An IS element-mediated synthetic biology approach for engineering multi-copy integration of large genetic constructs into bacterial genomes

Ph.D. Thesis

Ranti Dev Shukla

Supervisor: Tamás Fehér, MD, Ph.D.

Doctoral School of Biology

Faculty of Natural Sciences and Informatics, Szeged

Biological Research Centre of the Eötvös Loránd Research Network, Institute of Biochemistry,

Synthetic and Systems Biology Unit

and

University of Szeged, 2022

Contents

List of abbreviations

Ab	antibiotic
Amp	ampicillin
aTc	anhydroteracycline
AU	arbitrary unit, based on OD_{585} measurement
bp	basepairs
Cm	chloramphenicol
CmR	chloramphenicol resistance gene
CRISPR-Cas9	clustered regularly interspaced short palindromic repeats and CRISPR- associated protein 9.
ddPCR	droplet digital pcr
gDNA	genomic DNA
IS element	insertion sequence element
Km	kanamycin
KmR	kanamycin resistance gene
no-SCAR	scarless cas9 assisted recombineering
nt	nucleotide(s)
PCR	polymerase chain reaction
Sp	spectinomycin
SpR	spectinomycin resistance gene
TE	tris-EDTA buffer
wt	wild type
WGS	whole-genome sequencing

1. Introduction

1.1 Synthetic biology

The founding and the development of recombinant DNA technology have resulted in one of the most revolutionary shifts in paradigm within biological sciences. The ability to precisely modify the genetic content of living cells opened up novel avenues of research permitting the detailed analysis of biological processes from the level of genes to the level of the organismal phenotype. Referring to this technology as "genetic engineering" is not an exaggeration if considering the well-defined and accurately planned changes made at the level of the DNA sequence. However, when considering the predictability of changes in cellular function (resulting from genetic modifications), one quickly realizes that many of the projects in the golden age of genetic engineering were closer to "tinkering" instead: the genetic modifications were based on similarity to examples seen in nature, or were relying on a trial and error strategy. As a result, genetic engineering did not approach the quality demanded e.g. in aerospace or nuclear engineering, where unexpected outcomes can have devastating consequences. To initiate a change in this regard, Endy (Endy, 2005) recommended the incorporation of three engineering principles into molecular biology, and thereby create a novel entity of science called synthetic biology. These principles were decoupling, abstraction, and standardization. Decoupling is the division of a problem, an engineering task, or a system into subdivisions that can be solved, worked on, or investigated independently. Abstraction is the simplification of a problem or of the representation of a system by hiding details or other complicating properties. Abstraction is often applied by defining hierarchical levels of the system. This way, the system is divided into "parts". These parts can also be considered systems, which consist of parts, and so on. Applying abstraction hierarchy supports decoupling. Finally, standardization is defining required parameters for a certain part. If a sufficient number of parameters is dealt with, standardization can warrant the interchangeability of parts, similarly to the components of an electronic circuit.

Synthetic biology is sometimes defined as a field aiming to redesign existing living systems, or to *de novo* create new ones. This, however, does not demark it from genetic engineering, we, therefore, recommend using Tom Knight's definition of synthetic biology as "…an engineering technology based on living systems", emphasizing the incorporation of engineering principles

into biology. Although the necessity or the timeliness of founding synthetic biology was highly debated (Baker, 2014) (Kahn, 2011), this field has become established, and an increasing number of research groups declare their projects to belong here. The improvement in the predictability of genetic modifications seems to be a slow and gradual process, however, warranting the need for years or decades of further development to achieve levels acceptable in engineering sciences.

The work presented here aims to contribute to the standardization "pillar" of synthetic biology. As described below, propagating genetic constructs in bacterial cells using plasmids is often hampered by variation of plasmid copy numbers, spontaneous loss of plasmids, or frequent segregation of mutants. These phenomena all reduce the reproducibility of the experiments carried out with plasmid-based constructs and impair the predictability of the levels of transgene expression or the function of genetic circuits. By developing techniques for DNA integration into the bacterial chromosome and copy-amplification of the integrated gene(s), we wish to expand the toolbox available for circumventing the problems associated with the use of plasmids for transgene cloning and propagation in bacterial cells.

1.2 Plasmid- vs genome-based expression of heterologous DNA in bacteria

Extraction of a drug from its natural habitat can be time-consuming, tedious, and is often not economical. Microbial biosynthesis has been the alternate solution for years. *Escherichia coli* is one of the most convenient microbes for this purpose as evident from literature (Rosano & Ceccarelli, 2014). The United States Food and Drug Administration *(*FDA) has reported that 1/3rd of the FDA-approved proteinaceous drugs are produced in *Escherichiacoli*.(Baeshen et al., 2015).

Reprogramming *Escherichia coli* to perform the desired task is a multi-step process that usually requires heterologous gene expression, followed by the fine-tuning of transcript levels (Nielsen & Keasling, 2016). The most convenient way to reprogram *Escherichia coli* is to introduce the desired multi-gene operon into the bacterium *via* a plasmid vector. Besides their simplicity, however, plasmid-based systems have several drawbacks. There is cell to cell variation in the copy number of plasmid leading to variation in gene expression (Hajimorad & Gralnick, 2020). To maintain the plasmid inside the bacterial cell, continuous selection, usually the administration of antibiotics, is required. This way, the final product could get contaminated

with antibiotics, potentially supporting the evolution of antibiotic-resistant strains. Elimination of antibiotics necessitates an extra step of purification and exerts an economic burden on the biotech industry (Peubez et al., 2010) (Mignon et al., 2015).

Genes cloned on plasmids often acquire loss of function mutations, which inhibit the production of the target compound. As the bacterial cell divides, there is an unequal distribution of plasmids into daughter cells leading to overall reduction and eventual loss of expression (Ensley, 1986). The frequency of unequal distribution is higher for plasmids prone to form multimers by intermolecular recombination (Summers & Sherratt, 1984)

Genome-based expression systems offer a way to overcome the difficulties associated with plasmid-based cloning. Such systems rely on the direct insertion of transgenes or multi-gene operons into a bacterial genome for stable and selection-free production of the target compound, which is a highly valuable property for industrial production (Tyo et al., 2009). To date, several techniques have been used to integrate heterologous DNA into the bacterial chromosome. I classify these methods according to the circular or linear nature of the integrated DNA.

1.3 Approaches to plasmid DNA integration into the bacterial genome

Suicide plasmid-mediated genome editing was for some time a widely used method. Replication of suicide plasmids in bacteria is conditional. One type of suicide plasmids can replicate only in the presence of the π protein provided in trans, from the *pir* gene inserted either in a helper plasmid or in the bacterial genome. Another type of suicide plasmids are thermosensitive plasmids, which can replicate only at low temperatures, but are lost from the cell at higher temperatures. An example of its use for genome editing is the work of György Pósfai et al. who demonstrated the insertion of circular suicide plasmid into the *Escherichia coli* chromosome followed by removal of the inserted plasmid. In detail, a suicide plasmid was constructed with a gene of interest and the recognition site of the I-Scel meganuclease. The plasmid was transformed into host bacteria resulting in its homologous recombination-mediated integration resulting from crossing over between wild the type and the mutant allele. The inserted plasmid was removed from the genome in the second step leaving behind either the wt or the mutant allele of the targeted gene in the bacterial genome. The plasmid removal protocol involved a second crossover, which was initiated by a unique double-strand break (DSB) introduced into the chromosome at the I-Scel site of the plasmid (Pósfai et al., 1999).

Other plasmid-based methods depend on yeast-derived FRT/FLP systems or phage derived integrase systems. The FRT/FLP mediated method requires FRT sites in the genome (Huang et al., 1997). Bacteriophages encode integrases which help recombination between the attP (present in bacteriophage genome) and the attB site (present in the bacterial genome) resulting in the integration of the phage DNA into the bacterial genome. This phenomenon was exploited for bacterial genome editing where attP sites should be present in the desired DNA and attB site should be present in the bacterial genome for the process of recombination. This method, however, was limited by the availability of attB sites in the bacterial genome (Rutherford et al., 2013). In all these methods, the whole plasmid is integrated into the bacterial genome followed by its removal, leaving behind the gene of interest at the targeted locus of the bacterial genome.

1.4 Approaches to linear DNA integration into bacterial genomes

At the turn of the millennium, linear DNA fragment-mediated genome-editing methods became popular where linear DNA, flanked by appropriate homology boxes replaced the targeted locus in the bacterial genome by a double crossover. This method was depending upon bacterial RecA enzyme and genomic *recBC*- or *recD*- mutations; otherwise the linear DNA was quickly degraded by bacterial RecBCD exonuclease. This method was limited by the need for relatively long stretches of homologies and the low rate of integration.

The linear DNA-mediated gene-editing method, sometimes called recombineering became more feasible when applying λ-Red recombinase genes. The λ-Red system consists of three genes: *exo*, *bet*, and *gam*. The bacterial recBCD nuclease is inhibited by Gam to maintain the donor linear DNA while the exonuclease and strand invasion functions of Exo and Bet, respectively perform the recombination process resulting in the integration of the linear DNA at the target locus in the bacterial genome (Yu et al., 2000).

The λ-Red mediated method was popular for introducing point mutations, large deletions, and small insertions in bacterial genomes, but it was possible to insert large DNA (> 3 kbp) only if the target site in the genome was cleaved by an endonuclease during the recombineering (Kuhlman & Cox, 2010). The I-SceI homing endonuclease, an enzyme recognizing a unique 18-base-pair (bp) target site was often used for this purpose, but this required priorinsertion of an I-SceI cleavage site (landing pads) at the desired locus in the bacterial genome (Joshi et al., 2012). This problem was solved with the discovery of the CRISPR/Cas system which uses

guide RNA to direct the Cas9 enzyme to make a cleavage at the targeted locus without prior insertion of landing pads (Reisch & Prather, 2015). The capacity of transgene insertion however was limited to single copies using either of the linear DNA-mediated genome editing methods described here.

Concerning the copy-amplification of marked genetic constructs already inserted into the bacterial chromosome, a well-known strategy is to serially passage the strains in gradually increasing antibiotic concentrations, which leads to the RecA-mediated concatenation of the construct within the chromosome. The obtained genotype requires however the deletion of the *recA* gene for stabilization (Tyo et al., 2009). This strategy has also been successfully used by applying the Cre/Lox recombinase (Yin et al., 2015).

We aimed to develop a method where a natural landing pad is already present in the bacterial genome in multiple copies, which offers a possibility to achieve multi-copy integration of heterologous DNA. As an additional improvement, we sought techniques allowing the copy-amplification of the inserted genetic cargo in a second step. We believe that bacterial insertion sequence elements (IS elements) capable of replicative transposition can fulfill both requirements at the same time.

1.5 Insertion sequences and transposons

Insertion sequences (IS elements or ISes) are the smallest autonomous mobile genetic elements found in nature (Siguier et al., 2015). They comprise only one or two genes, which are responsible for the transposition process, and are surrounded by inverted repeats. To date, 29 families of ISes have been described (Siguier et al., 2015). ISes belong to prokaryotic transposable elements (TEs), along with more complex members like transposons, integrative conjugative elements, integrative mobilizable elements, type I and II introns, transposable and satellite prophages, mobile genomic islands, inteins, retrons, and IStrons (Roberts et al., 2008) (Piégu et al., 2015). Bacterial TEs have made a significant contribution to reshaping the genomes of their hosts over an evolutionary time scale (Vandecraen et al., 2017), and the effect of their mobility is readily observed in everyday medical microbiology. Examples of the latter phenomenon include altering the virulence of pathogens (Hammerschmidt et al., 1996), mobilizing antibiotic resistance genes. (Fekete et al., 2012) (Izumiya et al., 2011) or inducing antibiotic resistance.

Bacterial transposons differ from ISes in the sense that in addition to transposases, they also contain genes encoding functions unrelated to transposition. Over the decades, transposons have become popular tools of molecular biology, used for gene delivery, mutagenesis, and functional genomics studies (Choi & Kim, 2009). Perhaps the most generally known is Tn5, which has been applied to knock-in genes, create mutant libraries, study gene essentiality or create reduced genome (Goryshin et al., 2003)(Reznikoff, 2003). Examples utilizing further transposons are nevertheless countless. Random transposon insertions of Mu, Tn3, Tn5, and Tn1000 have been used for priming sequencing libraries (Boeke, 2002). Members of the Tn3 family have been employed in *Escherichia coli* for in-frame tagging of gene libraries (Ross-Macdonald et al., 1999) or pentapeptide-scanning mutagenesis of single genes cloned in plasmids. Due to its high target-selectivity, wild-type Tn7 is appropriate for gene transfer in-between replicons, but modified Tn7 lacking site-specificity is also available for random mutagenesis studies (Gwinn et al., 1997). Lately, two natural examples of Tn7-related transposons were identified in *Vibrio cholera* (Klompe et al., 2019) and *Scytonema hofmanni*, respectively, both of which carry various elements of the CRISPR-Cas machinery. These TEs allow RNA-directed insertion of the DNA-transposons, warranting the possibility of programmable gene delivery at predefined sites of bacterial genomes using transposition. More recently, *in vitro* application of the Tn5 transposase together with next-generation sequencing technology has opened multiple avenues in genomics studies, including the analysis of 3D genomic structure, detection of copy number variations, sequencing long fragments, and mapping DNA-methylation (reviewed by (N. Li et al., 2020).

As opposed to transposons, ISes are much less often used as molecular biology tools. Notable cases nevertheless exist: for example, IS21 has been used for linker-scanning mutagenesis of an enzyme-encoding gene cloned in *E. coli* (Seitz et al., 2000). Another exciting work described the reprogramming of IS608 (ISHp608) of *Helicobacter pylori* to allow predictable integration at chosen target sites, both *in vitro* and *in vivo*, years before the advent of the CRISPR-era (Guynet et al., 2009). Similarly, the fusion of the IS30 transposase to various specific DNA-binding proteins permitted the directed integration of the IS30 element both in *Salmonella enteritidis* (Imre et al., 2011) and in zebrafish (Szabó et al., 2003). A method relying on IS608 insertion by homologous recombination, followed by its precise transposase-mediated excision has also been developed for editing short genomic segments (Thakker et al., 2016)

Our work presented here provides a further example for the use of IS elements for bacterial genome editing. We test two elements of *Escherichia coli*, IS*1* and IS*3*, and report that they can

be used as natural landing pads to insert a marker gene, an unmarked gene, and a marked multigene operon in one or two copies into the chromosome. We continue with the copy number amplification of the inserted genetic cargo using the copy-paste transposition of the respective transposase. Finally, we demonstrate the stability of the genomic constructs and provide a protocol that potentially permits adapting the method to further bacterial strains.

2. Objectives:

We aimed to develop a natural landing pad-mediated synthetic biology toolkit that allows integration and amplification of heterologous DNA into the bacterial genome. We wished to answer the following questions:

1. Can we develop a system to clone foreign DNA in a bacterial genome in multiple copies using insertion sequences as targets?
2. Can we further increase the copy number of the inserted DNA by replicative transposition?
3. How does the stability of the multi-copy genomic construct compare to that of the plasmid-based construct?

3. Materials and methods

3.1. Strains, chemicals and media

The *E. coli* strains modified in this study are listed in Table 1. Bacteria were grown in liquid Luria–Bertani medium (LB), or on LB plates containing 1.5% agar. Components of the media were obtained from Molar Chemicals Kft., Halásztelek, Hungary. Antibiotics were obtained from Sigma Aldrich (St. Louis, MO, USA) and were used in the following concentrations: chloramphenicol (Cm): 25 μg/mL; ampicillin (Ap): 100 μg/mL; kanamycin (Km): 25 μg/mL; spectinomycin (Sp): 50 μg/mL; anhydrotetracycline (aTc): 50 ng/mL. Antibiotic-gradient plates were made applying the protocol of Szybalski and Bryson (Szybalski & Bryson, 1952) but using 60× higher Sp and 20× higher Km concentrations in the top layer, compared to the values listed above.

Table 1. *E. coli* strains modified in this study

E. coli Strain	IS type targeted	IS copy number	GenBank Acc. No.
MG1655	-	-	NC_000913.3
DH5α	-	-	NZ_CP080399
MDS16	IS*3*	3	this work
MDS27	IS*3*	2	this work
MDS30	IS*3*	1	this work
MDS39R2	IS*1*	2	this work
MDS42	-	-	AP012306
BL21(DE3)pLysE	-	-	NC_012947.1
BLK09	IS*3*	2 (active)	CP014641
BLK16	IS*3*	2 (inactive)	CP014642
MDS42IS*1*	IS*1*	1	this work

Plasmid preparations were made using the Zippy Plasmid Mini-Prep Kit (Zymo Research Ltd., Orange County, CA, USA). The horizontal electrophoresis of DNA was carried out using 1% Seakem LE agarose gels (Lonza, Basel, Switzerland). All cloning and molecular biology experiments were carried out according to established protocols (Sambrook, 2001) unless stated otherwise . All DNA-modifying enzymes were from the Thermo Fisher Scientific, Waltham, MA, USA. New England Biolabs and Viogene-Biotek Corporation were the source of rest enzymes as well as kits in this project.

3.2. Plasmids used in this study

Table 2. List of Plasmids

Plasmid	**Origin**	**Resistance**	**Function**	**GenBank**	**Reference**
pKDsg-IS1	pSC101*	Sp	expression of l-Red recombinases and a crRNA targeting IS1	–	this work
pKDsg-IS3	pSC101*	Sp	expression of l-Red recombinases and a crRNA targeting IS3	–	this work

pST76AIS1X	pSC101*	Ap	carrying two homology boxes of IS1, without inverted repeats	–	this work
pST76AIS3X	pSC101*	Ap	carrying two homology boxes of IS3, without inverted repeats	–	this work
pST76AIS1X::*vioABCDE*	pSC101*	Ap	carrying two homology boxes of IS1 (without inverted repeats), flanking the vioABCDE operon	–	this work
pST76AIS3X::*vioABCDE*	pSC101*	Ap	carrying two homology boxes of IS3 (without inverted repeats), flanking the vioABCDE operon	–	this work
pST76AIS1X::*vioABCDE*_SpR	pSC101*	Ap, Sp	carrying two homology boxes of IS1 (without inverted repeats), flanking the vioABCDE operon and the SpR gene	–	this work
pST76AIS3X::*vioABCDE*_SpR	pSC101*	Ap, Sp	carrying two homology boxes of IS3 (without inverted repeats), flanking the vioABCDE operon and the SpR gene	–	this work
pST76AIS1X::*vioABCDE*_KmR	pSC101*	Ap, Km	carrying two homology boxes	–	this work

			of IS1 (without inverted repeats), flanking the vioABCDE operon and the KmR gene		
pST76AIS3X::*vioABCDE*_KmR	pSC101*	Ap, Km	carrying two homology boxes of IS3 (without inverted repeats), flanking the vioABCDE operon and the KmR gene	–	this work
pZA31insAB'tetR	p15A	Cm	expression of IS1 transposase insAB', without the need of a tranlational frameshift	–	this work
pSTinsAB'tetR	pSC101*	Cm	expression of IS1 transposase insAB', without the need of a tranlational frameshift (heat-sensitive rep origin)	–	this work
pZA31tnp3tetR	p15A	Cm	expression of codon optimized IS3 transposase insEF', without the need of a tranlational frameshift	–	this work
pSTtnp3tetR	pSC101*	Cm	expression of codon optimized IS3 transposase insEF', without the need of a	–	this work

			tranlational frameshift (heat-sensitive rep origin)		
pSG78A_full_IS3::SpR	R6K	Ap, Sp	carrying two homology boxes of IS3 (including inverted repeats), flanking the SpR marker	–	this work
pCas9IS1	p15A	Cm	expression of the full CRISPR/Cas9 machinery targeting IS1	–	this work
pCas9IS3	p15A	Cm	expression of the full CRISPR/Cas9 machinery targeting IS3	–	this work
pCas9	p15A	Cm	expression of the full CRISPR/Cas9 machinery without guide RNA	–	Jiang et al., 2013
pSG76-CS	R6K	Cm	Carrying Cm marker	AF402780.1	Kolisnychenko et al., 2002
pST76-A	pSC101*	Ap	thermosensitive	Y09895.1	Pósfai et al., 1997
pST76-K	pSC101*	Km	thermosensitive	Y09897.1	Pósfai et al., 1997
pZA31CFPtetR	p15A	Cm	express CFP protein induced by tet promoter	–	Fehér et al., 2012
pORTMAGE2	pBBR1	Ap	expression of λ-Red recombinases and mutL mutator allele	–	Nyerges et al., 2016

pORTMAGE4	pBBR1	Cm	expression of λ-Red recombinases and mutL mutator allele	–	Nyerges et al., 2016
pSG78-A	R6K	Ap	can replicate in presence of π protein	Y09892.1	–
pCDM4	CloDF13	Sp	Carrying Sp marker	–	Xu et al., 2012
pUTLIQ_*vioABCDE*	pUC	Tc, Ap	Carrying vio*ABCDE* operon	gift from Prof. Kanji Nakamura	
pKDsg-ack	pSC101*	Sp	expression of λ-Red recombinases and crRNA targeting ack gene	–	Reisch & Prather, 2015
pCas9cr4	p15A	Cm	carrying SpCas9 gene	–	Reisch & Prather, 2015

pSC101*: stands for the thermosensitive version of the pSC101 origin of replication. Cm: chloramphenicol; Ap: ampicillin; Km: kanamycin; Sp: spectinomycin

3.3 Plasmid construction

pKDsg-IS1: the pKDsg-ack was amplified with primers pKD1 and IS1noScarFwd as well as with pKD2 + IS1noScarRev **(Table 3)** using Q5 DNA polymerase Both products were DpnI digested and purified with the Viogene Gel/PCR DNA Isolation Kit These fragments were assembled in a Circular Polymerase Extension Cloning (CPEC) (Quan & Tian, 2009) reaction using Phusion DNA polymerase and transformed into *E. coli* MDS42 and selected on an Sp plate at 30 °C. Colony PCR screening was performed with primers pKDSeq5 + IS1noScarCheckR. The plasmid was purified from a PCR-positive colony and the pKdSeq5 + pKDseq3 PCR fragment amplified from the plasmid was sequenced with primer pKDseq5 for verification. Plasmid pKDsg-IS*3* was made similarly, except that primers IS1noScarFwd,

IS1noScarRev and IS1noScarCheckR were replaced by IS*3*noScarFwd, IS*3*noScarRev and IS*3*noScarCheckR, respectively.

pST76AIS1X: two overlapping PCR fragments were amplified from the IS*1* element of the *E. coli* MDS42IS*1* genome with IS1_Ssp_up + IS1_Xho_Rev and IS1_Xho_Fwd + IS1_Ssp_Dn, respectively. Both fragments were purified with the Viogene Gel/PCR DNA Isolation Kit and then fused in an overlapping PCR to make an IS*1* fragment lacking the inverted repeats but having SspI sites on the ends and an XhoI site in the center. The PCR product was treated with T4 Polynucleotide Kinase for phosphorylation and was gel-extracted with the Viogene Gel/PCR DNA Isolation Kit. Plasmid pST76-A was cut with fast digest BsaI followed by phosphatase treatment. Linear pST76-A was ligated with the PCR fragment using T4 ligase overnight at 16 °C followed by heat inactivation at 65 °C for 10 min and dialysis for 1 h on a 50 nm pore-size membrane (Millipore). The ligation product was transformed into *E. coli* MDS42 and colonies were selected overnight at 30 °C on Ap plates. Colony PCR was performed using primers pSTA-F + pSTA-R to find positive clones. Two positive clones were used to make plasmid preparations, and the pSTA-F + pSTA-R fragments were PCR-amplified for sequencing using the T7 universal primer. Plasmid pST76AIS*3*X was made similarly, except that the IS*3* element of the *E. coli* MDS30 genome was amplified as two overlapping PCR fragments using primers IS3_Ssp_up + IS3_Xho_Rev and IS3_Xho_Fwd + IS3_Ssp_Dn.

pST76AIS*1*X::*vioABCDE:* pST76AIS*1*X was PCR-amplified using primers IS1VioFwd and IS1VioRev. The violacein operon was amplified as two overlapping PCR fragments from pUTLIQ_*vioABCDE* using primers pUTLIQfwd + VioCendRev2 and VioCendfw + pUTLIQRev7, respectively. All three fragments were assembled using the NEBuilder® HiFi DNA Assembly Cloning Kit and transformed into *E. coli* MDS42. Colony PCR screening was performed with primers pSTA-F + VioARev. Plasmids were prepared from positive clones and were verified by restriction digestion using the SspI fast digest enzyme. Plasmid pST76AIS3X::*vioABCDE* was made similarly, except that the initial PCR fragment was amplified from pST76AISX using primers IS3VioFwd and IS3VioRev.

pST76AIS*1*X::*vioABCDE*_SpR: It was generated by inserting the SpR gene into pST76AIS*1*X::*vioABCDE* by recombineering. This required the initial amplification of the SpR gene from pCDM4 using primers (SpRVioIntFw and SpRVioIntRev) that carried the

appropriate homology boxes to direct the integration of the PCR fragment into the region between the *vioE* gene and the right homology box, i.e. the 3′ fragment of IS*1*. This PCR fragment was transformed into *E. coli* MDS42/pORTMAGE4/pST76AIS*1*X::*vioABCDE* induced by a transient heat shock (see below) in order to express the λ-Red recombinases to facilitate homologous recombination. Cells were plated on Ap + Sp agar plates at 30 °C. Colony PCR screening was performed with primer pairs pSTA-R + SmRev as well as TetR1 + SmRev, respectively. Double positive colonies were replica-plated on Sp as well as Cm plates in parallel. Sp^RCm^S colonies were picked for plasmid preparation. The structure of plasmids was verified with double digestion using KpnI + NheI fast digest enzymes (Thermo Fisher Scientific, Waltham, MA, USA).

pST76AIS*1*X::*vioABCDE*_KmR: It was made in a way similar to pST76AIS*1*X::*vioABCDE*_SpR except that the KmR gene to be integrated was amplified from pSG76-K using primers KmRVioIntFw and KmRVioIntRev. Colony screening was performed with primer pairs pSTKF + pSTAR, pSTKF + TetR1, and Km_HindIII_Fwd + Km_HindIII_Rev. Triple positive clones were replica-plated on Km as well as Cm plates in parallel. Km^RCm^S colonies were picked for plasmid preparation. Plasmids were digested using the PvuII fast digest enzyme and compared with the digestion pattern of pST76AIS*1*X::*vioABCDE*.

pST76AIS*3*X::*vioABCDE*_SpR: It was made like pST76AIS*1*X::*vioABCDE*_SpR, except that the recombineering took place in *E. coli* MDS42/pORTMAGE4/pST76AIS*3*X::*vioABCDE*. Plasmid pST76AIS*3*X::*vioABCDE*_KmR was generated similarly to pST76AIS*1*X::*vioABCDE*_KmR, except that the recombineering took place in *E. coli* MDS42/pORTMAGE4/pST76AIS*3*X::*vioABCDE*. In the end, the PvuII-digested plasmid preparations were compared to PvuII-digested pST76AIS*3*X::*vioABCDE*.

pZA31insAB'tetR:- The *insA* ORF of IS*1* was PCR-amplified from the *E. coli* MG1655 genome with Q5 DNA polymerase using the InsAfw and InsArev primers. The *insB* of IS*1* was PCR-amplified similarly using the InsBfw + InsBrev primers. The two products were fused in an overlap-PCR to make the InsAB' linear fragment carrying the two ORFs of IS*1* within the

same frame. This fragment was digested with SphI + BamHI Fast Digest enzymes and the product was purified with the Viogene Gel/PCR DNA Isolation Kit. The pZA31CFPtetR vector (Fehér et al., 2012) was PCR-amplified with primers pZA31rev + pZA31fw, digested with SphI + BamHI, and ligated with the digested InsAB' fragment using the T4 ligase enzyme. The ligation product was transformed into *E. coli* MDS42 followed by selection on Cm agar plate. The obtained colonies were PCR-screened with primers tetP5 + IS1/1. Clones displaying a PCR product of 700 bp were used to make plasmid preparations which were sequenced using the tetP5 primer.

pSTinsAB'tetR (carrying a temperature-sensitive origin of replication): The fragment containing the pSC101* origin plus the *repA* gene was PCR-amplified from pST76-K using primers pSToriF + pSToriR followed by DpnI digestion and purification. Both the PCR product and pZA31insABtetR were digested with XmaJI + AseI, purified, and ligated using T4 ligase. The ligation product was transformed into *E. coli* MDS42 followed by selection on Cm agar plate at 30 °C. Colonies were PCR-screened with primers pSTA-R + tetP5. The temperature-sensitive nature of PCR-positive colonies was verified by their inability to grow at 37 °C.

pZA31tnp3tetR: The codon-optimized ORFs of IS*3* were obtained from Eurofins Genomics GmBH. The *insE* and *insF* ORFs were synthesized to be in the same reading frame, and the entire construct was flanked by SphI and BamHI sites. The synthetic *insEF'* gene was cloned between the SphI and BamHI sites of pZA31CFPtetR, as described above. The colony-PCR screening was carried out with the primer pairs tetP5 + tnp3rev and tetP5 + pZArev, respectively.

