

RESEARCH

Open Access

Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications

Kinga Umenhoffer¹, Tamás Fehér¹, Gabriella Balikó¹, Ferhan Ayaydin², János Pósfai³, Frederick R Blattner^{4,5} and György Pósfai*¹

Abstract

Background: Evolvability is an intrinsic feature of all living cells. However, newly emerging, evolved features can be undesirable when genetic circuits, designed and fabricated by rational, synthetic biological approaches, are installed in the cell. Streamlined-genome *E. coli* MDS42 is free of mutation-generating IS elements, and can serve as a host with reduced evolutionary potential.

Results: We analyze an extreme case of toxic plasmid clone instability, and show that random host IS element hopping, causing inactivation of the toxic cloned sequences, followed by automatic selection of the fast-growing mutants, can prevent the maintenance of a clone developed for vaccine production. Analyzing the molecular details, we identify a hydrophobic protein as the toxic byproduct of the clone, and show that IS elements spontaneously landing in the cloned fragment relieve the cell from the stress by blocking transcription of the toxic gene. Bioinformatics analysis of sequence reads from early shotgun genome sequencing projects, where clone libraries were constructed and maintained in *E. coli*, suggests that such IS-mediated inactivation of ectopic genes inhibiting the growth of the *E. coli* cloning host might happen more frequently than generally anticipated, leading to genomic instability and selection of altered clones.

Conclusions: Delayed genetic adaptation of clean-genome, IS-free MDS42 host improves maintenance of unstable genetic constructs, and is suggested to be beneficial in both laboratory and industrial settings.

Background

Molecular mechanisms generating genetic variation, coupled with selection due to the changing environment, provide the basis for the evolution of a population. While the ability to evolve is generally a necessity for survival on the long run, it might be an undesirable feature in synthetic biological (SB) applications [1]. The SB approach is ideally based on the notion that biological parts can be precisely designed and fabricated, and their interactions can be fully predicted [2]. Newly emerging, evolved features can result in undesired genotypic and phenotypic diversions of painstakingly engineered cells. Reducing the evolutionary potential of a cell, used as an SB chassis in a controlled environment, is therefore a valid goal.

In a typical bacterial cell, IS elements generate a significant share of genetic variation. Depending on the genetic and physiological context, their contribution to gene inactivation ranges from 3.9% [3] to 98% [4]. In addition to insertional inactivation, mobile elements are also capable of enhancing deletions, inversions or duplications by providing substrates for homologous recombination [5]. Under specific conditions, inhibition or stimulation of gene expression caused by genomic relocation of IS elements can help the population to overcome nutritional stress [6,7], antibiotic inhibition [8,9] or osmotic stress [10]. Moreover, in biotechnological applications, IS-mediated genomic rearrangements can cause instability of clones carrying engineered genes. Such cloning failures are by no means rare [11-17], and examples of IS-contaminated sequence artefacts obtained from shotgun sequencing projects [18-21] also indicate that the magni-

* Correspondence: posfai@brc.hu

¹ Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, 62 Temesvári krt, H6726 Szeged, Hungary
Full list of author information is available at the end of the article

tude of the problem is not negligible. *E. coli* host strains in common use in research and biotechnology (e.g., MG1655, DH10B, BL21), as well as various natural isolates carry a large number of genomic ISes [22]. MG1655, the prototype K-12 strain, carries 45 ISes of 11 different kinds [23]. In contrast, MDS42, a reduced-genome derivative of MG1655, engineered to lose most of its genes irrelevant for laboratory applications, is free of active ISes [24]. It is noted, that a copy of IS609 is still present in MDS42. However, due to mutation resulting in a stop codon in its transposase gene (*ynck*), it displays no transposition activity (data not shown). MDS42 also lacking IS609 has recently become available commercially from Scarab Genomics.

In an earlier study [24], it was shown that a chimeric gene, constructed for vaccine development, and composed of a synthetic gene of the structural capsid protein VP60 of rabbit haemorrhagic disease (RHD) virus [25] fused to the gene of the B subunit of cholera toxin (CTX) was very unstable in conventional *E. coli* hosts. All recovered plasmids contained mutations in the *ctxvp60* ORF, virtually all of which were IS insertions. In contrast, the recombinant plasmid was stable in IS-less MDS42 host; normal yields of plasmid DNA were obtained.

We analyse here the molecular details of this extreme case of instability. We identify a hydrophobic protein as the toxic byproduct of the *ctxvp60* clone, and show that IS elements spontaneously landing in the clone relieve the cell from the stress by blocking transcription of the toxic gene. Bioinformatic analysis of raw data from various shotgun sequencing projects, where *E. coli* served as host for clone libraries, demonstrates that such IS-mediated inactivation of ectopic genes inhibiting the growth of the *E. coli* host cannot be regarded as irrelevant and rare events, but might happen in many cases, leading to genomic instability and selection of altered clones.

We show here that removal of all IS elements from the genome of a bacterial cell results in a significant increase in genomic stability and phenotypic uniformity, yielding an improved cellular chassis with reduced evolutionary potential.

Methods

Strains, plasmids, and media

Experiments were done using *E. coli* K-12 MG1655 [23] and various derivatives of it. MDS12, MDS30, and MDS42 are reduced-genome versions of MG1655, carrying 12, 30, and 42 genomic deletions, respectively [24,26]. T7 polymerase-expressing MG1655-T7 and MDS42-T7 strains will be described in detail elsewhere. Briefly, they carry an IPTG-inducible lac operator/T7 polymerase cassette replacing the *yahA-yaiL* genomic region of the parental strain. MDS42+IS1 carries a single IS1 in *yeaJ*.

Plasmids pCTX and pCTXVP60 were originally constructed by Günther M. Keil [24] (Additional file 1). Plasmid pSG1144, carrying an F-type of replicational origin, a *trfA-oriV* replicon from plasmid R2 [27], a chloramphenicol resistance gene and a T7 promoter, was obtained from Scarab Genomics (Madison, WI, USA). Inducible variants of *orf238* and *ctxvp60* were constructed by cloning the genes in the multiple cloning site of pSG1144 (Additional file 1). Genes *ctxvp60opt* and *ctxvp60dezo* were synthesized by GenScript Corporation (Piscataway, NJ, USA). Plasmids were prepared using IS-free, MDS42 host strain. Standard laboratory media LB and LB-agar plates were used during the cultivations [28]. Chloramphenicol and ampicillin were added at 25 µg/ml and 50 µg/ml final concentrations, respectively.

PCR primers

Table 1 lists the oligonucleotides used in our experiments.

Analysis of mutations

To analyze the mutational spectrum of the *ctxvp60* gene, a 2481-bp segment was amplified from the mutant cells using the primer pair T7/chi3. The amplified fragments were resolved on a 1% agarose gel and compared to a fragment generated from the wild-type template. Identical size indicated mutations affecting only one or a few nucleotides, a decrease in size or failure of amplification indicated a deletion, and an increase in size suggested an IS insertion. In case of insertion mutants, the identity and the localization of ISes were determined by PCR using IS- and plasmid-specific primer pairs. For 12 mutants, insertions sites were further analyzed by sequencing.

Growth measurements and statistics

Growth characteristics were monitored in liquid medium in 100-well Honeycomb 2 plates (Oy Growth Curves Ab, Helsinki, Finland). Growth curves were taken by following the optical densities (O.D.) at 540nm in each well using the Bioscreen C Automated Microbiology Growth Analysis System (Oy Growth Curves Ab, Helsinki, Finland). Ten parallels of resuspended colonies for each plasmid-host combination were grown to saturation in ten wells at 37°C using continuous shaking in 200 µl LB medium supplemented with 25 µg/ml chloramphenicol, if required. The *orf238* containing plasmids were grown in the presence and in the absence of IPTG. IPTG was added to induce the cultures at 0.35 O.D. in 0.6 mM final concentration. When testing the effect of pCTXVP60 on the growth of strains harboring various numbers of IS elements, the median O.D. value of the ten parallel cultures corresponding to each strain (MG1655, MDS12, MDS30 and MDS42) was calculated and plotted for every time point.

