

Defensive Function of Transposable Elements in Bacteria

Catherine Fan,^{†,‡} Yin-Hu Wu,^{‡,†,‡} Christoph M. Decker,[§] Reza Rohani,[†] Manuela Gesell Salazar,[§] Hua Ye,[†] Zhanfeng Cui,[†] Frank Schmidt,^{§,||} and Wei E. Huang^{*,†,‡}

[†]Department of Engineering Science, University of Oxford, Parks Road, OX1 3PJ Oxford, United Kingdom

[‡]Environmental Simulation and Pollution Control State Key Joint Laboratory, State Environmental Protection Key Laboratory of Microorganism Application and Risk Control (SMARC), School of Environment, Tsinghua University, Beijing 100084, PR China

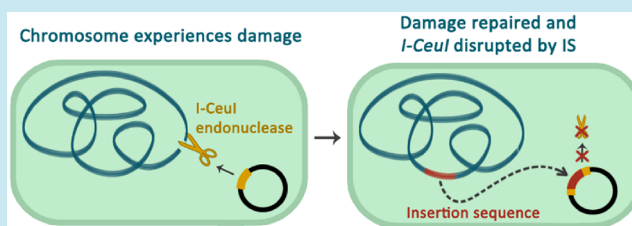
[§]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Felix-Hausdorff-Str. 8, 17475 Greifswald, Germany

^{||}Proteomics Core, Weill Cornell Medicine-Qatar, Education City, PO 24144 Doha, Qatar

S Supporting Information

ABSTRACT: It has been widely debated whether transposable elements have a positive or a negative effect on their host cells. This study demonstrated that transposable elements, specifically insertion sequences (ISs), can adopt a defensive role in *Escherichia coli*. In three different *E. coli* strains (S17, DH5 α , and Nissle 1917), IS1 and IS10 rapidly disrupted the *I-CeuI* gene (encoding I-CeuI endonuclease) on the plasmid pLO11-I-CeuI as early as the first generation, despite the gene-circuit being under control of an arabinose promoter. Proteomics analysis showed that the protein abundance profile of *E. coli* DH5 α with pLO11-I-CeuI in the fifth generation was nearly opposite to that of control strain (*E. coli* with pLO11, no I-CeuI). The DNA damage caused by the leaky expression of *I-CeuI* was enough to trigger a SOS response and alter lipid synthesis, ribosomal activity, RNA/DNA metabolism, central dogma and cell cycle processes in *E. coli* DH5 α . After the ISs disrupted the expression of *I-CeuI*, cells fully recovered by the 31st generation had a protein abundance profile similar to that of the control strain. This study showed that ISs readily mutated a harmful gene which subsequently restored host fitness. These observations have implications for the stability of designed gene circuits in synthetic biology.

KEYWORDS: transposable elements, insertion sequence, IS1, IS10, *Escherichia coli*, stress, proteomics, synthetic biology



Transposable elements (TEs) are mobile DNA sequences with the ability to change their positions within a genome.¹ TEs and the associated transposase genes are the most abundant and ubiquitous genes in nature and are essential for adaptation and biological diversification.² It is estimated that 40% of the human genome is comprised of TEs or their remnant sequences,³ conceivably playing a role in human disease.⁴ RNA or DNA-mediated transposition is responsible for the movement of TEs within the genome. DNA transposition involves a “cut-and-paste” mechanism to integrate DNA sequences into a new site within the genome.⁵ Insertion sequences (ISs) are the simplest TE and contain a transposase gene, flanked by inverted terminal repeats (IRs), that is responsible for catalyzing the movement of the IS via DNA cleavage and strand transfer. Other TEs, such as transposons, differ from ISs as they usually carry additional “cargo” genes such as antibiotic resistance genes.⁶

ISs are often overlooked in spite of the fact that ISs are broadly present in bacterial genomes and more than 4000 different ISs have been identified by 2015.^{7,8} Unlike other mobile genetic elements, there is no strict requirement for homology between the DNA sequences at the ends of TEs and the site at which it wishes to integrate (the target site). Due to

this flexibility, DNA transposition of TEs is suited for two major types of applications: to induce random mutagenesis^{9,10} and to introduce genes from a vector into a genome.^{11,12} It was previously believed that transposition is usually neutral or even harmful to the host because of the disruptive effects on the regulatory or coding regions of the host’s genes.^{13,14} However, studies have shown that ISs are capable of conferring a growth advantage by activating cellular processes such as modulation of metabolism,¹⁷ DNA repair,¹⁸ virulence, and antimicrobial resistance.^{19,20} Therefore, the influence of TEs on shaping genes and driving genome evolution is significant.^{15,16} Despite the beneficial effects mentioned above, the expression of TEs is usually under very tight control.^{21,22} TEs are considered to be “domesticated” by their host through gene inactivation. Additionally, the frequency of transposase expression and transposition activity are considered low²³ with the transposition rate of 10^{-6} to 10^{-4} per cell per generation, depending on the type of IS.^{24,25}

Received: May 16, 2019

Published: August 2, 2019

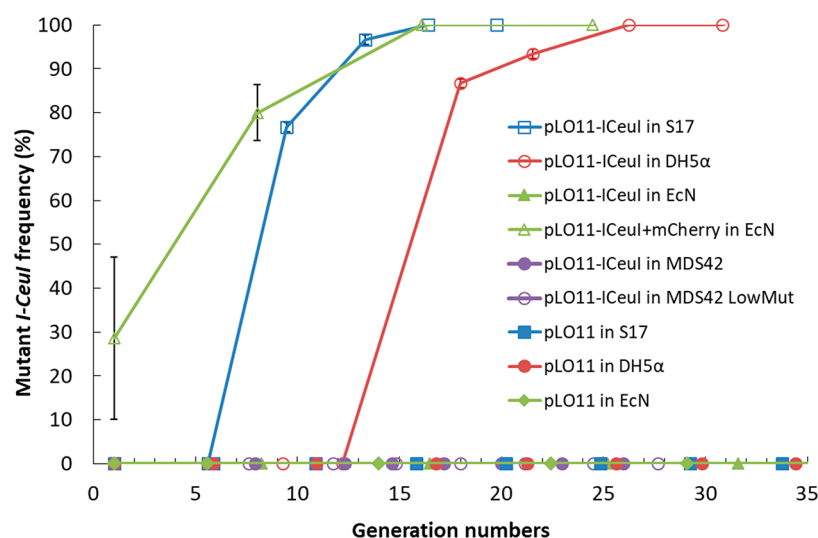


Figure 1. Mutation frequency of the plasmid with and without I-CeuI gene in *E. coli* S17 and *E. coli* DH5α and Nissle 1917 in reference to generation numbers. The vector backbone of pLO11 was used as a negative control. Strains with pLO11-I-CeuI quickly gained mutations, some as early as generation 1.

Synthetic biology usually involves the introduction of a designed gene circuit into a host, which may unintentionally activate ISs and end up disabled. However, the possibility of this happening is hard to predict *in silico*. Other studies have observed inactivation of cloned genes such as lipoprotein *lpp*²⁶ and PvuII methylase²⁷ by ISs. In this study, the *I-CeuI* gene was rapidly disrupted by the insertion of IS1 and IS10 in three different *Escherichia coli* (*E. coli*) strains: S17, DH5α, and Nissle 1917, even though the gene was under control of an arabinose regulatory promoter. We demonstrated that ISs can have a defensive mechanism in *E. coli* by making “jumps” from the genome onto a vector to inactivate a gene. The movement of ISs into the vector benefited the host strain by restoring its growth fitness. Proteomics analysis yielded an abundance profile of 1228 proteins which showed leaky expression of the *I-CeuI* gene had completely changed protein expression of the host cells; including ribosomal activity, SOS stress response, central dogma and cell division processes, membrane synthesis and RNA/DNA metabolism. However, after 31 generations (disruption of *I-CeuI*), the host returned to pattern of protein expression that resembled the control conditions (no *I-CeuI*). Our findings highlight the potential roles TEs can play in the destabilization of designed gene circuits used in synthetic biology.

RESULTS

Leaky Expression of *I-CeuI* Efficiently Triggered the Gene Disruption Activity of Transposable Elements. *I-CeuI* endonuclease recognizes a 26-base pair sequence (5'-TAACTATAACGGTCCTAAGGTAGCGA-3')²⁸ which is naturally present in chassis strains commonly used in synthetic biology (Table S1). *I-CeuI* endonuclease has 7 recognition sites in the genomes of *E. coli* S17, DH5α, MDS42, and MG1655; and 6 sites in *E. coli* Nissle 1917. However, the *I-CeuI* recognition sequence rarely, if ever, appears in common plasmids, as the probability of having a match to the 26-bp sequence is about 1 site in 4.5×10^{15} bp sequences.

