corresponding to the REase product (37 kDa) obtained only if at least one rare codon was replaced (pRF; pFR; pFF) or pRARE plasmid carrying the genes for rare tRNAs was present (B). The exquisite REase production (indicated by an arrow) was also confirmed by the pulse chase assay, as described in Fig. 2. The full blot is presented at Figure S1 of the Supplementary Material.

however the rare arginine codons outside the cluster, as well as other rare codons, such as isoleucine and glycine present in a high number throughout both REase and MTase genes also affect their expression (Fig. 1A, Supplementary Table S1).

We examined the cell morphology for the R–M system variants and the microscopic images revealed the presence of cell filamentation, a typical manifestation of SOS response induction, for the variants with at least one rare codon replaced with a higher frequency variant (pRF, pFR or pFF), but not for WT R–M system alone

Plasmid	Genotype	Plaque-forming units	^a Efficiency of plaque formation	Restriction relative to WT (R + M+)
pT7-3	R ⁻ M ⁻ (vector)	$(4.5 \pm 0.02) \times 10^6$	1	
pRR0	R ⁻ M ⁺	$(4.3 \pm 0.02) \times 10^6$	0.9	
pRR	R ⁺ M ⁺ (WT) AGG_AGG ^b	$(3.5 \pm 0.04) \times 10^5$	0.078	1
pFR	R ⁺ M ⁺ CGC_AGG ^b	$(5.3 \pm 0.02) \times 10^4$	0.012	6.5
pRF	R ⁺ M ⁺ AGG_CGC ^b	$(4.9 \pm 0.02) \times 10^4$	0.011	7.1
pFF	R ⁺ M ⁺ CGC_CGC ^b	$(4.4 \pm 0.03) \times 10^4$	0.0097	8.0
pT7-3; pACYC184		$(4.5 \pm 0.03) \times 10^6$	1	
pT7-3; pRARE		$(3.7 \pm 0.04) \times 10^6$	0.82	
pRR; pACYC184		$(3.4 \pm 0.04) \times 10^5$	0.075	1
pRR; pRARE		$(8.9 \pm 0.03) \times 10^3$	0.002	38

Table 2. Substitution of rare arginine codons in REase gene increases restriction activity of EcoVIII R–M system. R, restriction; M, modification; WT, wild-type. ^aEfficiency of plaque formation = plaque-forming units on strains carrying tested plasmid divided by plaque-forming units on pT7-3 vector. ^bArg codons at position 16 and 17 of *ecoVIII*R. All plasmids having R⁺ M⁺ genotype carry R–M system with intact amino acid sequence for EcoVIII REase and MTase. The host bacterium was *E. coli* MG1655. Standard deviation from four measurements is indicated.

(pRR) (Fig. 5C vs. D). However, when the plasmid carrying the WT EcoVIII R–M system was supplemented with pRARE plasmid

extensive cell filamentation occurred (Fig. 5B vs. A,C). These data are in agreement with the restriction activity test, as the elevated restriction of phage DNA correlated with SOS response induction.

Lack of rare arginine codon cluster in REase gene impairs the cell fitness. We also questioned whether or not the WT EcoVIII REase or its variant with cluster of arginine codons replaced by high frequency codons for *E. coli*, would show any differences in viability or fitness under long-term culture conditions. New plasmid constructs were prepared carrying the R-M system variants or vectors control, where the chloramphenicol resistance cassette was inserted into *bla* gene in order to restore the ampicillin sensitivity (pT7-3cm - vector, pFF0cm - R⁻M⁺ and pFFcm - R⁺M⁺) making antibiotic resistance the cell marker to monitor the numbers of competing strains over the entire growth course. For the competitions, cells carrying the appropriate plasmids (pRR WT, R⁺M⁺ ampR vs. pFFcm, R⁺M⁺, cmR) were mixed 1:1 and inoculated into the same flask (with replicates). The cultures were prepared in minimal media and kept over 220 generations without the antibiotic pressure with sub-culturing every 20/21 generations (CFU were counted at these time points). For controls, parallel flasks containing the restriction negative variants (pRR0, R⁻M⁺ ampR vs. pFF0cm; R⁻M⁺ cmR) and compatible vectors to normalize the data (pT7-3; R⁻M⁻ ampR vs. pT7-3cm; R⁻M⁻ cmR) were grown under the same conditions. We observed a significant and continuous loss of cells with the frequent codon clusters (pFFcm) when cultured with the strain bearing the WT version of the rare codon cluster (pRR) indicating a loss of fitness that likely due to higher REase levels (Fig. 6).

Moreover, we tested also the presence of restriction activity for cells derived from the last counting (about 220 generation) and found that some fraction of cells from pFFcm pool, but not pRR, had changed their initial highly-restrictive phenotype into restriction-negative. We isolated these plasmids and observed some changes using fragmentation of plasmid DNA with various restriction enzymes. For the control co-cultures with restriction-negative variants, a difference in relative fitness was not detected (pRR0 vs pFF0cm). The number of the competing cells stayed comparable for the entire course of the experiment (Fig. 6).