To make pSTtnp3tetR (carrying a temperature-sensitive origin of replication), the pSC101* origin was inserted exactly as described for pSTinsAB'tetR above.

pSG78A_full_IS*3*::SpR: The IS*3*::SpR cassette was PCR-amplified from the *E. coli* MDS30_IS*3*::SpR genome (generated by recombineering, see below) with primers Ecob1028fw + Bamb1028rev. The PCR fragment was DpnI digested, purified, and sequenced from both ends. After confirming the correct sequence, the PCR fragment and the pSG78-A vector were both digested with EcoRI + BamHI, purified, and ligated together using T4 ligase.

The ligation product was heat-inactivated at 65 °C, dialyzed on a 50 nm pore-size membrane (Millipore), and transformed into MDS42-π (expressing the Pir protein) followed by overnight selection on Ap + Sp plate at 37 °C. Colonies were PCR-screened with pSGR + SmFwd primers. Two positive clones were used for plasmid preparation followed by plasmid re-transformation into MDS42-π to ensure purity. Three colonies from each plate were picked, cultured in 5 mL of LB + Ap + Sp overnight at 37 °C, then the insertion of the IS*3*:SpRcassette was PCR-amplified with three primer pairs (Ap1 + SmRev, pSGR + SmFwd, and Ap1 + pSGR). Clones displaying correct products were picked for plasmid preparation and sent for sequencing with T7 + pSGR primers.

pCas9IS*1*: This plasmid is capable of directing Cas9 to cleave IS*1*. It was generated from pCas9 by annealing IS1-SPC1 and IS1-SPC2 single-stranded phosphorylated oligonucleotides and ligating the duplex into the BsaI site, as previously described in the pCas9 protocol (Addgene #42876) (Jiang et al., 2013). Colony PCR screening was performed with the pCas9Fwd and IS1_SPC_Chek_Rev primer pair. The plasmid was prepared from positive clones and plasmid PCR was made using primers pCas9Fwd + pCas9Rev for sequence verification of the spacer.

Table 3. Primers used in this study

name	**sequence**
Km13fw	GTCAAGAAGGCGATAGAAGGCG
Km96rev	GGCGAATGGGCTGACCG
Sp347fw	GCTGGACCTACCAAGGCAAC
Sp439rev	TGCAGGTATCTTCGAGCCAG
lacZ110fw	CCGTCGATATTCAGCCATGTGC
lacZ200rev	TCAGCCGCTACAGTCAACAG
Sp110fw	CTAGCTTCAAGTATGACGGGCTG
IS1D4	AAGCCTTACGCGAAGCGA
YeaE3	ACGCTCCTGAGATTACA
IS1ASpF	*GATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTC*TATTTAACGACCCTGCCCTG
IS1ASpR	*TGACTTTGTCATGCAGCTCCACCGATTTTGAGAACGACAGC*AAGGGAGAAAGGCGGA

	ACAG
IS1ACmR_F	*ATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTAT*AACAGGTTGAACTGCG GATC
IS1ACmR_R	*TGACTTTGTCATGCAGCTCCACCGATTTTGAGAACGACAG*CCTGTGACGGAAGA TCACT TCG
IS1Sp100F	*CCAGTGGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGG TGCGT AACGGCAAAAGCACCGCCGGACATCAGCGCTATCTCTGCT*TATTTAACGACCCT GCCCTG
IS1Sp100R	*ACTTTGTCATGCAGCTCCACCGATTTTGAGAACGACAGCGACTTCCGTCCCAGCC GTGCCA GGTGCTGCCTCAGATTCAGGTTATGCCGCTCAATTCGC*AACTTTGTATGTGTCC GCAGC
IS3SpF	*TCAGAGGTGACTCACATGACAAAAACAGTATCAACCAGTAC*TATTTAACGACCCT GCCCTG
IS3SpR	*CACACCAACTGTGCCGCCGCCACCGATTGTAATCACATT*CGACCGAGTGAGCTG GCTATT TG
IS3flanking1	CCTAAATTAGCGCCCGTTCC
IS3flanking2	TTAATCGTGGTGTGGTAGAAGG
pKdSeq5	CAGTGAATGGGGGTAAATGGC
pKDseq3	ACCACCGCGCTACTGC
pKD1	TTTATAACCTCCTTAGAGCTCGA
IS1noScarFwd	*GTGCGTACCGGGTTGAGAAG*GTTTTAGAGCTAGAAATAGCAAG
pKD2	CCAATTGTCCATATTGCATCA
IS1noScarRev	*CTTCTCAACCCGGTACGCAC*GTGCTCAGTATCTCTATCACTGA
IS1noScarCheckR	ACCTTCTCAACCCGGTACG
IS3noScarFwd	*CTGCGACAGCGTTGTCCTCG*GTTTTAGAGCTAGAAATAGCAAG
IS3noScarRev	*CGAGGACAACGCTGTCGCAG*GTGCTCAGTATCTCTATCACTGA
IS3noScarCheckR	ACCTTCTCAACCCGGTACG
IS1_Ssp_up	aatattTAGTGTATGATGGTGTTTTTG
IS1_Xho_Rev	TCAAACTCGAGTCGCCATAGTGCG
IS1_Xho_Fwd	GGCGACTCGAGTTTGACGTGGTGATATGG
IS1_Ssp_Dn	aatattGATAGTGTTTTATGTTCAG
pSTA-F	CCAGTTCGATGTAACCCACT
pSTA-R	GGATCTGAGGTTCTTATGGCT
T7 universal	TAATACGACTCACTATAGGG
IS3_Ssp_up	aatattCGCCTGAATTTCGC

IS3_Xho_Rev	cctcgagCGCAGATTATGCCGC
IS3_Xho_Fwd	tctgcgCTCGAGGTTGCTGCTACG
IS3_Ssp_Dn	aatattGTTCCGGACTGAG
IS1VioFwd	CAATCTGCTCTGATGCCGCACGAGTTTGACGTGGTGATATGG
IS1VioRev	CAGCTCATTTCTTAAGTCTCCCGACTGCGGCCTGAG
pUTLIQfwd	GAGACTTAAGAAATGAGCTG
VioCendRev2	GCGCCTGTTTCAGTTGGCTGG
VioCendfw	GGTACAAGATAGGGAGG GTCAAC
pUTLIQRev7	AGTCTTTCGACTGAGCCTTTC
VioARev	GTGAACGGATACACCTC GCTC
IS3VioFwd	CAATCTGCTCTGATGCCGCACGAGGTTGCTGCTACGATAATG
IS3VioRev	CAGCTCATTTCTTAAGTCTCATCCCGTTCTGCCAGC
SmFw	CTTACGTTGTCCCGCATTTGG
SmRev	ATATCACTGTGTGGCTTCAGG
TetR1	CACTAGAGAACATACTGGCTA
pSTKF	GTCTGTTGTGCCCAGTCATAG
Km_HindIII_Fwd	GGAAGCTT-AGGATCGTTTCGCATGATTG
Km_HindIII_Rev	GGAAGCTT-GACGGTATCGAACCCCAGAG
pCas9Fwd	CAGCTAGGAGGTGACTGAAG
pCas9Rev	GGACGATCACACTACTCTTC
IS1_SPC_Chek_Rev	CTTCTCAACCCGGTACGCAC
IS1-SPC1	AAACTGATTTTCTGGTGCGTACCGGGTTGAGAAGG
IS1-SPC2	AAAACCTTCTCAACCCGGTACGCACCAGAAAATCA
SPC-IS3-Fwd-long	AAACTCCGCCAACACTGCGACAGCGTTGTCCTCGg
SPC-IS3-Rev-long	AAAACCGAGGACAACGCTGTCGCAGTGTTGGCGGA
SPC-IS3-Fwd-short	AAACCTGCGACAGCGTTGTCCTCGg
SPC-IS3-Rev-short	AAAACCGAGGACAACGCTGTCGCAG
IS3_SPC_Chek_Rev	GCTCTAAAACCGAGGACAACG
InsAfw	CGCATGCTCGTGGCTTCTGTTTCTATCAG
InsArev	GCTTCTTCAAGTGACGTAAAATCGTG
InsBfw	CGTCACTTGAAGAAGCTCAGGCCGCAG
InsBrev	TTCGGATCCTTATTGATAGTGTTTTATGTTCAG
pZA31rev	TACGCATGCGGTACCTT
pZA31fw	CTAGGATCCATAAGCTTAATTAGCTGAGTCTAGAG

tetP5	ACTGAGCACATCAGCAGGAC
IS1/1	TGAGAACGACAGCGAC
pSToriF	ACCTAGGCTTGGCACTGGCTGATC
pSToriR	GATTAATACAGCGTTTGCGACATCC
tnp3rev	ACGCAGAACACGGCACATAG
pZArev	TCTAGGGCGGCGGATTTGTC
Ecob1028fw	tctgaattcACCGGGCGTAATAAGGTGG
Bamb1028rev	tgtggaTCCAGTCCGGCAAGATAATCG
pSGR	CACATACGATTTAGGTGACACT
Ap1	CCTCCATCCAGTCTATTAATTGTT
CEKG2B:	GGCCACGCGTCGACTAGTACNNNNNNNNNNACGCC
CEKG4:	GGCCACGCGTCGACTAGTAC
SpRVioIntFw	*AGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGC*CTATTTAACGACCCTGCCCTG
SpRVioIntRev	*TTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCA*CCGAGTGAGCTGGCTATTTG
KmRVioIntFw	*AGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGC*AGCCATGAGGGTTTAGTTCG
KmRVioIntRev	*TTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCA*CCTGTTATCCCTAGCGGATC

Italic font: overhangs, Lower case: restriction sites, Upper case: primer binding site

pCas9_IS*3*: This plasmid is capable of directing Cas9 to cleave IS*3*. It was generated similarly to pCas9IS*1*, except that the spacer was made by annealing the following pairs of phosphorylated oligonucleotides: SPC-IS3-Fwd-long and SPC-IS3-Rev-long (for 30 bp-long spacers) or SPC-IS3-Fwd-short and SPC-IS3-Rev-short (for 20 bp-long spacers). Colony PCR screening was performed with the pCas9Fwd and IS3_SPC_Chek_Rev primer pair.

Maps of pSG78A_full_IS*3*::SpR, pST76AIS*1*X::*vioABCDE and* pST76AIS*3*X::*vioABCDE* are shown in **appendix** of this thesis (**S1, S2, S3**).

3.4. Electroporation of *E. coli* cells

In our fast protocol to generate electrocompetent cells, *E. coli* cultures were grown in 10 mL of LB medium supplemented with the appropriate antibiotic to an OD600 value of 0.45–0.55.

Cells were pelleted at 10,000 g for 2 min and resuspended in 2 × 1.5 mL of ice-cold sterile Milli-Q water three times. The final resuspension took place in 40 μL of water. For electroporation, the cell suspension was mixed with the DNA, and the mixture was transferred into electroporation cuvettes harboring a 1 mm gap (Cell Projects Ltd., Herrietsham, UK). Electroporation was carried out in a MicroPulser electroporator (BioRad, Hercules, CA, USA) set to a voltage of 1.8 kV. The cells were recovered in 1 mL of LB and shaken for 1–2 h at 30 or 37°C as was required by the experiment. Finally, 10–100% of the recovery culture was plated on LB agar plates supplemented with the appropriate antibiotic.

3.5. Integration of resistance genes into genomic IS elements by recombineering

For linear DNA-mediated recombineering, the fragments targeting IS*1* were made by PCR employing the spectinomycin-resistant pCDM4 (Xu et al., 2012) as a template and IS1ASpF and IS1ASpR (40 bp long homologies) as primers. For a fragment encoding chloramphenicol-resistance, pSG76CS was used as a template, and, IS1ACmR_F and IS1ACmR_R as primers. To test the effect of long (100 bp) homologies primers IS1Sp100F and IS1Sp100R were used to amplify pCDM4. PCR fragments targeting IS*3* were made using primers IS3SpF and IS3SpR using pCDM4 as a template. For primer sequences, see **Table 3**.

A fully grown overnight culture of the *E. coli* host strain harboring pORTMAGE2 (Nyerges et al., 2016) was diluted 100-fold into 50 mL LB + Ap. The culture was grown to an OD600 of 0.45–0.55, then expression of the recombinase genes was then induced with a transient heat shock (42°C, 15 min) followed by 10 min cooling on ice. This culture was used to make fast electrocompetent cells that were electroporated with 100 ng of linear PCR fragment (antibiotic resistance gene flanked by homology boxes corresponding to the respective IS element). After 2 h of recovery at 30°C, 100 μL was spread on either an Sp or Km plate (depending on the resistance gene) and incubated overnight at 37°C. The next day, single co-integrants were screened by colony-PCR using pair of primers that hybridize to the resistance gene and to the chromosomal segment neighboring the targeted IS elements, respectively.

To obtain double co-integrants, Cas9 selection was included. The protocol began as described above, but the recovery culture was grown overnight at 30 °C. The next day, the culture was diluted 50-fold in LB, grown to an OD600 of 0.45–0.55, and was used to make fast competent cells. The cells were electroporated with 100 ng of pCas9_IS*1* or pCas9_IS*3*. After a 2 h recovery in LB at 30°C, Cm was added to select for the pCas9-derived plasmid and the culture

was shaken overnight at 30°C. On the third day, 100 µL was spread on either Sp + Cm or Km + Cm plates (depending on the resistance of the linear cassette) and incubated overnight at 37 °C. On the fourth day, double co-integrants were screened for by colony PCR targeting all the potential genomic IS elements.

3.6. Integration of a resistance gene or the *vioABCDE* operon into genomic IS elements using the NO-SCAR method

A fully grown overnight culture of the *E. coli* host strain harboring pCas9-Cr4 and either pKDsg-IS1 or pKDsg-IS3 (targeting IS*1* and IS*3*, respectively) was diluted 100-fold into 10 mL LB + Cm + Sp. The culture was grown at 30 °C to an OD600 of 0.45–0.55, then 0.2% L-arabinose was added (inducing the λ-Red recombinases), followed by further growth at 30 °C for 15 min and cooling on ice for 10 min. This culture was used to prepare electrocompetent cells (see above), which were electroporated with 100 ng of linear PCR fragment (KmR gene flanked by homology boxes corresponding to the respective IS element). After 2 h of recovery at 30 °C, 900 µL was spread on a Km + aTc plate (to induce the CRISPR/Cas system) and incubated overnight at 37 °C. The next day, single co-integrants were screened for by colony-PCR using a pair of primers that hybridize to the resistance gene and the chromosomal segment neighboring the targeted IS elements, respectively.

To integrate the *vioABCDE* operon into genomic IS elements, linear fragments containing the *vioABCDE* genes and the resistance marker gene flanked by homology boxes corresponding to the targeted IS elements, were generated by linearizing pST76AIS3::*vioABCDE*_KmR by SspI digestion or linearizing pST76AIS3::*vioABCDE*_SpR by KpnI+MunI+NheI co-digestion. The integration protocol was the same as above, except that 0.4% L-arabinose was used for the first induction, 600–800 ng of the linear fragment was electroporated, and the recovery culture was spread on a Km+Sp+Cm+aTc plate and incubated overnight at 30 °C.

3.7. Integration of non-selectable genes into genomic IS elements using the NO-SCAR method

A fully grown overnight culture of the *E. coli* host strain harboring pCas9-Cr4 and either pKDsg-IS1 or pKDsg-IS3 (targeting IS*1* and IS*3*, respectively), was diluted 100-fold into 10 mL LB + Cm + Sp. The culture was grown at 30 °C to an OD600 of 0.45–0.55 and 0.4% L-

arabinose was added (inducing the λ-Red recombinases), followed by further growth at 30 °C for 15 min and cooling on ice for 10 min. This culture was used to prepare electrocompetent cells (see above), which were electroporated with 100 ng of linear PCR fragment (*gfp* gene flanked by homology boxes corresponding to the respective IS element). After 2 h of recovery at 30 °C, 0.5 µL was spread on a Cm + Sp + aTc plate (to induce the CRISPR/Cas system), and incubated overnight at 37 °C. The next day, single co-integrants were screened for by colony-PCR using a pair of primers that hybridize to the gfp gene and the chromosomal segment neighboring the targeted IS elements, respectively.

3.8. Integration of marked IS elements into the genome using transposition

A fully grown overnight culture of the *E. coli* host strain (MDS30 or MDS42) harboring the pSTtnp3tetR was diluted 100-fold into 5 mL LB + Cm + aTc and was grown at 30 °C for 2 h. Then, a total of 5 mL of this culture was used to make fast electrocompetent cells, which were electroporated with 150–200 ng of pSG78A_full_IS3::SpR. After 1 h of recovery at 30 °C, 100 µL was spread on an Sp plate and incubated overnight at 37 °C. The next day, several colonies were replica-plated on Cm, Ap, and Sp plates at 37 °C. One day later, $Sp^{R}Cm^{S}Ap^{S}$ colonies (i.e., those that have lost both plasmids but retained the IS*3*::SpR) were inoculated into liquid cultures to generate glycerol stocks. The genomic positions of IS*3*::SpR insertions were determined using ST-PCR (see below).

3.9. Semi-random two-step PCR (ST-PCR)

To localize a known sequence (IS*3*::SpR in our case) within a bacterial genome, we applied semi-random two-step PCR (ST-PCR) (Chun et al., 1997) with some modifications. Briefly, genomic DNA was PCR-amplified from the colonies of interest first using primer pairs Sp439Rev + CEKG2B and Sp110Fwd + CEKG2B, applying the following program: 94 °C, 2 min, followed by 6 cycles of 94 °C 30 s denaturation; 53 °C, 30 s, annealing with 1 °C decrease each cycle; and 72 °C, 3 min elongation. The DNA product generated by the first PCR was

diluted 5x with TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0), and subsequently used as template for nested PCR with primer pairs SmFwd + CEKG4 and Sp347Fwd + CEKG4 corresponding to the first two PCRs, respectively. The applied program was 30 cycles of 94 °C, 30 s; 65 °C, 30 s; and 72 °C, 3 min. The products were separated on 1% agarose gels and bands of interest were extracted using Viogene Gel/PCR DNA Isolation Kit, and sequenced with SmFwd or Sp347Fwd (depending on which was used for the nested PCR). Relative positions of the primers used for ST-PCR are shown in **Figure 34**.

3.10. Amplification of IS elements marked with antibiotic resistance genes

A fresh overnight culture of the *E. coli* strain harboring the marked IS element was grown at 37 °C. Electrocompetent cells were transformed with the respective transposase-expressing plasmid (pSTinsAB'tetR or pSTtnp3tetR), spread on a Sp+Cm or Km+Cm plate, and incubated overnight at 30 °C. A colony was picked and grown as an overnight starter culture in LB using the same antibiotics. The grown culture was diluted 10,000-fold and grown again at 30 °C in LB with antibiotics and the inducer aTC (an uninduced control was grown in parallel). Dilution and growth were repeated after 24 h. The plasmids were then cured by growing the culture at 37 °C in LB without antibiotic selection for five days, applying a 10,000-fold dilution during each transfer. From the 5^{th} fully grown serial culture, 1 µL was spread on 60× Sp gradient plate or 20× Km gradient plate (Szybalski & Bryson, 1952) and incubated at 37 °C for 24 h. The next day, 15 colonies from the high antibiotic-concentration area of the gradient plate were picked and replica-plated on Cm at 30 °C and Sp or Km at 37 °C. After overnight incubation, $Cm^{S}Sp^{R}$ or $Cm^{S}Km^{R}$ colonies were chosen and grown in liquid LB without selection at 37 °C. The fully grown cultures were saved as glycerol stocks and/or used for violacein quantification or to prepare genomic DNA for droplet-digital dPCR analysis. This process, defined as one round of transposase induction, was carried out up to three times.

3.11. Quantification of violacein production

To quantify the violacein production of liquid *E. coli* cultures, we applied the modified protocol of Zhu et al. (Zhu et al., 2011). The investigated strain was grown in LB with or without antibiotic selection for 24 h at 37 °C. The cells were sedimented from one ml of the culture by centrifugation at 13,000 rpm for 10 min. After discarding the culture supernatant, 1 mL of

DMSO (>99%) (Molar Chemicals Kft., Halásztelek, Hungary) was added to the pellet. The solution was vigorously vortexed for 30 s to completely solubilize violacein and was centrifuged again at 13,000 rpm for 10 min to remove the cell debris. Then, 200 μL aliquots of the violacein-containing supernatant were transferred to a 96-well flat-bottomed microplate (Greiner Bio-One International, Kremsmünster, Austria), making three technical replicates, and the absorbance was recorded at 585 nm using a Synergy2 microplate reader (BioTek, Winooski, VT, USA).

3.12. Monitoring the stability of violacein production

The stability of violacein production was assessed in two different experiments. The first type of experiment monitored the ratio of purple colonies in the lack of antibiotic selection. Overnight starter cultures of the investigated strains were grown in LB with selection (Sp or Km for the genomic co-integrants, Tc for pUTLIQ_vioABCDE-carrying strains) at 37 °C. The culture was pelleted and resuspended twice in water to dispose of the antibiotics, then plated in appropriate dilutions on solid LB medium to obtain individual colonies. The number of purple and total colonies was counted on each plate, and their ratio yielded the "day 0" value of cells expressing violacein. The cultures were then diluted 10,000-fold in LB medium and were fully grown without selection at 37 °C. Plating and enumerating the ratio of purple colonies was repeated to yield the ratio of cells expressing violacein for days 1 through 4. Each strain was investigated in three biological replicates.

The second type of experiment monitored the violacein production of liquid cultures grown in the lack of selection. Overnight starter cultures of the investigated strains were grown in LB with selection (Sp or Km for the genomic co-integrants, Tc for pUTLIQ_*vioABCDE* -carrying strains) at 37 °C. The culture was then diluted 1000-fold in LB medium, and was fully grown without selection at 37 °C. This dilution/growth cycle was repeated three more times. Each day, 1 mL of the fully grown culture was analyzed for the violacein content, as described above. Each strain was investigated in three biological replicates.

3.13. Droplet digital PCR

To determine the copy numbers of marked IS elements within the bacterial chromosome, droplet digital PCR (ddPCR) experiments were performed using the EvaGreen protocol of BioRad QX200 Droplet DigitalPCR system (BioRad, Hercules, CA, USA). The template

genomic DNA from *E. coli* strains waspurified with the NucleoSpin Microbial DNA kit (Macherey Nagel) and digested with SspI (Thermo Fisher Scientific, Waltham, MA, USA). Final concentrations in the reaction mixtures were the following: 1× QX200 EvaGreen Digital PCR Supermix (BioRad, Hercules, CA, USA), 3 pg digested *E. coli* genomic DNA, and 200 nM combined primer mix in 25 μL final volume. Twenty microliter reaction mixtures were used for droplet generation using the QX200 droplet generator. After partitioning, the samples were transferred into a 96-well plate, sealed, and put in a T100 Thermal cycler (BioRad, Hercules, CA, USA). The following cycling protocol was used: 95 °C for 10 min, followed by 40 cycles of 94 °C for 30 s and 60 °C for 1 min followed by 5 min at 4 °C, 5 min at 95 °C, and finally at 4 °C infinite hold. The droplets were then read in the FAM channels and analyzed using the QX200 reader (Bio-Rad, Hercules, CA, USA). Primers specific for the Km-resistance gene (Km13fw and Km96rev) and the Sp-resistance gene (Sp347Fwd and Sp439rev) were designed using the ApE software (https://jorgensen.biology.utah.edu/wayned/ape) and ordered from MWG Eurofins Genomics Gmbh (Ebersberg, Germany). Final copy numbers were normalized to the bacterial single-copy *lacZ* gene, amplified with primers lacZ110fw and lacZ200rev. All primer sequences are listed in **Table 3**.

3.14. Whole genome sequencing of bacteria

For whole-genome shotgun sequencing, a genomic library was prepared from four strains (B0, B1, B2, and B3) using the Nextera XT Library Preparation Kit (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. The libraries were sequenced with Illumina NextSeq 500 sequencers using 2×150 PE sequencing. Reads from all samples were aligned to the *E. coli* BLK09IS3::*vioABCDE*_SpR reference genome with BWA (H. Li & Durbin, 2009). To find all insertion sites, we used the Smith-Waterman algorithm (Smith & Waterman, 1981) by filtering all reads where the IS*3* element's 3′ 25 nt (TGATCCTACCCACGTAATATGGACA) or 5′ 25 nt (TGTCCACTATTGCTGGGTAAGATCA) sequence or their reverse complements could be found with a maximum of 2 mismatches. Then we used the algorithm to select all those reads which could only be aligned with more than 10 mismatches to the +/− 200 nt region of the original single insertion site. Using the same algorithm, we trimmed all nt from each read which could be aligned to the insertion element. The resulting fastq file contained all reads marking the neighboring sequences of each insertion. This fastq file was remapped by BWA to the *E.*

coli BLK09IS3::*vioABCDE*_SpR reference genome. To mark independent insertion sites, we searched for peak positions different from those of sample B0. To calculate the overall number of independent insertions and repeated re-insertions, relative coverage was measured in the complete alignments by calculating the ratio of reads at +/− 10 nts from peak borders.