To estimate the insertion frequency caused by ISs, the plasmid pLO11-I-CeuI was introduced into *E. coli* strain DH5α, S17, EcN, and IS-free strain MDS42. We verified the sequence

of *I-CeuI* within the constructed plasmid, pLO11-I-CeuI, was correct each time before it was introduced to each *E. coli* strain. Although *I-CeuI* was under control of an arabinose promoter, the introduction of pLO11-I-CeuI significantly affected cell growth in arabinose-free media and the *I-CeuI* gene was completely disrupted within a few generations in the cultures (Figure 1). We tracked the integrity of the *I-CeuI* sequence over several generations using colony PCR (Figure S2–S5). A larger (or nonexistent) PCR product than the normal 657 bp encoding *I-CeuI* indicated the presence of an insertion. Sequencing results revealed that the insertion within *I-CeuI* was either IS1 or IS10. The frequency of IS-mediated insertions in *E. coli* S17 and DH5α reached 100% after 16 and 26 generations, respectively (Figure 1 and S2–S3). Interestingly, in EcN, the pLO11-I-CeuI plasmid was initially not mutated with an IS (Figure S4), until a second plasmid (pJKR-O-mphR) was introduced (Figure S5). In EcN with both pLO11-I-CeuI and pJKR-O-mphR, the *I-CeuI* gene was disrupted as early as the initial transformation or the first generation, and all checked colonies were disrupted after 16 generations (Figure 1 and S5). It could be that the second plasmid (pJKR-O-mphR) applied additional metabolic burden to the host cell and triggered the insertion mutation.

To confirm that the mutations were the result of IS activity in the host strain, pLO11-I-CeuI was also introduced to the IS-free strain *E. coli* MDS42 and no insertion was observed in the strains after 27 generations (Figure S6). To verify the appearance of ISs was due to the presence of *I-CeuI*, we also tested the pLO11 vector backbone (Figure S1b) in *E. coli* DH5α, S17 and EcN. After more than 30 generations in these strains, nothing was found to be inserted or deleted on the plasmids (Figure 1). In all randomly selected colonies (Figure 1 and S2–S5), sequencing results showed that colonies had either an IS1 or IS10 insertion rather than a point or deletion mutation. In the IS-free *E. coli* strain MDS42, the deletion and point mutations only occurred after 27 generations, while nearly all cells of *E. coli* S17 and DH5α acquired IS transposition before 25 generations (Figure 1 and S2–S3). Therefore, transpositions were dominant over other mutations during early stages of cultivation.

Table 1. Insertion Patterns of IS1 and IS10 within *I-CeuI* Gene

| insertion sequence | target sequence on <i>I-CeuI</i> ^a | reference |
|--------------------|---|---------------|
| IS1 | GACTTAAG IS1 GACTTAAGA | this study |
| | TGACTCAAC IS1 TGACTCAAC | this study |
| | ACAACTAG IS1 ACAACTAG | this study |
| | AAGCTTCTC IS1 AAGCTTCTC | this study |
| | CAGGGCGT IS1 CAGGGCGTA | this study |
| | ACAAAATCC IS1 ACAAAATCC | this study |
| IS10 | random | ²² |
| | CTCAAGACTTTGCTC IS10 CTCAGGACTTTGCTC | this study |
| | NGCTNAGCN | ²⁹ |

^aIS1 and IS10 in bold are the IS insertion point.

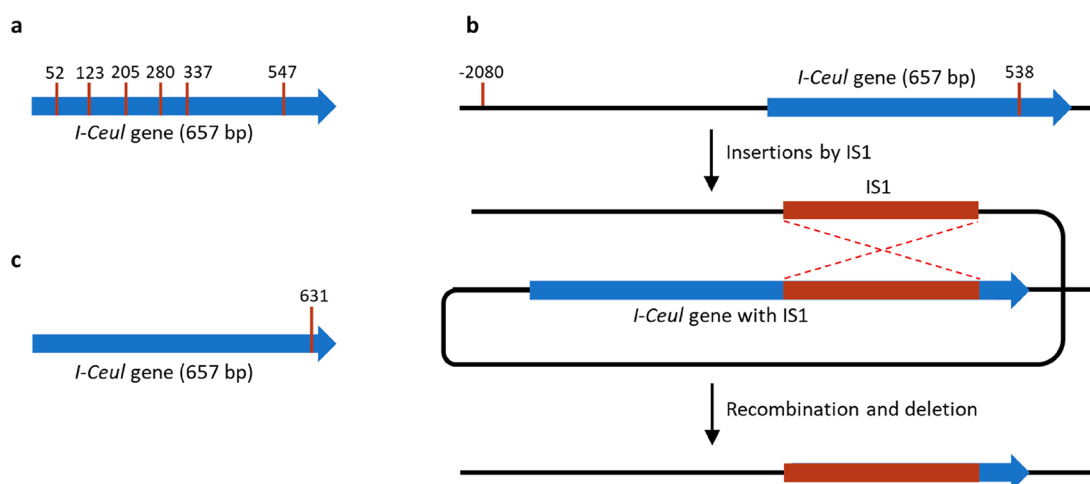


Figure 2. Process of three insertion cases that occurred in pLO11-ICeuI. (a) The insertion points of IS1 in the *I-CeuI* gene. For IS1, the insertion point is random. A total of 6 different insertion points were found in the 10 sequenced plasmids from 10 different colonies of *E. coli* S17. (b) The deletion-insertion process of IS1. Two IS1 insertion sequences attacked the plasmid at two different points (538 bp on the *I-CeuI* gene and 2080bp upstream of the *I-CeuI* gene). Then, a recombination event occurs between these two IS1s, and causing the plasmid section in between to be deleted, leaving only one IS1. This event occurred in 10 plasmids from *E. coli* S17. (c) The insertion point of IS10 in the *I-CeuI* gene. For IS10, the insertion point was only at one specific position. This event occurred in 10 plasmids from *E. coli* S17, 6 plasmids from *E. coli* DH5 α and 3 plasmids from *E. coli* Nissle 1917.

Therefore, the activation of ISs to disable the *I-CeuI* gene in pLO11-ICeuI occurred in response to the double stranded breaks (DSBs) on the chromosome caused by the leaky expression of *I-CeuI* gene in *E. coli* DH5 α , S17 and EcN. Since 6–7 DSBs can be made simultaneously by *I-CeuI* endonuclease in *E. coli* DH5 α , S17 and EcN, the cells were under more stress than usual, which triggered response to the DNA damage. In the case of EcN, IS activity was observed after an additional plasmid (additional metabolic burden) was added to the strain. Therefore, it is possible that both DNA damage and its associated stress, as well as metabolic burden can activate IS activity.

Insertion Sequences IS1 and IS10 Have Different Mechanisms of Transposition. To identify the ISs, we sequenced the disrupted plasmid pLO11-ICeuI and found two different transposition mechanisms used by IS1 and one mechanism used by IS10 in the *E. coli* strains. The full sequences of these two ISs are shown in Table S2 and their insertion sites in the *I-CeuI* gene is shown in Table 1.

IS1-mediated disruptions were observed in *E. coli* S17, and demonstrated two transposition mechanisms: a straightforward random insertion and an insertion-deletion *via* homologous recombination (Figure 2a and 2b). For the straightforward insertion case, IS1 showed a pattern of random insertion

(Figure 2a and Table 1), which is in agreement with literature.²² IS1 duplicated a random 9 bp sequence within the *I-CeuI* gene and inserted itself (768 bp) in between, with the pattern of 123456789-IS1-123456789 (the Arabic numerals represent the order of nucleotide N). For the insertion-deletion case, IS1 was inserted into pLO11-ICeuI at two different sites (538 and –2080 relative to the *I-CeuI* gene). Subsequently, a recombination event occurred between these two IS1s, resulting in the deletion of the plasmid sequence flanked by these IS1s and the elimination of one of the IS1 copies (Figure 2b).