Rare codon clusters at other R–M system units. To determine if what we observed in the EcoVIII R-M system occurs in the other Type II systems, we surveyed the REBASE data, for “true” (not putative) Type II R-M systems, whose genes are cloned, sequenced and isolated from an *E. coli* host⁵⁴. Next, we analyzed the codons for the REase and MTase genes to determine if the rare codon clusters are also formed at the N-terminus of other similar REases. There are single rare codons in several R-M systems, but out of 22 R-M systems surveyed, we found only two examples, where a cluster of rare codons was detected: EcoRI and Eco571/Eco9272I. These codons were at position of 9 and 10, and coded for different amino-acids (arginine and leucine) in both cases: AGG_CUA (Table 3). In most cases, one to three rare codons within 25 first codons were detected at REase gene, and comparable number at MTase gene (Table 3).

Finally, we used the well-studied EcoRI R-M system as a test to see if changing of the rare codon cluster would significantly alter the gene expression and restriction activity in a manner similar to EcoVIII R-M system. We changed the rare codons AGG_CUA (pIM-RM) into the more frequent CGC_CUG (pIMEcoRI-FF) via site-directed mutagenesis. When compared to WT, the relative restriction level for the EcoRI R-M variant system (pIMRM vs. pIMEcoRIFF) are indeed higher (phage assay, Fig. 7A). However, the two-fold increase over WT is modest compared to the change observed for the EcoVIII R-M system (8-fold, Table 2) using the same phage assay with a comparable number of restriction sites on phage genome. This may be explained by the fact the EcoRI R-M system has cluster of two codons (arginine and leucine), that are delivered by two different tRNAs, and this affects translation of REase gene to lesser extent than in EcoVIII case. In addition, we also tested if the

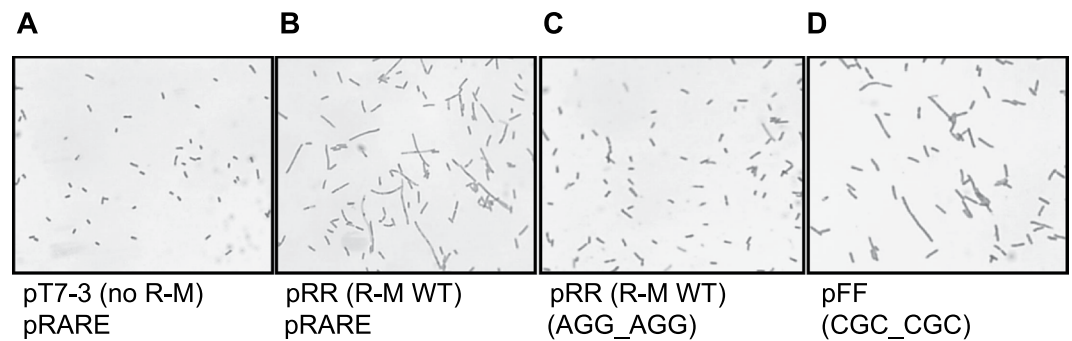


Figure 5. Microscopic images for morphology analysis of *E. coli* MG1655 cells carrying plasmids without (A) and with EcoVIII R-M system of WT (C) or its variant with two rare arginine codons replaced by high frequency codons, as exemplified by the pFF plasmid (D). Cells with two compatible plasmids of: WT (pRR) and pRARE (B). Cell filamentation appears if REase expression is upregulated - (B) and (D).

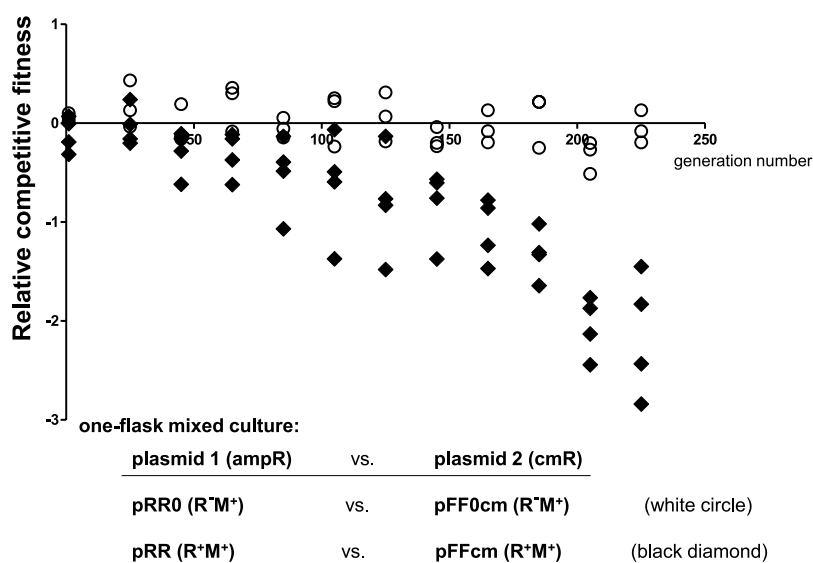


Figure 6. Relative fitness of cells with a variant of EcoVIII R-M system with high frequency codon cluster at its REase is heavily impaired. Mixed cultures were prepared by adding equal number of two types of competing *E. coli* cells into medium without any antibiotic (Material and Methods). Each type carried a plasmid with a specific R-M system variant and its distinct antibiotic marker (chloramphenicol or ampicillin), as indicated below the diagram. One flask co-cultures were diluted every 24 generations into fresh media and CFUs of competing cells were measured. Relative competitive fitness (W) was estimated individually for each mixed culture represented on the diagram as a single symbol, calculated as $W = \log(\text{CFU}_{\text{cm}}/\text{CFU}_{\text{amp}})$ and normalized to vector control (pT7-3 vs. pT7-3cm) (Materials and Methods). Black diamonds represent four separate co-cultures, where the WT R-M system in ampicillin resistant cells (pRR) competed with its R-M system variant with frequent codon cluster in REase gene (pFFcm) in chloramphenicol resistant cells. In control, parallel co-cultures of cells with plasmids with restriction-negative and modification-positive variants (pRR0 vs. pFF0cm; white circles) were used.

elevated restriction activity of EcoRI REase variant can induce the SOS response quantitatively using a *E. coli* YFP reporter strain ($P_{\text{sulA}}\text{-yfp}$). The results show that the altered EcoRI R-M system is comparable to the parental (WT) in inducing some level of SOS signal (Fig. 7B).