4. Results

4.1 IS*1* and IS*3* elements can be used as landing pads in the *E. coli* genome

The general strategy tested in this work is to use ISes as landing pads, i.e. recombination targets to integrate transgenes into bacterial genomes, followed by their copy-number amplification using transposition. Several considerations are supporting this idea: i) to the best of our knowledge, ISes have never been found to be essential, ii) ISes are near-ubiquitous, having been detected in 76% of bacterial genomes (Touchon & Rocha, 2007), iii) the type, number and exact position of ISes in each sequenced genome are well known, iv) copy-and-paste type of ISes have been identified and v) the mobility (transposition rate) of various IS types is often known, allowing the anticipation of potential transgene amplification. We chose to test two copy-and-paste types of IS elements, IS*1* and IS*3*, which are at the high- and the low-end of the transpositional activity scale in *E. coli*, respectively (Chalmers & Blot, 1999). Concerning the copy numbers of these elements, wild type (wt) *E. coli* K-12 MG1655 carries 8 copies of IS*1* and 5 copies of IS*3*, while *E. coli* BL21(DE3)pLysGold carries 28 and 4 copies, respectively. In the course of an earlier project carried out in our laboratory, however, we deleted all ISes from MG1655 (Pósfai et al., 2006), and saved intermediate strains carrying 2 or 1 copies of IS*1* (called MDS39R2 and MDS42IS*1*, respectively), or 3, 2 or 1 copies of IS*3* (called MDS16, MDS27, and MDS30, respectively). Similarly, during BL21(DE3) genome reduction (Umenhoffer et al., 2017), we generated BLK09, which carries 2 active copies of IS*3*, and BLK16, where the same 2 copies of IS*3* are inactivated by nonsense mutations within the transposase gene (**Table 1**). We were therefore in the fortunate situation that we could choose multi-deletion strains carrying limited copies of the targeted elements, thereby simplifying the analysis of transgene integration.

4.2 Targeting single IS-copy strains by selection marker using recombineering

In the first type of experiment, we asked whether targeting IS*1* and IS*3* were just as straightforward as targeting other nonessential genes of *E. coli*. We generated PCR fragments

that harbored an antibiotic resistance gene flanked by two appropriate homology boxes provided by the PCR primers. We expressed the λ-Red recombinases from the heat-inducible pORTMAGE2 plasmid to facilitate the double-crossover between the respective chromosomal IS element and the resistance-cassette. In parallel experiments, we targeted the single IS*1* copy of MDS42IS*1* and the single IS*3* copy of MDS30. As expected, the colonies obtained on selective plates were genomic co-integrants: using colony PCR, 10 out of 10 colonies tested positive in both the IS*1*-targeting and the IS*3*-targeting experiment.

These results showed that integrating selective markers into IS elements by recombineering is not different from targeting other non-essential genes. It is noteworthy, however, that SpR as a selection marker yielded 20-fold more recombinants than CmR (15.26 colony/ng DNA vs. 0.718 colony/ng DNA, respectively, as listed in **Figure 1**.

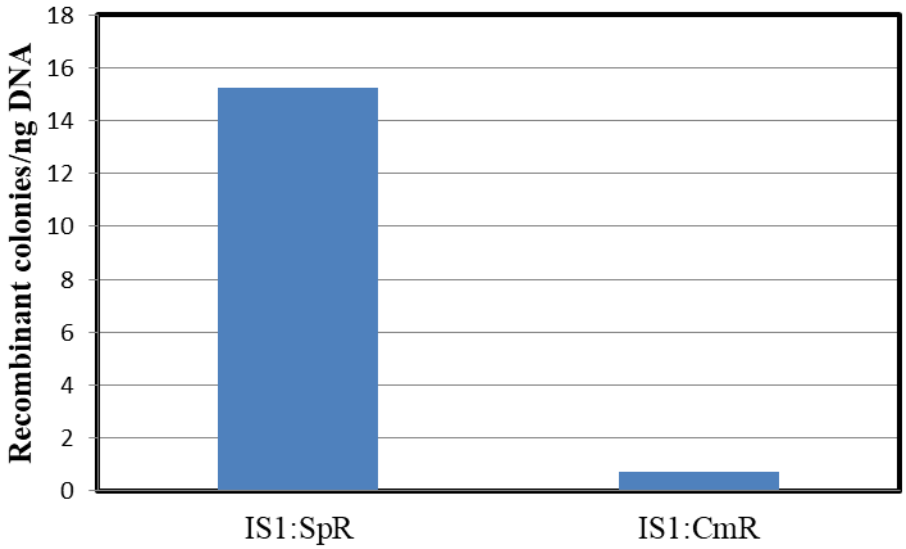


Figure 1. The effect of antibiotic choice on the yield of recombinants. The absolute yield of integration is ~20 times higher for SpR than for CmR when targeting a single copy of IS*1* in MDS42IS*1* strain by recombineering.

4.3 Targeting double IS-copy strains by selection marker using recombineering

In the second class of experiments, we repeated the recombineering process as above, but used strains carrying two copies of the targeted ISes: MDS39R2 for IS*1* and BLK09 or BLK16 for IS*3*. All colonies obtained on the selective plates were PCR-screened for integration at both loci. IS*1* targeting in MDS39R2 resulted in 13 of 25 colonies (52%) that carried the SpR gene

in either of the IS*1* elements. In the remaining 12 clones the SpR cassette integrated at an unknown locus. The majority of the known integrations occurred in the *yeaJ* gene, and the minority in the *ais gene* (11 vs 2, respectively (**Figure 2**).

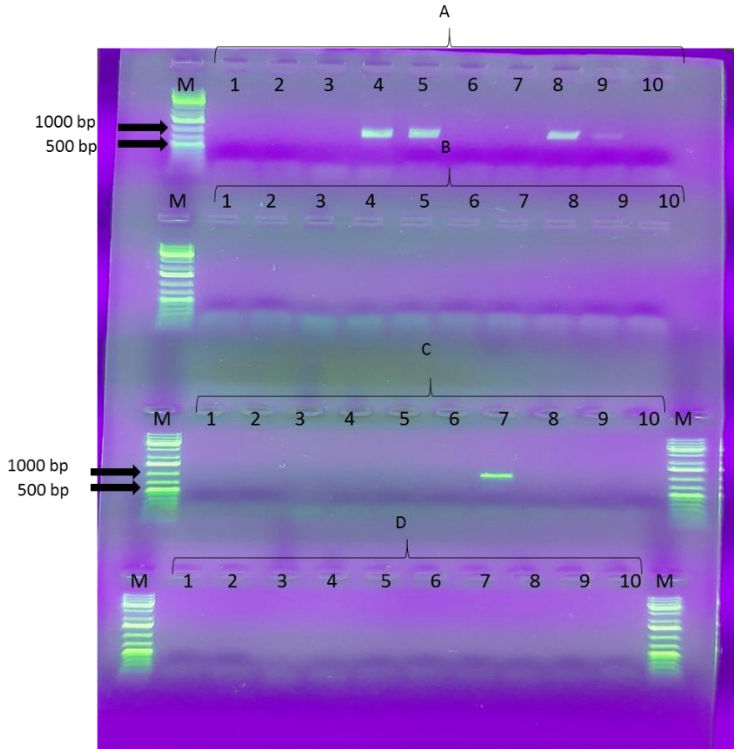


Figure 2. Targeting the two genomic IS*1* elements of MDS39R2 with the SpR cassette using pORTMAGE2-mediated recombineering (one of three parallel experiments). Ten colonies obtained upon the recombineering process are screened using four primer sets: A: SmFw + YeaE3 = 924 bp, B: SmRev + YeaE3, C: SmFw + aisE2=1000 bp, D: SmRev + aisE2. Four colonies were positive for integration at the *yeaJ* locus, and one colony was positive for integration at the *ais* locus. M: GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific)

In contrast to what we observed with IS1, targeting IS*3* in BLK09 and BLK16 always resulted in integration of the resistance cassette at either one of the two loci. Again, there was a bias favoring the integration into the IS*3* at locus 1 (gene ECBD2567) vs. locus 2 (gene ECBD0875) (7:3, respectively). The ratios of "PCR-verified co-integrants" at each locus are shown in **Figure 3**.

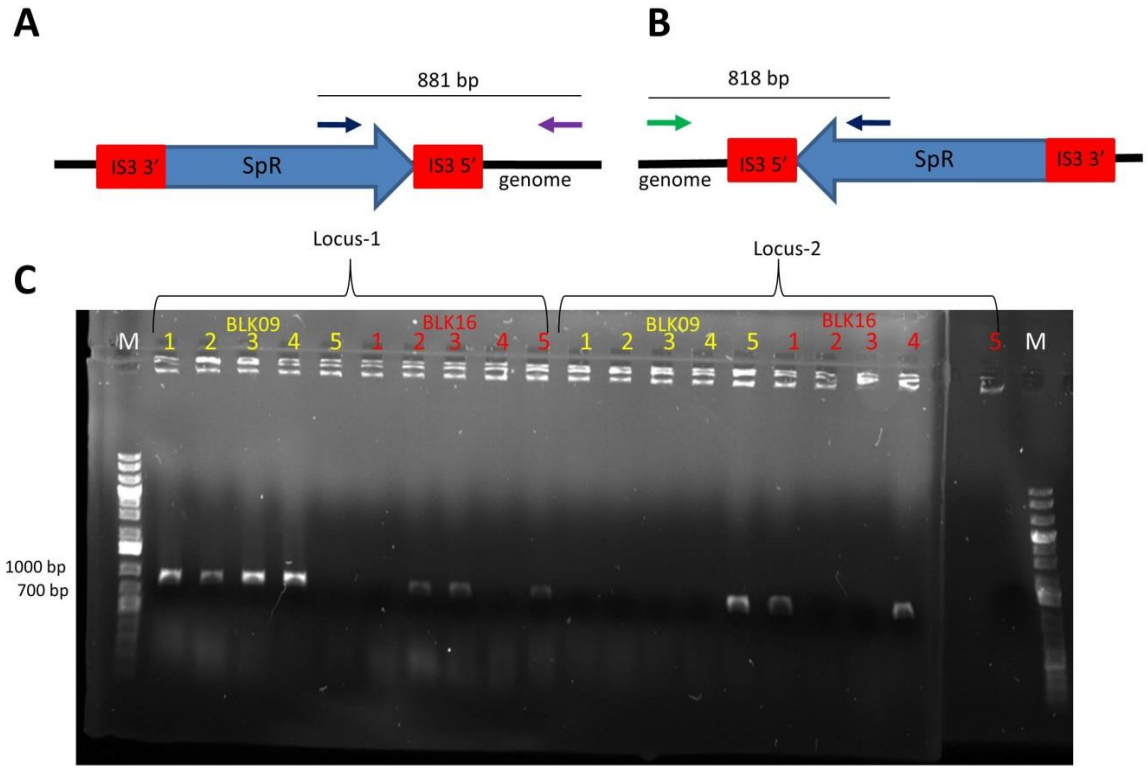


Figure 3. Colony PCR to verify chromosomal SpR integration targeting IS*3* after recombineering. (A) Expected genetic arrangement and PCR product in case of successful SpR integration at locus-1. (B) Expected genetic arrangement and PCR product in case of successful SpR integration at locus-2. Black arrow: primer SmFw; purple arrow: primer IS3flanking1; green arrow: primer IS3flanking2. SpR: spectinomycin resistance gene. Red boxes depict IS*3*-homologies. (C) Gel electrophoresis of colony-PCR reactions made for screening of 5 *E. coli* BLK09 and 5 *E. coli* BLK16 colonies obtained upon recombineering. Primer pairs either specific for locus 1-integration or locus 2-integration were used in both cases. M: GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific).

Table 4. Yields of recombinants obtained by recombineering of the SpR, CmR, *vioABCDE*:SpR, and *vioABCDE*:KmR cassettes into strains harboring single, double, and triple copies of the targeted IS element.

Strain	IS element tpe	no of IS elements in genome	linear DNA type	ng DNA used	total colonies per ml	PCR positive colonies	% of positive colonies	total correct colonies	Absolute efficiency of integration
MDS42IS*1*	IS1	1	S1::SpR/short homology	100	2064	10 out of 10	100	2064	20.6
MDS42IS*1*	IS1	1	IS1::SpR/long homology	100	26880	10 out of 10	100	26880	268.8
MDS39R2	IS1	2	S1::SpR/short homology	100	31300	5 out of 10	50	15650	156.5
MDS30	IS3	1	IS3::SpR	100	920	10 out of 10	100	920	9.2
BLK09	IS3	2	IS3::SpR	100	272	10 out of 10	100	272	2.7
MDS27	IS3	2	IS3::SpR	100	860	10 out of 10	100	860	8.6
MDS42IS*1*	IS1	1	IS1::VioAESpR	450	2400	6 out of 8	75	1800	4.0
MDS42IS*1*	IS1	1	IS1::VioAEKmR	400	1010	20 out of 20	100	1010	2.5
MDS39R2	IS1	2	IS1::VioAESpR	450	1690	7 out of 20	35	591.5	1.3
MDS39R2	IS1	2	IS1::VioAEKmR	400	3400	13 out of 20	65	2210	5.5
MDS30	IS3	1	IS3::VioAESpR	1000	10	3 out of 9	33.33	3.333	0.003
MDS30	IS3	1	IS3::VioAEKmR	750	4060	27 out of 30	90	3654	4.9
BLK09	IS3	2	IS3::VioAESpR	1000	5090	4 out of 20	20	1018	1.0
BLK09	IS3	2	IS3::VioAEKmR	750	630	13 out of 20	65	409.5	0.5
BLK16	IS3	2	IS3::VioAESpR	1000	3820	15 out of 20	75	2865	2.9
BLK16	IS3	2	IS3::VioAEKmR	750	3350	10 out of 20	50	1675	2.2
MDS27	IS3	2	IS3::VioAESpR	1000	450	5 out of 20	25	112.5	0.1
MDS27	IS3	2	IS3::VioAEKmR	750	900	9 out of 20	45	405	0.5
MDS42IS*1*	IS1	1	IS*1*::SpR	107.6	1642	10 out of 10	100	1642	15.3
MDS42IS*1*	IS1	1	IS*1*::CmR	123.8	89	10 out of 10	100	89	0.7
MDS39R2	IS1	2	IS1::SpR	107.6	2160	5 out of 6	83.33	1799.928	16.7
MDS39R2	IS1	2	IS1::CmR	123.8	81	3 out of 6	49.99	40.4919	0.3
MDS16	IS3	3	IS3::VioAEKmR	750	3200	11 out of 20	55	1760	2.3

Total correct colonies' = 'total colonies per ml' * '% of positive colonies'
'Absolute efficiency of integration' = 'total correct colonies'/ 'ng DNA used'

Double co-integrants have never been obtained this way in either the IS*1* or the IS*3*-targeting experiments. We confirmed the observation previously made with targeting single copy IS*1*

that that using SpR as a transgene yields more co-integrants than CmR (16.73 vs. 0.33 colony/ng, respectively, see **Table 4**, explaining our preference for SpR in further experiments.

4.4 Recombineering-mediated integration of violacein operon was not possible without antibiotic selection

In our third group of experiments, we attempted the integration of the 7.3 kbp-long *vioABCDE* cassette into the bacterial genome using λ-Red-mediated recombineering. The *vioABCDE* operon originates from *Chromobacterium violaceum* and is responsible for producing the purple pigment violacein (Durán et al., 2007). Since generating linear *vioABCDE* fragments flanked by appropriate homologies using PCR was unsuccessful, we cloned the operon flanked by homology arms corresponding to IS*1* or IS*3* elements into the temperature-sensitive plasmid vector pST76-A to yield pST76AIS1X::*vioABCDE* and pST76AIS3X::*vioABCDE*, respectively (Note that neither of these plasmids contains the inverted repeats of the targeted IS). Linear DNA fragments generated from these plasmids were substrates of recombineering using pORTMAGE2. These cassettes lacked an antibiotic selection marker, our rationale was to obtain a lawn of cells on nonselective plates, and visually screen for purple sectors. Such purple sectors were never visible by the naked eye, indicating that the rate of recombination is too low for a visual screening strategy.

4.5 Recombineering-mediated integration of violacein operon was achieved in single copies using antibiotic selection

Since the integration of the violacein operon into the genome of *E. coli* was not possible using visual screening, we inserted, by pORTMAGE4 mediated recombineering, an antibiotic resistance gene (SpR or KmR) into the cassette next to the *vio* operon, within the segment flanked by IS-specific homologies. Four plasmids were generated: pST76AIS1X::*vioABCDE*_SpR, pST76AIS1X::*vioABCDE*_KmR, pST76AIS3X::*vioABCDE*_SpR and pST76AIS3X::*vioABCDE*_KmR. Again, linear DNA cassettes for recombineering were generated from these plasmids by restriction digestion. When electroporating these cassettes, we obtained mixed results: on the one hand, if targeting genomic IS*1* in MDS42IS*1*, correct recombination events could be verified, but none of the co-integrant

colonies were purple. This was the case even though all four plasmids used to generate the cassettes granted their host a dark purple colony phenotype. On the other hand, when targeting the IS*3* of MDS30, the PCR-verified genomic co-integrants displayed a purple color: Three out of nine colonies (33%) were PCR-positive when using SpR, and 9/10 (90%) were positive when using KmR as a selection marker (**Figure 4**).

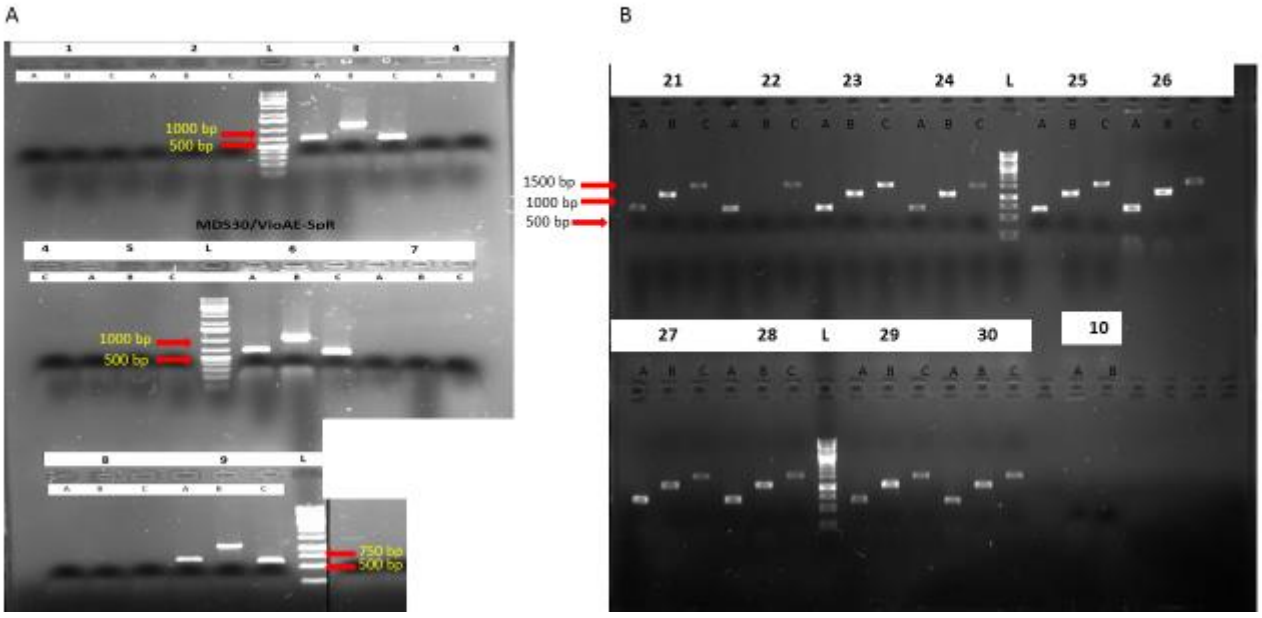


Figure 4. The pORTMAGE-mediated insertion of the *vioABCDE* operon into MDS30. (**A**) Engineering MDS30 with the *vioABCDE*_SpR cassette inserted into IS*3*. Nine colonies obtained upon the recombineering process were PCR-screened using three primer sets: A: VioDF1 + VioDR3 = 660 bp, B: VioAR2 + b1028rev = 1084 bp, C: b1025fwd + SmFw = 616 bp. Three colonies (colonies 3, 6, and 9) yield PCR fragment with the expected size. L: GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific) (**B**) Engineering MDS30 with the *vioABCDE*_KmR cassette inserted into IS*3*::. Ten colonies obtained upon the recombineering process were PCR-screened using three primer sets: A: VioDF1 + VioDR3, B: VioAR2 + b1028rev, C: b1025fwd + Km_HindIII_Rev. Successful PCR-amplification with primer pair B shows genomic integration of *vioABCDE* operon for colonies 21 and 23-30. M: GeneRuler 1 kb DNA Ladder (Thermo Scientific)

Integration of the *vioABCDE*_SpR cassette into IS*3* of *E. coli* strains carrying two genomic copies of IS*3* occurred with rates acceptable for practical purposes: the fractions of colonies

harboring single co-integrants in *E. coli* BLK09, BLK16 and MDS27 strains fell between 20 and 75% (**Table 5)**.

Table 5. The fraction of colonies that carried co-integrant at either of the two IS*3* loci, as verified by PCR. (ND = None detected). BLK09, BLK16 and MDS27 carried two copies of IS*3* element. MDS16 carried three copies of IS*3* element.

	Selection Marker used	
	SpR	KmR
BLK09	20%	65%
BLK16	75%	50%
MDS27	25%	45%
MDS16	ND	55%

Double co-integrants were never obtained this way. We have nevertheless demonstrated that the 9074 bp and 9359 bp-long *vioABCDE*-SpR and *vioABCDE*-KmR cassettes, respectively, could be targeted into genomic IS*3* elements by recombineering. In both experiments targeting IS*3* of BLK09 or BLK16, the majority (>90%) of the colonies displayed a purple color, indicating a relatively low rate of false-positive resistance. In strain MDS27 however, <50% of colonies were purple to the naked eye, and some of the white colonies were true co-integrants verified by PCR. This exemplifies the strain dependence of the efficiency of integration and of the functional expression of the inserted operon. The bias of the integration seen for resistance genes was also present when inserting the five-gene operon, although to a variable extent (**Figure 5**).

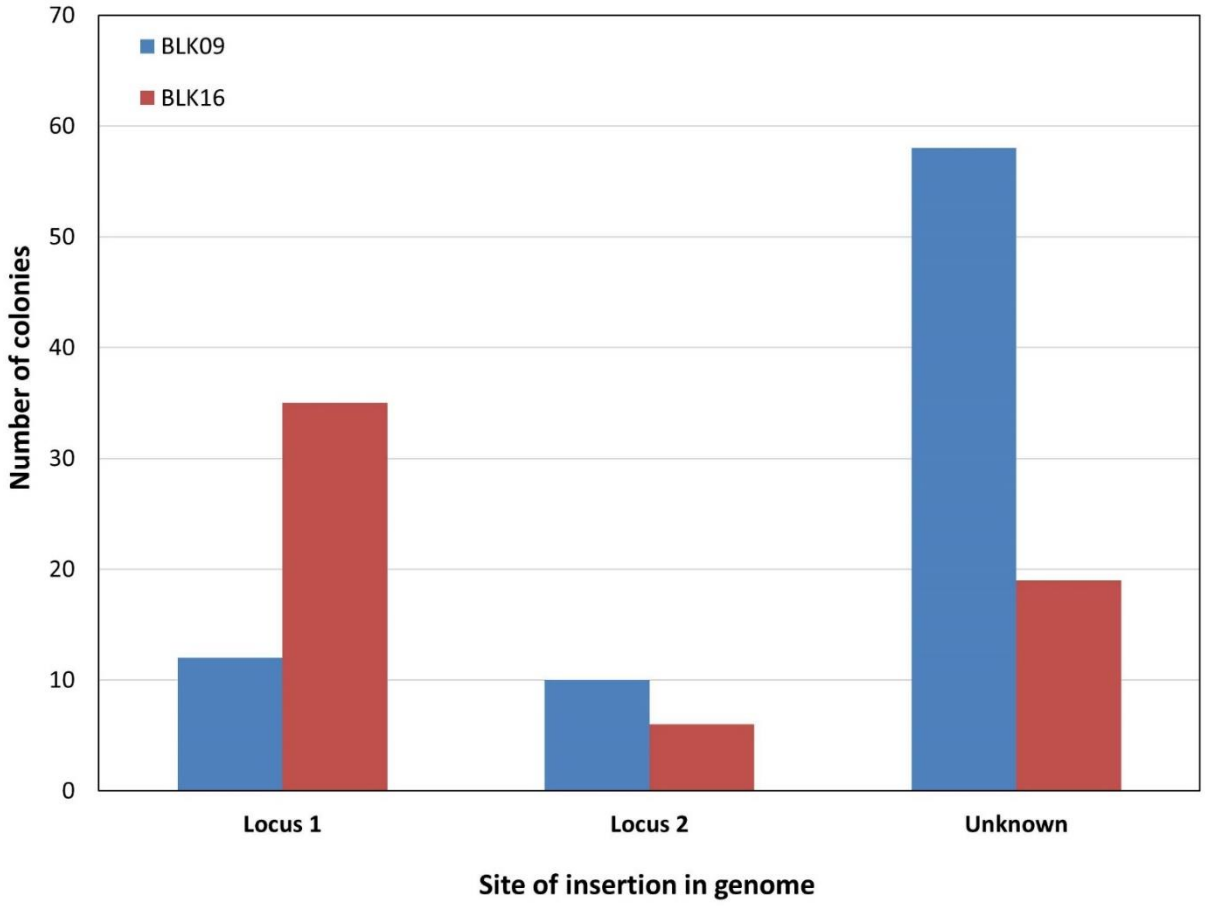


Figure 5. The combined results of recombineering experiments integrating the *vioABCDE*_Km or the *vioABCDE*_SpR cassette into a chromosomal IS*3*. The number of colonies harboring a PCR-verified integration at locus 1, locus 2, or displaying no PCR-verified integration ("unknown") are shown for *E. coli* BLK09 (blue) and *E. coli* BLK16 (red). The displayed results derive from a set of four similar experiments (two for each strain).