IS10 disruption was observed in all three *E. coli* strains, S17, DH5 α , and EcN, and was inserted after a specific 15 bp site CTCAAGACTTTGCTC↓ within the *I-CeuI* gene (Figure 2c and Table 1). IS10 recognized, duplicated this 15 bp sequence, and inserted itself (1323 bp) in between. The insertion pattern was CTCAAGACTTTGCTC-IS10-CTCAGGACTTTGCTC (the underlined G was identified as point mutation in more than ten replicates) (Table 1). While it has been previously suggested that IS10 preferentially inserts into hotspots containing a consensus target sequence of 9 bp (5'-NGCTNAGCN-3'),²⁹ our results and other reports suggest that IS10 could also recognize other target sequences.^{30–32}

These results show that there are different transposition mechanisms, even in closely related *E. coli* strains. The *I-CeuI* gene in *E. coli* S17 was disrupted by two different ISs (IS1 or IS10), while in *E. coli* DH5 α and EcN it was disrupted only by IS10. This could offer some explanation to the difference in the number of generations needed to acquire ISs.

The rates of IS1 and IS10 transposition are 10^{-6} and 10^{-4} per cell per generation, respectively.^{24,25} The probability of an IS targeting the *I-CeuI* sequence (657 bp) is 0.0029 per cell (given that the copy number of ColE1 ori plasmid pLO11-*I-CeuI* is 20 per cell), assuming it is a random event. Therefore, in a 1 mL *E. coli* culture with 10^8 cells/mL, there would be 0.29 cells that would arise with an *I-CeuI*:IS1 disruption and 29 cells with an *I-CeuI*:IS10 disruption after the first generation. This may explain why IS10-mediated disruptions were observed in all tested *E. coli* strains, while IS1-mediated disruptions were only seen in *E. coli* S17.

A BLAST analysis with the IS1 and IS10 sequences in this study (Table S2) yielded a 100% sequence match to the IS1 and IS10 in *E. coli*. Interestingly, the IS10 sequence also completely matched with sequences within many advanced organisms from different kingdoms including *Oryza sativa* (rice), *Pan troglodytes* (chimpanzee), and even *Homo sapiens* (human).³² The matches with the human genome are summarized in Table S3. Four different institutes found the complete IS10 sequence in seven different human samples, specifically in choriocarcinoma of the placenta (18) eye, retinoblastoma (18), chromosome 19 (19), and chromosome 21. However, these IS10 insertions were found in cDNA sequences. Given the high probability events we observed, and previous observations,^{32,33} it is likely that the IS10 in plant, animal, and human genomes might have been inserted into *E. coli* vectors as cloning artifacts, instead of being present in the original DNA sequences of those advanced organisms. These results suggest that IS transposition in *E. coli* is pervasive and could cause unexpected artifacts.

Insertion Sequences in *I-CeuI* Restore Host Fitness.

We hypothesized that the ISs not only acted as a defensive mechanism for the host cell, but also restored its fitness. To demonstrate this, the growth of strains with undisrupted or disrupted versions of *I-CeuI* were compared (Figure 3). In all cases, the wild type version of the strains without any plasmid had better growth than the strains with plasmids. Generally, strains with just the pLO11 vector backbone also grew well, similar to the wild type. However, when *I-CeuI* was introduced into any strain, the growth was adversely affected. This response is the most obvious in EcN, in which the strain with pLO11-*I-CeuI* in the eighth generation did not seem to replicate at all (Figure 3C). In contrast, after *I-CeuI* was disrupted by ISs, the growth of all of these strains considerably improved (Figure 3). In *E. coli* S17, the growth of the strain with pLO11-*I-CeuI* disrupted by IS10 was even better than the pLO11 vector backbone (Figure 3a).

The growth of EcN with two plasmids (pLO11-*I-CeuI* and pJKR-O-mphR) also recovered after the insertion of IS10 (Figure 3c). After the introduction of the two plasmids, the growth of the culture was impeded in the beginning due to the extra load, until the growth recovered after 11 h (Figure 3c). It is likely that a significant proportion of cells within the culture acquired an insertion mutation in the *I-CeuI* gene by that time. Although EcN with one plasmid pLO11-*I-CeuI* never acquired an IS, the growth of this strain after 52 generations closely resembled control groups (Figure 3c). Sequencing of this

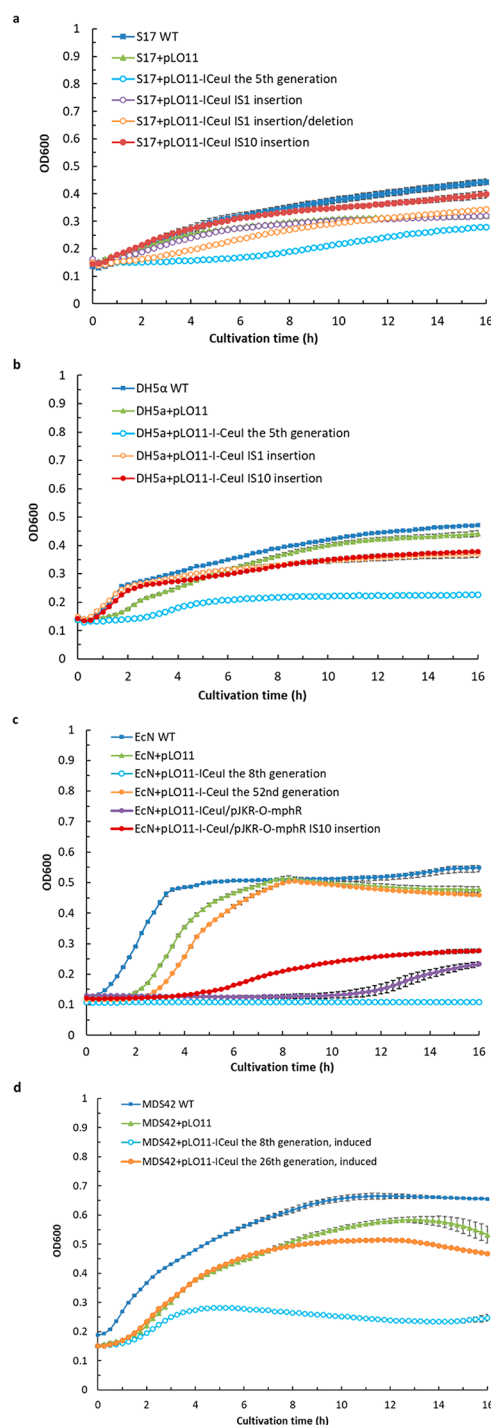
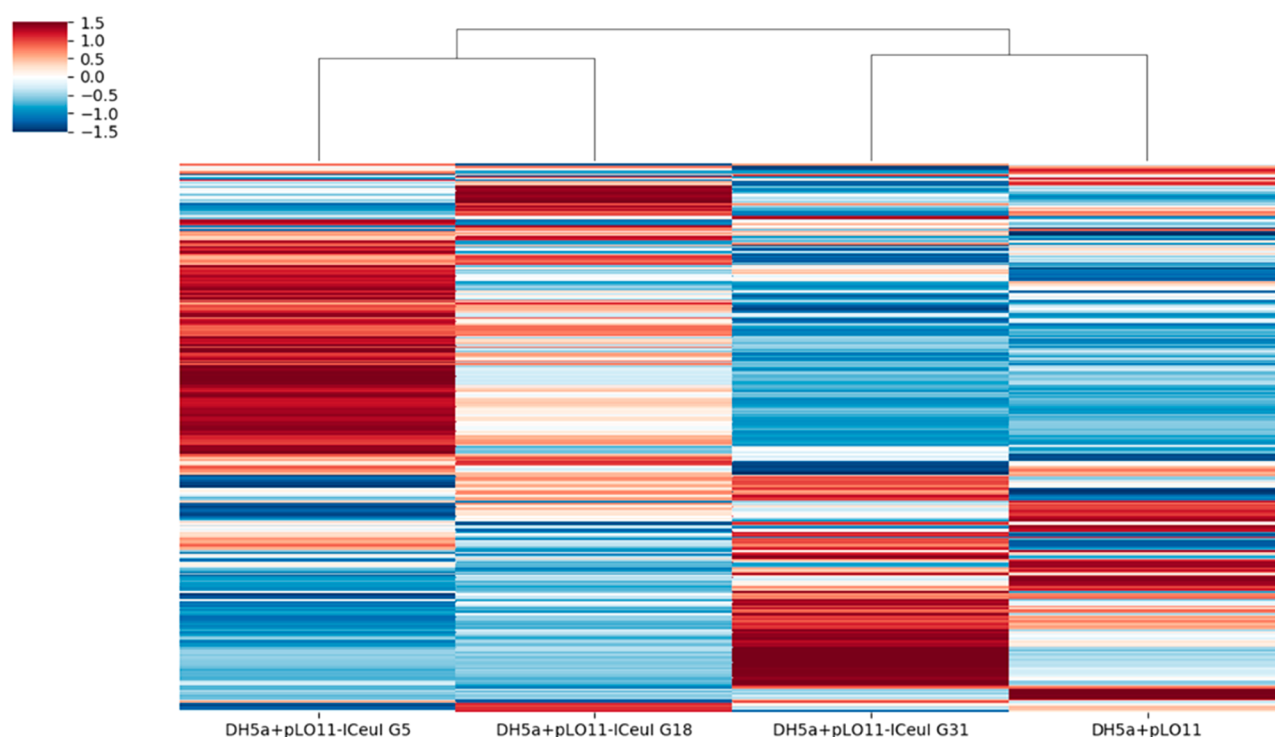