Table 1: Primers used in this study.

name	sequence (5'-3')	purpose
ORFstart	cgggatccatgctattgctgctatttc	cloning ORF238
ORFstop	cggaattctcatacaaaattcaaaatcttcag	cloning ORF238
CTXVP2	cctcctggaactggttga	amplifying/sequencing ctxvp60
CTXVP3	ggaggatctacttctgct	amplifying/sequencing ctxvp60
CVek3	gcctggtgtacgcctgaa	amplifying/sequencing ctxvp60
T7	taatacactcactataggg	amplifying/sequencing ctxvp60
chi3	gtgagttccaccagtttga	amplifying/sequencing ctxvp60
vpf1	cctgggctcattccatcagtagtagttaat	constructing ctxvp60 frame-shift
vpf2	tgatggaataggaccaggagtggtgct	constructing ctxvp60 frame-shift
IS1-1	tcgctgtcgttctca	IS1-specific PCR
IS1-2	aagccactggagcac	IS1-specific PCR
IS2-1	tcgcaggcataccatcaa	IS2-specific PCR
IS2-2	cagacgggtaacggca	IS2-specific PCR
IS3-1	agcggctgtgatacgtggt	IS3-specific PCR
IS3-2	tcatgctgtggcgacattga	IS3-specific PCR
IS5-1	gacagttcggcttctgta	IS5-specific PCR
IS5-2	gctcgtgactccacca	IS5-specific PCR
groL1	atggcgtgggtgaagaag	amplification of control <i>groL</i>
groL2	caatctgctgacggatctga	amplification of control <i>groL</i>

RNA isolation from colonies and RT PCR

After transformation with pCTXVP60, small and large colonies were collected separately from the transformation plate. Cells from small and large colonies were resuspended into fresh minimal medium and the O.D. of the suspensions were adjusted to 1.0. RNA was isolated from the suspensions using the RNEasy RNA isolation kit with RNAProtect Bacteria Reagent (Qiagen, Düsseldorf, Germany), supplemented with lysozyme from Chicken Egg White (Sigma, St. Louis, MO, USA). DNA contamination was digested using DNaseI from Bovine Pancreas (Boehringer Mannheim, Germany) for 45 min at 37°C in the presence of 10 mM Tris (Sigma, St. Louis, MO, USA) and 2.25 mM MgCl₂ (Fermentas, Vilnius, Lithuania). The reaction was stopped by adding 2.5 mM EDTA (Merck, Darmstadt, Germany) and incubation at 65°C for 10 min. cDNA was synthesized from purified- and DNase-treated RNA samples using the Transcriptor First Strand cDNA Synthesis Kit (Roche, Mannheim, Germany) with random hexanucleotide primers, according to manufacturer's instructions. The cDNA samples were used as templates in the following PCR using ORFstart and ORFstop primers, specific to the previously identified small ORF region of the *ctxvp60* transcript. Primers groL1 and groL2 were used as positive controls.

Cell imaging

Cells were incubated 30 mins. at 37°C with 10 μM nonyl acridine orange (NAO) and 10 μg/ml Ethidium bromide (EtBr). DNA staining was performed using 500 ng/ml 4',6-diamidino-2-phenylindole (DAPI) for 10 minutes after 70% ethanol fixation (30 mins) and PBS washes. Two percent agarose gel blocks (1 cm × 1 cm) were used to immobilize the cells on cover slips before microscopy observation. NAO, EtBr and DAPI were from Invitrogen, Carlsbad, CA, USA. Confocal laser scanning microscopy was performed using Olympus Fluoview FV1000 confocal laser scanning microscope (Olympus Life Science Europa GmbH, Hamburg, Germany) equipped with UPL-SAPO 60× (oil, NA:1.35) objective. Using Olympus Fluoview software (version 1.7.2.2), DAPI images were pseudo colored red.

Bioinformatic analysis of *E. coli* IS sequences in shotgun data

Shotgun sequencing data from early genome sequencing projects (Sanger method), where clone libraries of various genomes were constructed and maintained in *E. coli* host, provide a large dataset for checking the frequency of host IS contamination of the cloned sequences. Sequence reads were fetched from the NCBI Trace Archive. Bacterial and archaeal genomes were analyzed when individual

read sequences, assembled sequence, annotations, and experimental notes were available. We built two BLAST databases for each analyzed organism, one from the read sequences, and the other from the assembled genome. A representative set of *E. coli* K12 MG1655 (U00096) IS elements was compiled, and expanded by an IS10 sequence from DH10B (NC_010473). Database queries were done by nucleotide level BLAST searches. To produce a conservative estimate for the number of insertion events, stringent parameter settings (expectation value 0.1, minimum score 75, minimum run length 40 nucleotides) were used. Reads from possibly non-cloning sequencing, with "finishing", "primerwalk", "454" annotations were carefully removed.

The first pass of BLAST used IS sequence baits against shotgun read databases. This round of BLAST caught reads that contain some *E. coli* IS-like sequences. With these reads as baits, a second round of BLAST mapped the genomic sequence parts of the reads onto the assembled genomes. Using IS baits, genome segments with matches to *E. coli* IS sequences were identified next. Potentially, these segments contain endogenous IS elements (and not elements acquired in the cloning process), and were excluded from further analysis. Using the remaining segments as queries, the read databases were searched, and all the reads matching to IS-acquiring segments were identified. Finally, reads collected at this pass were BLAST-ed against the assembled genomes, to find and retain only the reciprocal best hits. This back-and-forth BLAST-ing identified genome segments from which some cloning inserts acquired *E. coli* IS elements during the shotgun sequencing process. Also, the procedure yielded the reads that genuinely map to these segments, including those that contain acquired *E. coli* IS sequences. To find the functions of the genes encoded in

these segments, GenBank annotations were parsed. The relationships of reads, IS elements, genome segments and gene annotations were overlaid in simple diagrams.

Supporting online material. Nucleotide sequences of pCTXVP60, pCTXVP60frameshift, pSG1144-*orf238*, pSG1144-*ctxvp60*, pSG1144-*ctxvp60opt*, and pSG1144-*ctxvp60dezo*, codon use of *ctxvp60opt* and *ctxvp60dezo*, as well as data related to the structure prediction of ORF238 can be found at <http://www.brc.hu/genome>, and at the end of this article (Additional files 1, 2, 3 and 4).

Results

Instability of pCTXVP60 clones propagated in wt *E. coli* MG1655

In an earlier study [24], it was found that an expression construct of a fusion gene, coding for adjuvant CTX B subunit fused to VP60, could not be stably propagated in MG1655. Upon transformation by plasmid pCTXVP60, heterogeneous colonies of transformants were obtained on agar plates (Fig. 1A). While the majority of the colonies were small, slow-growing, and translucent, a few percent showed normal growth and appearance, as compared to control pCTX clones (not shown). Upon prolonged incubation, eventually all small colonies started to develop sectors of normal growth. A preliminary analysis of the cells from normal, healthy colonies showed that the plasmids recovered had altered restriction digestion patterns, primarily due to IS element insertions in the fusion gene [24]. An extended analysis of 100 colonies by PCR spanning the fusion gene confirmed the extreme instability of the plasmid. A vast majority, 92%, of the large colonies carried plasmids with IS insertions, while 8% displayed deletions or no alteration detectable by PCR. Insertion events were due to IS1, IS3, and IS5 translocac-