When targeting *E. coli* MDS16, which harbors three IS*3* elements, we managed to detect single co-integrants at all three loci, although we only found one insertion at b0374::IS*3* out of 20 colonies tested (**Figure 6**). These data show that, by relying solely on recombineering, a single copy of relatively long (>9 kbp) operon can be inserted into *E. coli* strains that have one, two, or three IS elements within their genomes. We have no reason to think that IS elements could not be used as recombination targets if they were present in higher copy numbers on the chromosome, just the localization of the transgene insertion becomes more tedious in such cases. The absolute efficiencies of obtaining true co-integrants are shown in **Table 4.**

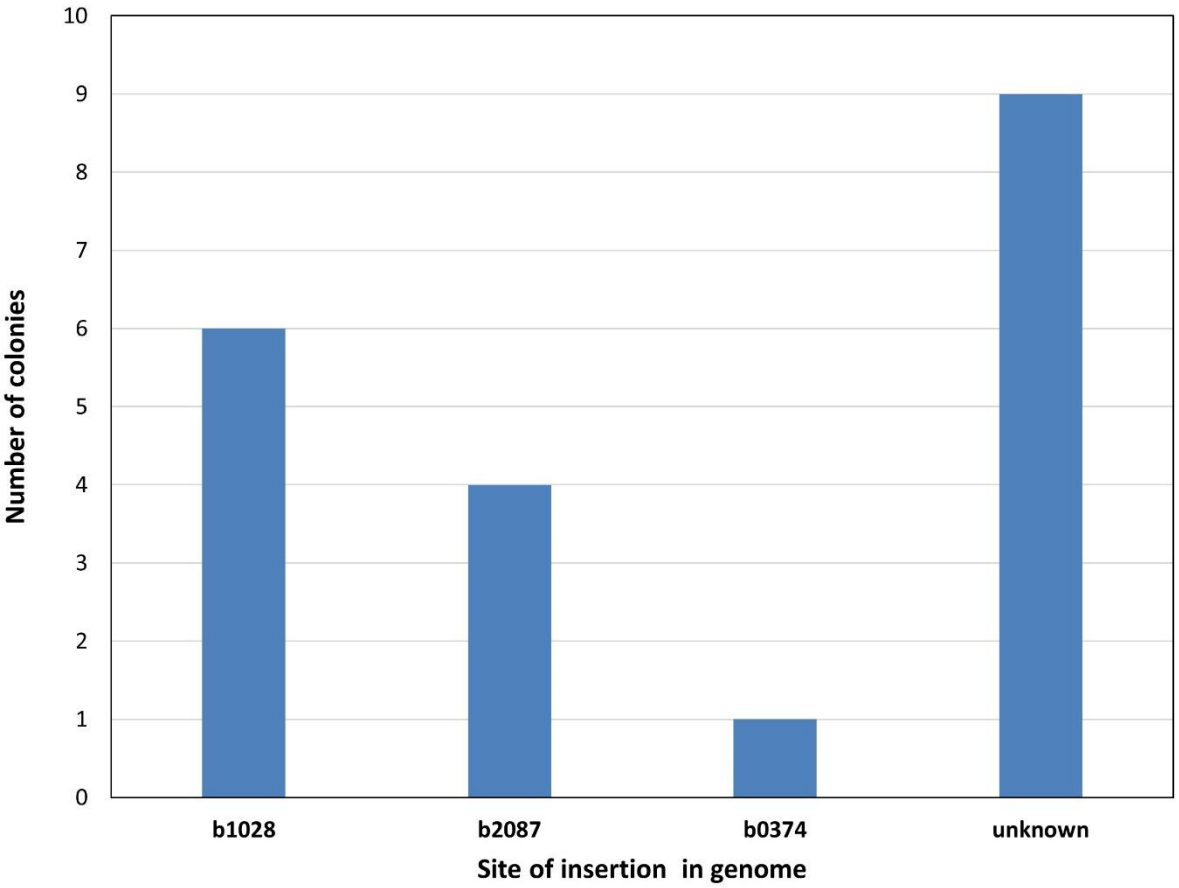


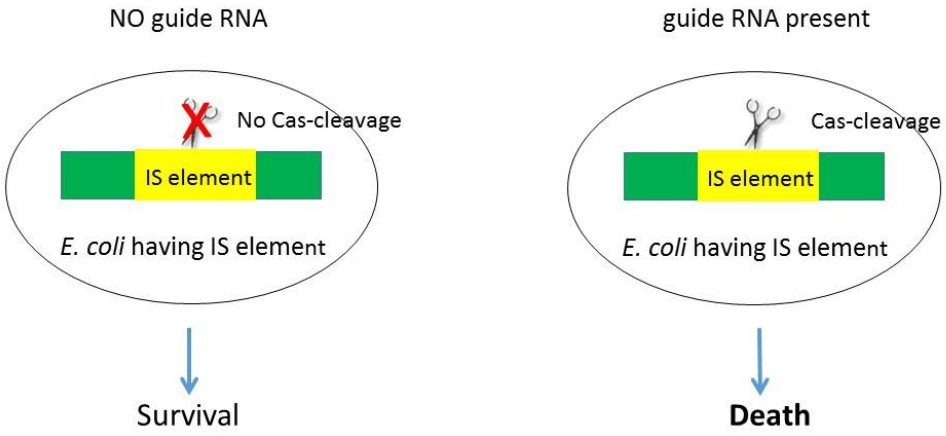
Figure 6. Results of recombineering experiments integrating the *vioABCDE*_KmR cassette into a chromosomal IS*3* of *E. coli* MDS16. The number of colonies harboring a PCR-verified integration at either of three loci (b1028, b2087, or b03741), or displaying no PCR-verified integration ("unknown") are shown. An exemplary experimental outcome is displayed from a set of two similar experiments.

4.6 Targeting ISes by CRISPR-mediated recombineering

4..6.1 Directing CRISPR/Cas9 to cleave a genomic IS element of *E. coli*

First, we tested the functionality of CRISPR/Cas9 by introducing pCas9IS*1* into a host strain (MG1655), which has multiple copies of the targeted IS element within its genome. In a control experiment, we transformed the empty pCas9 plasmid which did not encode any gRNA. We found that there were no survivors upon pCas9IS*1* transformation indicating CRISPR/Cas9 cleavage, while pCas9 could be readily transformed **(Figure 7).**

A.



B.

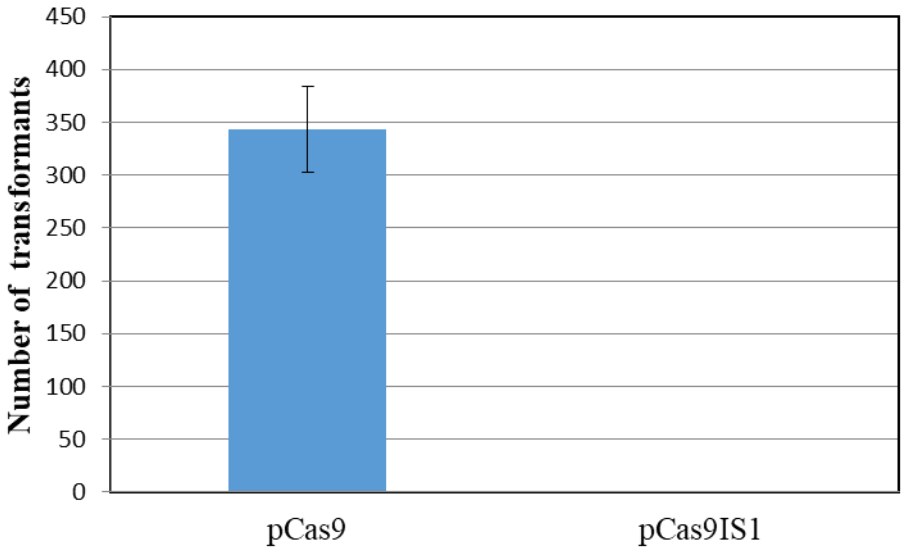


Figure 7. Testing the functionality of the CRISPR/Cas9 system targeting a genomic IS element. (A) The rationale of the experiment; successful cleavage of genomic IS elements should result in cell death, but the lack of gRNA spares the cell. (B) The colony counts observed upon pCas9 and pCas9IS*1* transformation into MG1655.

4.6.2. CRISPR/Cas-assisted recombineering permitted the genomic integration of a selective marker in two copies

Once we confirmed the efficiency of the CRISPR/Cas systemto select against bacteria containing the original targeted sequence, we applied it to enhance the efficiency of recombineering in our experimental system. We had multiple objectives such as to increase both the efficiency of recombineering and the maximum size of the insertable payload, to attempt the insertion of unselectable genes and to insert transgenes in multiple copies. Our CRISPR/Cas-mediated experiments could be divided into two strategies: in the first approach, we used CRISPR/Cas cleavage after the recombineering step as an additional tool of counterselection against the wild type (wt) genotype. In these experiments, the λ-Red recombinase enzymes and the CRISPR/Cas machinery were provided by the pORTMAGE2 and pCas9 plasmids, respectively. In the second approach we used concomitant CRISPR/Cas cleavage during the recombineering step to aid the recombination process by generating free DNA ends in addition to the selective effect mentioned previously. In this latter case, we used the NO-SCAR system (i.e. plasmids pKDsg and pCas9cr4), developed earlier by the Prather laboratory for a similar purpose (Reisch & Prather, 2015).

We started with testing the process of CRISPR/Cas-mediated recombineering of resistance genes into genomic IS elements. When targeting a single copy of IS*1* or IS*3* in *E coli* MDS42IS*1* or MDS30, respectively, subsequent CRISPR/Cas cleavage (with pCas9IS1 or pCas9IS3) was occasionally able to increase the absolute efficiency of recombination, but this effect was inconsistent, with certain cases even falling short of the recombination rates seen in simple recombineering (**Figure 8**: 2 cases improve, 2 cases worsen SpR integration into MDS42IS*1*).

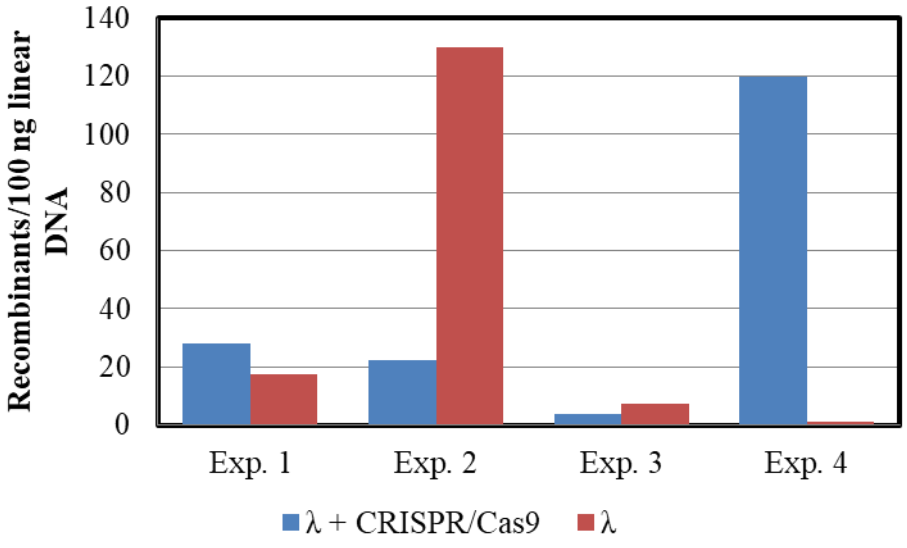


Figure 8. Testing the effect of CRISPR/Cas cleavage on the efficiency of λ-Red recombinase-mediated recombineering into the genomic IS*1* of MDS42IS*1*. The linear SpR cassette was transformed along with the pCas9IS1 plasmid into target cells harboring induced pORTMAGE2 plasmid. Red bars indicate the efficiency of recombineering alone, blue bars indicate the efficiency of CRISPR/Cas-mediated recombineering.

The advantage of subsequent CRISPR/Cas cleavage became readily apparent for targeting strains harboring two copies of the respective IS element, double co-integrants were routinely obtained using multiple protocols. In the end, we simplified the process to two transformations (described in the Methods): the linear DNA is transformed first for recombineering, and the appropriate pCas9 plasmid cleaving the wt form of the IS element is transformed the next day to enforce selection. The pCas9IS1-mediated counterselection in MDS39R2 resulted in 1/7 (14%) or 3/10 (30%) of the tested colonies harboring the resistance cassette at both loci of IS*1* (for short and long homologies, respectively). Counterselection using pCas9IS*3* was even better with 15/20 (75%) of the colonies proving to harbor double co-integrants in BLK16. Double IS*3* co-integrants were also successfully obtained in MDS27, displaying the portable nature of this technique. **Figure 9** shows the fraction of colonies harboring double co-integrants after one or two rounds of recombineering, using either short or long homologies in MDS39R2.

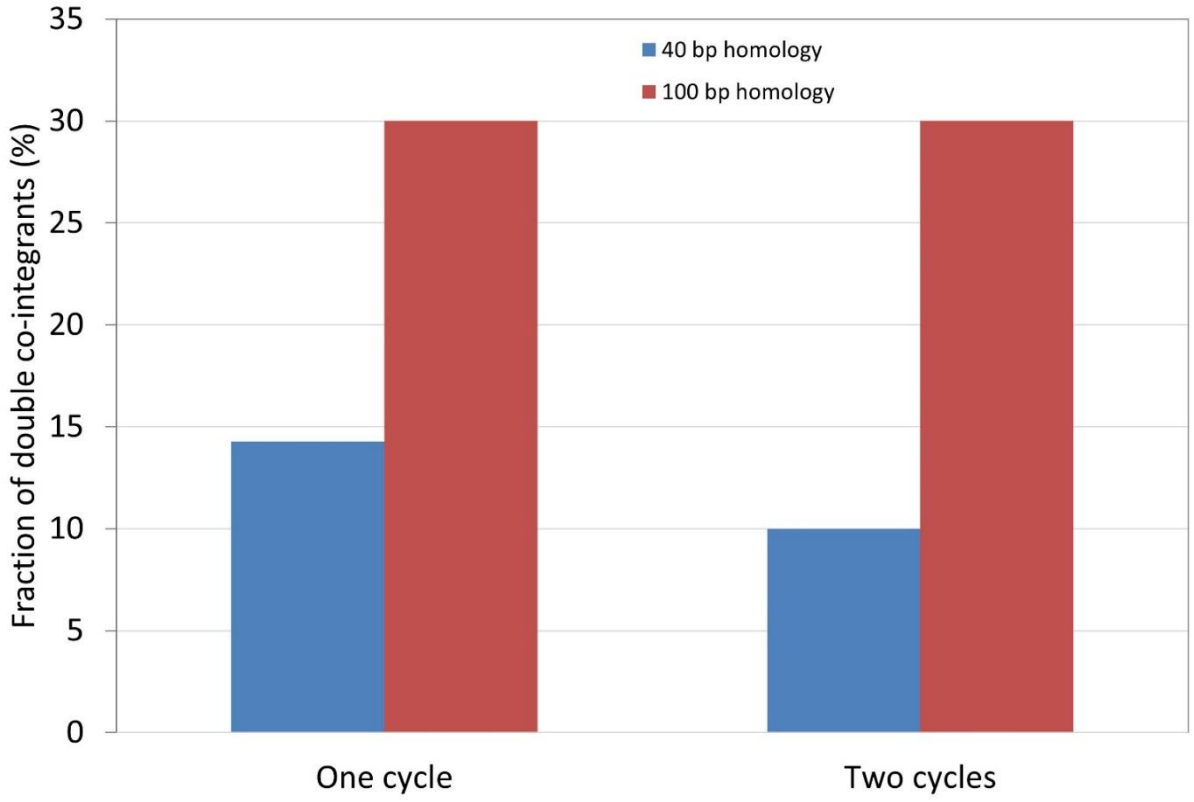


Figure 9. The efficiency of engineering double co-integrants. The IS*1* elements of *E. coli* MDS39R2 were targeted by recombineering followed by consecutive pCas9IS1-mediated counterselection. The fraction of colonies harboring an SpR cassette within both chromosomal IS*1* elements are shown after one or two cycles of recombineering and one round of counterselection. Blue bars depict the use of short (40 bp), red bars depict the use of long (100 bp) homologies. An exemplary experimental outcome is displayed from a set of two similar experiments.

4.6.3. CRISPR/Cas-assisted recombineering allowed the genomic integration of an unselectable gene in a single copy

Besides generating double co-integrants, the great value of applying CRISPR/Cas-based facilitation was unveiled upon the integration of unselectable genes into the chromosome. Peculiarly, we managed to insert the *gfp* gene into the IS*3* of *E. coli* MDS30 using recombineering and subsequent CRISPR/Cas selection. In the course of these experiments, we demonstrated that a 20 nt spacer is equal or superior to a 30 nt spacer in pCas9IS*3* concerning efficiency, and up to 20% (2/10) of colonies obtained proved to be positive by PCR. The integration of *gfp* into the IS*1* element of MDS42IS*1* was successfully achieved using recombineering and concomitant CRISPR/Cas cleavage provided by the NO-SCAR system (plasmids pKDsg-IS1 and pCas9cr4). In the latter case, 3 out of 30 colonies (10%) were PCR-positive. The elevated green fluorescence levels of both engineered strains were verified using a microplate reader (**Figure 10**).

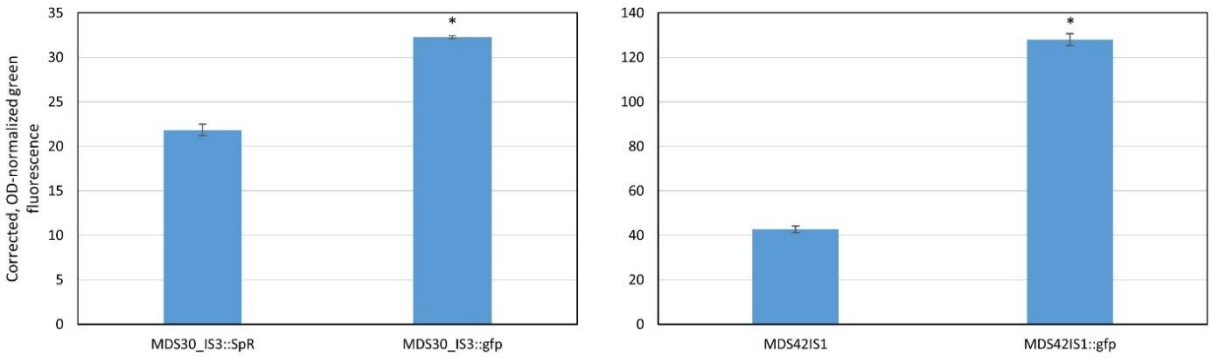


Figure 10. Effect of chromosomal *gfp* integration on the fluorescence of the host strain. The fluorescence of fully grown cultures are shown, recorded in a Synergy2 microplate reader

(Excitation: 500 nm; Emission: 540 nm). Fluorescence was corrected by subtracting the blanks and normalized to OD600. The means of three technical replicates are shown. * $p<2\times10^{-5}$ with a two-tailed, unpaired t-test

4.7. Genomic integration of a five-gene operon in single or double copies using CRISPR/Cas-assisted recombineering

Next, we tested the effect of CRISPR/Cas cleavage on the integration efficiency of long (>9 kbp) selectable DNA cassettes. Since integrating our *vio* operon into IS*1* targets by recombineering led to the loss of violacein production (see above), we focused on experiments targeting IS*3*. First, we tested the effect of subsequent Cas cleavage using pCas9IS3. We found that in most of the experiments targeting MDS27, BLK09, or BLK16 (using either SpR or KmR as selection markers), subsequent Cas cleavage not only failed to improve the efficiency of pORTMAGE-mediated recombineering but on many occasions the recombinant cells were missing altogether. In a few cases nevertheless, we did manage to obtain single *vioABCDE*_KmR insertions into either of the IS*3* elements of BLK09 (**Figure 11 and 12**) or BLK16 but double co-integrants were not obtained this way. Second, we tested the effect of concomitant Cas cleavage using the NO-SCAR system. We started by inserting the *vioABCDE*_KmR cassette into the single IS*3* of MDS30 using NO-SCAR (plasmids pKDsg-IS3 and pCas9cr4). We obtained single co-integrants with low absolute efficiency (on the order of 10^{-2} correct colonies/ng), the high ratio of correct colonies (6/10) nevertheless permitted the easy detection of true recombinants. The same system was used to target the two copies of IS*3* residing in MDS27 as well as two copies of IS*1* residing in MDS39R2. We observed 7.7% of the colonies in MDS27 and 5% of the colonies in MDS39R2 to harbor the operon at both loci (**Figure 13**). The absolute efficiencies of obtaining recombinants using CRISPR/Cas-assisted recombineering are listed in **Table 4**.

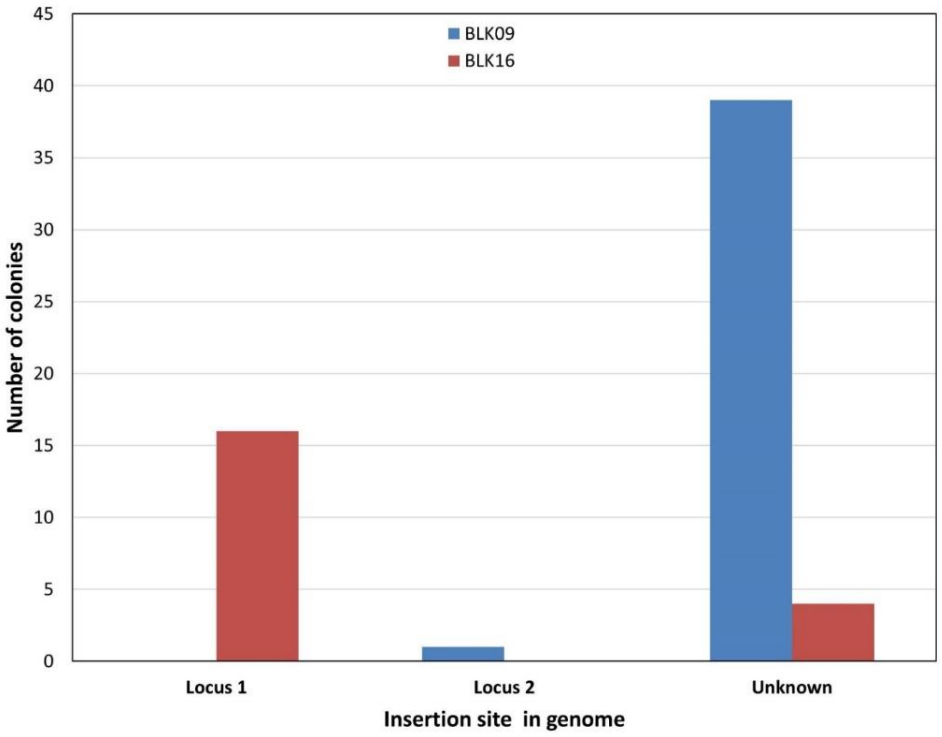


Figure 11. Integration of the *vioABCDE*_KmR cassette into chromosomal IS*3* elements using recombineering followed by pCas9IS3-mediated counterselection. The number of colonies harboring a PCR-verified integration at locus 1, locus 2, or displaying no PCR-verified integration ("unknown") are shown for *E. coli* BLK09 (blue) and *E. coli* BLK16 (red). The yields of recombinants were 4 and 30 recombinants/ng DNA for BLK09 and BLK16, respectively. An exemplary experimental outcome is displayed from a set of four similar experiments (two for each strain).

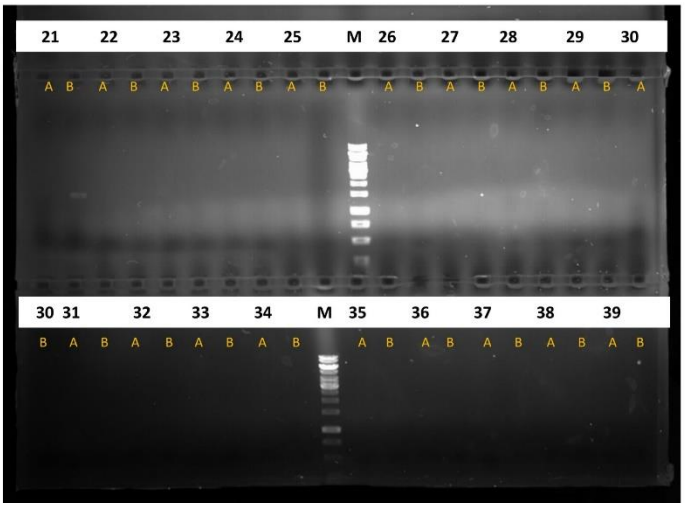


Figure 12. Targeting BLK09/pORTMAGE2 cells with IS*3*::*vioABCDE*_KmR cassettes, followed by pCas9IS3 transformation. Nineteen colonies obtained upon the CRISPR/Cas-

mediated recombineering process are PCR-screened using two primer sets: A: IS3flanking1 + VioAR2, B: IS3flanking2 + VioAR2. One positive colony is detected. M: GeneRuler 1 kb DNA Ladder (Thermo Scientific)

Table 6. Yields of recombinants obtained by CRISPR/Cas9-assisted recombineering of SpR, KmR, *gfp*, and *vioABCDE*:KmR cassettes into strains harboring single and double copies of targeted IS element

Strain	IS element type	No. of target IS elements in genome	System	Linear DNA type	ng DNA used	Total colonies per ml	PCR-verified single positive colonies	PCR-verified double positive colonies	% of positive colonies	Total correct colonies	Absolute yield of integration (recombinants /ng DNA)
MDS42.IS1	IS1	1	NO-SCAR-IS1	IS1:KmR	100	87	10 out of 10	NA	100	87	0.87
MDS42.IS1	IS1	1	NO-SCAR-IS1	IS1:KmR	100	97	10 out of 10	NA	100	97	0.97
MDS30	IS3	1	NO-SCAR-IS3	IS3:KmR	200	286	10 out of 10	NA	100	286	1.43
MDS30	IS3	1	NO-SCAR-IS3	IS3:KmR	200	1114	4 out of 10	NA	40	445.6	2.23
MDS42.IS1	IS1	1	pORTMAGE-2/pCas9-IS1	IS1:SpR	100	5360	10 out of 10	NA	100	5360	53.6
MDS42.IS1	IS1	1	pORTMAGE-2/pCas9-IS1	IS1:SpR	100	9160	10 out of 10	NA	100	9160	91.6
MDS30	IS3	1	pORTMAGE-2/pCas9-IS3	IS3:SpR	100	800	10 out of 10	NA	100	800	8
MDS30	IS3	1	pORTMAGE-2/pCas9-IS3	IS3:SpR	100	730	10 out of 10	NA	100	730	7.3
MDS42.IS1	IS1	1	NO-SCAR-IS1	IS1:GFP	100	636000	3 out of 30	NA	10	63600	636
MDS42.IS1	IS1	1	NO-SCAR-IS1	IS1:GFP	100	426000	2 out of 65	NA	3.1	13107.7	131.1
MDS42.IS1	IS1	1	NO-SCAR-IS1	IS1:GFP	100	410000	3 out 0f 50	NA	6	24600	246
MDS30	IS3	1	pORTMAGE-2/pCas9-IS3	IS3:GFP	100	5440	1 out of 20	NA	5	272	2.72
MDS30	IS3	1	pORTMAGE-2/pCas9-IS3	IS3:GFP	100	6820	2 out of 10	NA	20	1364	13.6
MDS30	IS3	1	pORTMAGE-2/pCas9-IS3	IS3:GFP	100	2760	3 out of 40	NA	7.5	207	2.07
MDS30	IS3	1	NO-SCAR-IS3	IS3:VioABCDE-KmR	600	62	6 out of 10	NA	60	37.2	0.06
MDS42.IS1	IS3	1	NO-SCAR-IS3	IS3:VioABCDE-KmR	750	2520	15 out of 20	NA	75	1890	2.52
MDS39R	IS1	2	pORTMAGE-2/pCas9-IS1	IS1:SpR	100	1780	NA	2 out of 10	20	356	3.56
MDS39R	IS1	2	pORTMAGE-2/pCas9-IS1	IS1:SpR	100	340	NA	3 out of 10	30	102	1.02
MDS27	IS3	2	pORTMAGE-2/pCas9-IS3	IS3:SpR	100	210	NA	6 out of 10	60	126	1.26
BLK16	IS3	2	pORTMAGE-2/pCas9-IS3	IS3:SpR	100	1330	NA	15 out of 20	75	997.5	9.98
MDS27	IS3	2	NO-SCAR-IS3	IS3:VioABCDE-KmR	750	400	NA	1 out of 13	7.6	30.4	0.04
MDS39R	IS3	2	NO-SCAR-IS3	IS3:VioABCDE-KmR	750	764	NA	1 out of 20	5.0	38.2	0.05

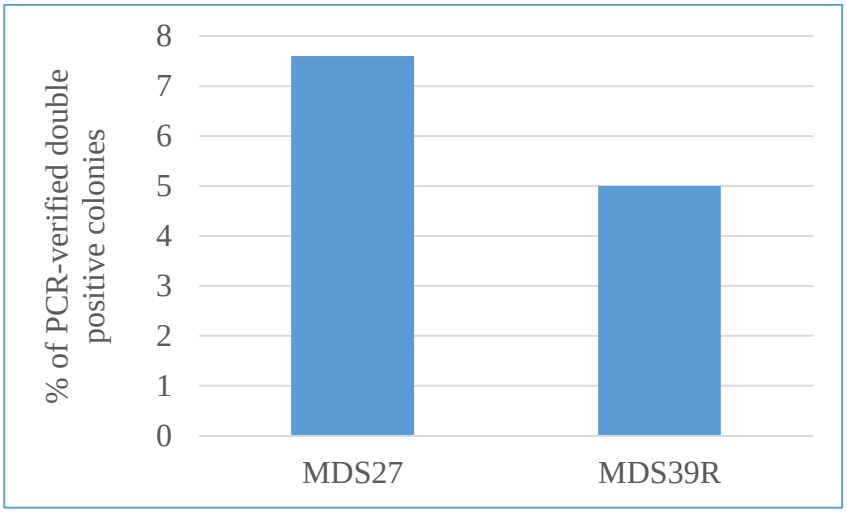


Figure 13. Integration of *vioABCDE*_KmR cassettes into both copies of chromosomal IS*3* or IS*1* elements of *E. coli* MDS27 and MDS39R2, respectively using recombineering and concomitant CRISPR/Cas9 cleavage provided by the NO-SCAR system. The ratios of colonies that harbor the operon at both targeted loci are shown.