Discussion

Horizontal gene transfer is a major force driving the microbial evolution resulting in genome plasticity and mosaic pattern of their genes⁵⁵⁻⁵⁸. In general, the genes to be acquired need to successfully overcome the wide range of barriers to sustain the viability of the accepting host and to be retained in the genome^{59,60}. The newly acquired DNA fragments with genes carrying the beneficiary function for the host have the highest chance to be integrated into the genome⁶¹. R-M systems certainly deliver the advantageous efficient anti-phage tools to their host⁵, but also once taken, remain utterly linked to the host as the addictive modules⁶². Losing the R-M system has

Type II R-M system	Position of rare codons within first 25 codons of gene coding for:	
	REase	MTase
EcoVIII	AGGAGG (16,17)	no
EcoRI	AGGCUA (9,10)	AGA (3)
Eco128I	CUA (10)	NA
EcoRII	CUA (8)	no
EcoT38I	AUA (15)	GGG (12,15)
Eco1524I	GGA (21)	no
Eco29kI	AGA (8,14)	AGA (2)
	CUA (19)	CUA (17)
	AUA (22)	GGA (9,21)
EcoGIII	AUA (15)	CUA (5)
	AGA (22)	
EcoO109I	AUA (7)	GGG (15)
	AGA (21)	
	CUA (25)	
EcoRV	CUA (11)	AUA (10)
		AGA (25)
EcoHK31I	AGA (12)	no
Eco47II	no	no
EcoGIII	AUA (15)	CUA (5)
	CCC (21)	CCC (17)
	AGA (22)	
Eco31I	GGA (4)	M1 AUA (4)
		M2 no
	REase/MTase fused into one polypeptide	
Eco4465II	no	
Eco57I/Eco9272I	AGACUA (9,10)	
Eco644I/Eco933II	no	
EcoE1140I	CUA (25)	
EcoMVII	no	
EcoPT54I	no	
EcoTWII	no	
Eco9276II	CUA (25)	

Table 3. Distribution of rare codons near start of translation for Type II R-M systems isolated from *E. coli* strains. Data based on available sequence source of REBASE collection on Nov. 2018⁵⁴. Rarest codons are: AGA/AGG – arginine; CUA – leucine; AUA – isoleucine; GGA – glycine; CCC – proline. NA – sequence not available.

the serious death consequences by mechanism of the post-segregational killing also typical for toxin-antitoxin systems²³. EcoVIII R-M genes, carried by a natural ColE1-type plasmid, displayed the features of the horizontally transferred genes^{36,37} by conferring the atypical codon usage and biased GC content⁶³. It is known that codon usage reflects an adaptation of genes to the translational machinery of their host, unlike the most recently acquired, alien genes⁶⁴. It also seems that the successful lateral gene transfer is most likely to occur if the codon usage of entering genes is compatible with the accepting genome^{65–67}. Moreover, if accepted, the genes also tend to adapt to the host tRNA pool for their better expression to maintain their function⁶⁸. However, this process may not occur for certain genes, as for tested here EcoVIII R-M system, which maintained the unusual overrepresentation of rare codons for *E. coli*. Both EcoVIII R-M genes display a strong codon bias, which seems to be favored by the host because there is a strong selective pressure to keep the rare codons and provide an adequately low amount of protein to perform its function satisfactorily, for fitness and to minimize the R-M system toxicity, as reported by others^{65,69,70}. Many rare codon cluster positions are conserved within homologous coding sequences across diverse bacteria, suggesting they result from a positive selection and have a functional role⁷¹. In particular, we tested the role of rare arginine cluster near the start codon of REase and found its potential regulatory function in protein synthesis, R-M system maintenance and impact on the host cell fitness. We found that the cluster of rare arginine codon alone negatively affects the REase synthesis, as the delivery of tRNAs for the rare arginine codons highly increased its expression, in accord with other reports related to production of heterologous proteins^{72,73}. Indeed, the tRNAs' availability seems to be the limiting factor for the protein synthesis, when the rare codon cluster is located close to translational start of the gene. We tested whether such cluster affects EcoVIII REase (positions 16 and 17) and MTase (positions 122 and 123) synthesis by pulse chase assay with a T7 promoter system. Our results confirmed the cluster position is detrimental for REase synthesis due to its proximity to translational start, but not for MTase, whose synthesis occurred efficiently. The REase translational suppression could