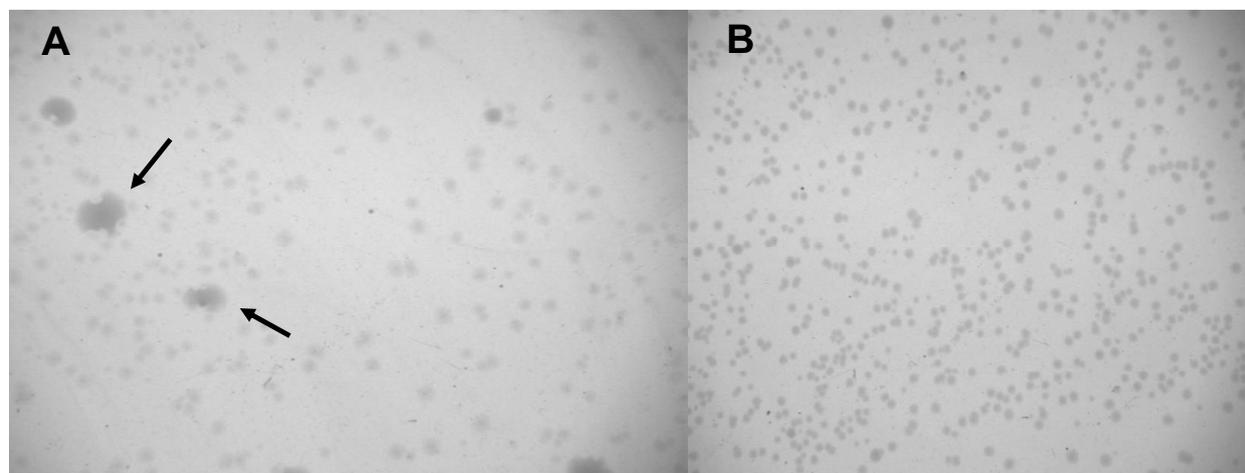


Figure 1 Transformants of pCTXVP60/MG1655 (A) and pCTXVP60/MDS42 (B) on LB plates containing chloramphenicol. Arrows indicate sectors of colonies that have arisen by outgrowth of cells harboring mutated plasmids.

tions, as detected by using IS-specific PCR primers. Insertion sites were determined by sequencing 12 clones. All IS insertions occurred in the 5' third of the fusion gene, in the *ctx* part or near the 5' end of *vp60*. Results of the analysis are summarized in Fig. 2.

Similar heterogeneity of pCTXVP60 transformants was observed in other regular *E. coli* hosts, including DH5 α , DH10B or C600. In contrast, transformants of IS-less MDS42 [24] displayed a nearly uniform morphology (Fig. 1B) on LB plates. Colonies generally displayed moderately retarded growth, and only < 0.1% of the colonies was larger, showing normal growth. Generally, plasmid DNA recovered from cultures using MDS42 host yielded unaltered restriction digestion pattern and nucleotide sequence (data not shown).

Growth retardation effect of pCTXVP60 is due to an artificial byproduct, a Leu-rich ORF

The synthetic fusion gene *ctxvp60* contains a large number of rare Arg (1 AGG, 21 AGA) codons. It was assumed that translation of *ctxvp60* exerts a toxic effect on the cell by exhausting the rare tRNA_{Arg} pool and thus rendering it unavailable for the synthesis of essential proteins. To test this hypothesis, two new versions of *ctxvp60* were synthesized. Version *ctxvp60*opt was codon-optimized for *E. coli* and carried no rare codons, while *ctxvp60*dezo_{opt} contained the rare AGG codons for all the 22 arginines (Additional files 2 and 3). Both versions were cloned in

pSG1144 (Additional file 1), using MG1655-T7 host. Surprisingly, induced expression of both constructs, optimized and deoptimized, had only a minor negative effect on the growth of the cell, comparable to that seen with the vector plasmid (Fig. 3).

On the other hand, a frame-shift mutation of *ctxvp60*, introduced in a position near the 5' end of the *vp60* part (Fig. 2; Additional file 1), did not relieve the growth retardation effect of the plasmid (data not shown). All these results indicate that neither the CTXVP60 protein *per se*, nor the supposed rare tRNA depletion phenomenon is the cause of the toxic effect of pCTXVP60.

A thorough inspection of the original *ctxvp60* gene revealed the presence of a 238 aa ORF (ORF238) out of the original reading frame. ORF238 spans the joint between *ctx* and *vp60* and extends into the 5' part of *vp60* (for coordinates, see Fig. 2). Translation of ORF238 results in an extremely Leu-rich (102 Leu residues) protein (Additional file 4). The hydrophobic nature of the protein and its four putative transmembrane domains (predicted with HMMTOP [29] and DAS-TMfilter [30,31]) strongly suggest that it integrates in the cell membranes, and might be the cause of the toxic effect (Additional file 4).

To test the hypothesis that ORF238 was the culprit of the growth defect, it was cloned as an inducible construct in plasmid pSG1144. Upon induction, expression of ORF238 resulted in a phenotype identical to that of the

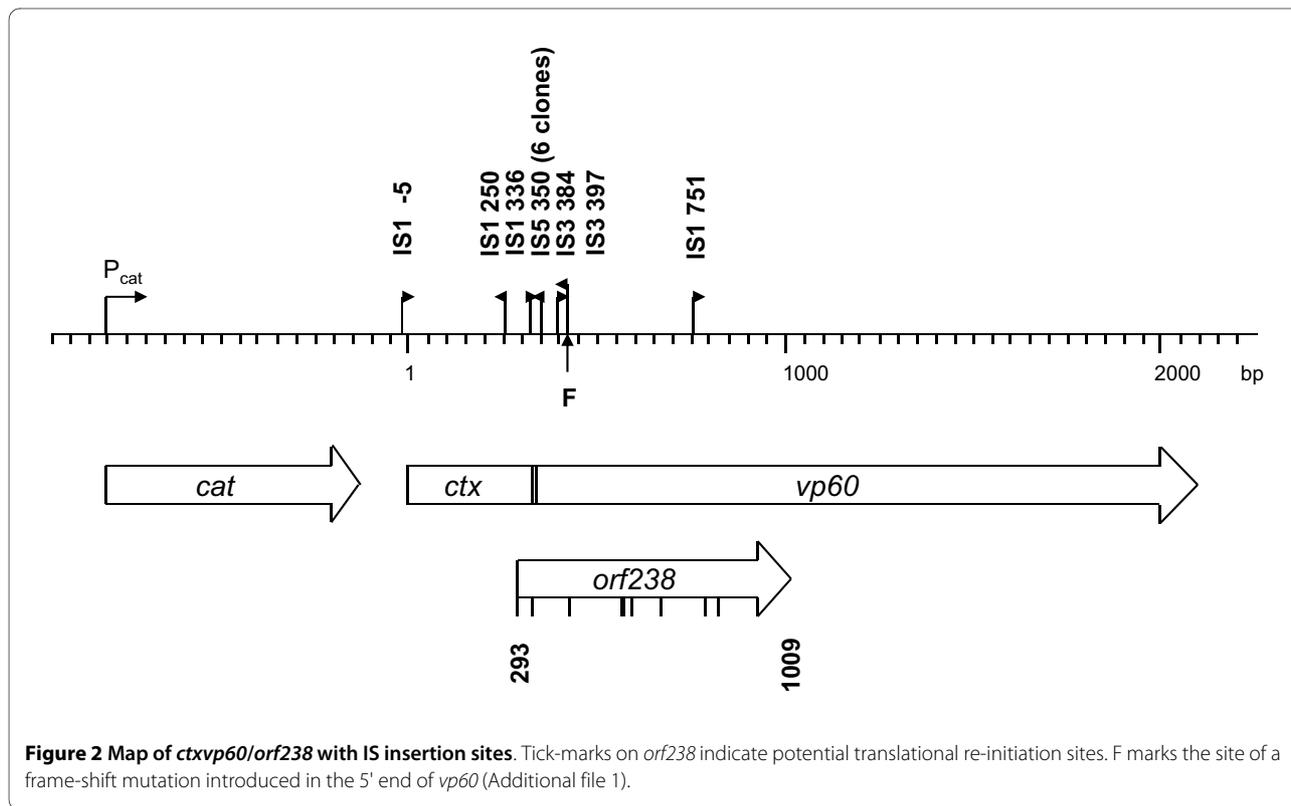
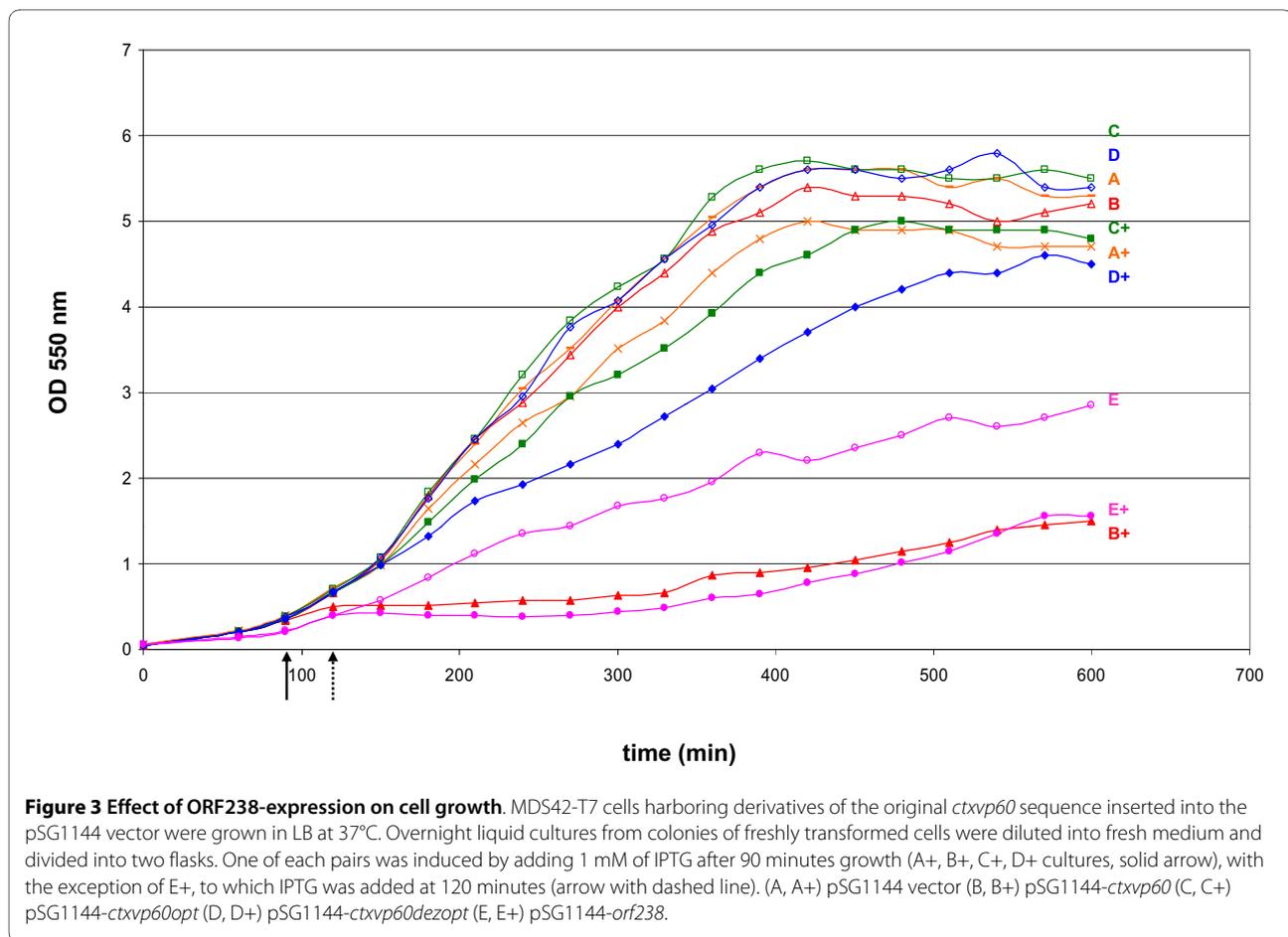