4.8. Copy number amplification of cargo genes using copy-paste transposition of IS elements

4.8.1. The majority of the detected copy-amplification derives from the transposition of the resistance marker into the transposase-expressing plasmid

After verifying that genes and large operons can be integrated into IS elements, our next goal was to achieve copy-number amplification of these co-integrants *via* copy/paste transposition. Our initial experiments dealt with the amplification of single resistance genes (*e.g.* SpR) inserted into IS*1* or IS*3*. The respective transposase genes were expressed from a plasmid in trans, and host cells carrying copy-number amplified resistance genes were selected on antibiotic-gradient plates. Colonies picked from the high-end of the gradient plates went through genomic DNA preparation and ddPCR to quantify the copy number of the resistance gene. These pilot experiments revealed that both IS*1* and IS*3* could be used to yield highly-resistant colonies upon transposase induction (**Figures 14, 15**).

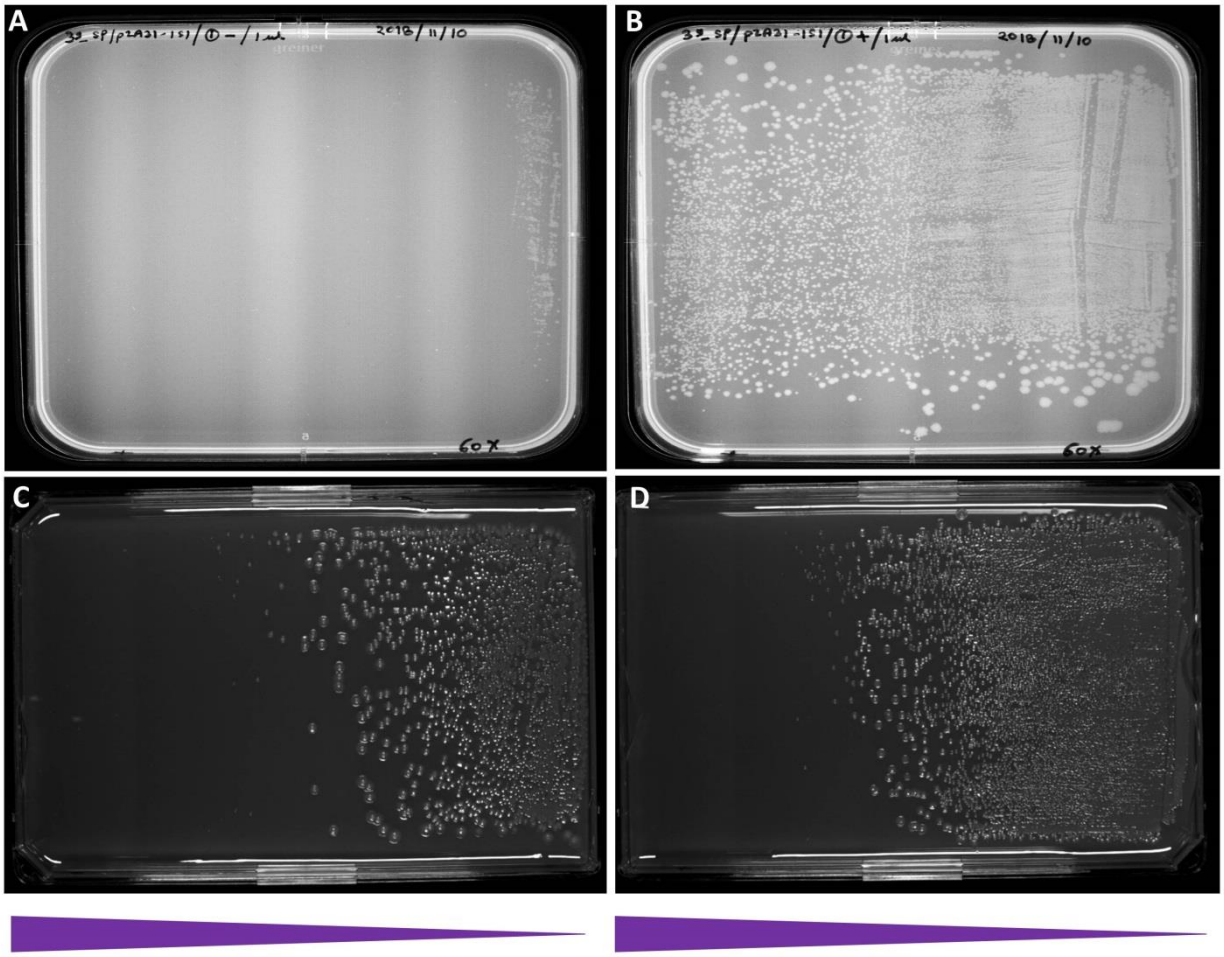


Figure 14. The effect of IS*1* transposase induction on Sp-resistance. (A) Uninduced MDS39R2IS*1*::SpR cells, (B) MDS39R2IS*1*::SpR cells induced with aTc. (C) Uninduced MDS42IS*1*::*vioABCDE*_KmR cells, (D) MDS42IS*1*::*vioABCDE*_KmR cells induced with aTc. The triangles represent the gradient of Sp concentration (A, B) or Km concentration (C, D) within the medium.

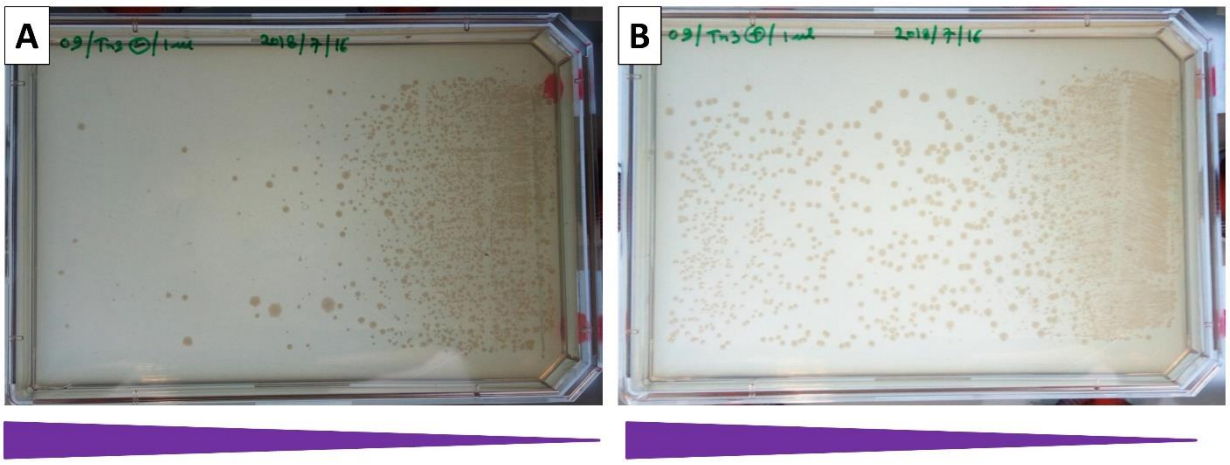


Figure 15. The effect of IS*3* transposase induction on the Sp-resistance of BLK09IS*3*::SpR colonies. (A) Uninduced cells, (B) cells induced with aTc. The triangles represent the gradient of Sp concentration within the medium.

IS*1* experiments, however, tended to be less reproducible, with unknown factors influencing the observable increase in resistance (**Figure 14C, D**). Copy number estimation by ddPCR in the initial amplification experiments indicated a major increase in copy numbers of the resistance gene, rising to 20 copies/cell in the first round (**Figure 16, left**).

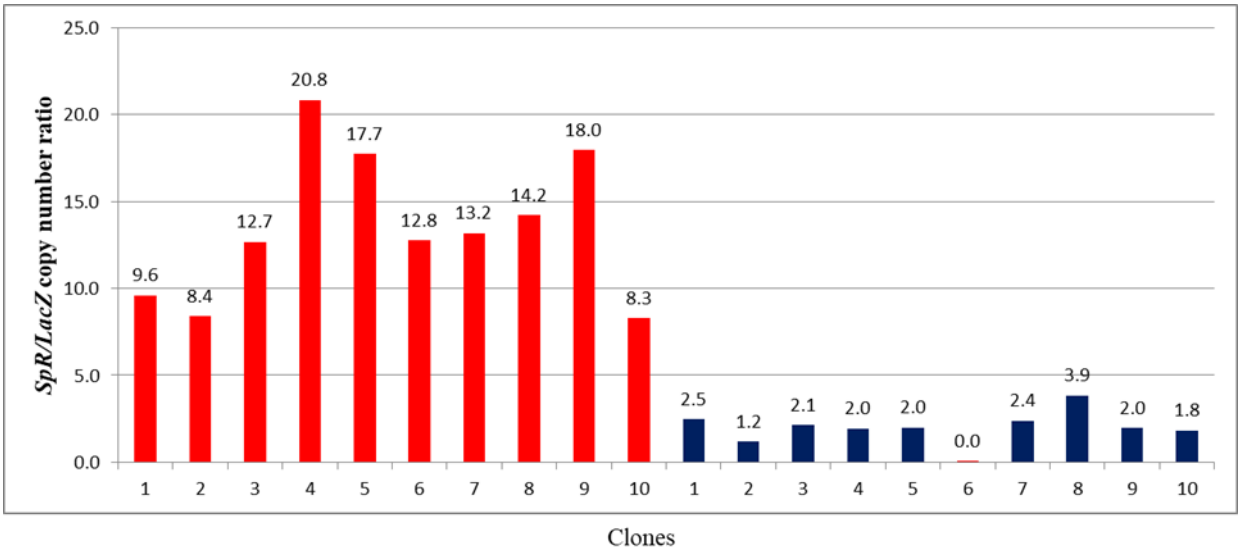


Figure 16. Quantification of the *SpR* gene in the MDS39R2IS*1*::SpR genome by ddPCR after one round of IS*1* transposase induction. Red bars indicate clones analyzed before and blue bars after plasmid curing.

To confirm 20 copies/cell, we isolated the plasmid from multi SpR copy strains, digested it with BamHI as well as tested inverse PCR where outward primers were binding to the SpR cassette. Restriction digestion pattern and inverse PCR followed by sequencing confirmed that SpR was transposed from the bacterial genome onto the transposase expressing plasmid (**Figure 17**).

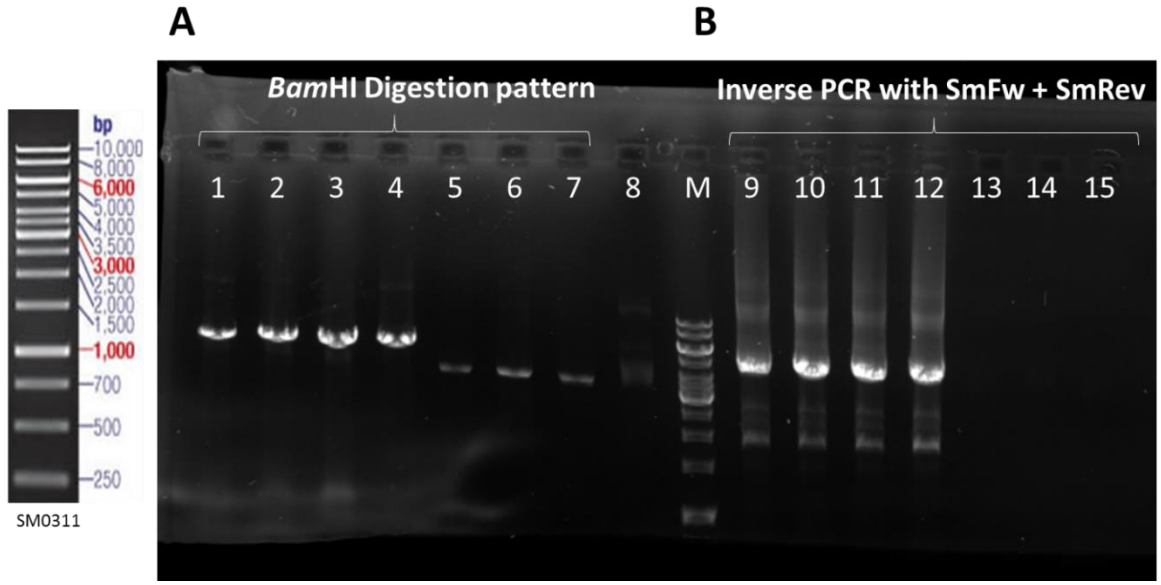


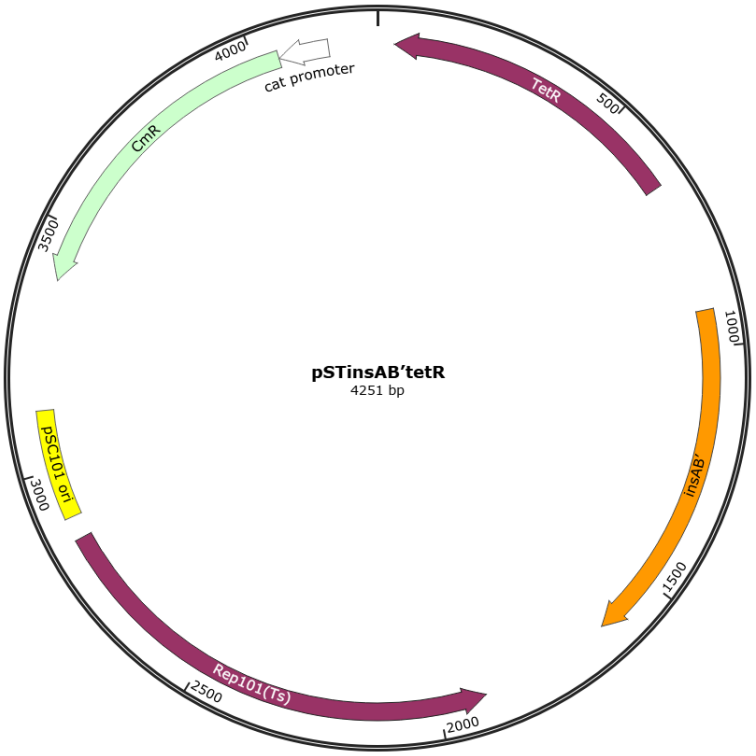
Figure 17. Verification of transposition of spectinomycin resistance gene (SpR) from genome to plasmid by restriction digestion (A) and inverse PCR (B). Lane 1-4 shows transposition positive clones (clone-1-4), lane 5,6 (clone-5,6) shows transposition negative clones, lane -7 shows BamHI digested original Tnp1 expressing plasmid, lane - 8 shows undigested original plasmid. Lane 9-12 shows inverse PCR on transposition positive clones (clone-1-4). Lane 13-14 shows inverse PCR on transposition negative clones (clone-5,6), Lane 15 shows inverse PCR on original Tnp1 expressing plasmid (negative control).

We repeated the ddPCR on the same cells after curing them from the transposase-expressing plasmids and dramatically, the detectable copy numbers of SpR fell to the range of 2-4 (**Figure 16, right**). This indicated that most of the high resistance was caused by the transposition of the resistance cassette into the transposase-expressing plasmid.

4.9. True multi-copy recombinant strains achieved with the thermo-sensitive version of transposase-expressing plasmids

To allow the easy elimination of high-resistance cells resulting from transposition of the marked IS into plasmids, we engineered temperature-sensitive transposase plasmids (pSTinsAB'tetR and pSTtnp3tetR for IS*1* and IS*3*, respectively) **(Figure 18A, B).**

A.



B.

Figure 18 A, B. Maps of thermo-sensitive versions of transposase expressing plasmids having pSC101 ori. (A) pSTinsAB'tetR for IS*1* and (B) pSTtnp3tetR for IS*3* mobilization.

We also modified the copy-amplification protocol so that after transposase expression, the cells are plated on gradient plates only after the plasmids had been cured from the cells (as described in the Methods section). This guaranteed that only chromosomal co-integrants of the resistance gene are selected after transposase expression.

Due to the low reproducibility of IS*1*-mediated amplification and to the fact that violacein expression was not observable upon IS*1*-targeted integration (described in the previous section), we focused primarily on amplification of the violacein operon (*vioABCDE*) using IS*3* elements. Strains of *E. coli* BLK16_IS*3*::*vioABCDE*_SpR, generated by recombineering (see above) were used as starting points. These carried a single copy of the violacein operon inside the IS*3* element residing at locus 1, along with an SpR marker. After the first round of transposase expression, we analyzed five colonies by ddPCR (targeting the resistance marker) from the high-end of the gradient plates and found the mean copy number to be 5.50 (±1.22) (**Figure 19**). Choosing a colony displaying 7.59 copies and re-expressing the IS*3* transposase led to a further significant elevation of the mean copy number within colonies displaying high resistance levels (8.03±1.21, n=4) (**Figure 19**). Again, we chose the colony displaying the highest copy number (9.10) and expressed the IS*3* transposase for the third time. The mean copy number of the colonies analyzed in the third round did not significantly differ from those of the second round (**Figure 19**). However, by 'cherry picking', we were still able to find individual clones displaying further elevated copy numbers (see below).

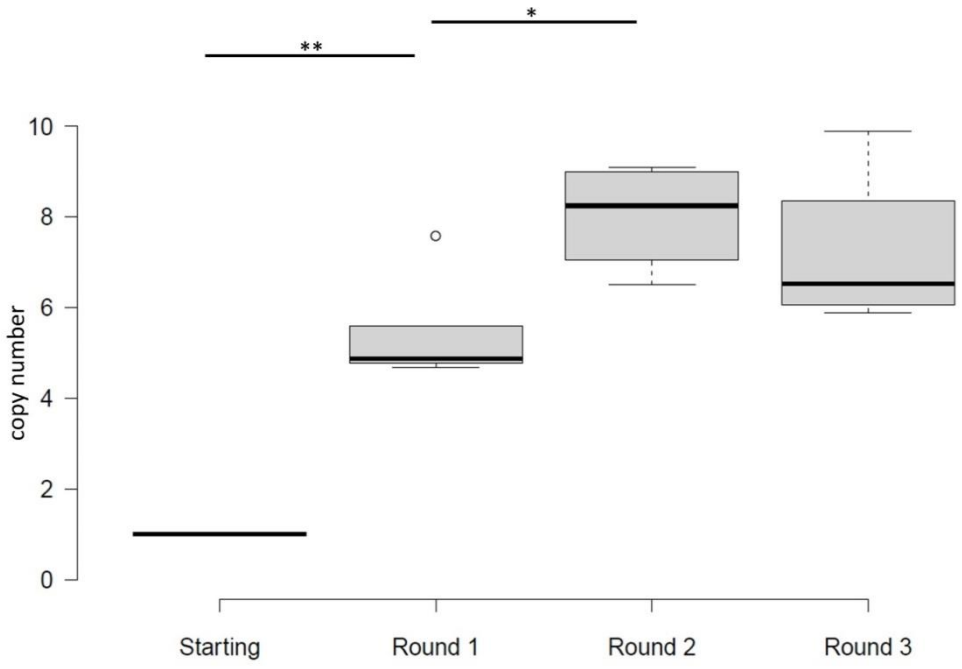


Figure 19. Copy numbers of the SpR gene of *E. coli* BLK16_IS*3*::*vioABCDE*_SpR detected by ddPCR after 1, 2, or 3 rounds of IS*3* transposase induction. Centerlines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software (R Core Team, 2020); whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots. n = 1, 5, 4, 4 sample points. * p<.05 with two-tailed, unpaired t-test; ** p<.002 with two-tailed, one-sample t-test.

4.10. Whole-genome sequencing confirmed the multiple copies of the *vioABCDE* operon in the genome of engineered strains

Next, we sent four strains obtained from the copy-amplification experiment described above (the starting strain and one clone from each round of transposase induction) for whole-genome sequencing. Our aims were i) to verify the copy numbers of the amplified IS*3*::*vioABCDE* _SpR elements, ii) to analyze the fraction of mutant or truncated *vioABCDE* operons, iii) to test whether amplification leads to tandem repeats or a random scatter of the loaded IS*3* element within the genome and iv) to check for unexpected genomic rearrangements. The sequence of the starting strain (B0) confirmed the presence of the *vioABCDE* operon, along with the SpR marker within the IRs of the IS*3* element residing at locus 1 (genomic coordinates 649,839-659,074). The sequence of strain B1, obtained in the first round of transposase induction displayed a 5.71-fold elevated relative coverage of the loaded IS*3*, and unexpectedly, of a 14 kbp segment directly downstream of the IS (amplified region: 649,839-673,907). This stands in good agreement with the 5.59 copy numbers measured by ddPCR. Sequencing of strain B2, which is not a direct descendant of B1, revealed an 11.32-fold increased relative coverage of a similarly large genomic segment delimited on the left side by the left IR of the loaded IS*3* (coordinates: 649839-677086). Interestingly, it also displayed a further 5.44 copies (totaling 16.76) of an interior segment (659074-667565), which is directly to the right of the manipulated IS. The approx. 11-fold increase in the copy number of the violacein operon roughly supports the 9.1 copies indicated by the ddPCR. Finally, strain B3, a descendant of B2, was expected to harbor 26.11 copies of the resistance marker, based on ddPCR. Sequencing did not confirm this, the relative coverage of the genomic region encompassing the loaded IS*3* (649839-677086) nevertheless increased from 11.32 to 12.51, and the number of additional copies of the interior segment on the right side of the IS (659074-667565) decreased from 5.44 to 3.54. We, therefore, concluded that although the third round of induction did not increase the mean copy number of the inserted operon, screening a low number of colonies (<10) permitted the identification of a

clone that displayed further amplification of the transgenes. We note that in another clone of the same series of induced strains we detected 9.89 copies, and in another induction series done in parallel, we measured 10.71 and 11.83 copies of SpR using ddPCR. We also observed that in a small fraction of analyses (1-5%), the ddPCR results were misleading (data not shown), underlining the importance of result verification by sequencing. The sequenced strains did not harbor the loaded IS*3* scattered throughout the genome, but rather amplified in tandem repeats. However, the IS copies did not form a back-to-back array since, rather unexpectedly, they were amplified together with a large (14-18 kbp) DNA segment lying directly downstream to them.

4.11. Violacein production was proportional to the copy number of the operon encoding its production

Violacein content of the starting strain and its derivatives obtained after various rounds of transposase induction were routinely measured. The amount of violacein released from the BLK16IS*3*::*vioABCDE*_SpR derivatives displayed a good correlation (R^2=0.98) with the copy number of the *vioABCDE* operon inferred from its relative sequencing coverage (**Figure 20**).

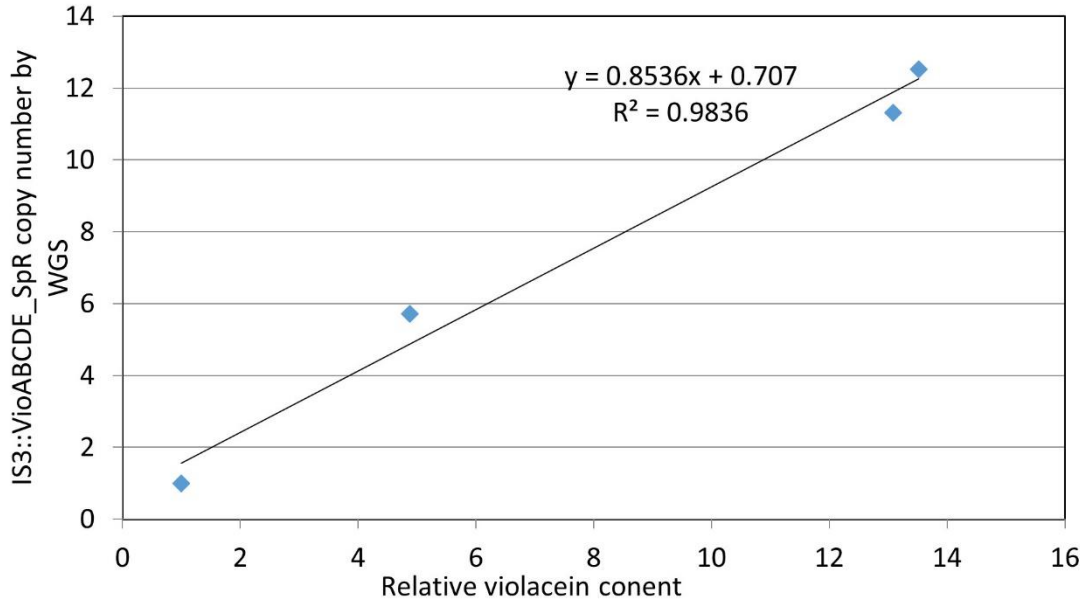


Figure 20. The correlation between the relative violacein content of various BLK16_IS*3*::*vioABCDE*_SpR strains obtained after induction of IS*3* transposase, and their

corresponding copy number of the genomic IS*3*::*vioABCDE*_SpR cassette detected by whole-genome sequencing (WGS).

Copy numbers of the SpR marker determined by ddPCR, the relative violacein content, and the relative sequence coverage of the violacein operon acquired by whole genome sequencing are displayed in **Table 7** for each sequenced strain of the IS*3*::*vioABCDE*_SpR-amplification process done in BLK16.

Table7. Properties of *E. coli* BLK16IS*3*::*vioABCDE*_SpR derivatives after the indicated rounds of IS*3* transposase induction.