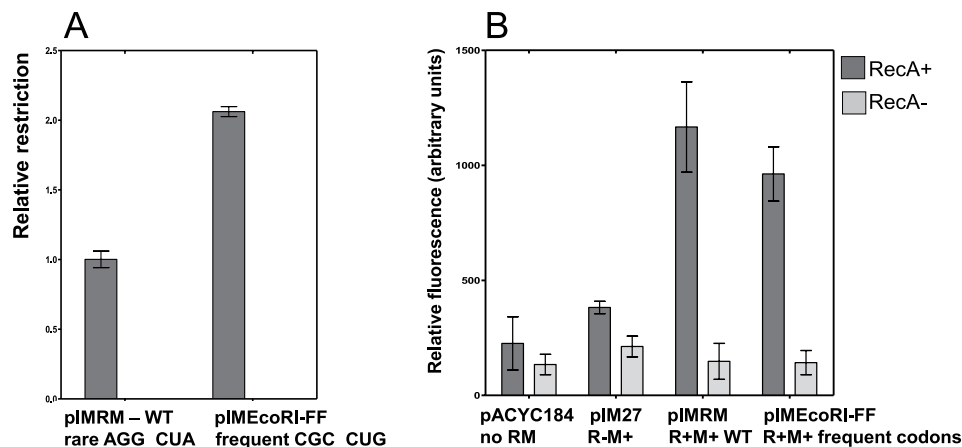


Figure 7. Effect of cluster of rare codons on REase from EcoRI R-M system. For WT and variant of EcoRI R-M system with frequent codons (FF) relative restriction as measured (A) and the SOS response induction in the context of *recA*⁻ and *recA*⁺ cells (B) was assayed. Standard deviation from at least three measurements is indicated.

be relieved in two ways: if more tRNA^{Arg} was delivered or at least one rare arginine codon within the cluster was replaced by a high frequency codon. The data from the REase overproduction in T7 promoter system matched well the natural promoter context, where we measured the level of relative restriction, which is the function of two active enzymes: REase and MTase. We also show that if only one rare arginine codon was replaced for the high frequency codon, the relative restriction due to REase synthesis was significantly augmented (by 7-fold). Such exchange seems to be deleterious to cell fitness. Using direct competition fitness assay, we show that the cells carrying the REase gene variant with cluster of high frequency arginine codons were outcompeted by the cells with WT REase. Outcompeted cells with elevated relative restriction (about 7-fold; REase with CGC cluster) were eliminated from co-cultures due to possible autorestriction of its host genomes. Clearly, loss of rare arginine codon cluster in REase gene also shifted the balance in REase/MTase activities and subsequently triggered the SOS response, manifested in cell filamentation. This phenomenon is frequently observed whenever R-M system balance is lost^{12-14,74,75}.

There are many reports that link the position of translationally non-optimal codons near the start of the gene with modulation of its gene expression at the translational stage⁷⁶⁻⁷⁹. The presence of consecutive rare codons may lead to translational pausing, frameshifting and/or amino-acid misincorporation in the growing polypeptide chain due to delay in appropriate tRNA delivery^{47,49-52,80,81}. The local pauses in translation may be beneficial for some protein synthesis and co-translational folding⁸²⁻⁸⁵. In particular, the NGG codons located near the initiation codon (+2 – +5 of codons) have a severe reducing effect on gene expression, but if shifted slightly downstream, the expression proceeds at normal level⁸⁶. However, some reports also indicate that the effect of rare codons' cluster is related rather to the mRNA secondary structure and its stability, and sometimes it may even enhance gene expression⁸⁷⁻⁹⁰. Genome-wide analysis also supports this observation, i.e. the relaxed mRNA structure at the beginning of gene is favored, not the codon usage⁹¹ and the non-optimal codons tend to destabilize mRNA⁹². When we made the RNA structure prediction for different arginine cluster variants using the RNAfold software, the result showed comparable structures (Supplementary Fig. S3).

The regulatory role of rare codon cluster at 5' gene termini has been shown for bacteriophage lambda integrase⁹³. In addition, even a single synonymous change for rare codon for certain genes could exert a strong phenotypic effect. Such a change at the *hfq* gene, a global RNA regulator of *S. typhimurium*, caused only a two-fold decrease of *hfq* expression, but showed a distinct phenotype of the strain virulence with reduced biofilm formation, motility and survival in macrophages⁹⁴. For some genes, the presence of several rare codons in regulatory genes is required to keep their low expression and sustain the optimal operon coordination, like for FimB recombinase⁹⁵ or global regulator BldA in *Streptomyces*^{96,97}. These schemes of regulation rely on the accessibility of tRNAs for rare codons as a limiting factor for gene expression, as it also seems true for the reported here EcoVIII restriction endonuclease, whose upregulation may be toxic for the host strain. We screened more Type II R-M systems isolated from *E. coli* to find whether the rare codon-dependent regulation may be a more general phenomenon. We found several examples of REase genes with only a single rare codon present near the 5' terminus, on average. In two cases, the cluster of two rare codons for different amino-acids was detected, for the EcoRI and Eco57I R-M system. We tested the effect of a synonymous change for the EcoRI REase, and found the relative restriction changed only by two-fold, much less than for the EcoVIII R-M system (7-fold).

For mobilizable R-M genes with toxic potency (also toxin and antitoxin systems), the flexible level of expression should be avoided, thus such a wide repertoire of regulatory mechanisms exist, where REase and MTase expression is controlled at the transcription step²³. However, in this report, we indicate a novel aspect, never raised before, which is associated with translation of the REase. Its expression seems to be dependent on the pool of host tRNAs for arginine, which may be limiting in some bacteria, such as in *E. coli*, or not at all in bacteria which use the AGG codon frequently. In the latter case, the installation and optimal maintenance of EcoVIII R-M

system may be problematic. During bacterial growth the tRNAs' level also slightly changes⁹⁸, so the R-M system is challenged to adapt to such conditions and not to kill its host. The example of the Type II EcoVIII R-M system shows the potency of the toxic gene expression, the restriction endonuclease, to be tuned to the host's genetic context, that may help to alleviate the lethality of unbalanced R-M system gene expression.