Figure 2 Map of *ctxvp60/orf238* with IS insertion sites. Tick-marks on *orf238* indicate potential translational re-initiation sites. F marks the site of a frame-shift mutation introduced in the 5' end of *vp60* (Additional file 1).



original *ctxvp60* cloned in the same plasmid. Growth retardation of the culture, due to induction of ORF238, is demonstrated in Fig. 3.

Confocal laser scanning microscopy imaging of the cells expressing either CTXVP60 or ORF238 confirmed its inhibitory role. Nucleic acid staining (Ethidium bromide or DAPI) combined with differential interference contrast (DIC) imaging and membrane staining (nonyl acridine orange) revealed that cell division is impaired, resulting in slow-growing, abnormally long cells with aberrant nucleoids (Fig. 4B, E, F). Untransformed (MG1655) or uninduced ORF238-harboring cells, on the other hand, displayed normal phenotype and size distribution (Fig 4A, C, D).

ISes integrated in *ctxvp60* block transcription of the Leu-rich ORF

All IS integrations, found in pCTXVP60, occurred either upstream or in the 5' end region of ORF238. It was assumed that ISes landing in this region relieve the cell from stress by blocking transcription of ORF238. To prove this, we showed that RNA prepared from cells harboring pCTXVP60 contained transcripts of the toxic gene detectable by RT-PCR. However, when picking a

mutant with an IS1 or IS5 inserted into the 5' end of *ctxvp60*, transcripts of downstream sequences could not be detected (Fig. 5).

Growth dynamics and evolution of strains harboring pCTXVP60 is related to the number of genomic ISes

Growth characteristics of transformants of toxic and mutant plasmids were monitored in liquid medium. Ten parallels originating from 10 colonies were grown for each plasmid-host combination, and automatic O.D.-readings were taken in a Bioscreen C instrument (Fig. 6).

Initial growth of MG1655/pCTXVP60 stopped at ~O.D. = 0.2-0.5. However, after prolonged incubation, individual cultures resumed growth at various time points, and reached a final density comparable to the MG1655/pCTX control (Fig. 6A, B). Plasmids isolated from these outgrown cultures almost invariably contained ISes inserted in pCTXVP60. Reinoculation of the outgrown cultures into fresh medium resulted in uninterrupted, normal growth (Fig. 6C). The results were consistent with the observed growth characteristics of transformants on agar plates: i) pCTXVP60 causes growth retardation, ii) in a stochastic manner, a fraction of the cells picks up an IS insertion in *ctxvp60*, due to the