Figure 3. A comparison of the growth of strains with an inactivated *I-CeuI* gene versus strains with the normal *I-CeuI* gene. Wild type, wild type control of the strains; pLO11, the pLO11 vector backbone; IS1 case, *I-CeuI* gene with an IS1 insertion; IS10 case, *I-CeuI* gene with an IS10 insertion; IS1 deletion case, two IS1 insertions on pLO11-*I-CeuI* caused a sequence deletion via recombination. Strains with an inactivated version of *I-CeuI* have restored growth fitness close to control conditions. (a) *E. coli* S17, (b) *E. coli* DH5 α , (c) *E. coli* Nissle1917, (d) *E. coli* MDS42 induced with arabinose.

fitness-restored strain revealed point mutations resulting in the change of two amino acids: histidine to tyrosine (94th amino acid) and alanine to threonine (147th amino acid). These changes could have altered the secondary structure of *I-CeuI* endonuclease and its ability to bind and function.

a



b

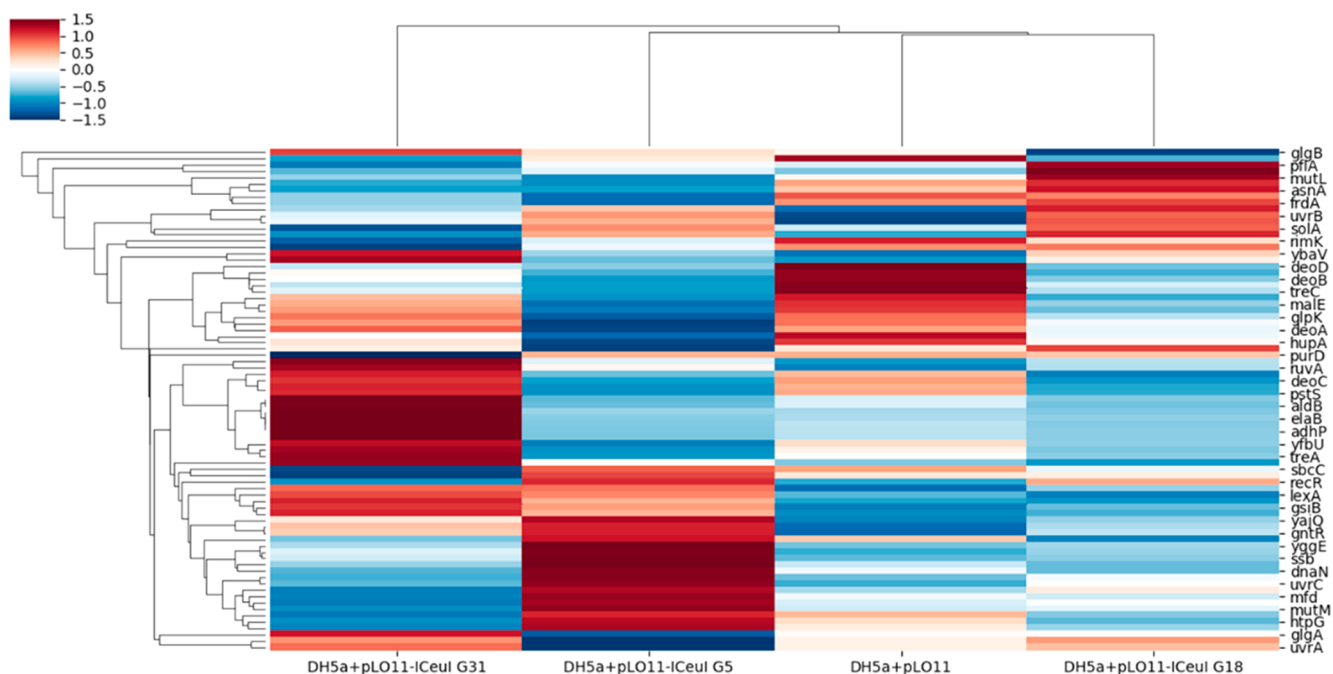


Figure 4. Dendrograms and heatmaps of the relative protein abundance profile that were detected and quantified with proteomics in *E. coli* DH5 α strains. A cluster analysis was done to yield a dendrogram with its corresponding heatmap. The data was row scaled. DH5a+pLO11, pLO11 vector backbone; DH5a+pLO11-ICeuI G5, 18, and 31 correspond to days 1, 4, and 7. The relative protein abundance profile of pLO11-ICeuI G5, the normal version of *I-CeuI*, was distinct from the other conditions but was closest to G18. (a) All proteins. (b) Proteins associated with DNA stress response.

A restoration in fitness was also observed for the MDS42 strain with pLO11-ICeuI, but it was due to mutations not

related to ISs. Initially, it seemed generation 8 and 26 had very similar growth kinetics, but the arabinose promoter had tighter

Table 2. Genes Associated with the Proteins That Experienced the Highest Fold Change between pLO11-ICeuI G5 and Wild Type with pLO11 in *E. coli* DH5 α in Which These Proteins Had Higher Abundance in G5

| gene name | pLO11 | ICeuI G5 | ICeuI G18 | ICeuI G31 | system | biological process |
|-----------|--------|----------|-----------|-----------|-----------------------------------|--|
| mog | −1.196 | 1.252 | −0.055 | −0.0002 | Cofactor biosynthesis | Mo-molybdopterin cofactor biosynthesis |
| fnt | −1.259 | 1.188 | 0.024 | 0.047 | Translation | translational initiation |
| ftsY | −1.274 | 1.171 | 0.046 | 0.057 | Plasma membrane proteins | protein targeting to membrane |
| kdsB | −1.141 | 1.298 | −0.066 | −0.090 | Carbohydrate biosynthesis | lipopolysaccharide biosynthesis |
| bamE | −1.152 | 1.285 | 0.011 | −0.144 | Sigma factor regulons | outer membrane protein assembly |
| ispE | −1.274 | 1.164 | 0.149 | −0.039 | Secondary metabolite biosynthesis | isoprenoid biosynthesis |
| cmk | −1.116 | 1.317 | −0.081 | −0.120 | Sigma factor regulons | nucleobase-containing small molecule interconversion |
| envC | −1.229 | 1.204 | −0.130 | 0.155 | Transcription factor regulons | peptidoglycan-based cell wall biogenesis |
| ybgC | −1.108 | 1.320 | −0.168 | −0.043 | Other | lipid metabolic process |
| ydcI | −1.094 | 1.331 | −0.164 | −0.073 | RNA metabolism | negative regulation of transcription, DNA-templated |
| menC | −1.165 | 1.258 | −0.214 | 0.121 | Cofactor biosynthesis | menaquinone biosynthesis |
| cutC | −1.338 | 1.084 | 0.113 | 0.141 | Other | cellular copper ion homeostasis |
| trkA | −1.324 | 1.095 | 0.011 | 0.217 | Plasma membrane proteins | potassium ion transmembrane transport |
| argR | −1.344 | 1.071 | 0.208 | 0.065 | DNA/RNA metabolism | positive regulation of DNA-templated transcription, initiation |
| ygiB | −1.065 | 1.349 | −0.091 | −0.194 | Sigma factor regulons | biofilm formation |