Strain code	**Strain name**	**SpR copy number by ddPCR**	**Violacein content (AU)**	**Relative violacein content[1]**	**Relative sequence coverage of *vioABCDE* by WGS[2]**
B0	BLK16IS3::*vioABCDE*_SpR, starting	0.98	0.06	1	1
B1	BLK16IS3::*vioABCDE*_SpR, 1st round	5.59	0.2926	4.88	5.71
B2	BLK16IS3::*vioABCDE*_SpR, 2nd round	9.10	0.7847	13.08	11.32
B3	BLK16IS3::*vioABCDE*_SpR, 3rd round	26.11	0.8113	13.52	12.51

[1] Relative to strain B0, [2] Ratio of sequence reads obtained +/- 10 nt of peak border.

To verify the general applicability of our strategy, which we call inPOSE (integration followed by transposition), we tested the amplification using IS*3* elements residing elsewhere in the same

strain, employing further strains of *E. coli* and also using a different selection marker. Inducing transposition of IS*3*::*vioABCDE*_SpR inserted at locus 2 in strain BLK16 for one round also yielded purple colonies with significantly elevated copy numbers of the operon (3.28±0.15) (**Figure 21**).

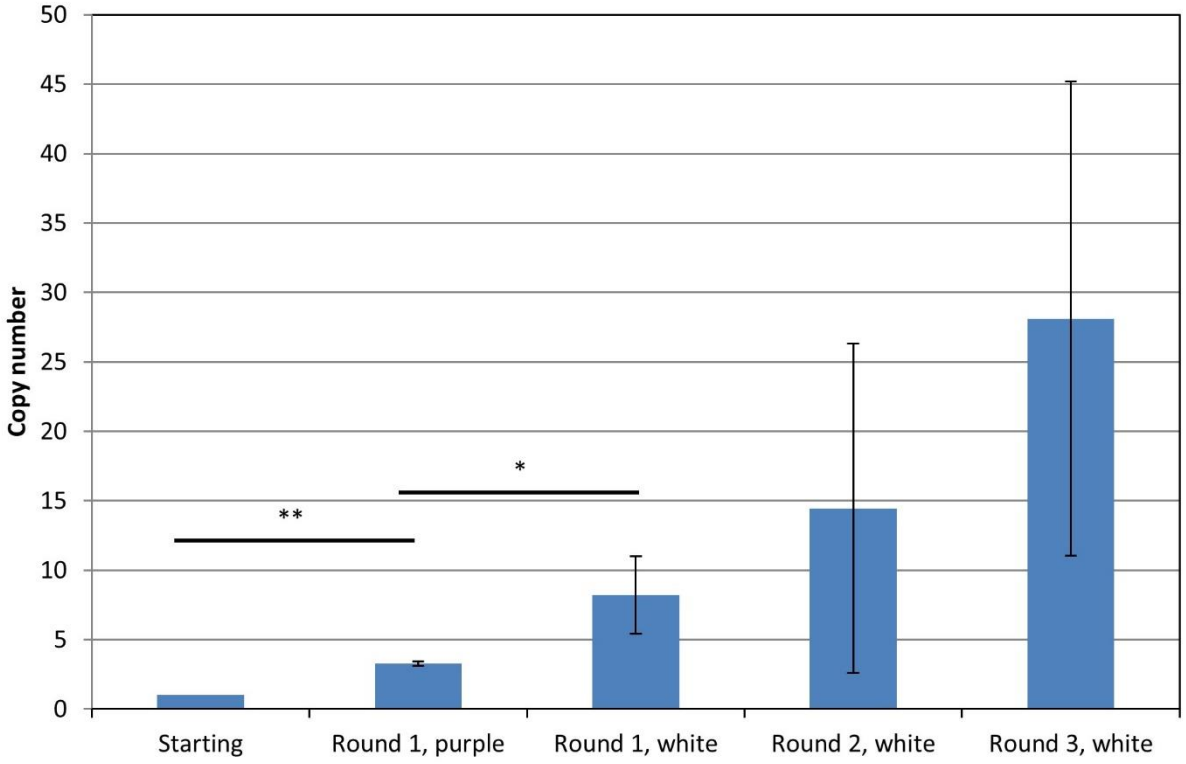


Figure 21. Copy numbers of the SpR gene of *E. coli* BLK16_IS*3*::*vioABCDE*_SpR (at locus 2) detected by ddPCR after 1, 2, or 3 rounds of IS*3* transposase induction. Note that after round 1 of transposase induction, the purple and the white colonies are displayed in separate bars. n = 1, 3, 3, 4,6 sample points. * $p<.05$ with two-tailed, unpaired t-test; ** $p<.002$ with two-tailed, one-sample t-test.

Importantly, when analyzing white colonies by ddPCR, the mean copy number was even higher (8.20±2.79). In this experiment, we induced a white colony for two further rounds and identified white colonies harboring ever higher copy numbers of the SpR marker (31.70 and 49.14 in round 2 and 3, respectively), the increase of the means, however, were not significant (**Figure 21 and Table 8**). One round of transposase induction also worked in BLK09IS3::*vioABCDE*_SpR (locus 2) (**Figure 22 and Table 8**), the obtained mean copy

number was 3.88 (±1.68). Interestingly, using KmR in the same locus, we only found a maximum copy number of 2.27, with the mean not being significantly elevated (1.51±0.52). (**Figure 23 and Table 8**).

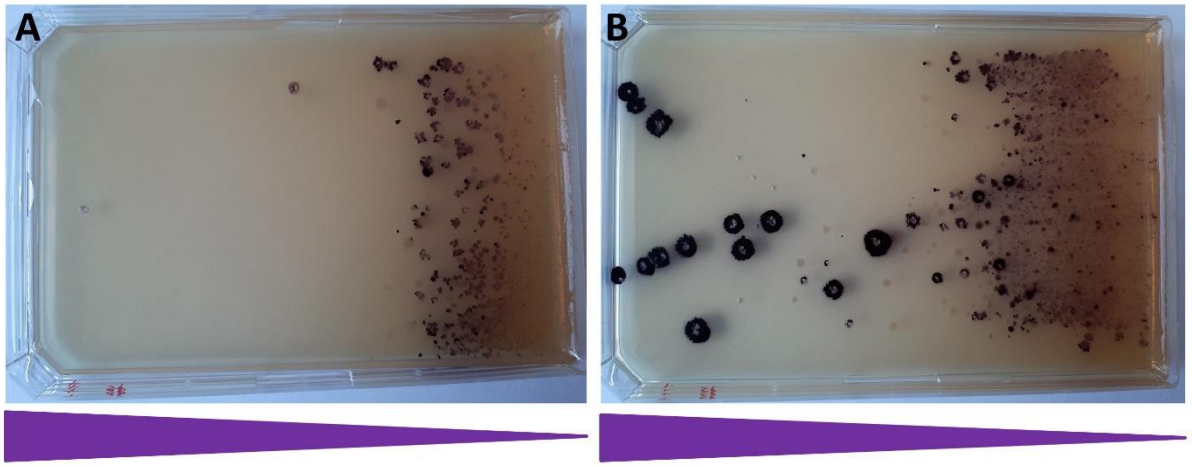


Figure 22. The effect of one round of IS*3* transposase induction on the Sp-resistance of BLK09IS*3*::*vioABCDE*_SpR (locus 2) colonies. Note the intense purple colors of the colonies caused by violacein. (A) Uninduced cells, (B) cells treated with aTc to induce transposase expression. The triangles represent the gradient of Sp concentration within the medium.

To test a completely different genetic background, amplification was also tested in strains MDS27 and MDS30, derivatives of *E. coli* K-12. For *E. coli* MDS30IS*3*::*vioABCDE*_KmR, one round of transposase induction boosted the number of highly resistant colonies compared to the uninduced (**Figure 23**).

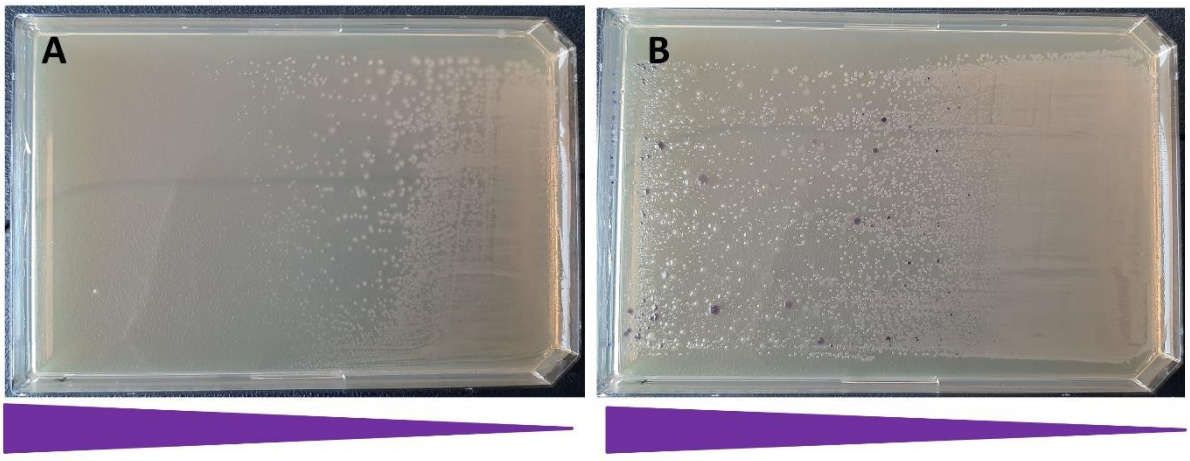


Figure 23. The effect of one round of IS*3* transposase induction on the Km-resistance of MDS30_IS*3*::*vioABCDE*_KmR colonies. (A) Uninduced cells, (B) cells induced with aTc.

Note that the majority of colonies are white. The triangles represent the gradient of Km concentration within the medium.

The mean copy number of the KmR marker was also significantly elevated (7.57±2.39), however, only a small fraction of the colonies growing on the plate were purple. For *E. coli* MDS27doubleIS3::*vioABCDE*_KmR, a strain that had both copies of IS*3* loaded with *vioABCDE*_KmR at the starting point, we carried out three rounds of induction. Similarly to that seen for MDS30IS*3*::*vioABCDE*_KmR, the first round of induction elevated the copies of KmR marker to 4.78 (±0.76), but most of the colonies were white. The second and third rounds of induction, both of which were carried out choosing a purple colony, could not significantly increase the mean copy number of the marker (5.7±2.66 and 4.0±0.64, respectively). Again by cherry-picking, we could find colonies carrying up to 7.8 copies. Importantly, the second and third rounds of this experiment revealed that in the lack of induction, the majority of the colonies remain purple (**Figure 24** and **25**), but the obtainable mean copy numbers are somewhat limited (4.07±1.96 and 2.95±1.16, respectively).

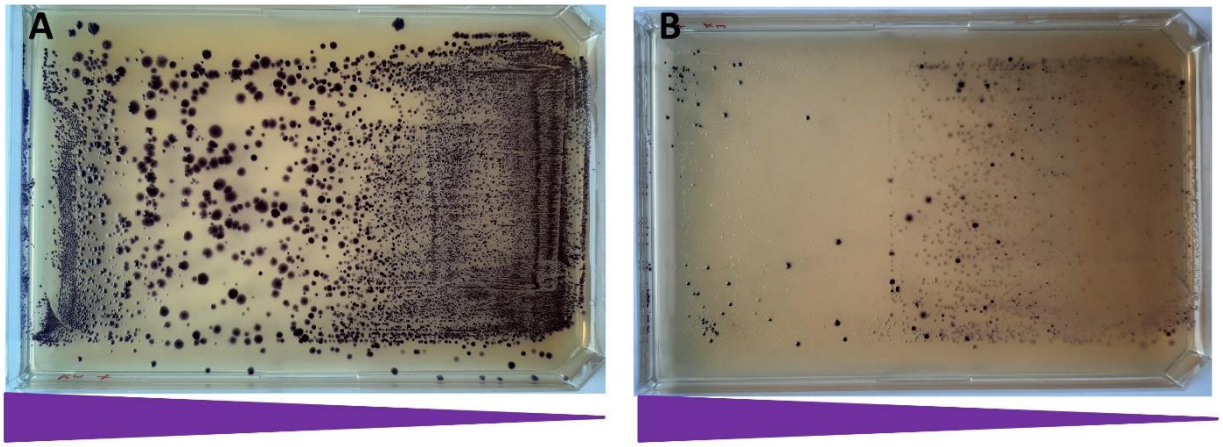


Figure 24. The effect of the second round of IS*3* transposase induction on the Km-resistance of MDS27_doubleIS*3*::*vioABCDE*_KmR colonies (starting copy number: 5.8 copies). (A) Uninduced cells, (B) cells induced with aTc. Note that the majority of induced colonies are white. The triangles represent the gradient of Km concentration within the medium.

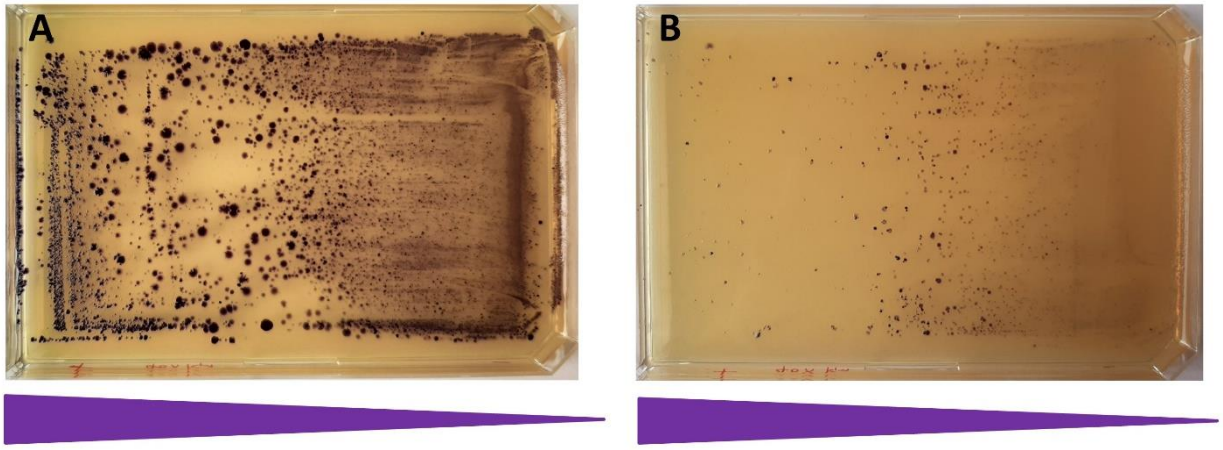


Figure 25. The effect of the third round of IS*3* transposase induction on the Km-resistance of MDS27_doubleIS*3*::*vioABCDE*_KmR colonies (starting copy number: 5.7 copies). (A) Uninduced cells, (B) cells induced with aTc. Note that the majority of induced colonies are white. The triangles represent the gradient of Km concentration within the medium.

Table 8. Increase in the copy number of *vioABCDE* in the genome of respective strains after the first, second and third round of IS*3* transposase induction.

Strain	Round 1	Round 2	Round 3
BLK16_IS*3*::*vioABCDE*_SpR, locus 1	5.50 ± 1.22	8.03 ± 1.21	9.1
BLK16_IS*3*::*vioABCDE*_SpR, locus 2, PURPLE	3.28 ± 0.15	–	–
BLK16_IS*3*::*vioABCDE*_SpR, locus 2, WHITE	8.20 ± 2.79	31.7	49.14
BLK09_IS*3*::*vioABCDE*_SpR, locus 2	3.88 ± 1.68	–	–
BLK09_IS*3*::*vioABCDE*_KmR, locus 2	1.51 ± 0.52	–	–
MDS27_doubleIS3::vioABCDE_KmR	4.78 ± 0.76	5.7 ± 2.66	4.0 ± 0.64
MDS30_IS3::vioABCDE_KmR, INDUCED	7.57 ± 2.39	–	–

These results indicate that transposase expression, together with the genetic background of the MDS strains apparently elevates the inactivation rate of the *vio* operon, without necessarily inactivating the resistance gene. The reason for the strain-dependence of this cargo-instability is unknown.

4.12. Phenotypic stability of genomic co-integrants was superior to their plasmid-based counter-parts under selection-free conditions

A key phenotypic trait of engineered strains is the stability of expressing the inserted transgenes in the lack of selection. We compared strains harboring chromosomal *vioABCDE* operons to those carrying the same operon on multi-copy plasmids using two different tests. Our first test monitored the fraction of cells displaying violacein production. We chose three well-known strains, *E. coli* BL21(DE3), DH5α, and MG1655 to be our controls, all transformed with the pUTLIQ_*vioABCDE* plasmid. From our collection of genome-engineered strains we tested three versions of BLK16IS*3*::*vioABCDE*_SpR (carrying 1, 5.59 or 10.5 copies of IS*3*::*vioABCDE*_SpR, respectively), and strain MDS27IS*3*::*vioABCDE*_KmR, carrying 5.8 copies, according to our ddPCR measurements. As apparent in **Figure 26**, all three control strains displayed a steep decrease in the ratio of purple colonies and completely lost the purple phenotype by generation 40. On the contrary, all four genomic co-integrants tested in this experiment retained purple color in >96% of their colonies, with the single copy strain not displaying loss of function at all.

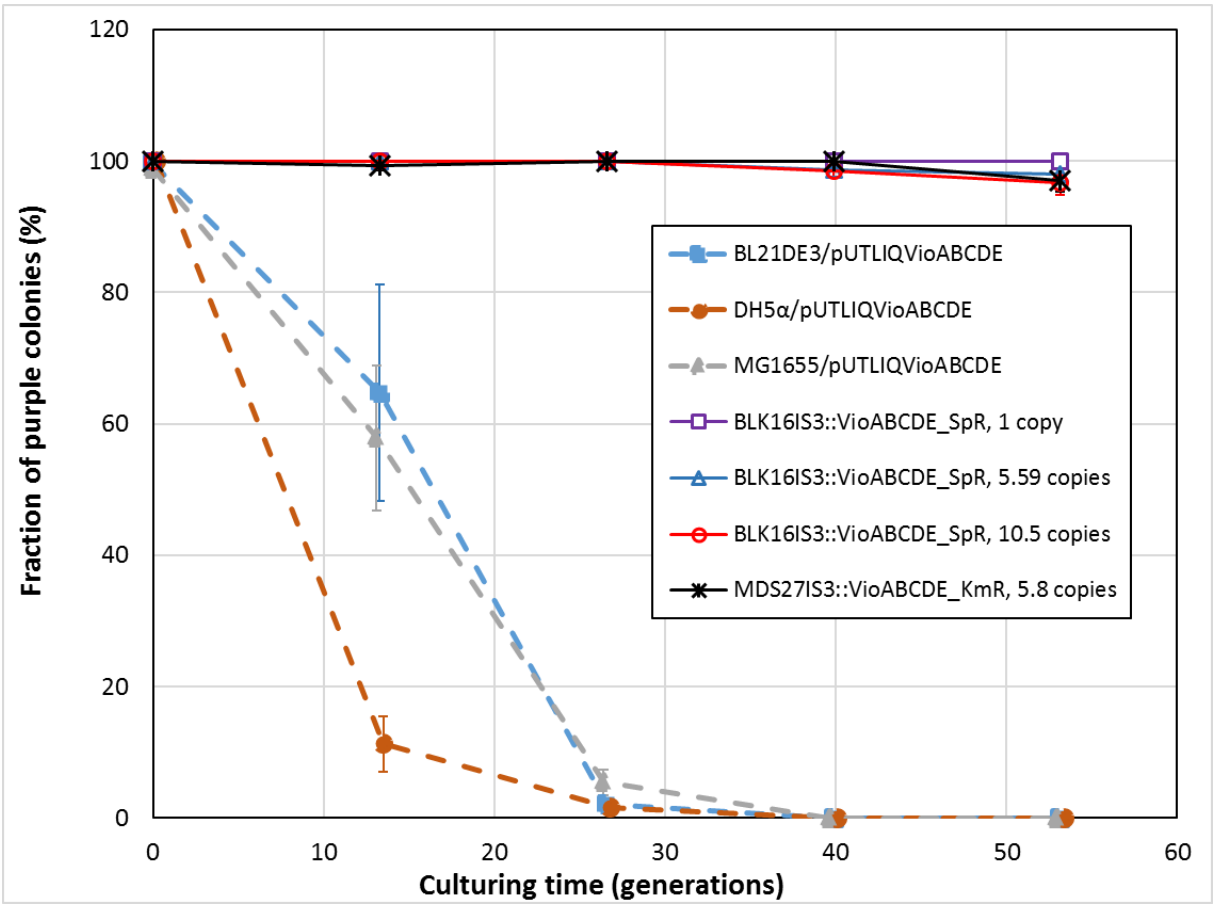


Figure 26. Fractions of violacein-producing cells within bacterial cultures grown in the lack of antibiotic selection. Dashed lines mark strains carrying plasmid pUTLIQ_*vioABCDE*, solid lines mark strains carrying an IS*3*::*vioABCDE* cassette on their chromosomes, as indicated on the legend. All values are means of three biological replicates. Error bars mark the standard error of the means.

We note that i) multi-deletion strains (MDS42 and BLK16) of *E. coli* displayed complete plasmid loss even faster than their conventional, non-genome reduced counterparts, and ii) single and the double copy-harboring strains [BLK16IS*3*::*vioABCDE*_SpR (1 copy, locus 2) and MDS27IS*3*:: *vioABCDE* _KmR, 2 copies, respectively] did not display any loss of function either (**Figure 27**). We, therefore, conclude that our genomic co-integrants can be used for the prolonged expression of transgene arrays in the lack of selection with a negligible fraction of cells displaying a complete loss of production.

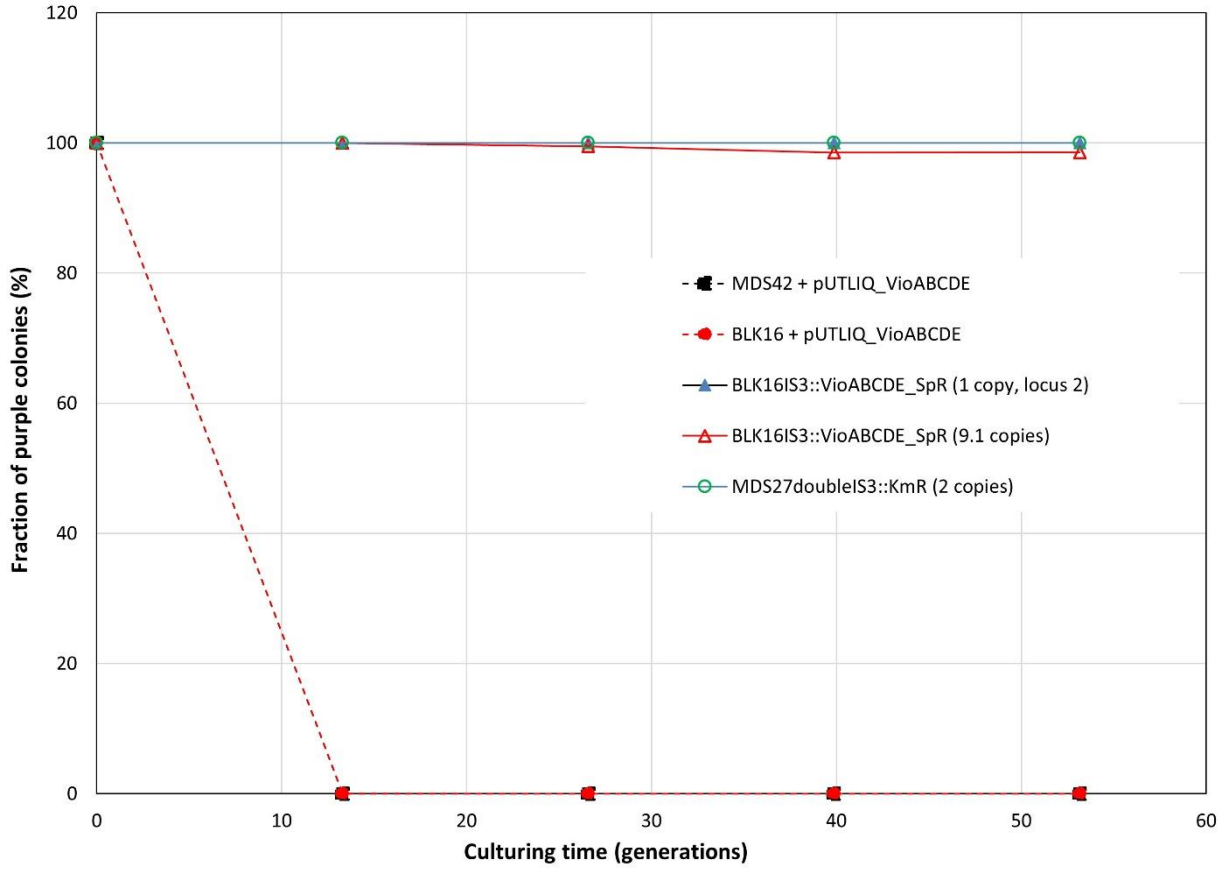


Figure 27. Fractions of violacein-producing cells within bacterial cultures grown in the lack of antibiotic selection. Dashed lines mark strains carrying the pUTLIQvio_ABCDE plasmid, solid

lines mark strains carrying an IS*3*::*vioABCDE* cassette on their chromosomes, as indicated on the legend. All values are means of three biological replicates.

Our second test assaying the stability of violacein production in the lack of selection measured the levels of violacein extracted from the cultures each day. As expected, we saw the rapid decline of violacein levels in the case of *E. coli* BL21(DE3), MG1655, and MDS42 strains initially harboring the pUTLIQ_*vioABCDE* plasmid (**Figure 28**). *E. coli* DH5 α displayed a somewhat smaller, but still significant loss of 35% (P=0.012 with a two-tailed, unpaired t-test). In contrast, no significant decrease in the violacein levels was observed on day 4 (compared to day 1) for strains BLK16IS*3*::*vioABCDE*_SpR carrying 5.59, 9.1, and 10.5 copies of the violacein operon, respectively, and strains MDS27IS*3*::*vioABCDE*_KmR carrying 2 and 5.8 copies, respectively. Furthermore, in this test, the violacein levels produced by MDS27IS*3*::*vioABCDE*_KmR carrying 5.8 copies significantly surpassed that of all other strains, including the four strains carrying the pUTLIQ_*vioABCDE* plasmid (**Figure 28**). Possibly there are two reason for it; 1) MDS27 is K-12 derived, BLK16 is B-derived *E. coli*. In addition, the prior strain has more deletions than the latter, their genetic background is therefore totally different. 2) Location of gene integration is different, and the genetic environment can influence gene expression.