References

- Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Research* **41**, 4360–4377, <https://doi.org/10.1093/nar/gkt157> (2013).
- Oliveira, P. H., Touchon, M. & Rocha, E. P. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res* **42**, 10618–10631, <https://doi.org/10.1093/nar/gku734> (2014).
- Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE - a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* **43**, D298–299, <https://doi.org/10.1093/nar/gku1046> (2015).
- Loenen, W. A., Dryden, D. T., Raleigh, E. A., Wilson, G. G. & Murray, N. E. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res* **42**, 3–19, <https://doi.org/10.1093/nar/gkt990> (2014).
- Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* **8**, 317–327, <https://doi.org/10.1038/nrmicro2315> (2010).
- Abedon, S. T. Bacterial 'immunity' against bacteriophages. *Bacteriophage* **2**, 50–54, <https://doi.org/10.4161/bact.18609> (2012).
- Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **11**, 675–687, <https://doi.org/10.1038/nrmicro3096> (2013).
- Vasu, K. & Nagaraja, V. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev* **77**, 53–72, <https://doi.org/10.1128/MMBR.00044-12> (2013).
- Rocha, E. P., Danchin, A. & Viari, A. Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. *Genome Res* **11**, 946–958, <https://doi.org/10.1101/gr.153101> (2001).
- Jeltsch, A. Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? *Gene* **317**, 13–16 (2003).
- Oliveira, P. H., Touchon, M. & Rocha, E. P. Regulation of genetic flux between bacteria by restriction-modification systems. *Proc Natl Acad Sci USA* **113**, 5658–5663, <https://doi.org/10.1073/pnas.1603257113> (2016).
- Pleška, M. *et al.* Bacterial Autoimmunity Due to a Restriction-Modification System. *Curr Biol* **26**, 404–409, <https://doi.org/10.1016/j.cub.2015.12.041> (2016).
- Katna, A., Boratynski, R., Furmanek-Blaszk, B., Zolcinska, N. & Sektas, M. Unbalanced Restriction Impairs SOS-induced DNA Repair Effects. *Journal of Microbiology and Biotechnology* **20**, 30–38, <https://doi.org/10.4014/jmb.0907.07005> (2010).
- Handa, N., Ichige, A., Kusano, K. & Kobayashi, I. Cellular responses to postsegregational killing by restriction-modification genes. *Journal of Bacteriology* **182**, 2218–2229, <https://doi.org/10.1128/JB.182.8.2218-2229.2000> (2000).
- Leon, L. M., Mendoza, S. D. & Bondy-Denomy, J. How bacteria control the CRISPR-Cas arsenal. *Curr Opin Microbiol* **42**, 87–95, <https://doi.org/10.1016/j.mib.2017.11.005> (2017).
- Bikard, D. & Marraffini, L. A. Control of gene expression by CRISPR-Cas systems. *F1000Prime Rep* **5**, 47, <https://doi.org/10.12703/P5-47> (2013).
- Goldberg, G. W. & Marraffini, L. A. Resistance and tolerance to foreign elements by prokaryotic immune systems - curating the genome. *Nat Rev Immunol* **15**, 717–724, <https://doi.org/10.1038/nri3910> (2015).
- Werbowsky, O. & Kaczorowski, T. Plasmid pEC156, a Naturally Occurring *Escherichia coli* Genetic Element That Carries Genes of the EcoVIII Restriction-Modification System, Is Mobilizable among *Enterobacteria*. *PLoS One* **11**, e0148355, <https://doi.org/10.1371/journal.pone.0148355> (2016).
- Mruk, I. & Blumenthal, R. M. Real-time kinetics of restrictionmodification gene expression after entry into a new host cell. *Nucleic Acids Research* **36**, 2581–2593, <https://doi.org/10.1093/nar/gkn097> (2008).