Table 3. Genes Associated with the Proteins That Experienced the Highest Fold Change between pLO11-ICeuI G5 and Wild Type with pLO11 in *E. coli* DH5 α in Which These Proteins Had Lower Abundance in G5

| gene name | pLO11 | ICeuI G5 | ICeuI G18 | ICeuI G31 | system | biological process |
|-----------|-------|----------|-----------|-----------|-----------------------------------|--|
| kbl | 1.207 | −1.242 | −0.0003 | 0.035 | Sigma factor regulons | L-threonine catabolic process to glycine |
| ftsZ | 1.252 | −1.192 | −0.104 | 0.044 | Cell cycle and division | FtsZ-dependent cytokinesis, division septum assembly |
| rihC | 1.271 | −1.172 | −0.112 | 0.013 | Nucleoside/nucleotide degradation | ribonucleoside catabolic process, cellular response to DNA damage stimulus |
| rpmH | 1.151 | −1.291 | 0.048 | 0.092 | Protein metabolism | translation |
| nanE | 1.133 | −1.297 | 0.184 | −0.021 | Sigma factor regulons | carbohydrate metabolic process |
| gldA | 1.318 | −1.110 | −0.032 | −0.176 | Transcription factor regulons | glycerol metabolic process |
| ytjC | 1.168 | −1.253 | 0.220 | −0.136 | Other | glycolytic process |
| sucD | 1.160 | −1.255 | 0.239 | −0.144 | Protein metabolism | succinyl-CoA metabolic process, TCA |
| rpmF | 1.063 | −1.351 | 0.190 | 0.097 | Translation | ribosomal large subunit assembly |
| zapE | 1.112 | −1.301 | 0.258 | −0.069 | Cell cycle and division | cell division protein |
| hisD | 1.064 | −1.348 | 0.213 | 0.071 | Amino acid biosynthesis | histidine biosynthetic process |
| hupA | 1.045 | −1.355 | 0.050 | 0.261 | Transcription factor regulons | DNA-binding protein, cellular response to DNA damage stimulus |
| rpmG | 1.146 | −1.253 | 0.288 | −0.181 | Translation | ribosomal large subunit assembly |
| tyrB | 1.038 | −1.361 | 0.067 | 0.256 | Amino acid biosynthesis | tyrosine biosynthesis |
| ltaE | 1.031 | −1.358 | 0.025 | 0.303 | Amino acid degradation | threonine catabolic process |

control in this strain, so all cultures were grown with arabinose (Figure 3d). Even though no ISs were detected by colony PCRs, the growth of generation 8 was impeded when induced with arabinose, while after 26 generations the culture closely resembled the wild type versions of MDS42 that did not possess *I-CeuI* gene. Sequencing revealed a base pair deletion in *I-CeuI* in generation 26, effectively disabling the endonuclease.

Therefore, several different kinds of mutations could be occurring but ISs are the easiest to detect and quantify. Whether it be point mutations or ISs, the ultimate consequence is a dysfunctional version of *I-CeuI* endonuclease and the restoration of the cell's fitness.

Proteomics Analysis Links Stress Response Pathways and Activation of ISs. We observed that ISs can be rapidly activated to disrupt the expression of the *I-CeuI* gene in order to restore cell fitness. To understand how *E. coli* responded to the leaky expression of *I-CeuI* from pLO11-ICeuI, a proteomics experiment and LC-MS/MS analysis was carried out using *E. coli* DH5 α as a model strain. The wild type strain (no vector) and wild type with the pLO11 vector backbone

acted as the controls while pLO11-ICeuI at generations (G) 5, 18, and 31 were the experimental conditions. G5 was the earliest point at which *I-CeuI* was undisrupted and had poor growth kinetics while by G31 *I-CeuI* had been disrupted and the strain regained fitness. G18 was chosen as it was the first day ISs occurred in the colony PCRs (Figure S3).

Figure 4a shows a clustered heatmap of the relative abundance of all proteins that were identified and quantified. Visually, the overall protein abundance pattern of DH5 α with pLO11-ICeuI at G5 greatly contrasted with DH5 α with pLO11-ICeuI at G31 and the pLO11 control. The dendrogram clustered pLO11-ICeuI G5 and 18 together, distinct from G31 and the control (wild type with pLO11). G18 seems to be a transitional point between G5 and G31, as well as the control (Figure 4a). Principal component analysis (PCA) shows a similar pattern (Figure S7), showing that the overall protein pattern of pLO11-ICeuI G5 differed from the other conditions (Figure S7). In contrast, pLO11-ICeuI G31 is more related to wildtype DH5 α (Figure S7).

In *E. coli*, the presence of DSBs can trigger an SOS response to repair the DNA damage.³⁴ *I-CeuI* endonuclease can make 7

DSBs on the chromosome of *E. coli* DH5 α ; IS1 and IS10 transposases also make DSBs themselves during transposition,^{23,35} any occurrence of these activities would generate an SOS response.³⁶ Figure 4b shows the abundance of proteins associated DNA damage stress response. In this case, the abundance profile of SOS-related proteins differed between all the conditions. Temporal variations could be due to the activation and subsequent inactivation of the SOS response after attempted DNA repair, and this was reflected as fluctuations in different genes.

We sought to quantify the abundance of IS10 transposase as IS10 had disrupted *E. coli* DH5 α pLO11-ICeuI in all cases. However, IS10 transposase was not picked up by the LC–MS/MS, perhaps due to low abundance of the protein. The only transposase that was detected was a peptide sequence belonging to IS3 transposase, which belongs to a different transposase family and uses a different mechanism of disruption compared to IS10.³⁷ Although the abundance of IS3 transposase was correspondent to the integrity of *I-CeuI* gene (Figure S8), it is uncertain why IS3 was not found during our sequencing of the plasmids. The peptide intensity significantly increased in pLO11-ICeuI G5 and 18 when compared to wild type with pLO11. At G31 however the peptide intensity dropped below wild type levels. This was an indication that IS activity is correlated to the presence of cell stressors such as DNA damage; certain proteins like transposases become active as a response. Then, when the culture's fitness is restored there is no longer a need for these defense mechanisms so the intensities recalibrate to wild type levels. An increase in transposition endangers the quantity of ISs in the chromosome, so other cellular mechanisms turned on to tamper this activity.³⁸ This study showed that ISs can be rapidly activated to disarm genes causing cell stress. The results also suggest that ISs are efficient and effective in re-establishing the cell's original objectives of growth and propagation.

Finally, Tables 2 and 3 list the genes associated with the proteins that experienced the highest fold change in pLO11-ICeuI G5 versus wild type with pLO11. In broad terms, cell division is halted, likely because it is energy-consuming and not essential. This results in the increase of lipopolysaccharide synthesis as the cell expands to make room for other components being synthesized. It was observed that cells with pLO11-ICeuI are indeed longer, at least 10 μm , opposed to a normal *E. coli* cell which only about 2 μm in length. Overall, the energy and metabolic pathways (e.g., the TCA cycle) are less active. There is an increase transcription but less translation via methylation of rRNA. There seems to be an activation of genes related to biofilm formation. The sheer density of changes to cellular processes suggests that the appearance of ISs is only a small part of the cellular response to DNA damage caused by I-CeuI endonuclease.

DISCUSSION

Transposable Elements as a Host Defensive System.

The impact that ISs can have on its host is still controversial, but our findings support the argument that ISs could be utilized by the cell to defend against stress caused by harmful genes such as I-CeuI endonuclease. In our study, TEs, such as ISs in *E. coli*, were triggered to disrupt the introduced harmful gene *I-CeuI*, despite being under control of the arabinose regulation system. This selection pressure is likely similar to what cells experience when they are under stress from antibiotic selection. As shown with proteomics, the DSBs on

the chromosome made by leaky expression of I-CeuI triggered a response from many pathways associated with stress and DNA damage (Figure 4, Tables 2 and 3). It is very likely that these pathways (e.g., small RNA, DNA modification, sequence-specific repressors, etc.) are responsible for TE expression. Notably, many proteins related to DNA stress response increased in abundance in the presence of a functional I-CeuI but then reduced when the cell recovered its fitness (Figure 4). This corresponded to the appearance of transposition-mutated *I-CeuI* sequences, which led to the ultimate restoration of growth kinetics. TEs have a self-interest to propagate and they accomplish this by protecting their host.³⁹ Therefore, it is not surprising that TEs become more active when the cell is under stress, especially from DNA damage.^{40,41}

The cell protected itself by disabling the I-CeuI endonuclease gene with ISs in a shorter period of time than what was reported in literature.²² In the cases where ISs did not appear, point mutations/deletions effectively disabled *I-CeuI* instead. The disruption of the harmful gene considerably improved cell fitness, and thus mutants within cultures were highly selected for. This allowed the experimental strains to grow as well as the wild type strains. Among the strains of *E. coli* tested, different ISs were expressed and they displayed the ability to insert in additional hotspots to what was previously reported in literature.²⁹ ISs are usually categorized based on the transposases that facilitate their movement. However, as they are being studied more widely, it becoming apparent that it is difficult to characterize ISs as they have a variety of preferential target sites in terms of plasmids, chromosomes, DNA sequences, transcribed regions, and secondary structures.⁸