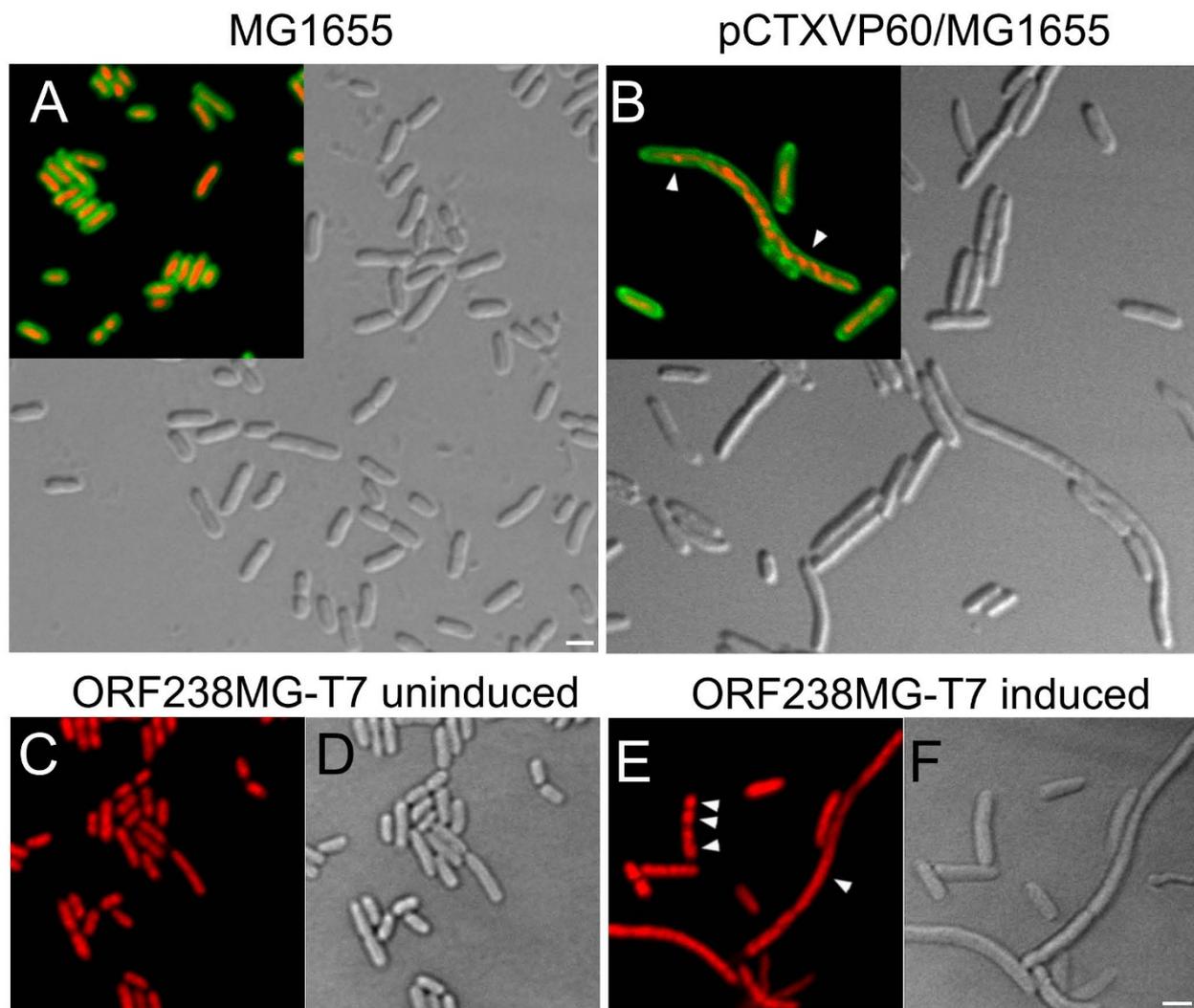


Figure 4 Fluorescence microscopy images (A-F) of cells harboring control and toxic plasmids. Plasmids and host cells are indicated for each panel. While toxic constructs pCTXVP60 (B) or induced ORF238 (E,F) give rise to abnormally long cells, untransformed MG1655 cells (A) and uninduced ORF238 (C,D) cells displayed no defects. Insets of (A) and (B) show nonyl acridine orange-stained (green) membranes superimposed onto ethidium bromide stained (red) nucleoids. DAPI stained nucleoids (pseudocolored red) are shown in (C) and (E). Arrowheads in (B) and (E) indicate abnormal, stretched or multiple nucleoids in undivided cells. Scalebars: 2 μ m.

mobility of genomic ISes, iii) cells harboring an IS-inactivated ORF238 resume normal growth and quickly become dominant in the culture.

In contrast, growth of MDS42/pCTXVP60 displayed a nearly uniform pattern. Initial growth of the cultures stopped at \sim O.D. = 0.3, and only one culture of 10 parallels grew eventually to high density (Fig. 6D). Plasmid isolated from this dense culture revealed a deletion in *ctxvp60*. Results indicate that the primary route to inactivation of the toxic gene is via IS translocation.

Next, strains representing the various stages of the genome reduction process (MG1655, MDS12, MDS30, MDS42), and thus harboring various numbers of ISes (44, 25, 5, and 0, respectively) were transformed with

pCTXVP60. The median growth curves observed for the transformants seemed to support the notion that the time needed for evolution of a non-toxic plasmid variant in the culture correlated with the number of ISes in the host genome (Fig. 6E). Even a single *IS1*, present in the host genome, had a marked effect on growth dynamics, and accelerated the evolution of fast-growing variants (Fig. 6F), as compared to the IS-less, essentially isogenic host (MDS42) (Fig. 6D).

Bioinformatic analysis of *E. coli* IS sequences in shotgun sequencing data

To test whether IS-mediated instability of cloned sequences could be more wide-spread than anticipated, a

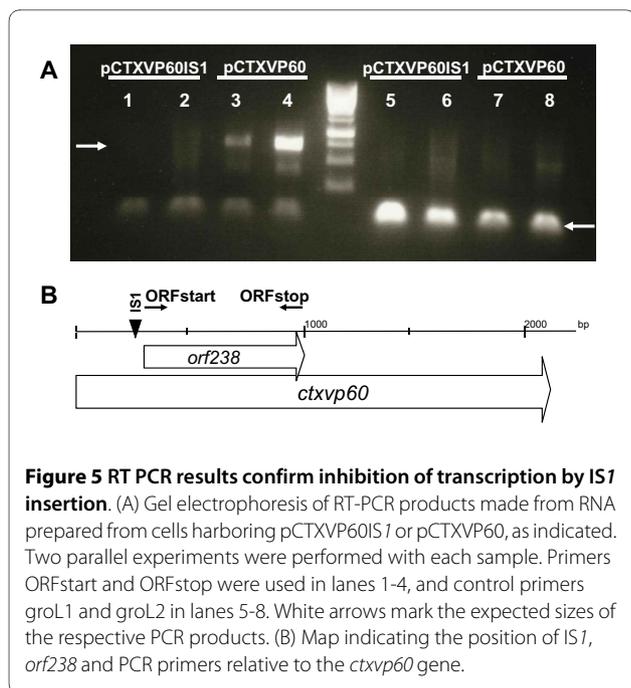


Figure 5 RT-PCR results confirm inhibition of transcription by IS1 insertion. (A) Gel electrophoresis of RT-PCR products made from RNA prepared from cells harboring pCTXVP60IS1 or pCTXVP60, as indicated. Two parallel experiments were performed with each sample. Primers ORFstart and ORFstop were used in lanes 1-4, and control primers groL1 and groL2 in lanes 5-8. White arrows mark the expected sizes of the respective PCR products. (B) Map indicating the position of IS1, *orf238* and PCR primers relative to the *ctxvp60* gene.

bioinformatic analysis of raw data from early genome sequencing projects was performed. Protocols for shotgun genome sequencing by the Sanger method started by the insertion of fragments from a target genome into a plasmid vector, followed by transformation into *E. coli* cells. Next, transformed cells were grown first on solid medium, then in liquid cultures, prior to processing. Enriched target fragments were then sequenced, and overlapping sequence reads were assembled into contigs. The middle, cell growing steps can be viewed upon as some biological test of *E. coli*'s tolerance to the introduction of foreign DNA, consequently, read data from the thousands of clones can be mined for signs of the host's reaction to foreign DNA segments [32]. In this study we searched the read sets for the appearance of *E. coli* IS elements in raw (prior to their assembly into target genome sequences) sequence reads.