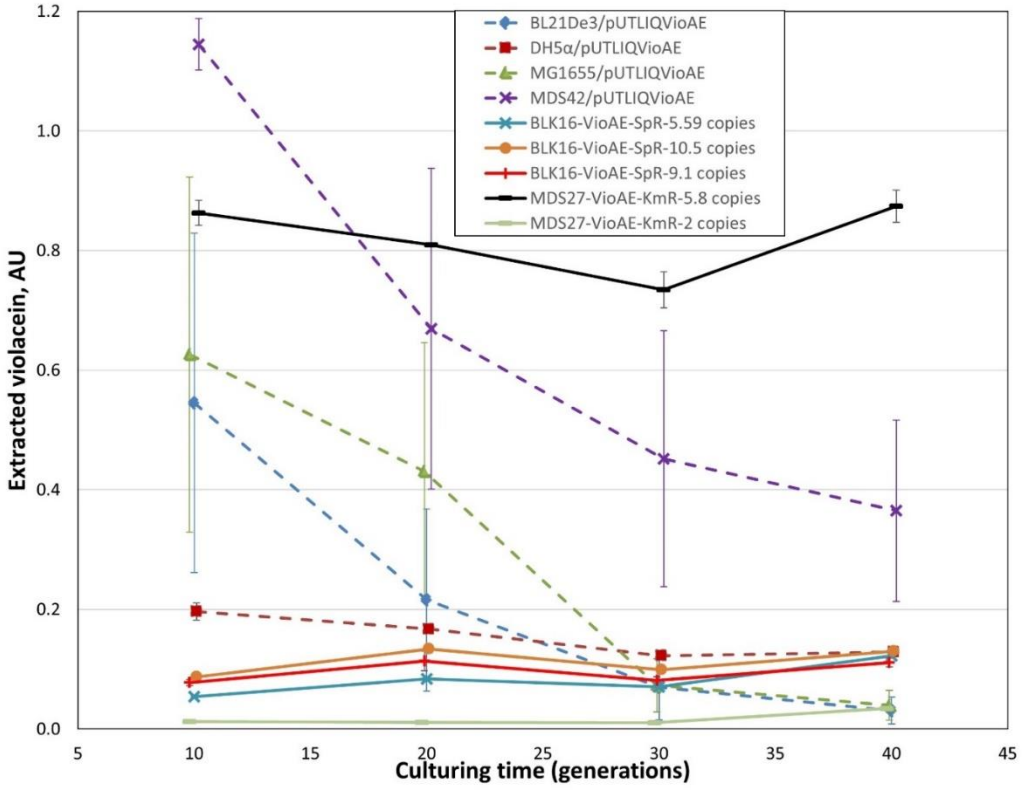


Figure 28. Violacein levels of liquid bacterial cultures grown in the lack of antibiotic selection. Dashed lines mark strains carrying plasmid pUTLIQ_*vioABCDE*, solid lines mark strains carrying an IS3::*vioABCDE* cassette on their chromosomes, as indicated on the legend. All values are means of three biological replicates. Error bars mark the standard error of the mean.

4.13. Transposition of marked IS*3* elements into the *E. coli* chromosome

Finally, we asked whether IS*3* can be transferred into a bacterial strain of choice by controlled transposition. This process, which can be referred to as a potential "Step 0" of the inPOSE protocol (**Figure 31**) could be valuable for future users who wish to utilize IS*3* for chromosomal cloning and amplification inside a strain that has no copies of this element at all. For the process of IS*3* entry into the chromosome, we decided to use transposition, catalyzed by the plasmid-encoded transposase used in the transgene copy-amplification experiments above. To make the IS*3* entry selectable, a marked version of IS*3* was required, we, therefore, used the IS*3*::SpR allele generated in the course of SpR recombineering in MDS30 (see the section on "Targeting ISes by recombineering" above). The IS*3*::SpR cassette was PCR-amplified from the chromosome and cloned into a suicide plasmid that is unable to replicate in strains lacking the *pir* gene. This donor plasmid, pSG78A_full_IS*3*::SpR was transformed into *E. coli* MDS42 carrying the transposase-expressing pSTtnp3tetR plasmid, followed by plating on Sp plates. Replica plating revealed that 10% (2/20) of the obtained colonies were $Sp^{R}Cm^{S}Ap^{S}$, indicating that they had lost both the donor plasmid and the transposase plasmid, but retained the SpR gene, presumably due to its transposition into the genome. To verify transposition and identify the locus of entry, we carried out ST-PCR on the chromosomal DNA prepared from two colonies. Two pairs of PCR reactions were carried out, as described in the Methods (**Figure 29**).

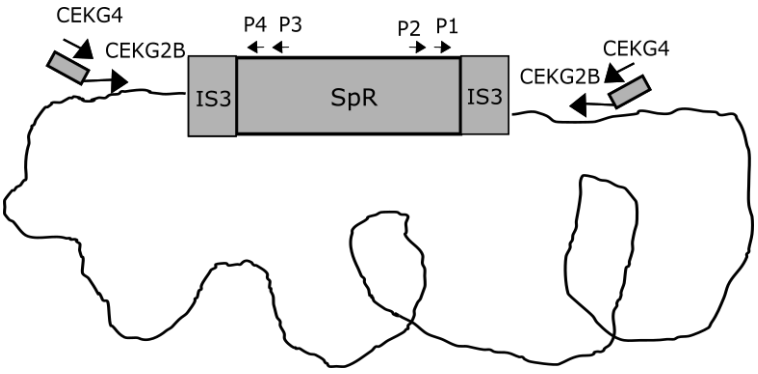


Figure 29. Relative positions of primers used for ST-PCR. P2 (Sp439Rev) and P3 (Sp110Fwd) primers were used for the 1st PCR, and P1 (SmFwd) and P4 (Sp347Fwd) for the 2nd PCR.

Sequencing a unique product (from SmFwd + CEKG4 PCR on colony 11) confirmed IS*3*::SpR insertion into the *pqiA* gene at position 1,012,519 (using NC000913.3 coordinates). In a similar experiment, the transposition of IS*3*::SpR into MDS30 (not into the original IS*3*) was induced and selected, and this time, 33% (10/30) of the colonies displayed a $Sp^{R}Cm^{S}Ap^{S}$ phenotype. Most of the ST-PCR reactions carried out on the genomic DNA of ten such colonies generated unique PCR patterns, probably indicating different points of insertion (**Figure 30**). Sequencing a PCR product from one of the least complex patterns (Sp347Fwd + CEKG4 PCR product of colony 8) revealed IS*3*::SpR integration into the 5' untranslated region of the *waaL* gene (position 3,796,945, NC000913.3 coordinates). We, therefore, demonstrated the transposition-mediated entry of IS*3*::SpR into two different strains of *E. coli* (one IS-free, the other harboring a single copy of IS*3*), with IS*3*::SpR entering the chromosome at two distinct loci.

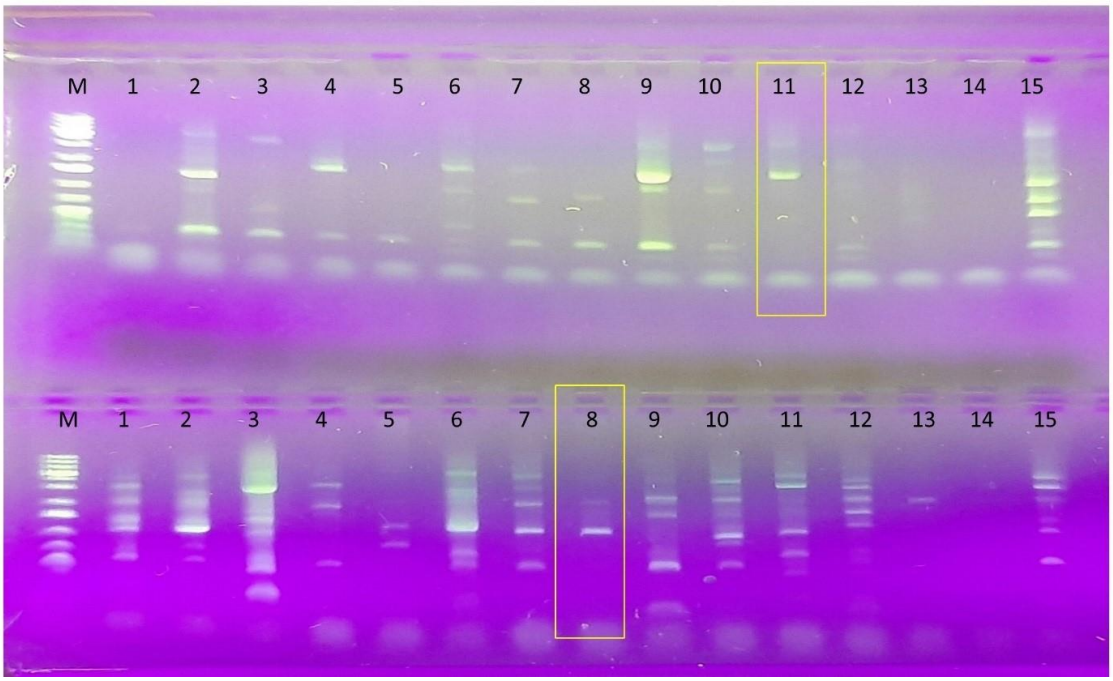


Figure 30. **Verifying the transposition of IS*3*::SpR from pSG78A_full_IS*3*::SpR into the genome by ST-PCR.** The figure shows gel electrophoresis of ST-PCR products generated with primers SmFwd + CEKG4 (top row) or Sp347Fwd + CEKG4 (bottom row), as described in the Methods. In both rows, lanes 1-10 represent ST-PCR reactions from 10 colonies of MDS30 after IS*3*::SpR transposition. Lanes 11 and 12 represent ST-PCR reactions from two colonies of MDS42 after IS*3*::SpR transposition. Lanes 13 and 14 are negative control ST-PCRs, made

on MDS30 and MDS42 colonies, and lane 15 is a positive control ST-PCR made on MDS30_IS*3*::SpR. Yellow rectangles mark the ST-PCR products chosen for sequencing. M: GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific)

5. inPOSE technology:

We were in a lucky situation to have multi-deletion strains harboring a defined number of IS elements at known locations. But if other labs want to use our method, they need to introduce the marked respective IS element (we recommend to use IS*3* element) first into the bacterial genome (Step 0), and then the marked gene can be replaced with multi-gene cargo (step 1) followed by copy number amplification of inserted cargo (step 2). Graphical representation of inPOSE technology is shown in **Figure 31.**

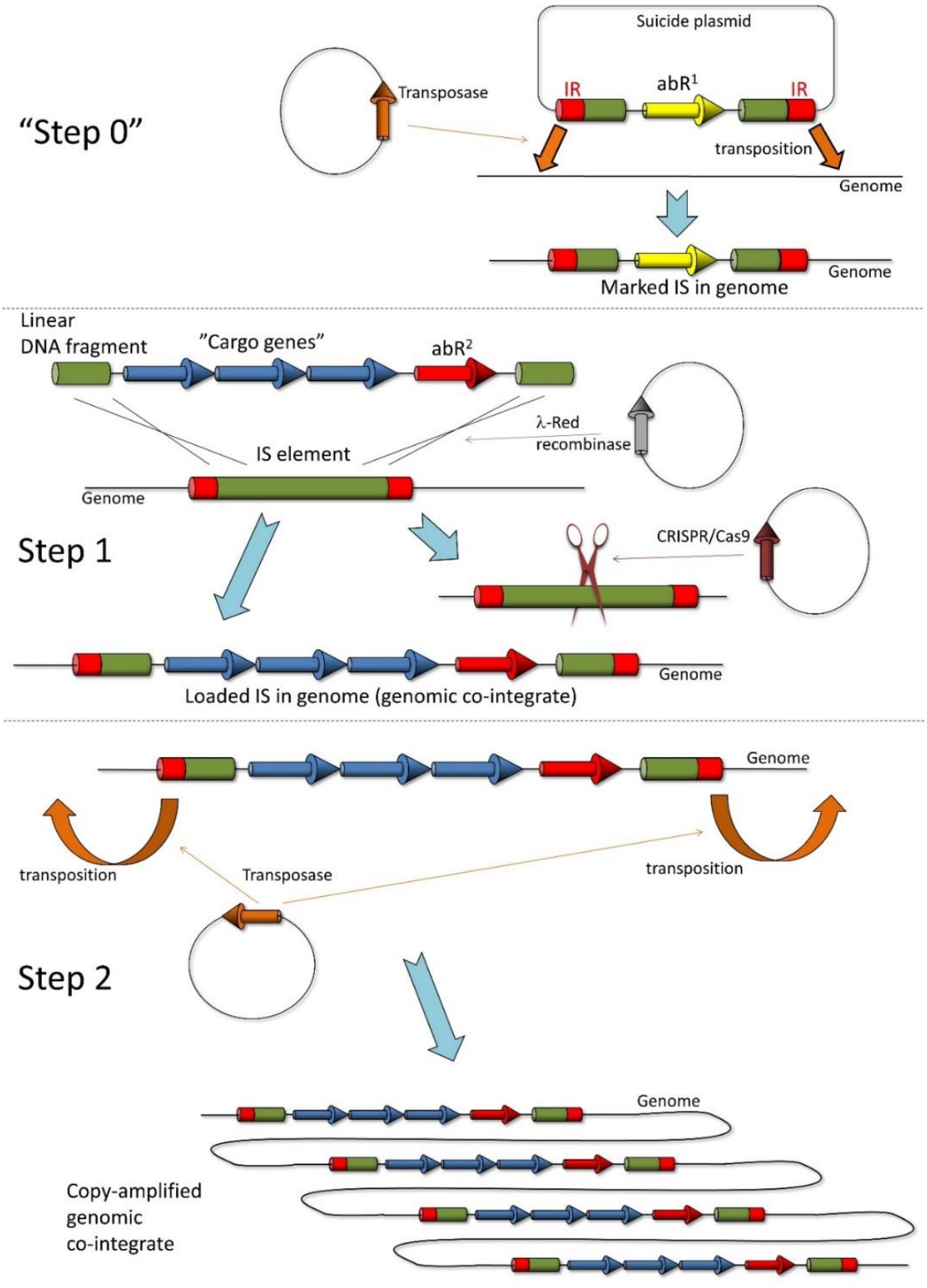


Figure 31. The inPOSE protocol. Step 1 is the entry of a gene or operon of interest into an IS element residing on the host genome by recombineering, facilitated by the λ-Red recombinase enzymes. CRISPR/Cas-mediated cleavage of the wt IS element(s) enforces

selection for the recombinants and, if applied concomitantly, facilitatesf the recombination event. In Step 2, the genomic co-integrant (i.e. the loaded IS) is copy-amplified by the transposase corresponding to the IS, expressed in trans. “Step 0” is an optional accessory step that can transpose the marked IS element into the genome of the host cells chosen for chromosomal transgene cloning. Green boxes: the targeted IS element (or segments thereof); red boxes: inverted repeats (IRs) of the IS; yellow and red arrows: two different antibiotic resistance genes (abR^1 and abR^2, respectively); blue arrows: the transgenes to be integrated; orange arrows: transposase gene of the targeted IS; gray arrow: λ-Red recombinase genes; brown arrow: CRISPR/Cas genes; open thin black lines: host bacterial genome; closed thin black lines: circular plasmids.

5.1 Protocol of the inPOSE technology

Step 0. Integrating an IS*3* element into the chromosome (optional)

a) Transform the targeted bacterial strain with pSTtnp3tetR. Pick a colony, and grow a fresh overnight culture in LB+Cm at 30°C.

b) Dilute the culture 100-fold into 5 mL LB+Cm+aTc (50 ng/mL), and grow at 30°C for 2 h.

c) Make electrocompetent cells. (Pellet and resuspend cells in sterile Milli-Q water three times. The final resuspension should take place in 40 μl of water.)

d) Electroporate with 150–200 ng of pSG78A_full_IS3::SpR.

e) After 1 h of shaking at 30°C, spread 100 μl on Sp plate and incubate overnight at 37°C.

f) The next day, replica-plate several colonies on Cm, Ap, and Sp plates, incubate overnight at 37°C.

g) Find $Sp^RCm^SAp^S$ colonies (i.e., those that have lost both plasmids but retained the IS*3*::SpR) and inoculate into liquid cultures for overnight growth to generate glycerol stocks.

h) The putative genomic insertions of IS*3*::SpR can be localized using inverse PCR, ST-PCR, or whole-genome sequencing.

Step 1. Integrating a large genetic construct into a chromosomal IS*3* element

a) Design primers to amplify your genetic construct. The 5′ overhangs to be added are the following:

b) Forward primer:
TCAGAGGTGACTCACATGACAAAAACAGTATCAACCAGTA

c) Reverse primer:
CACACCAACTGTGCCGCCGCCACCGATTGTAATCACATTCGA

d) Note: If inserting a marked gene construct, the use of Km resistance is recommended.

e) PCR-amplify your genetic construct and purify the PCR product using any commercial kit.

f) Transform the IS*3*-harboring bacterial strain with pKDsg-IS3 and pCas9cr4.

g) Grow an overnight starter culture of the double-plasmid harboring strain in LB+Sp+Cm at 30 °C.

h) Dilute the culture 100-fold in LB+Sp+Cm and grow at 30 °C to OD600 of 0.45–0.55.

i) Add 0.2% L-arabinose and grow at 30 °C for 15 min.

j) Cool on ice for 10 min.

k) Make electrocompetent cells (as in Step 0/C).

l) Electroporate 100–200 ng of your linear PCR fragment, add 1 mL of LB, and shake at 30 °C for 2 h (recovery time).

m) Spread 900 µl on a Cm+Km+aTc (50 ng/mL) plate and incubate overnight at 37 °C.

n) The next day, screen for co-integrants by colony-PCR using a pair of primers that hybridize to the cargo and the chromosomal segment neighboring the targeted IS elements, respectively.

Step 2. Amplification of the inserted genetic construct

a) Grow a fresh overnight culture of the *E. coli* strain harboring the IS*3* element (marked with SpR or KmR) at 37 °C.

b) Make electrocompetent cells (as in Step 0/C).

c) Electroporate the transposase-expressing plasmid (e.g., pSTtnp3tetR), spread on a Sp+Cm or Km+Cm plate, and incubate overnight at 30 °C.

d) Pick a colony and grow as an overnight starter culture in liquid LB using the same antibiotics at 30 °C.

e) Dilute the grown culture 10,000-fold and grow again overnight at 30°C in LB with antibiotics and 50 ng/mL aTC (an uninduced control can be grown in parallel).

f) Repeat dilution and growth after 24 h.

g) Cure the plasmids by growing the 10,000-fold diluted culture at 37 °C in LB without antibiotic selection for five days, applying a 10,000-fold dilution every day.

h) From the 5^{th} fully grown serial culture, spread 1 µl on a 60x Sp gradient plate or a 20× Km gradient plate and incubate at 37°C for 24 h.

i) The next day, pick 15 colonies from the high antibiotic-concentration area of the gradient plate and replica-plate on Cm at 30 °C and Sp or Km, grow overnight at 37°C.

j) Choose $Cm^{S}Sp^{R}$ or $Cm^{S}Km^{R}$ colonies and grow in liquid LB overnight without selection at 37°C.

k) The fully grown cultures can be saved as glycerol stocks and/or used for product quantification or to prepare genomic DNA for droplet-digital dPCR analysis.

l) (This process of Step 2 is defined as one round of transposase induction, and can be repeated with the best-producing colonies.)

6. Discussion

When using bacteria to express transgenes from plasmids, allelic segregation can quickly result in the loss of plasmid-encoded functions. Mathematical models have shown, however, that ordered inheritance, e.g. encoding transgenes on chromosomes can prolong their expression ten-fold (Corchero & Villaverde, 1998)(Hajimorad & Gralnick, 2020). Several methods have been developed for the integration of genetic constructs into the bacterial genome. Initial techniques were dependent on homologous recombination-mediated plasmid integration followed by resolution of co-integrants for marker-less integration of transgenes (Martinez-Morales et al., 1999). Other strategies used site-specific recombinases, e.g. the FLP/FRT system derived from yeast to deliver the desired plasmid into the genome by site-specific recombination (Huang et al., 1997). This method however needs prior insertion of FRT, the FLP recombinase target site into the host chromosome. Later linear DNA cassette-mediated genomic allele exchange methods became popular which used λ–Red recombinases to insert transgenes into the chromosome (Jasin & Schimmel, 1984). To integrate >3 kb fragments, target cleavage by endonucleases was usually required, and as the size of the insert increased, efficiency of recombination decreased (Pósfai et al., 1999).

Besides the mentioned drawbacks, these methods were capable of engineering single-copy transgene integration only. In the year 2009 Tyo et al. showed that transgenes inserted into a bacterial genome in single copy along with a resistance gene can be amplified to multiple copies by growing the strain in the presence of gradually increasing concentration of antibiotic (Tyo et al., 2009). In this work, we describe a method based on transgene entry into chromosomal IS elements using homologous recombination, gene amplification using transposition, and selection of high-copy lines using antibiotic gradient plates.

We choose bacterial IS elements as a tool for this project because they are the simplest transposable elements, not essential for the survival of bacteria, present in multiple copies in

bacterial genomes, and are well-studied (Touchon & Rocha, 2007). Within our project, we demonstrated the feasibility of the technique using IS*1* and IS*3* elements. The Tn5 transposon is a well-known tool used for the transfer of DNA from a plasmid to the bacterial genome (Reznikoff, 2003). It however relies on its transposition function for gene transfer, the gene entry takes place at an unknown locus, and it does not allow gene amplification, it is therefore quite different from our strategy. We started with loading IS elements already present in the genome with linear heterologous DNA by λ-Red mediated homologous recombination. In our approach IS elements serve as natural landing pads for genomic targeting. We observed that targeting IS elements by homologous recombination was not different fromtargeting any non-essential gene. We started with the integration of single selectable genes (spectinomycin- as and chloramphenicol resistance genes in parallel) into the genome of strains that have a single copy of either IS1 or IS3 in their chromosome. The yields of recombinants were mostly in the range of 0.5-20 recombinants/ng linear DNA transformed. This was in good agreement with values found in the literature (Kuhlman & Cox, 2010), and could be increased further by using longer homologies which is also a well-known phenomenon (Liu et al., 2003). We observed a 13-fold increase in the efficiency of recombination when increasing the length of the homologies just 2.5 fold (**Table 4 and figure 9**).

Once we achieved the integration of selectable genes, we moved on to the integration of a multi-gene operon. For this purpose, we chose the ~9.5 kb violacein operon consisting of five genes (*vioABCDE*), because it produces a purple dye and co-integrates can easily be detected visually. Even single copy and multi-copy strains can be differentiated visually by the intensity of purple color. The first question was how to make the linear fragment. Amplifying such a long fragment by PCR was not easy, so we assembled it into a thermosensitive plasmid, where the cargo was flanked by homology regions corresponding to the targeted IS sequence, then the fragment used for transformation was prepared the plasmid by restriction digestion. Despite our low expectations based on earlier reports, we easily achieved single co-integrants irrespective of whether the host strain had 1,2 or 3 copies of the targeted IS element. PCR-based genotyping verified the formation of co-integrants into both IS*1* as well as IS*3*. Of the two IS elements IS3 proved more efficient and reproducible with regard to targeting as well as expression of the cargo gene. Investigation of the worse performance of IS1 was beyond the scope of our project, so we dropped the use of IS*1* and focused on IS*3*.

Next, we turned to the integration of an un-selectable gene (*gfp*). Relying solely on recombineering, we could not obtain co-integrants because *gfp* expressing cells cannot be easily recognized visually in the background of wild-type cells. We concluded that a further level of selection, provided by CRISPR/Cas cleavage of the wt allele would be needed to isolate recombinants (see below).

When inserting a single selectable gene into strains harboring a single copy of IS element, we observed that 100% of the selected colonies were carrying heterologous DNA at the correct locus. When we targeted strains harboring 2 or 3 copies of IS element, however, we found colonies that carried the integrated DNA at genomic loci other than the targeted IS element. Cases of unknown integration were more frequent for IS*1* in comparison to IS*3* targets. We obtained similar results when inserting a long DNA cassette (*vioABCDE*), but unknown integrations were seen even in strains harboring a single copy of the respective IS element.

Our next goal was to integrate a transgene or the multi-gene *vioABCDE* operon) into the IS elements of bacteria whose genome carried more than one copy of the element. Using recombineering alone, only integration of single copies could be achieved. To select for double insertions we used the CRISPR/Cas system to cleave the unloaded IS elements. We expected this cleavage to i) increase the efficiency of recombination by providing free DNA ends, and ii) increase the selective pressure by removing the unedited cells from the mixed population. We constructed a pCas9 plasmid derivative for each IS element, one for IS*1* and the other for IS*3*. We observed that CRISPR/Cas-assisted recombineering did not enhance the efficiency of recombination when targeting a single IS in the genome, but permitted achieving single co-integrants of an un-selectable gene (*gfp*) in IS*1* as well as IS*3*. When we targeted bacterial strains having 2 copies of the respective IS element, it became possible to achieve double co-integrates of the selectable marker as well as of the marked *vioABCDE* operon. We attribute the results to the fact that Cas9 will keep on cleaving all the wild-type loci until all are occupied by heterologous DNA, or in other words, removes unedited cells from the mixed population. When we tested this technique with the strain MDS16 having 3 copies of IS*3* element, we could obtain triple co-integrants neither with short genes (SpR, *gfp*) nor with bigger cargo (*vioABCDE* operon). The possible reason for the lack of triple cointegrants can be the frequency of triple recombination being lower than that of double-strand break repair. Perhaps multiple cycles of linear DNA transformation (≥3) followed by CRISPR/Cas selection could help to attain triple co-integrants but we did not test this.

We observed that CRISPR/Cas assisted recombineering could not show a prominent effect when inserting single genes into the genome but, integrating the >9 kbp DNA cassette in multiple copies into the *E. coli* chromosome was only possible using the NO-SCAR system. This system applies CRISPR/Cas cleavage and recombination at the same time. We believe this is due to the facilitating effect of the cleavage on the efficiency of recombination.

Once we achieved single and double co-integrants of the SpR gene as well as of the *vioABCDE* operon, our next goal was to amplify the inserted cargo to obtain multi-copy strains. Insertion of the transgene at multiple pre-designed landing pads in repetitive cycles has been reported but the achievement was limited to obtaining four copies of the *lacZ* gene in the *E. coli* chromosome (St-Pierre et al., 2013)(Peredelchuk & Bennett, 1997). A study reported the use of serial passaging cells in a medium containing increasing antibiotic concentration as a method to select strains carrying multi-copy cointegrates of an antibiotic resistance gene. The underlying mechanism was a RecA-mediated gene duplication arising from unequal crossovers between the daughter chromosomes, which permitted achieving up to ~40 copies (Tyo et al., 2009). The major disadvantage of the method was that the engineered strains were not stable to hold the copies, and required deletion of the *recA* gene to maintain stability.

Our method developed in this project is a bit similar to the latter one, except that we used the replicative transposition of an IS element for the amplification of already inserted single or double copies of the transgene. We provided the transposase expressing plasmid (inducible and thermo-sensitive) to the host cell carrying heterologous DNA at known loci. We used the antibiotic gradient plate for screening of the multi-copy strains so antibiotic gradient was required for "readout" but not responsible for copy number amplification.

We observed the prominent difference in the distribution of resistance levels in induced and non-induced conditions after the first round of transposase expression **(Figure 14, 15)**. This difference was smaller after the second and third rounds of induction (**Figures 24, 25**). This may be explained by two alternative mechanisms. First, at higher starting copy numbers of the marked IS, even the leaking expression of the transposase was sufficient to facilitate enough transpositions to generate highly-resistant cells in the uninduced population. Second, the spontaneous RecA-mediated duplication of marked ISes most likely takes place at all copy numbers; the duplication of many copies (at later rounds) however probably has a more dramatic effect onresistance than the duplication of a single copy (in the first round).