TEs Could Be Significant to Synthetic Biology and Evolution. Synthetic biology is the application of engineering principles to design cells to execute novel functions.^{42–45} Gene circuits are built using characterized and standardized biobricks,⁴⁶ which are supposed to instruct the biological system to perform predicted functions. The gene circuits, once introduced into a host, will inevitably interact with the native cellular environment, attempt to rewire native gene networks, and rely on the host machinery to express the genes. One important difference between synthetic biology and other engineering disciplines is that it engages with living organisms, which are constantly evolving and usually contain TEs. ISs, the simplest TE, are responsible for failure of gene circuits in many engineered bacterial cells.^{47–50} The majority of studies observed the “jumping” of TEs was constrained to within the chromosome or from a plasmid into the chromosome of the host (as a genetic tool). In our study, it was demonstrated that the TEs on the chromosome of the host could jump onto an engineered plasmid to inactivate certain genes, which could indicate that the host's chromosome is not as stable as previously thought, and the jumping of TEs can be bidirectional. Genome sequencing of the *E. coli* DH5 α strains used for proteomics analysis found no changes in the location of ISs on the genome, which means ISs exclusively jumped to the plasmid in this study.

The results of this study highlights the fact that synthetic biology has to consider evolution. The conflict of interest between organisms and synthetic biology could lead to a failure expression of designed gene-circuits, and TEs may be responsible⁵¹ (as in this study). Once gene circuits are inactivated, the cells with dysfunctional gene circuits could dominate the population in as early as the first generation. In

Table 4. Bacterial Strains and Plasmids Used in This Study

| bacterial strains or plasmids | description | reference |
|-------------------------------|--|--|
| Strains | | |
| <i>E. coli</i> DH5 α | F [−] Φ 80lacZ Δ M15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17 (rK [−] , mK ⁺) phoA supE44 λ − thi-1 gyrA96 relA1 | Lab collection |
| <i>E. coli</i> S17 | TpR SmR recA, thi, pro, hsdR-M+RP4:2-Tc:Mu:Km Tn7 λ pir | Lab collection |
| <i>E. coli</i> MDS42 | MG1655 multiple-deletion strain | Scarab Genomics, LLC ⁵⁹ |
| <i>E. coli</i> Nissle 1917 | | Ardeypharm GmbH (Herdecke, Germany) |
| Plasmids | | |
| pLO11 | Expression vector; Tet ^r | A gift from Oliver Lenz, Berlin, Germany |
| pGeneArt-ICeuI | Vector with synthesized I-CeuI | ThermoFisher, GeneArt |
| pLO11-ICeuI | pLO11 containing the DNA fragment coding for the endonuclease, I-CeuI | this study |
| pJKR-O-mphR | Plasmid with mCherry production controlled by MphR | ⁶⁰ |
| pLO11-ICeuI-IS1 | pLO11 containing the DNA fragment coding for the endonuclease, I-CeuI with an IS1 insertion | this study |
| pLO11-ICeuI-IS1D | pLO11 containing the DNA fragment coding for the endonuclease, I-CeuI with an IS1 deletion-insertion | this study |
| pLO11-ICeuI-IS10 | pLO11 containing the DNA fragment coding for the endonuclease, I-CeuI with an IS10 insertion | this study |

Table 5. Primers Used in This Study

| primer | sequence (5' → 3') | description |
|-------------------|--|-----------------------------|
| ICeuINcoI.FOR | cattagccatggatgtcaaaccttataacttaaac | <i>I-CeuI</i> amplification |
| ICeuISacI.REV | caaggagctccctcctcttaactcttataaccttttta | |
| Sequence primer F | tttctccatacccggttttttgggc | From the backbone of pLO11 |
| Sequence primer R | aactcgggtggctccaagc | |
| Sequence 2F | gcttaacgttggttgccac | From the middle of IS10 |
| Sequence 2R | tccgtgtcgagcgctgatt | |
| Sequence 3R | gagcgccataacctgaatct | From the middle of IS1 |

this system, the frequency of IS-mediated gene inactivation was very high (much higher than point-mutation-mediated inactivation) because the ISs inactivated the *I-CeuI* gene to counteract the inhibition of growth and restored host fitness. Thus, the mutants dominated the whole culture after natural selection. These results indicated that TEs can inactivate exogenous, harmful genes to protect the host.

The importance of TEs might be underestimated as they had previously been thought of as “selfish genes” and that they did not contribute much to the survival of its host. However, TEs are ubiquitous and abundant in genomes across all kingdoms of life,² and the observations of its positive contributions that were made in this study and others^{17–20} serve as a testament to how important TEs are in cell response and regulation. This work is a reminder that cells are living systems; if fitness and the ability to replicate is compromised, the cell will fight back with its arsenal of tools developed and acquired over billions of years of evolution.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Growth Conditions.

Bacterial strains and plasmids used in this study are listed in Table 4. The strain *E. coli* Nissle 1917 (EcN) was acquired from Ardeypharm GmbH (Herdecke, Germany). MDS42 was from Scarab Genomics Ltd. (US). All *E. coli* strains: DH5 α , S17, Nissle 1917, and MDS42 were grown in LB medium with the corresponding antibiotics at 37 °C. The plasmid pGeneArt-ICeuI synthesized by ThermoFisher was selected for with 50 μ g·mL^{−1} carbenicillin. For the strains with pLO11 and its derivatives, 20 μ g·mL^{−1} tetracycline was used. For pJKR-O-mphR, 50 μ g·mL^{−1} spectinomycin was used. The plasmid pJKR-O-mphR was from Addgene (Addgene plasmid #

62570). For induction of *I-CeuI* endonuclease in MDS42, 0.2% arabinose was used.

Cloning of the *I-CeuI* Gene into pLO11. The *I-CeuI* endonuclease gene was synthesized by ThermoFisher Ltd. (GeneArt) to yield pGeneArt-ICeuI (Figure S1a). In pGeneArt-ICeuI, an extra 15 bp in the pBAD promoter region⁵² was added to form a hairpin loop with the corresponding reverse complementary sequence on the *I-CeuI* gene, effectively halting gene expression and preventing synthesis. The *I-CeuI* sequence was then cloned into pLO11 (kindly provided by Dr. Oliver Lenz, Berlin, Germany), under the control of pBAD promoter and arabinose regulatory protein AraC. The *I-CeuI* gene was amplified with the primers ICeuINcoI-for and ICeuISacI-rev (Table 5) from pGeneArt-ICeuI. This PCR product and the pLO11 vector (Figure S1b) were digested with NcoI and SacI, and the digested products were ligated with T4 DNA ligase. The ligated plasmids were then transferred into *E. coli* DH5 α . Tetracycline-resistant colonies were screened with colony PCR using the ICeuINcoI.FOR and ICeuISacI.REV primers. Plasmids isolated from positive clones were sequenced to verify the sequence of *I-CeuI* and the construct was named pLO11-ICeuI (Figure S1c).

Identification of Insertion Event and Frequency of Transposition. To estimate the mutation frequency caused by insertion sequences (ISs), the pLO11-ICeuI construct was introduced to *E. coli* DH5 α , S17, EcN, and MDS42 (Figure S1b and c). Positive colonies of *E. coli* with pLO11-ICeuI after transformation were designated as generation 1 (G1). Three positive replicates were inoculated into three sets of liquid culture with 5 mL LB medium with 20 μ g·mL^{−1} tetracycline incubated at 37 °C overnight. These day 1 cultures were

designated as the cumulative generation calculated from OD change (eq 1). The next day, a diluted sample from this culture was spread on an LB plate with 20 $\mu\text{g}\cdot\text{mL}^{-1}$ tetracycline to obtain single colonies. From the overnight liquid culture, 500 μL was transferred into 4.5 mL fresh LB medium supplemented with 20 $\mu\text{g}\cdot\text{mL}^{-1}$ tetracycline. This culture was designated by the new generation number calculated plus the number of generations from the previous day. This pattern continued until all of the colonies screen demonstrated a mutation or 7 days of culturing was reached. These experiments were conducted in triplicate.