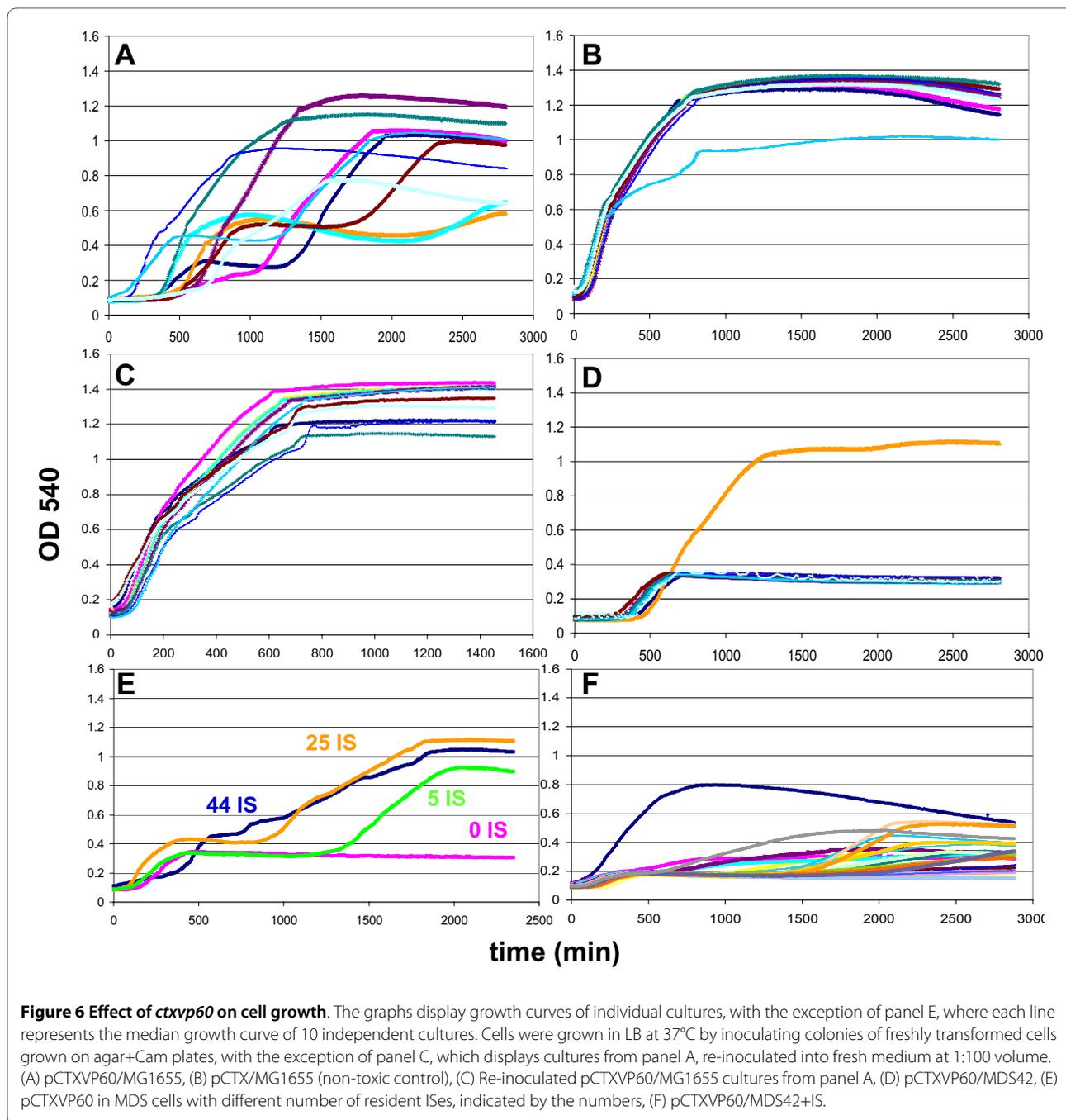
A typical bacterial shotgun library for Sanger sequencing would comprise $2-4 \times 10^4$ clones, which are grown for about 20 generations on solid medium, followed by another 10 generations of growth in liquid cultures, prior to processing. Considering the typical spontaneous, combined transposition rate of 10^{-8} /gene/generation of resident ISes [9], the chances of sequencing a spontaneously arisen insertion mutant is about 1.2×10^{-2} ($=30 \times 4 \times 10^4 \times 10^{-8}$). This means that approximately one in 100 sequencing projects would yield a gene interrupted by an IS of the cloning host. Higher incidence of host ISes in the sequence reads would likely indicate selection events preferring the IS-inserted clones during cultivation.

Data from 295 shotgun genome sequencing projects were downloaded and analyzed. The sets typically contain 50-70 thousand reads. Of the ~ 18 million reads, a total of 22 thousand match some *E. coli* IS sequence within the strict expectation, score, and run length thresholds. While 166 sets contain one or more reads with IS sequences, 129 sets are free of such reads. Of the 295 analyzed organisms 62 have complete (assembled) genome sequence. The number of organisms possessing both IS- containing reads and an assembled genome (intersection of the 166 set and the 62 set) is 30. In some IS-containing reads chromosomal sequence was not detectable, while in others IS elements seemed to have cloning vector origins (e.g., are spliced with *E. coli lacZ*), these have been weeded out. To avoid the difficulties of distinguishing true cloning related *E. coli* IS insertions from genomic self-rearrangements, organisms which have their own IS elements were excluded. After these filtering steps 14 shotgun sequencing sets remain, which possess a total of 109 IS-containing reads. IS10 and IS1 are the dominating invaders (they show up 68 and 30 times, respectively), but six other IS types were also found (IS2, IS3, IS4, IS5, IS150, IS186). Examples of IS elements appearing in sequence reads are shown in Fig. 7 and Additional file 5.

Discussion

Synthetic biology is envisioned as the art of predictable assembly of off-the-shelf biological parts and devices with standard connections in a robust biological chassis (such as *E. coli*), using engineering principles [33,34]. To achieve such a goal, a major challenge to overcome is the intrinsic genetic instability of the living systems. Cells engineered to generate useful biological products almost invariably suffer a fitness loss, compared to wt. This can lead to selected accumulation of mutations, produced by the cell's mutation-generating mechanisms, inactivating the disabling, engineered parts, and compromising productivity. A possible solution to the problem is rendering certain metabolic pathways essential both for the desired product and for survival by interrupting all pathways that would allow the cellular biochemical network to operate with decreased production yields [35,36]. A complementing approach can be the reduction of the cell's evolutionary potential by genetic disabling of the mutation-generating mechanisms.

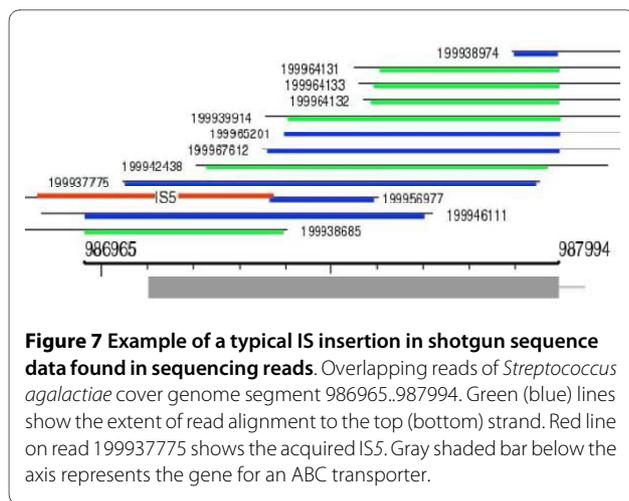
In *E. coli*, a significant part of the mutation spectrum is due to genomic rearrangements, caused by ISes. In an earlier study, we have shown that in unstressed cells, 20-25% of all mutations in an unselected gene (*cycA*) are due to IS hopping [24]. In stressed cells, the share of ISes in the total mutational load can even be higher. It was found that overexpression of CAT resulted in increased IS



translocation frequency, manifested in a 4-fold elevated level of IS-mediated inactivation of *cycA*. The share of IS-generated mutations in adaptation can be critical in some settings. It was shown that genetic adaptation to salicin carbon source proceeded primarily via IS mutagenesis [4,6], and was significantly reduced in an IS-less host [24].

Developed for animal vaccine production, fusion gene *ctxvp60* was inserted downstream of the vector's cloramphenicol resistance gene and transcribed from the *cam* promoter. This proved to be extremely unstable in ordinary IS-containing *E. coli* hosts. Almost invariably, clones

with ISes in the 5' end of *ctxvp60* were obtained from cells transformed with the plasmid. Analysis of the clones revealed that the toxic component of the clone was a Leu-rich peptide, the translated product of a secondary reading frame. Interestingly, *orf238* coding for the Leu-rich peptide is an artificial gene, obtained unintentionally by designing and synthesizing a plant host-optimized variant of *ctxvp60*. The cause of instability lies in the dynamic process of stress, mutagenesis and selection. The translated product of *orf238* interferes with host cell division and caused severe growth retardation. In this slow-grow-



ing population, random IS hopping, normally at frequencies $\sim 10^{-8}$ /gene/generation [9] but potentially elevated several-fold during stress [24], can produce clones with ISes interrupting sequences upstream of *orf238*. IS-inserted clones that stopped expressing ORF238, resumed normal growth, and rapidly became dominant in the culture. Insertions of *IS1* and *IS3* were found at multiple locations in both orientations, while *IS5* was only seen at a single locus, corresponding to its integration hotspot "CWAR" [37], in opposite orientation only. Why is IS insertion the favored mode of inactivation of ORF238? We propose that ISes, due to their relatively large size and complex composition, are particularly effective at disrupting genes, either directly or indirectly. In a fast-growing pCTXVP60 mutant, we showed that *IS1* acted by blocking transcription of *orf238*. Although in some cases insertion of ISes can generate outward-pointing transcription, expressing downstream genes [38], in this instance, apparently, it is the transcription-disrupting role that comes into play.