- Mruk, I., Rajesh, P. & Blumenthal, R. M. Regulatory circuit based on autogenous activation-repression: roles of C-boxes and spacer sequences in control of the PvuII restriction-modification system. *Nucleic Acids Res* **35**, 6935–6952, <https://doi.org/10.1093/nar/gkm837> (2007).
- Mruk, I. & Blumenthal, R. M. Tuning the relative affinities for activating and repressing operators of a temporally regulated restriction-modification system. *Nucleic Acids Res* **37**, 983–998, <https://doi.org/10.1093/nar/gkn1010> (2009).
- Morozova, N. *et al.* Temporal dynamics of methyltransferase and restriction endonuclease accumulation in individual cells after introducing a restriction-modification system. *Nucleic Acids Res* **44**, 790–800, <https://doi.org/10.1093/nar/gkv1490> (2016).
- Mruk, I. & Kobayashi, I. To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res* **42**, 70–86, <https://doi.org/10.1093/nar/gkt711> (2014).
- Protsenko, A., Zakharova, M., Nagornyykh, M., Solonin, A. & Severinov, K. Transcription regulation of restriction-modification system Ecl18kI. *Nucleic Acids Res* **37**, 5322–5330, <https://doi.org/10.1093/nar/gkp579> (2009).
- Burenina, O. Y. *et al.* Peculiarities of the Regulation of Gene Expression in the Ecl18kI Restriction-Modification System. *Acta Naturae* **5**, 70–80 (2013).
- Fedotova, E. A. *et al.* SsoII-like DNA-methyltransferase Ecl18kI: interaction between regulatory and methylating functions. *Biochemistry (Mosc)* **74**, 85–91 (2009).
- Nagornyykh, M. *et al.* Regulation of gene expression in restriction-modification system Eco29kI. *Nucleic Acids Res* **39**, 4653–4663, <https://doi.org/10.1093/nar/gkr055> (2011).
- Mruk, I., Liu, Y., Ge, L. & Kobayashi, I. Antisense RNA associated with biological regulation of a restriction-modification system. *Nucleic Acids Res* **39**, 5622–5632, <https://doi.org/10.1093/nar/gkr166> (2011).
- Rezulak, M., Borsuk, I. & Mruk, I. Natural C-independent expression of restriction endonuclease in a C protein-associated restriction-modification system. *Nucleic Acids Research* **44**, 2646–2660, <https://doi.org/10.1093/nar/gkv1331> (2016).
- Semenova, E. *et al.* Transcription regulation of the EcoRV restriction-modification system. *Nucleic Acids Res* **33**, 6942–6951, <https://doi.org/10.1093/nar/gki998> (2005).
- Bogdanova, E. *et al.* Transcription regulation of the type II restriction-modification system AhdI. *Nucleic Acids Res* **36**, 1429–1442, <https://doi.org/10.1093/nar/gkm1116> (2008).
- Bogdanova, E. *et al.* Transcription regulation of restriction-modification system Esp1396I. *Nucleic Acids Res* **37**, 3354–3366, <https://doi.org/10.1093/nar/gkp210> (2009).
- Kita, K., Tsuda, J. & Nakai, S. Y. C. EcoO109I, a regulatory protein for production of EcoO109I restriction endonuclease, specifically binds to and bends DNA upstream of its translational start site. *Nucleic Acids Res* **30**, 3558–3565 (2002).
- Vijesurrier, R. M., Carlock, L., Blumenthal, R. M. & Dunbar, J. C. Role and mechanism of action of C. PvuII, a regulatory protein conserved among restriction-modification systems. *J Bacteriol* **182**, 477–487 (2000).
- Jeltsch, A. & Pingoud, A. Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J Mol Evol* **42**, 91–96 (1996).
- Mruk, I. & Kaczorowski, T. Genetic organization and molecular analysis of the EcoVIII restriction-modification system of *Escherichia coli* E1585-68 and its comparison with isospecific homologs. *Applied and Environmental Microbiology* **69**, 2638–2650, <https://doi.org/10.1128/AEM.69.5.2638-2650.2003> (2003).