Generation number calculation: Ten-milliliter cultures were grown in 50-mL tubes shaken at 150 rpm at 37 °C. After a day of growth, cells were diluted 1:100 into a fresh tube.⁵³ The OD600 of the LB culture was measured before and after overnight incubation. The generation number was calculated using the following equation:

$$N = \log_2 \left(\frac{\text{OD}_t}{\text{OD}_0} \right) \quad (1)$$

where N is the generation number after overnight culture; OD_t is the OD600 value after overnight cultivation; OD_0 is the OD600 before overnight cultivation.

An initial colony PCR using the primers ICeuINcoI.FOR and ICeuISacI.REV was done with up to six replicates of liquid culture on each day to screen for an appearance of any mutations. If such an event occurred, the liquid culture would be plated to yield single colonies. Then, ten single colonies would then be checked by PCR for mutations to calculate the inactivation frequency of the *I-CeuI* gene. The number of colonies checked from G1 depended on the yield from the initial transformation of the *E. coli* strain. After PCR, gel electrophoresis was used to check the size of the *I-CeuI* gene. For the normal *I-CeuI* gene, the size of the PCR product should be 657 bp. If the size of the PCR product deviated from this number, it was an indication the gene was inactivated.

Growth Kinetics. The colonies of the desired strain were transferred into 15 mL fresh LB medium with 20 $\mu\text{g}\cdot\text{mL}^{-1}$ tetracycline, and then incubated overnight at 37 °C. The growth of these mutated strains and control strains were tested in a 96 well black-sided, clear-bottomed plate (BD Falcon) in quadruplicate. Each well contained 180 μL LB medium with the appropriate antibiotic and was inoculated with 20 μL of overnight culture. The plate was then read with a BioTek Synergy HT microplate reader (BioTek Corporation, UK) maintained at 37 °C with reads of optical density (OD) at 600 nm.

Sequencing the Insertion Sequence and Nucleotide Sequence Analysis. The plasmid pLO11-ICeuI was extracted from the overnight 15 mL cultures used for growth kinetics, and then sequenced with the primers listed in Table 5. The sequencing results were aligned with the original nucleotide sequence of *I-CeuI* in pLO11-ICeuI using SnapGene, and the insertion sequences were identified.

The nucleotide sequences of the insertion sequences were used to search for homologous sequences in GenBank, using the BLAST⁵⁴ service on the network server at blast@ncbi.nlm.nih.gov.

Sample Preparation for Proteomics and LC–MS/MS Analysis. The pellets of *E. coli* DH5 α wildtype, with pLO11, with pLO11-ICeuI Generations 5, 18, and 31 were resuspended in a solution containing 8 mol/L urea (Merck,

Darmstadt, Germany) and 2 mol/L thiourea (Merck, Darmstadt, Germany). For cell disruption the solution was 5 times frozen by immersion in liquid nitrogen followed by 10 min of incubation at 30 °C. Finally, medium intensity ultrasonic pulses were applied for 30 s. Cell lysates were centrifuged for 1 h at 20 000g at room temperature. Resulting supernatant was collected.

Protein concentration was estimated following the Bradford approach.⁵⁵ Protein solutions containing 4 μg protein were incubated overnight with 160 ng Sequencing grade Modified Trypsin (Promega, Madison, WI, USA). After 16 h incubation was stopped by application of acetic acid (Carl Roth, Karlsruhe, Germany) at a final concentration of 1%.

Tryptic peptides were desalted using C18 PureSpeed LTS tips (Mettler-Toledo, Gießen, Germany) with 20 μL electric multichannel pipet (Mettler-Toledo, Gießen, Germany). After freeze-drying peptides were resolved in Buffer A consisting of 2% acetonitrile (J.T. Baker, part of Fisher Scientific, Waltham, MA, USA), 0.1% acetic acid (Carl Roth, Karlsruhe, Germany) in water (J.T. Baker, part of Fisher Scientific, Waltham, MA, USA), and subjected to LC–MS/MS analysis. For data independent acquisition (DIA) HRM (Biognosys AG, Schlieren, Switzerland) peptide standard was spiked in.

Data analysis was performed using a Q Exactive mass spectrometer (Thermo-Fisher Scientific, Idstein, Germany) coupled to an UltiMate 3000 RSLC (Thermo-Fisher Scientific, Idstein, Germany) in combination with a Nanospray Flex Ion Source (Thermo-Fisher Scientific, Idstein, Germany). To construct an *E. coli* ion library, each sample was analyzed in data-dependent acquisition (DDA) and DIA mode. Further mass spectrometry settings are listed in Supporting Information. All experiments were done in triplicate.

Raw Data Analysis and Protein Identification. The *E. coli* ion library was created using DDA data with Spectronaut Pulsar X (v 12.0.20491.0.25470; Biognosys AG, Schlieren, Switzerland). A FASTA database containing 4288 identical proteins of *E. coli* DH5 α (ISNDC; Date 2018/11/09) was used as protein sequence library. An ion library was generated using digestion rule trypsin/P with at maximum 2 missed cleavages. Oxidation (M) was considered as variable modification while no fixed modification was set. Only the 7–52 AA peptides were preselected and those with more than 6 transitions were finally considered for the library. The resulting *E. coli* library contained 24 246 peptides and 2274 proteins, of which 1228 were found in DIA data set. The proteomics data has been uploaded to MassIVE with DOI number: 10.25345/C5HD3F.

DIA raw data was further analyzed with (v 12.0.20491.0.25470; Biognosys AG, Schlieren, Switzerland). Resulting data was analyzed in an R-environment (v 3.5.1). MS2-peak-areas were median normalized on replicate level and conditions were compared on peptide level using reproducibility-optimized peptide change averaging method (ROPECA), formerly introduced by Suomi and Elo (2017).^{56,57} Adjusted P -values (Benjamini and Hochberg) < 0.05 and fold change of ± 1.5 was considered as significantly changed. All used R packages are listed in supplemental Table S6.

The heatmaps were generated using the seaborn Python data visualization library.⁵⁸ For the dendrograms, correlation was used as the distance metric with single-linkage clustering. Protein abundances were row scaled to show the fold change.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.9b00218.

Supplementary Figures S1–S8; Supplementary Tables S1–S6 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +44 1865 283786. E-mail: wei.huang@eng.ox.ac.uk.

ORCID

Manuela Gesell Salazar: 0000-0002-6727-1978

Wei E. Huang: 0000-0003-1302-6528

Author Contributions

#C.F. and Y.-H.W. contributed equally to the work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

W. E. Huang acknowledges support from EPSRC (EP/M002403/1 and EP/N009746/1). C. Fan acknowledges support from China Regenerative Medicine International Ltd.