In principle, point-mutations (such as promoter mutations, active-site mutations, or frame-shifts) can also inactivate a gene. In fact, being generally more frequent than IS translocations, they might eventually be dominantly selected. In this particular case, however, point-mutations are rarely selected, because i) the promoter must be preserved to activate the selection marker, *cam*, and ii) single point-mutations are unlikely to eliminate the membrane-disrupting effect of the relatively large hydrophobic Leu-rich protein, iii) frame-shift mutations are presumably relatively ineffective, due to the presence of several potential in-frame re-initiation sites with AUG codons and ribosome-binding sites in ORF238 (Fig. 2). Deletions, another type of potentially effective, disruption mutation, are also rarely selected, presumably due to the absence of deletion-stimulating repeat sequences in pCTXVP60.

Sequence reads from shotgun sequencing libraries provide a convenient, large database for checking how widespread the emergence of ISes in cloned sequences is. We have calculated that, considering typical shotgun library construction/processing steps and background transposition rates of resident ISes of the cloning host, one would expect practically no host ISes to emerge in gene-sized shotgun sequence reads. However, the survey of a large set of shotgun sequence reads revealed quite a number of host ISes in the cloned segments. This elevated rate could be explained by processes analogous to those seen when cloning *ctxvp60*. In cases the cloned segment imposes stress on the host, individual clones grown for DNA preparation for sequencing can become dominated by IS-inserted variants, due to selection of the faster-growing mutants. This in turn could allow the IS elements of the cloning host to appear in the sequencing reads.

The very stringent criteria, by which we identify host ISes in shotgun sequence reads, probably underestimates the real number of IS transpositions. Furthermore, in addition to insertions, other sequence alterations (deletions and rearrangements) can also be induced by ISes. The occurrence and identification of IS-positive segments may depend on other factors as well, that are related to experimental procedures and to data collection protocols. Genes from taxonomically distant organisms may not get expressed in an *E. coli* host and would be less likely to cause growth retardation, thus, their IS-inserted mutants will not be subjects of selection. Information regarding the handling of clones and the construction of databases is sometimes unavailable, in other times incongruous (annotation is not uniform, same data types may refer to different content). It seems likely, that most datasets have probably undergone some computational pre-filtering, thereby IS hopping is masked in a fraction of the datasets. These considerations suggest, that *E. coli* IS-mediated events are not rare in cloning experiments, and underscore the significance of using IS-free MDS host in large-scale cloning projects.

The contribution of various IS elements to the detected transposition events do not necessarily reflect their relative transposition-activities. For example, despite the fact that *IS150* was shown to be responsible for the majority of rearrangements over a 10,000-generation evolutionary experiment [39], as well as during repeated freeze-thaw cycles [40], its highly specific target site most probably does not allow its frequent occurrence in mutant-selection mechanisms described above. On the other hand, the promiscuous target selection of *IS1*, in addition to its high transposition-activity explains its high ratio among the insertion-mutants of *orf238*, *ebgR* [7], *bglR* [4], *cycA* [9], *tonB* [8] in evolutionary experiments [39,40] as well as in the output of our bioinformatic analysis. The frequent occurrence of *IS10* could seem surprising, since it

is not native to most K-12 strains. It was introduced accidentally in the 1970's through the use of Tn10 as a genetic transfer agent [41]. Considering our findings, together with its pronounced transposition activity described earlier [8,42-44], one might call it an invasive species of IS elements.

It is tempting to speculate that ISes, besides their parasitic and occasionally beneficial roles, could have a protective effect in natural settings (e.g., alleviating the negative consequences of horizontal gene transfer, HGT, on the host). However, this is an unlikely scenario: it does not seem economical to maintain a general, IS-based system (potentially deleterious itself) to alleviate harmful gene transfer events. Cells suffering a HGT event with negative consequences may simply become extinct, while the rest of the population would survive.

Conclusions

Engineered genetic constructs, toxic for the *E. coli* K-12 host cell, may accumulate neutralizing, IS-mediated mutations, and the altered clones can be rapidly selected by their faster growth, leading to undesired genotypic and phenotypic changes. We show here that reducing the evolutionary potential of the host cell by removal of all genomic IS elements can result in significant stabilization of otherwise unstable clones. It is interesting to note, moreover, that reduced-genome MDS42 appears to possess a generally lowered propensity for recombination [45], further reducing the evolutionary potential of the strain.

Delaying the genetic adaptation in clean-genome MDS42, serving as host in SB applications, is suggested to be beneficial in both laboratory and industrial settings.

Additional material

Additional file 1 Map and sequences of pCTXVP60, pCTXVP60frameshift, pSG1144-orf238, pSG1144-ctxvp60, pSG1144-ctxvp60opt, pSG1144-ctxvp60dezoft.

Additional file 2 Translation and codon usage of ctxvp60opt gene.

Additional file 3 Translation and codon usage of ctxvp60dezoft gene.

Additional file 4 Structure predictions for ORF238.

Additional file 5 Relationship of reads, IS elements, genome segments and gene annotations.

Competing interests

FRB has a financial interest in Scarab Genomics LLC.

Authors' contributions

KU carried out most cloning, transformation and overexpression experiments, as well as growth measurements. TF participated in cloning, strain development, database and literature searches, acquisition of funding and completion of the manuscript. GB performed DNA and protein sequence analysis of ctxvp60, database searches and growth rate and protein overexpression studies. FA carried out all the experiments concerning imaging. JP was responsible for the bioinformatic analysis of raw sequencing data. FRB provided intellectual help, as well as critical reading and correction of the manuscript. GP conceived the research, obtained funding, coordinated the experiments and drafted the

manuscript. All authors have read and approved the final version of the manuscript.

Acknowledgements

We thank Csaba Pál for critical reading of the manuscript and Bálint Csörgő, Gábor Draskovits, Zsuzsa Gyrrfy, Ildikó Karcagi, Agnes Szalkanovics and Edit Timár for technical assistance. This project was supported by the Hungarian Research Fund (OTKA K43260 and OTKA PD72719) and the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 212894. TF is supported by the Bolyai Foundation.

Author Details

¹Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, 62 Temesvári krt, H6726 Szeged, Hungary, ²Cellular Imaging Laboratory, Biological Research Center of the Hungarian Academy of Sciences, 62 Temesvári krt, H6726 Szeged, Hungary, ³New England Biolabs, Inc, 240 County Rd, Ipswich, MA, 01938, USA, ⁴Scarab Genomics LLC, 1202 Ann Str, Madison, WI 53713, USA and ⁵Department of Genetics, University of Wisconsin, 425-G Henry Mall, Madison, WI 53706, USA