37. Mruk, I., Sektas, M. & Kaczorowski, T. Characterization of pEC156, a ColE1-type plasmid from *Escherichia coli* E1585-68 that carries genes of the EcoVIII restriction-modification system. *Plasmid* **46**, 128–139, <https://doi.org/10.1006/plas.2001.1534> (2001).
38. Mise, K. & Nakajima, K. Isolation of restriction enzyme EcoVIII, an isoschizomer of HindIII, produced by *Escherichia coli* E1585-68. *Gene* **30**, 79–85 (1984).
39. Casadaban, M. J. & Cohen, S. N. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J Mol Biol* **138**, 179–207 (1980).
40. Novy, R., Drott, D., Yaeger, K. & Mierendorf, R. Overcoming the codon bias of *E. coli* for enhanced protein expression. *InNovations* **12**, 1–2 (2001).
41. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113–130 (1986).
42. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. (1986).
43. Diederich, L., Roth, A. & Messer, W. A versatile plasmid vector system for the regulated expression of genes in *Escherichia coli*. *Biotechniques* **16**, 916–923 (1994).
44. Miller, J. H. *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., (1972).
45. Sharp, P. M. & Li, W. H. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**, 1281–1295 (1987).
46. Jansen, R., Bussemaker, H. J. & Gerstein, M. Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Res* **31**, 2242–2251 (2003).
47. Goldman, E., Rosenberg, A. H., Zubay, G. & Studier, F. W. Consecutive low-usage leucine codons block translation only when near the 5' end of a message in *Escherichia coli*. *J Mol Biol* **245**, 467–473 (1995).
48. Moffatt, B. A. & Studier, F. W. T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* **49**, 221–227, 0092-8674(87)90563-0 (1987).
49. Rosenberg, A. H., Goldman, E., Dunn, J. J., Studier, F. W. & Zubay, G. Effects of consecutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system. *J Bacteriol* **175**, 716–722 (1993).
50. Spanjaard, R. A., Chen, K., Walker, J. R. & van Duin, J. Frameshift suppression at tandem AGA and AGG codons by cloned tRNA genes: assigning a codon to argU tRNA and T4 tRNA(Arg). *Nucleic Acids Res* **18**, 5031–5036 (1990).
51. Spanjaard, R. A. & van Duin, J. Translation of the sequence AGG-AGG yields 50% ribosomal frameshift. *Proc Natl Acad Sci USA* **85**, 7967–7971 (1988).
52. Gurvich, O. L., Baranov, P. V., Gesteland, R. F. & Atkins, J. F. Expression levels influence ribosomal frameshifting at the tandem rare arginine codons AGG-AGG and AGA-AGA in *Escherichia coli*. *J Bacteriol* **187**, 4023–4032, <https://doi.org/10.1128/JB.187.12.4023-4032.2005> (2005).
53. Chen, G. F. & Inouye, M. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. *Nucleic Acids Res* **18**, 1465–1473 (1990).
54. Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* **43**, D298–299, <https://doi.org/10.1093/nar/gku1046> (2015).
55. Vos, M., Hesselman, M. C., Te Beek, T. A., van Passel, M. W. J. & Eyre-Walker, A. Rates of Lateral Gene Transfer in Prokaryotes: High but Why? *Trends Microbiol* **23**, 598–605, <https://doi.org/10.1016/j.tim.2015.07.006> (2015).
56. Treangen, T. J. & Rocha, E. P. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* **7**, e1001284, <https://doi.org/10.1371/journal.pgen.1001284> (2011).
57. Martin, W. Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *Bioessays* **21**, 99–104, 10.1002/(SICI)1521-1878(199902)21:2<99::AID-BIES3>3.0.CO;2-B (1999).
58. Boto, L. Horizontal gene transfer in evolution: facts and challenges. *Proc Biol Sci* **277**, 819–827, <https://doi.org/10.1098/rspb.2009.1679> (2010).
59. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3**, 711–721, <https://doi.org/10.1038/nrmicro1234> (2005).
60. Popa, O. & Dagan, T. Trends and barriers to lateral gene transfer in prokaryotes. *Curr Opin Microbiol* **14**, 615–623, <https://doi.org/10.1016/j.mib.2011.07.027> (2011).
61. Davids, W. & Zhang, Z. The impact of horizontal gene transfer in shaping operons and protein interaction networks—direct evidence of preferential attachment. *BMC Evol Biol* **8**, 23, <https://doi.org/10.1186/1471-2148-8-23> (2008).
62. Naito, T., Kusono, K. & Kobayashi, I. Selfish behavior of restriction-modification systems. *Science* **267**, 897–899, <https://doi.org/10.1126/science.7846533> (1995).
63. Médigue, C., Rouxel, T., Vigier, P., Hénaut, A. & Danchin, A. Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J Mol Biol* **222**, 851–856 (1991).
64. Ikemura, T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol* **146**, 1–21 (1981).
65. Medrano-Soto, A., Moreno-Hagelsieb, G., Vinuesa, P., Christen, J. A. & Collado-Vides, J. Successful lateral transfer requires codon usage compatibility between foreign genes and recipient genomes. *Mol Biol Evol* **21**, 1884–1894, <https://doi.org/10.1093/molbev/msh202> (2004).
66. Tuller, T. *et al.* Association between translation efficiency and horizontal gene transfer within microbial communities. *Nucleic Acids Res* **39**, 4743–4755, <https://doi.org/10.1093/nar/gkr054> (2011).
67. Tuller, T. Codon bias, tRNA pools and horizontal gene transfer. *Mob Genet Elements* **1**, 75–77, <https://doi.org/10.4161/mge.1.1.15400> (2011).
68. Lawrence, J. G. & Ochman, H. Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol* **44**, 383–397 (1997).
69. Hershberg, R. & Petrov, D. A. Selection on codon bias. *Annu Rev Genet* **42**, 287–299, <https://doi.org/10.1146/annurev.genet.42.110807.091442> (2008).
70. Yannai, A., Katz, S. & Hershberg, R. The Codon Usage of Lowly Expressed Genes Is Subject to Natural Selection. *Genome Biol Evol* **10**, 1237–1246, <https://doi.org/10.1093/gbe/evy084> (2018).
71. Chaney, J. L. *et al.* Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLoS Comput Biol* **13**, e1005531, <https://doi.org/10.1371/journal.pcbi.1005531> (2017).
72. Kirienko, N. V., Lepikhov, K. A., Zheleznyaya, L. A. & Matvienko, N. I. Significance of codon usage and irregularities of rare codon distribution in genes for expression of BspLU11III methyltransferases. *Biochemistry (Mosc)* **69**, 527–535 (2004).
73. McNulty, D. E. *et al.* Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. *Protein Expr Purif* **27**, 365–374 (2003).
74. Nagamalleswari, E., Rao, S., Vasu, K. & Nagaraja, V. Restriction endonuclease triggered bacterial apoptosis as a mechanism for long time survival. *Nucleic Acids Res* **45**, 8423–8434, <https://doi.org/10.1093/nar/gkx576> (2017).
75. Handa, N. & Kobayashi, I. Post-segregational killing by restriction modification gene complexes: Observations of individual cell deaths. *Biochimie* **81**, 931–938 (1999).
76. Bhattacharyya, S. *et al.* Accessibility of the Shine-Dalgarno Sequence Dictates N-Terminal Codon Bias in *E. coli*. *Mol Cell* **70**, 894–905, <https://doi.org/10.1016/j.molcel.2018.05.008> (2018).