■ REFERENCES

- (1) Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvak, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., and Feschotte, C. (2018) Ten things you should know about transposable elements. *Genome Biol.* 19, 199.
- (2) Aziz, R. K., Breitbart, M., and Edwards, R. A. (2010) Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res.* 38, 4207–4217.
- (3) International Human Genome Sequencing, C (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860.
- (4) Kazazian, H. H., Jr., and Moran, J. V. (2017) Mobile DNA in Health and Disease. *N. Engl. J. Med.* 377, 361–370.
- (5) Craig, N. L., Craigie, R., Gellert, M., and Lambowitz, A. M. (2002) *Mobile DNA II*, American Society for Microbiology, Washington D.C.
- (6) Siguier, P., Gournayre, E., and Chandler, M. (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.* 38, 865–891.
- (7) Siguier, P., Gournayre, E., Varani, A., Bao, T. H., and Chandler, M. (2015) Everyman's Guide to Bacterial Insertion Sequences. *Microbiol. Spectr.* 3, 555.
- (8) Siguier, P., Gournayre, E., and Chandler, M. (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.* 38, 865–891.
- (9) Simon, R., Priefer, U., and Pühler, A. (1983) A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Bio/Technology* 1, 784–791.
- (10) Kawakami, K., Largaespada, D. A., and Ivics, Z. (2017) Transposons As Tools for Functional Genomics in Vertebrate Models. *Trends Genet.* 33, 784–801.
- (11) Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., and Xu, T. (2005) Efficient Transposition of the piggyBac (PB) Transposon in Mammalian Cells and Mice. *Cell* 122, 473–483.
- (12) Kebraie, P., Izsvák, Z., Narayanavari, S. A., Singh, H., and Ivics, Z. (2017) Gene Therapy with the Sleeping Beauty Transposon System. *Trends Genet.* 33, 852–870.
- (13) Doolittle, W. F., Kirkwood, T. B., and Dempster, M. A. (1984) Selfish DNAs with self-restraint. *Nature* 307, 501–502.
- (14) Werren, J. H. (2011) Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10863–10870.
- (15) Kazazian, H. H. (2004) Mobile Elements: Drivers of Genome Evolution. *Science* 303, 1626–1632.
- (16) Feschotte, C. (2008) Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405.
- (17) Hall, B. G. (1998) Activation of the *bgl* operon by adaptive mutation. *Mol. Biol. Evol.* 15, 1–5.
- (18) Hartl, D. L., Dykhuizen, D. E., Miller, R. D., Green, L., and de Framond, J. (1983) Transposable element IS50 improves growth rate of *E. coli* cells without transposition. *Cell* 35, 503–510.
- (19) Vandecraen, J., Chandler, M., Aertsen, A., and Van Houdt, R. (2017) The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol.* 43, 709–730.
- (20) Jangam, D., Feschotte, C., and Betrán, E. (2017) Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. *Trends Genet.* 33, 817–831.
- (21) Chuong, E. B., Elde, N. C., and Feschotte, C. (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18, 71–86.
- (22) Mahillon, J., and Chandler, M. (1998) Insertion sequences. *Microbiol. Mol. Biol. Rev.* 62, 725.
- (23) Nagy, Z., and Chandler, M. (2004) Regulation of transposition in bacteria. *Res. Microbiol.* 155, 387–398.
- (24) Shen, M. M., Raleigh, E. A., and Kleckner, N. (1987) Physical analysis of Tn10- and IS10-promoted transpositions and rearrangements. *Genetics* 116, 359–369.
- (25) Sousa, A., Bourgard, C., Wahl, L. M., and Gordo, I. (2013) Rates of transposition in *Escherichia coli*. *Biol. Lett.* 9, 20130838.
- (26) Nakamura, K., and Inouye, M. (1981) Inactivation of the *Serratia marcescens* gene for the lipoprotein in *Escherichia coli* by insertion sequences, IS1 and IS5; sequence analysis of junction points. *Mol. Gen. Genet.* 183, 107–114.
- (27) Blumenthal, R. M., Gregory, S. A., and Cooperider, J. S. (1985) Cloning of a restriction-modification system from *Proteus vulgaris* and its use in analyzing a methylase-sensitive phenotype in *Escherichia coli*. *J. Bacteriol.* 164, 501–509.
- (28) Liu, S. L., Hessel, A., and Sanderson, K. E. (1993) Genomic mapping with I-Ceu I, an intron-encoded endonuclease specific for genes for ribosomal RNA, in *Salmonella* spp., *Escherichia coli*, and other bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 90, 6874–6878.
- (29) Halling, S. M., and Kleckner, N. (1982) A symmetrical six-base-pair target site sequence determines Tn10 insertion specificity. *Cell* 28, 155–163.
- (30) Bender, J., and Kleckner, N. (1992) IS10 transposase mutations that specifically alter target site recognition. *EMBO J.* 11, 741–750.
- (31) Gonzalez-Prieto, C., Agundez, L., and Llosa, M. (2015) Chloramphenicol Selection of IS10 Transposition in the *cat* Promoter Region of Widely Used Cloning Vectors. *PLoS One* 10, e0138615.
- (32) Kovarik, A., Matzke, M. A., Matzke, A. J. M., and Koukalova, B. (2001) Transposition of IS10 from the host *Escherichia coli* genome to a plasmid may lead to cloning artefacts. *Mol. Genet. Genomics* 266, 216–222.
- (33) Binns, M. (1993) Contamination of DNA database sequence entries with *Escherichia coli* insertion sequences. *Nucleic Acids Res.* 21, 779.
- (34) Motamedi, M. R., Szigety, S. K., and Rosenberg, S. M. (1999) Double-strand-break repair recombination in *Escherichia coli*: physical evidence for a DNA replication mechanism in vivo. *Genes Dev.* 13, 2889–2903.
- (35) Morisato, D., and Kleckner, N. (1984) Transposase promotes double strand breaks and single strand joints at Tn10 termini in vivo. *Cell* 39, 181–190.
- (36) Lane, D., Cavaille, J., and Chandler, M. (1994) Induction of the SOS response by IS1 transposase. *J. Mol. Biol.* 242, 339–350.
- (37) Curcio, M. J., and Derbyshire, K. M. (2003) The outs and ins of transposition: from mu to kangaroo. *Nat. Rev. Mol. Cell Biol.* 4, 865–877.

- (38) Wu, Y., Aandahl, R. Z., and Tanaka, M. M. (2015) Dynamics of bacterial insertion sequences: can transposition bursts help the elements persist? *BMC Evol. Biol.* 15, 288.
- (39) Venner, S., Feschotte, C., and Biemont, C. (2009) Dynamics of transposable elements: towards a community ecology of the genome. *Trends Genet.* 25, 317–323.
- (40) Horvath, V., Merenciano, M., and Gonzalez, J. (2017) Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response. *Trends Genet.* 33, 832–841.
- (41) Eichenbaum, Z., and Livneh, Z. (1998) UV light induces IS10 transposition in *Escherichia coli*. *Genetics* 149, 1173–1181.
- (42) Benner, S. A., and Sismour, A. M. (2005) Synthetic biology. *Nat. Rev. Genet.* 6, 533–543.
- (43) Brophy, J. A. N., and Voigt, C. A. (2014) Principles of genetic circuit design. *Nat. Methods* 11, 508–520.
- (44) Khalil, A. S., and Collins, J. J. (2010) Synthetic biology: applications come of age. *Nat. Rev. Genet.* 11, 367–379.
- (45) Purnick, P. E. M., and Weiss, R. (2009) The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* 10, 410–422.
- (46) Shetty, R. P., Endy, D., and Knight, T. F., Jr. (2008) Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* 2, 5–5.
- (47) Geng, P., Leonard, S. P., Mishler, D. M., and Barrick, J. E. (2019) Synthetic Genome Defenses against Selfish DNA Elements Stabilize Engineered Bacteria against Evolutionary Failure. *ACS Synth. Biol.* 8, 521–531.
- (48) Rugbjerg, P., Myling-Petersen, N., Porse, A., Sarup-Lytzen, K., and Sommer, M. O. A. (2018) Diverse genetic error modes constrain large-scale bio-based production. *Nat. Commun.* 9, 787.
- (49) Sleight, S. C., Bartley, B. A., Lieviant, J. A., and Sauro, H. M. (2010) Designing and engineering evolutionary robust genetic circuits. *J. Biol. Eng.* 4, 12–12.
- (50) Sleight, S. C., and Sauro, H. M. (2013) Visualization of Evolutionary Stability Dynamics and Competitive Fitness of *Escherichia coli* Engineered with Randomized Multigene Circuits. *ACS Synth. Biol.* 2, 519–528.
- (51) Kwok, R. (2010) Five hard truths for synthetic biology. *Nature* 463, 288–290.
- (52) Cramer, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* 14, 315–319.
- (53) Dekel, E., and Alon, U. (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436, 588–592.
- (54) Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- (55) Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry.* 72, 248–254.
- (56) Suomi, T., and Elo, L. L. (2017) Enhanced differential expression statistics for data-independent acquisition proteomics. *Sci. Rep.* 7, 5869.
- (57) Suomi, T., Seyednasrollah, F., Jaakkola, M. K., Faux, T., and Elo, L. L. (2017) ROTS: An R package for reproducibility-optimized statistical testing. *PLoS Comput. Biol.* 13, No. e1005562.
- (58) Waskom, M., Botvinnik, O., O’Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., Augspurger, T., Halchenko, Y., Cole, J. B., Warmenhoven, J., Ruiter, J. d., Pye, C., Hoyer, S., Vanderplas, J., Villalba, S., Kunter, G., Quintero, E., Bachant, P., Martin, M., Meyer, K., Miles, A., Ram, Y., Yarkoni, T., Williams, M. L., Evans, C., Fitzgerald, C., Brian, F., Fonesbeck, C., Lee, A., and Qalieh, A. (2017) *mwaskom/seaborn*, v0.8.1.
- (59) Posfai, G., Plunkett, G., Feher, T., Frisch, D., Keil, G. M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S. S., de Arruda, M., Burland, V., Harcum, S. W., and Blattner, F. R. (2006) Emergent properties of reduced-genome *Escherichia coli*. *Science* 312, 1044–1046.
- (60) Rogers, J. K., Guzman, C. D., Taylor, N. D., Raman, S., Anderson, K., and Church, G. M. (2015) Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic Acids Res.* 43, 7648–7660.