Received: 15 February 2010 Accepted: 21 May 2010

Published: 21 May 2010

References

1. Keasling JD: **Synthetic biology for synthetic chemistry.** *Acs Chemical Biology* 2008, **3**:64-76.
2. Endy D: **Foundations for engineering biology.** *Nature* 2005, **438**:449-453.
3. Halliday JA, Glickman BW: **Mechanisms of spontaneous mutation in DNA repair-proficient Escherichia coli.** *Mutat Res* 1991, **250**:55-71.
4. Hall BG: **Activation of the bgl operon by adaptive mutation.** *Mol Biol Evol* 1998, **15**:1-5.
5. Zeyl C, Bell G: **Symbiotic DNA in eukaryotic genomes.** *Trends in Ecology & Evolution* 1996, **11**:10-15.
6. Reynolds AE, Felton J, Wright A: **Insertion of DNA Activates the Cryptic bgl Operon in E. coli K-12.** *Nature* 1981, **293**:625-629.
7. Hall BG: **Spectra of spontaneous growth-dependent and adaptive mutations at ebgr.** *Journal of Bacteriology* 1999, **181**:1149-1155.
8. Kitamura K, Torii Y, Matsuoka C, Yamamoto K: **DNA sequence changes in mutations in the tonB gene on the chromosome of Escherichia coli K12: insertion elements dominate the spontaneous spectra.** *Jpn J Genet* 1995, **70**:35-46.
9. Fehér T, Cseh B, Umenhoffer K, Karcagi I, Posfai G: **Characterization of cycA mutants of Escherichia coli. An assay for measuring in vivo mutation rates.** *Mutat Res* 2006, **595**:184-190.
10. Stoebel DM, Hokamp K, Last MS, Dorman CJ: **Compensatory evolution of gene regulation in response to stress by Escherichia coli lacking RpoS.** *PLoS Genet* 2009, **5**:e1000671.
11. Muller J, Reinert H, Malke H: **Streptokinase mutations relieving Escherichia coli K-12 (prlA4) of detriments caused by the wild-type skc gene.** *J Bacteriol* 1989, **171**:2202-2208.
12. Rawat P, Kumar S, Pental D, Burma PK: **Inactivation of a transgene due to transposition of insertion sequence (IS136) of Agrobacterium tumefaciens.** *J Biosci* 2009, **34**:199-202.
13. Blumenthal RM, Gregory SA, Cooperider JS: **Cloning of a Restriction-Modification System from Proteus vulgaris and its Use in Analyzing a Methylase-Sensitive Phenotype in Escherichia coli.** *Journal of Bacteriology* 1985, **164**:501-509.
14. Fernandez C, Larhammar D, Servenius B, Rask L, Peterson PA: **Spontaneous Insertions Into Cosmid Vector - A Warning.** *Gene* 1986, **42**:215-219.
15. Nakamura K, Inouye M: **Inactivation of the Serratia marcescens gene for the lipoprotein in Escherichia coli by insertion sequences, IS1 and IS5; sequence analysis of junction points.** *Mol Gen Genet* 1981, **183**:107-114.
16. Rood JJ, Sneddon MK, Morrison JF: **Instability in tyrR strains of plasmids carrying the tyrosine operon: isolation and characterization of plasmid derivatives with insertions or deletions.** *J Bacteriol* 1980, **144**:552-559.
17. Prather KL, Edmonds MC, Herod JW: **Identification and characterization of IS1 transposition in plasmid amplification mutants of E. coli clones producing DNA vaccines.** *Appl Microbiol Biotechnol* 2006, **73**:815-826.

18. Astua-Monge G, Lyznik A, Jones V, Mackenzie SA, Vallejos CE: **Evidence for a prokaryotic insertion-sequence contamination in eukaryotic sequences registered in different databases.** *Theor Appl Genet* 2002, **104**:48-53.
19. Binns M: **Contamination of DNA database sequence entries with Escherichia coli insertion sequences.** *Nucleic Acids Research* 1993, **21**:779.
20. Kovarik A, Matzke MA, Matzke AJ, Koulakova B: **Transposition of IS10 from the host Escherichia coli genome to a plasmid may lead to cloning artefacts.** *Mol Genet Genomics* 2001, **266**:216-222.
21. Hill F, Gemund C, Benes V, Ansorge W, Gibson TJ: **An estimate of large-scale sequencing accuracy.** *Embo Reports* 2000, **1**:29-31.
22. Sawyer SA, Dykhuizen DE, DuBose RF, Green L, Mutangadura-Mhlanga T, Wolczyk DF, Hartl DL: **Distribution and abundance of insertion sequences among natural isolates of Escherichia coli.** *Genetics* 1987, **115**:51-63.
23. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, et al.: **The complete genome sequence of Escherichia coli K-12.** *Science* 1997, **277**:1453.
24. Pósfai G, Plunkett G, Fehér T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, et al.: **Emergent properties of reduced-genome Escherichia coli.** *Science* 2006, **312**:1044-1046.
25. Meyers G, Wirblich C, Thiel HJ: **Rabbit hemorrhagic disease virus--molecular cloning and nucleotide sequencing of a calicivirus genome.** *Virology* 1991, **184**:664-676.
26. Kolisnychenko V, Plunkett G, Herring CD, Feher T, Posfai J, Blattner FR, Posfai G: **Engineering a reduced Escherichia coli genome.** *Genome Res* 2002, **12**:640-647.
27. Wild J, Hradecna Z, Posfai G, Szybalski W: **A broad-host-range in vivo pop-out and amplification system for generating large quantities of 50- to 100-kb genomic fragments for direct DNA sequencing.** *Gene* 1996, **179**:181-188.
28. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning. A Laboratory Manual* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1987.
29. Cserzr M, Eisenhaber F, Eisenhaber B, Simon I: **On filtering false positive transmembrane protein predictions.** *Protein Eng* 2002, **15**:745-752.
30. Tusnády GE, Simon I: **Principles governing amino acid composition of integral membrane proteins: application to topology prediction.** *J Mol Biol* 1998, **283**:489-506.
31. Tusnády GE, Simon I: **The HMMTOP transmembrane topology prediction server.** *Bioinformatics* 2001, **17**:849-850.
32. Zheng Y, Posfai J, Morgan RD, Vincze T, Roberts RJ: **Using shotgun sequence data to find active restriction enzyme genes.** *Nucleic Acids Res* 2009, **37**:e1.
33. Heinemann M, Panke S: **Synthetic biology - putting engineering into biology.** *Bioinformatics* 2006, **22**:2790-2799.
34. Canton B, Labno A, Endy D: **Refinement and standardization of synthetic biological parts and devices.** *Nature Biotechnology* 2008, **26**:787-793.
35. Pharkya P, Burgard AP, Maranas CD: **Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock.** *Biotechnol Bioeng* 2003, **84**:887-899.
36. Trinh CT, Unrean P, Srien F: **Minimal Escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses.** *Applied and Environmental Microbiology* 2008, **74**:3634-3643.
37. Engler JA, van Bree MP: **The nucleotide sequence and protein-coding capability of the transposable element IS5.** *Gene* 1981, **14**:155-163.
38. Galas DJ, Chandler M: **Bacterial Insertion Sequences.** In *Mobile DNA* Edited by: Berg DE, Howe MM. Washington, D.C.: American Society for Microbiology; 1989:109-162.
39. Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M: **Long-term experimental evolution in Escherichia coli. IX. Characterization of insertion sequence-mediated mutations and rearrangements.** *Genetics* 2000, **156**:477-488.
40. Sleight SC, Orlic C, Schneider D, Lenski RE: **Genetic basis of evolutionary adaptation by Escherichia coli to stressful cycles of freezing, thawing and growth.** *Genetics* 2008, **180**:431-443.
41. Durfee T, Nelson R, Baldwin S, Plunkett G, Burland V, Mau B, Petrosino JF, Qin X, Muzny DM, Ayele M, et al.: **The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse.** *J Bacteriol* 2008, **190**:2597-2606.
42. Szeverényi I, Hodel A, Arber W, Olasz F: **Vector for IS element entrapment and functional characterization based on turning on expression of distal promoterless genes.** *Gene* 1996, **174**:103-110.
43. Skaliter R, Eichenbaum Z, Shwartz H, Ascarelli-Goell R, Livneh Z: **Spontaneous transposition in the bacteriophage lambda cro gene residing on a plasmid.** *Mutat Res* 1992, **267**:139-151.
44. Bender J, Kuo J, Kleckner N: **Genetic evidence against intramolecular rejoining of the donor DNA molecule following IS10 transposition.** *Genetics* 1991, **128**:687-694.
45. Chakiath CS, Esposito D: **Improved recombinational stability of lentiviral expression vectors using reduced-genome Escherichia coli.** *Biotechniques* 2007, **43**:466-470.

doi: 10.1186/1475-2859-9-38

Cite this article as: Umenhoffer *et al.*, Reduced evolvability of Escherichia coli MDS42, an IS-less cellular chassis for molecular and synthetic biology applications *Microbial Cell Factories* 2010, **9**:38

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