77. Eyre-Walker, A. & Bulmer, M. Reduced synonymous substitution rate at the start of enterobacterial genes. *Nucleic Acids Res* **21**, 4599–4603 (1993).
78. Agashe, D., Martinez-Gomez, N. C., Drummond, D. A. & Marx, C. J. Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. *Mol Biol Evol* **30**, 549–560, <https://doi.org/10.1093/molbev/mss273> (2013).
79. Quax, T. E., Claessens, N. J., Söll, D. & van der Oost, J. Codon Bias as a Means to Fine-Tune Gene Expression. *Mol Cell* **59**, 149–161, <https://doi.org/10.1016/j.molcel.2015.05.035> (2015).
80. Zamora-Romo, E., Cruz-Vera, L. R., Vivanco-Dominguez, S., Magos-Castro, M. A. & Guarneros, G. Efficient expression of gene variants that harbour AGA codons next to the initiation codon. *Nucleic Acids Res* **35**, 5966–5974 (2007).
81. Kerrigan, J. J. *et al.* Frameshift events associated with the lysyl-tRNA and the rare arginine codon, AGA, in *Escherichia coli*: a case study involving the human Relaxin 2 protein. *Protein Expr Purif* **60**, 110–116, <https://doi.org/10.1016/j.pep.2008.02.016> (2008).
82. Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* **16**, 274–280, <https://doi.org/10.1038/nsmb.1554> (2009).
83. Komar, A. A., Lesnik, T. & Reiss, C. Synonymous codon substitutions affect ribosome traffic and protein folding during *in vitro* translation. *FEBS Lett* **462**, 387–391 (1999).
84. Clarke, T. F. & Clark, P. L. Increased incidence of rare codon clusters at 5' and 3' gene termini: implications for function. *BMC Genomics* **11**, 118, <https://doi.org/10.1186/1471-2164-11-118> (2010).
85. Clarke, T. F. & Clark, P. L. Rare codons cluster. *PLoS One* **3**, e3412, <https://doi.org/10.1371/journal.pone.0003412> (2008).
86. Gonzalez de Valdivia, E. I. & Isaksson, L. A. Abortive translation caused by peptidyl-tRNA drop-off at NGG codons in the early coding region of mRNA. *FEBS J* **272**, 5306–5316, <https://doi.org/10.1111/j.1742-4658.2005.04926.x> (2005).
87. Goodman, D. B., Church, G. M. & Kosuri, S. Causes and effects of N-terminal codon bias in bacterial genes. *Science* **342**, 475–479, <https://doi.org/10.1126/science.1241934> (2013).
88. Fredrick, K. & Ibb, M. How the sequence of a gene can tune its translation. *Cell* **141**, 227–229, <https://doi.org/10.1016/j.cell.2010.03.033> (2010).
89. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–258, <https://doi.org/10.1126/science.1170160> (2009).
90. Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc Natl Acad Sci USA* **110**, 14024–14029, <https://doi.org/10.1073/pnas.1301301110> (2013).
91. Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z. & Blüthgen, N. Efficient translation initiation dictates codon usage at gene start. *Mol Syst Biol* **9**, 675, <https://doi.org/10.1038/msb.2013.32> (2013).
92. Presnyak, V. *et al.* Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–1124, <https://doi.org/10.1016/j.cell.2015.02.029> (2015).
93. Zahn, K. & Landy, A. Modulation of lambda integrase synthesis by rare arginine tRNA. *Mol Microbiol* **21**, 69–76 (1996).
94. Behera, P., Kutty, V. H., Kumar, A. & Sharma, B. Changing the Codon Usage of *hfq* Gene has Profound Effect on Phenotype and Pathogenicity of *Salmonella Typhimurium*. *Curr Microbiol* **72**, 288–296, <https://doi.org/10.1007/s00284-015-0951-2> (2016).
95. Ritter, A. *et al.* The Pai-associated leuX specific tRNA^{5(Leu)} affects type 1 fimbriation in pathogenic *Escherichia coli* by control of FimB recombinase expression. *Mol Microbiol* **25**, 871–882 (1997).
96. Hou, B. *et al.* Global regulator BldA regulates morphological differentiation and lincomycin production in *Streptomyces lincolnensis*. *Appl Microbiol Biotechnol* **102**, 4101–4115, <https://doi.org/10.1007/s00253-018-8900-1> (2018).
97. Chandra, G. & Chater, K. F. Evolutionary flux of potentially bldA-dependent *Streptomyces* genes containing the rare leucine codon TTA. *Antonie Van Leeuwenhoek* **94**, 111–126, <https://doi.org/10.1007/s10482-008-9231-5> (2008).
98. Emilsson, V., Näslund, A. K. & Kurland, C. G. Growth-rate-dependent accumulation of twelve tRNA species in *Escherichia coli*. *J Mol Biol* **230**, 483–491, <https://doi.org/10.1006/jmbi.1993.1165> (1993).
99. Reece, K. S. & Phillips, G. J. New plasmids carrying antibiotic-resistance cassettes. *Gene* **165**, 141–142 (1995).
100. Mruk, I. & Kaczorowski, T. A rapid and efficient method for cloning genes of type II restriction-modification systems by use of a killer plasmid. *Applied and Environmental Microbiology* **73**, 4286–4293, <https://doi.org/10.1128/aem.00119-07> (2007).
101. Tabor, S. & Richardson, C. C. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci USA* **82**, 1074–1078 (1985).

Acknowledgements

We thank Natalia Biliska for technical assistance and Dr. Marian Sektas for critical reading of the manuscript and helpful comments. We also acknowledge Dr. Robert Lintner for his great help at the stage of manuscript editing. This work was supported by a grant from National Science Center Poland (7241/B/P01/2011/40) to I.M.

Author Contributions

I.M. and T.K. conceived the study and analyzed the data. I.M. designed the experiments. I.M. and A.W. contributed experimental data. I.M. wrote the manuscript and prepared all figures and tables. All authors reviewed and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-42311-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019