Irrespective of the mechanism of emergence, these spontaneously arising clones of elevated resistance could in certain cases harbor just as many copies as the induced clones.

The induction of the IS*3* transposase was repeated 3 times. After each round of transposase induction, 5-6 colonies were picked from the high antibiotic concentration region, and the copy-number of the resistance gene was determined by droplet digital PCR. We observed that the mean copy numbers of the colonies displayed a significant increase up to the second round. In the third round of induction, the overall copy number was unchanged, but we could identify instances of further copy-number increase, and using whole-genome sequencing we managed to identify a lineage carrying 12 intact copies of the violacein operon.

From the biotechnological point of view, a key trait for engineered strains is the ability to maintain the stable expression of the desired function under selection-free conditions. Our next goal was therefore to test the selection free expression of the violacein operon in engineered strains and compare them with plasmid-based counterparts. We set up two experiments for such a comparative analysis. Our first experiment was based on monitoring the fractions of purple and non-purple colonies obtained during growth under selection-free conditions by spreading samples on LB plates. We observed a rapid loss of violacein production in plasmid-based strains, but their genome-engineered counterparts were able to maintain the production in >96% of the cells (**Figure 26**).

In the second type of experiment we monitored the extracted violacein levels during selection-free growth in liquid culture. We observed that plasmid-based expression strains showed a significant decrease in the violacein levels after 40 generations of growth, as opposed to the unchanged levels of violacein production displayed by genome-engineered strains (**Figure 28**). But the strain-dependent expression of the violacein operon was visible in both plasmid-based as well as in genomic co-integrant strains, this phenomenon however was not investigated further.

Homologous recombination-based integration followed by replicative transposition to increase the copy number was successfully demonstrated by our experiments using IS*3*. We do not declare however that IS*3* is the best element for this purpose, nor do we have information about its applicability in other species. We, therefore, believe that a) IS*3*-mediated integration and amplification should be tested in other bacterial species, and b) other bacterial IS elements

should be tested in *E. coli* and possibly further species if one wishes to identify a broad spectrum transposable elements for application in inPOSE.

We have further shown that a marked IS*3* element can be inserted into the genome of bacterial strains which lack the IS*3* element. We call it "step 0", where a marked element can transpose to an unknown locus of the bacterial genome from a conditionally replicating plasmid. We verified the locus of genomic integration by ST-PCR followed by sequencing. Although this step is similar to chromosomal cloning of transgenes by Tn5-mediated transposition (Reznikoff, 2003), we emphasize that we recommend this step, not for transgene entry, but to integrate a landing pad that can warrant further transgene entry by homologous recombination.

Our technique, inPOSE (integration followed by transposition) looks to be a promising tool among the methods currently available for genome editing; we note however that it has certain disadvantages that could pose challenges in the future. For example, at present, we cannot explain, why IS*1* displayed suboptimal performance. Therefore, we cannot predict how other IS elements will perform. To choose optimal IS element for inPOSE, we will likely follow the trial and error method.

We verified by whole genome sequencing that the distribution of copies is not based on random scattering but instead, observed a concatenation just on the right side of the first copy. We found that not only the modified IS but 14-18 kbp-long genomic segment lying directly downstream of the IS was also amplified within the tandem repeats. Possibly the transposase enzyme was not acting on the right IR of the ISes, but on an unidentified pseudo-IR lying further downstream, as seen earlier (Mendiola et al., 1994). This could be caused by the fact that the central 827 bp long segment of the IS was replaced with the transgene, perhaps resulting in the elimination of an unknown factor required for correct transposition. The transposase gene of IS*3*, expressed from a plasmid in-trans was codon-optimized, which can potentially eliminate any unknown ORF or regulatory sequence overlapping with the transposase ORF. Identifying and providing this potential missing factor may result in a correct, randomly scattered transposition of the loaded IS.

We also observed that the expression level of the inserted operon was not easily predictable from its copy number. For example, MDS27IS*3*::*vioABCDE*_KmR strains carrying 5.8 copies of the KmR gene (measured by ddPCR of the KmR gene) produce un-proportionally more violacein than the same type of strain carrying 2 copies (**Figure 28)**. On the contrary, BLK16IS*3*::*vioABCDE*_SpR strains displayed violacein production that was more or less

proportional to the copy number of the *vio* operon, as apparent in (**Figure 28)**. On the other hand, the MDS27IS*3*::*vioABCDE*_KmR strain carrying 5.8 copies of the KmR gene was a much higher producer than the BLK16IS*3*::*vioABCDE*_SpR strain having 12 copies, possibly indicating that the locus of transgene integration is also an important factor concerning its level of expression. We cannot rule out the possibility of strain and locus-dependent factors influencing the expression of the inserted transgene, either. We note that the ddPCR was based on antibiotic marker detection, which may have a copy number different than the fully functional operon. The possible solution to avoid the locus dependent expression could be flank the cargo inserted into IS element with transcription and translation terminators (Sibley & Raleigh, 2012), thereby improve reproducibility of the inPOSE system.

Another potential drawback of our method is the possibility of a homologous recombination-mediated loss of tandem repeats. The initial slight decrease in violacein production, seen for multi-copy co-integrant lines in **Figure 28** may be attributable to this phenomenon. Hopefully, complementing the missing factor of the IS*3* element will eliminate this caveat by avoiding concatenation. We assume that removal of recA gene from final strain (Tyo et al., 2009) or using the inPOSE process in a recA- as starting strain could also help to avoid these issues, these solutions, however, require experimental testing. We are looking forward to receiving feedback from other labs concerning the use of IS*3* or other IS elements for inPOSE and using the obtained experience to further optimize the process.

7. Conclusions

In this work we demonstrated the recombineering mediated integration of an antibiotic resistance gene and of a five-gene operon marked with an antibiotic resistance gene into the *E. coli* genome using the IS*1* or the IS*3* insertion elements as landing pads. Facilitating the recombineering step with CRISPR/Cas cleavage (applied either concomitantly or subsequently) aided the integration of long (>9 kbp) cassettes and was essential to insert unselectable genes (e.g., *gfp*), or to obtain double co-integrants of selectable gene cassettes in a single step. We have shown that the copy number of IS*3*-carried transgenes can be increased by expressing the transposase in trans. Our work also demonstrated the increased stability of chromosomal transgenes in the lack of selection, compared to their plasmid-borne counterparts. The flexibility of this strategy, called inPOSE, is demonstrated using the IS*3* element but is most likely

applicable to other bacterial ISes as well. For those who wish to use IS*3* for this purpose in strains harboring no IS*3*, we developed tools to rapidly introduce a marked copy of IS*3* into the chromosome by transposition. In conclusion, our work demonstrated that IS elements can be used as tools for bacterial genome engineering. Marked IS elements can be transferred into the genome of IS element-free bacterial strains by transposition. The IS elements already present in the bacterial genome can be loaded with desired heterologous DNA by recombineering. The heterologous DNA, once inserted into the bacterial genome can be further amplified to up to >12 functional copies by trans-expression of the corresponding transposase.

8. Personal contribution

The results presented in this Ph.D. dissertation originated from the work I carried out in the Tamás Fehér lab, the laboratory of Bacterial Physiology and Strain Engineering at the Institute of Biochemistry, Biological Research Centre, Szeged, Hungary. I performed the majority of the experiments, such as the construction of transposase expressing system, violacein producing plasmid, genome engineering experiments, strain screening, violacein quantification, and outcrossing experiments to determine the rate of transposition of the marker from genome to plasmid. Preparation of plasmid and gDNA samples for ddPCR and whole-genome sequencing was done by me. PCR-based genotyping was done by me along with Ákos Avramucz and Alyona Biketova. Whole-genome sequencing was done by Delta Bio 2000 Ltd., Szeged. All the plasmid and gene sequencing was done in the central DNA sequencing facility of the BRC by Nikoletta Borák and all the primers were synthesized by Eurofins Genomics Germany GmbH. ddPCR experiments were performed by Ágnes Zvara (Laboratory for Functional Genomics, Biological Research Centre, Szeged, Hungary). Fluorescence-activated cell sorting (FACS) experiments were performed by Edit Kotogány (Laboratory for Functional Genomics, Biological Research Centre, Szeged, Hungary). Ákos Nyerges (Department of Genetics, Harvard Medical School, Boston, MA) helped with his suggestions.

9. Acknowledgments

First and foremost I am extremely grateful to my supervisor Dr. Tamás Fehér. He provided an outstanding international scientific environment in his lab in Szeged. His invaluable advice and continuous support during my PhD study helped me to accomplish this goal. It has been really a great experience working with Tamás. Thanks for the opportunity!

I would like to thank Viktor Vernyik for scientific discussions in coffee breaks. Thanks to Alona Yu. Biketova, Ákos Avramucz and Ágnes Zvara for contributions during the project. I'm very thankful to Ildikó Karcagi, Gabriella Balikó, Edit Tímár, Zsuzsanna Györfy, Nikolett Zsibrita, Mihály Koncz and Pál Albert for their great help and support. Thanks are due to Ákos Nyerges for his valuable suggestions and the pORTMAGE plasmids.

I would like to thank Dr György Pósfai for asking thought provoking questions at group meetings and for proof reading my 1st author article. I am thankful to Kristala Prather, Mattheos Koffas and Kanji Nakamura for gifting the respective plasmids. I am thankful to Walliyulahi Ajibola for his kind help. It is the kind help and support of the Pósfai and the Fehér group members that made my study and life in Szeged a wonderful time. I would also like to thank all my friends in Hungary who made my life enjoyable. I thank Pramod Kumar Jangid, Imran Baba, Gyula and Alena for making my time great in Szeged.

Finally, I would like to express my gratitude to my parents and family, without their love and support, it would be impossible for me to go ahead in my career. I would like to thank all those who directly or indirectly contributed in my PhD journey.

I am really thankful to my thesis reviewers for their kind time and helping me to improve my thesis by constructive comments and suggestions.

This work was supported by the GINOP-2.3.2-15-2016-00001 grant and the National Research, Development, and Innovation Office of Hungary (NKFIH) grant No. K119298.

10. References

Baeshen. (2015). *J. Microbiol. Biotechnol*. *25*(7), 953–962.

Baker, B. (2014). *Synthetic Biology*. CQ Researcher. http://library.cqpress.com/cqresearcher/cqresrre2014042500

Boeke, J. D. (2002). Putting mobile DNA to work. In A. M. L. (Editor) Nancy L. Craig (Editor), Robert Cragie (Editor), Martin Gellert (Editor) (Ed.), *Mobile DNA II.* (II, pp. 24–37). American Society for Microbiology (ASM Press). https://www.wiley.com/en-uz/Mobile+DNA+II%2C+2nd+Edition-p-9781555812096

Chalmers, R., & Blot, M. (1999). Insertion Sequences and Transposons. In *Organization of the Prokaryotic Genome* (pp. 151–169). https://doi.org/https://doi.org/10.1128/9781555818180.ch9

Choi, K.-H., & Kim, K.-J. (2009). Applications of transposon-based gene delivery system in bacteria. *Journal of Microbiology and Biotechnology*, *19*(3), 217–228. https://doi.org/10.4014/jmb.0811.669

Chun, K. T., Edenberg, H. J., Kelley, M. R., & Goebl, M. G. (1997). Rapid amplification of uncharacterized transposon-tagged DNA sequences from genomic DNA. *Yeast, 13*(3), 233–240. https://doi.org/10.1002/(SICI)1097-0061(19970315)13:3<233::AID-YEA88>3.0.CO;2-E

Corchero, J. L., & Villaverde, A. (1998). Plasmid maintenance in Escherichia coli recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnology and Bioengineering*, *58*(6), 625–632. https://doi.org/https://doi.org/10.1002/(SICI)1097-0290(19980620)58:6<625::AID-BIT8>3.0.CO;2-K

Durán, N., Justo, G. Z., Ferreira, C. V., Melo, P. S., Cordi, L., & Martins, D. (2007). Violacein: properties and biological activities. *Biotechnology and Applied Biochemistry*, *48*(3), 127. https://doi.org/10.1042/ba20070115

Endy, D. (2005). Foundations for engineering biology. *Nature, 438*(7067), 449–453. https://doi.org/10.1038/nature04342

Ensley, B. D. (1986). Stability of recombinant plasmids in industrial microorganisms. *Critical*

Reviews in Biotechnology, *4*(3), 263–277. https://doi.org/10.3109/07388558609150796

Fehér, T., Bogos, B., Méhi, O., Fekete, G., Csörgő, B., Kovács, K., Pósfai, G., Papp, B., Hurst, L. D., & Pál, C. (2012). Competition between transposable elements and mutator genes in bacteria. *Molecular Biology and Evolution*, *29*(10), 3153–3159. https://doi.org/10.1093/molbev/mss122

Fekete, P. Z., Brzuszkiewicz, E., Blum-Oehler, G., Olasz, F., Szabó, M., Gottschalk, G., Hacker, J., & Nagy, B. (2012). DNA sequence analysis of the composite plasmid pTC conferring virulence and antimicrobial resistance for porcine enterotoxigenic Escherichia coli. *International Journal of Medical Microbiology*, *302*(1), 4–9. https://doi.org/10.1016/j.ijmm.2011.07.003

Goryshin, I. Y., Naumann, T. A., Apodaca, J., & Reznikoff, W. S. (2003). Chromosomal deletion formation system based on Tn5 double transposition: Use for making minimal genomes and essential gene analysis. *Genome Research*, *13*(4), 644–653. https://doi.org/10.1101/gr.611403

Guynet, C., Achard, A., Hoang, B. T., Barabas, O., Hickman, A. B., Dyda, F., & Chandler, M. (2009). Resetting the Site: Redirecting Integration of an Insertion Sequence in a Predictable Way. *Molecular Cell*, *34*(5), 612–619. https://doi.org/10.1016/j.molcel.2009.05.017

Gwinn, M. L., Stellwagen, A. E., Craig, N. L., Tomb, J. F., & Smith, H. O. (1997). In vitro Tn7 mutagenesis of Haemophilus influenzae Rd and characterization of the role of atpA in transformation. *Journal of Bacteriology*, *179*(23), 7315–7320. https://doi.org/10.1128/jb.179.23.7315-7320.1997

Hajimorad, M., & Gralnick, J. A. (2020). Plasmid copy number variation of a modular vector set in Shewanella oneidensis MR-1. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2020-July*, 2372–2375. https://doi.org/10.1109/EMBC44109.2020.9176029

Hammerschmidt, S., Hilse, R., Van Putten, J. P. M., Gerardy-Schahn, R., Unkmeir, A., & Frosch, M. (1996). Modulation of cell surface sialic acid expression in Neisseria meningitidis via a transposable genetic element. *EMBO Journal*, *15*(1), 192–198. https://doi.org/10.1002/j.1460-2075.1996.tb00347.x

Huang, L. C., Wood, E. A., & Cox, M. M. (1997). Convenient and reversible site-specific

targeting of exogenous DNA into a bacterial chromosome by use of the FLP recombinase: The FLIRT system. *Journal of Bacteriology*, *179*(19), 6076–6083. https://doi.org/10.1128/jb.179.19.6076-6083.1997

Imre, A., Olasz, F., & Nagy, B. (2011). Site-directed (IS30-FljA) transposon mutagenesis system to produce nonflagellated mutants of Salmonella Enteritidis. *FEMS Microbiology Letters*, *317*(1), 52–59. https://doi.org/10.1111/j.1574-6968.2011.02210.x

Izumiya, H., Sekizuka, T., Nakaya, H., Taguchi, M., Oguchi, A., Ichikawa, N., Nishiko, R., Yamazaki, S., Fujita, N., Watanabe, H., Ohnishi, M., & Kuroda, M. (2011). Whole-genome analysis of Salmonella enterica serovar typhimurium T000240 reveals the acquisition of a genomic island involved in multidrug resistance via IS1 derivatives on the chromosome. *Antimicrobial Agents and Chemotherapy*, *55*(2), 623–630. https://doi.org/10.1128/AAC.01215-10

Jasin, M., & Schimmel, P. (1984). Deletion of an essential gene in Escherichia coli by site-specific recombination with linear DNA fragments. *Journal of Bacteriology*, *159*(2), 783–786. https://doi.org/10.1128/jb.159.2.783-786.1984

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, *31*(3), 233–239. https://doi.org/10.1038/nbt.2508

Joshi, R., Ho, K. K., Tenney, K., Chen, J., Golden, B. L., & Gimble, F. S. (2012). *Recognition Site Specificity*. *405*(1), 185–200. https://doi.org/10.1016/j.jmb.2010.10.029.Evolution

Kahn, J. (2011). *Valparaiso University Law Review Synthetic Hype : A Skeptical View of the Promise of Synthetic Biology SYNTHETIC HYPE : A SKEPTICAL VIEW OF*. *45*(4), 29–46.

Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., & Sternberg, S. H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature*, *571*(7764), 219–225. https://doi.org/10.1038/s41586-019-1323-z

Kuhlman, T. E., & Cox, E. C. (2010). Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Research*, *38*(6). https://doi.org/10.1093/nar/gkp1193

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

https://doi.org/10.1093/bioinformatics/btp324

Li, N., Jin, K., Bai, Y., Fu, H., Liu, L., & Liu, B. (2020). Tn5 transposase applied in genomics research. *International Journal of Molecular Sciences*, *21*(21), 1–15. https://doi.org/10.3390/ijms21218329

Liu, P., Jenkins, N. A., & Copeland, N. G. (2003). A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Research*, *13*(3), 476–484. https://doi.org/10.1101/gr.749203

Martinez-Morales, F., Borges, A. C., Martinez, A., Shanmugam, K. T., & Ingram, L. O. (1999). Chromosomal integration of heterologous DNA in Escherichia coli with precise removal of markers and replicons used during construction. *Journal of Bacteriology*, *181*(22), 7143–7148. https://doi.org/10.1128/jb.181.22.7143-7148.1999

Mendiola, M. V., Bernales, I., & De La Cruz, F. (1994). Differential roles of the transposon termini in IS91 transposition. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(5), 1922–1926. https://doi.org/10.1073/pnas.91.5.1922

Mignon, C., Sodoyer, R., & Werle, B. (2015). Antibiotic-free selection in biotherapeutics: Now and forever. *Pathogens*, *4*(2), 157–181. https://doi.org/10.3390/pathogens4020157

Nielsen, J., & Keasling, J. D. (2016). Engineering cellular metabolism. *Cell*, *164*(6), 1185–1197. https://doi.org/10.1016/j.cell.2016.02.004

Nyerges, Á., Csörgő, B., Nagy, I., Bálint, B., Bihari, P., Lázár, V., Apjok, G., Umenhoffer, K., Bogos, B., Pósfai, G., & Pál, C. (2016). A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(9), 2502–2507. https://doi.org/10.1073/pnas.1520040113

Peredelchuk, M. Y., & Bennett, G. N. (1997). A method for construction of E. coli strains with multiple DNA insertions in the chromosome. *Gene*, *187*(2), 231–238. https://doi.org/10.1016/S0378-1119(96)00760-3

Peubez, I., Chaudet, N., Mignon, C., Hild, G., Husson, S., Courtois, V., De Luca, K., Speck, D., & Sodoyer, R. (2010). Antibiotic-free selection in E. coli: new considerations for optimal design and improved production. *Microbial Cell Factories*, *9*(1), 65. https://doi.org/10.1186/1475-2859-9-65

Piégu, B., Bire, S., Arensburger, P., & Bigot, Y. (2015). A survey of transposable element classification systems - A call for a fundamental update to meet the challenge of their diversity and complexity. *Molecular Phylogenetics and Evolution*, *86*, 90–109. https://doi.org/10.1016/j.ympev.2015.03.009

Pósfai, G., Kolisnychenko, V., Bereczki, Z., & Blattner, F. R. (1999). Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome. *Nucleic Acids Research*, *27*(22), 4409–4415. https://doi.org/10.1093/nar/27.22.4409

Pósfai, G., Plunkett, G. 3rd, Fehér, T., Frisch, D., Keil, G. M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S. S., de Arruda, M., Burland, V., Harcum, S. W., & Blattner, F. R. (2006). Emergent properties of reduced-genome Escherichia coli. *Science (New York, N.Y.)*, *312*(5776), 1044–1046. https://doi.org/10.1126/science.1126439

Quan, J., & Tian, J. (2009). Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE*, *4*(7). https://doi.org/10.1371/journal.pone.0006441

Reisch, C. R., & Prather, K. L. J. (2015). The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli. *Scientific Reports*, *5*(September), 1–12. https://doi.org/10.1038/srep15096

Reznikoff, W. S. (2003). Tn5 as a model for understanding DNA transposition. *Molecular Microbiology*, *47*(5), 1199–1206. https://doi.org/10.1046/j.1365-2958.2003.03382.x

Roberts, A. P., Chandler, M., Courvalin, P., Guédon, G., Mullany, P., Pembroke, T., Rood, J. I., Jeffery Smith, C., Summers, A. O., Tsuda, M., & Berg, D. E. (2008). Revised nomenclature for transposable genetic elements. *Plasmid*, *60*(3), 167–173. https://doi.org/10.1016/j.plasmid.2008.08.001

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, *5*(APR), 1–17. https://doi.org/10.3389/fmicb.2014.00172

Ross-Macdonald, P., Coelho, P. S. R., Roemer, T., Agarwal, S., Kumar, A., Jansen, R., Cheung, K. H., Sheehan, A., Symonlatis, D., Umansky, L., Heldtman, M., Nelson, F. K., Iwasaki, H., Hager, K., Gerstein, M., Miller, P., Roeder, G. S., & Snyder, M. (1999). Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature*, *402*(6760), 413–418. https://doi.org/10.1038/46558

Rutherford, K., Yuan, P., Perry, K., Sharp, R., & Van Duyne, G. D. (2013). Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Research*, *41*(17), 8341–8356. https://doi.org/10.1093/nar/gkt580

Sambrook, J. (n.d.). *Molecular cloning : a laboratory manual*. Third edition. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, [2001] ©2001. https://search.library.wisc.edu/catalog/999897924602121

Seitz, T., Berger, B., Nguyen, V. T., Tricot, C., Villeret, V., Schmid, S., Stalon, V., & Haas, D. (2000). Linker insertion mutagenesis based on IS21 transposition: Isolation of an AMP-insensitive variant of catabolic ornithine carbamoyltransferase from Pseudomonas aeruginosa. *Protein Engineering*, *13*(5), 329–337. https://doi.org/10.1093/protein/13.5.329

Sibley, M. H., & Raleigh, E. A. (2012). A versatile element for gene addition in bacterial chromosomes. *Nucleic Acids Res.*, *40*, e19.

Siguier, P., Gourbeyre, E., Varani, A., Ton-Hoang, B., & Chandler, M. (2015). Everyman's guide to bacterial insertion sequences. *Mobile DNA III*, 555–590. https://doi.org/10.1128/9781555819217.ch26

Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology*, *147*(1), 195–197. https://doi.org/https://doi.org/10.1016/0022-2836(81)90087-5

St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., & Shearwin, K. E. (2013). One-Step Cloning and Chromosomal Integration of DNA. *ACS Synthetic Biology*, *2*(9), 537–541. https://doi.org/10.1021/sb400021j

Summers, D. K., & Sherratt, D. J. (1984). Multimerization of high copy number plasmids causes instability: Cole 1 encodes a determinant essential for plasmid monomerization and stability. *Cell*, *36*(4), 1097–1103. https://doi.org/10.1016/0092-8674(84)90060-6

Szabó, M., Müller, F., Kiss, J., Balduf, C., Strähle, U., & Olasz, F. (2003). Transposition and targeting of the prokaryotic mobile element IS30 in zebrafish. *FEBS Letters*, *550*(1–3), 46–50. https://doi.org/10.1016/S0014-5793(03)00814-7

Szybalski, W., & Bryson, V. (1952). Genetic Studies on Microbial Cross Resistance To Toxic Agents I. ,. *Journal of Bacteriology*, *64*(4), 489–499. https://doi.org/10.1128/jb.64.4.489-

499.1952

Thakker, C., Lin, K., Martini-Stoica, H., & Bennett, G. N. (2016). Use of transposase and ends of IS608 enables precise and scarless genome modification for modulating gene expression and metabolic engineering applications in Escherichia coli. *Biotechnology Journal*, *11*(1), 80–90. https://doi.org/10.1002/biot.201500205

Touchon, M., & Rocha, E. P. C. (2007). Causes of insertion sequences abundance in prokaryotic genomes. *Molecular Biology and Evolution*, *24*(4), 969–981. https://doi.org/10.1093/molbev/msm014

Tyo, K. E. J., Ajikumar, P. K., & Stephanopoulos, G. (2009). Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nature Biotechnology*, *27*(8), 760–765. https://doi.org/10.1038/nbt.1555

Umenhoffer, K., Draskovits, G., Nyerges, Á., Karcagi, I., Bogos, B., Tímár, E., Csörgö, B., Herczeg, R., Nagy, I., Fehér, T., Pál, C., & Pósfai, G. (2017). Genome-Wide Abolishment of Mobile Genetic Elements Using Genome Shuffling and CRISPR/Cas-Assisted MAGE Allows the Efficient Stabilization of a Bacterial Chassis. *ACS Synthetic Biology*, *6*(8), 1471–1483. https://doi.org/10.1021/acssynbio.6b00378

Vandecraen, J., Chandler, M., Aertsen, A., & Van Houdt, R. (2017). The impact of insertion sequences on bacterial genome plasticity and adaptability. *Critical Reviews in Microbiology*, *43*(6), 709–730. https://doi.org/10.1080/1040841X.2017.1303661

Xu, P., Vansiri, A., Bhan, N., & Koffas, M. A. G. (2012). EPathBrick: A synthetic biology platform for engineering metabolic pathways in E. coli. *ACS Synthetic Biology*, *1*(7), 256–266. https://doi.org/10.1021/sb300016b

Yin, J., Wang, H., Fu, X. Z., Gao, X., Wu, Q., & Chen, G. Q. (2015). Effects of chromosomal gene copy number and locations on polyhydroxyalkanoate synthesis by Escherichia coli and Halomonas sp. *Applied Microbiology and Biotechnology*, *99*(13), 5523–5534. https://doi.org/10.1007/s00253-015-6510-8

Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., & Court, D. L. (2000). An efficient recombination system for chromosome engineering in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(11), 5978–5983. https://doi.org/10.1073/pnas.100127597

Zhu, H., He, C. C., & Chu, Q. H. (2011). Inhibition of quorum sensing in Chromobacterium violaceum by pigments extracted from Auricularia auricular. *Letters in Applied Microbiology*, *52*(3), 269–274. https://doi.org/10.1111/j.1472-765X.2010.02993.x

Number of peer-reviewed scientific publications: 3

Number of citations: 35

H-index: 2

i10 Index: 1

Total impact factor: 10.23

MTMT identification number: 10082382

11. List of publications:

1. **Shukla, R.D**[#].; Zvara, Á.; Avramucz, Á.; Biketova, A.Yu.; Nyerges, A.; Puskás, L.G.; Fehér, T. inPOSE: A Flexible Toolbox for Chromosomal Cloning and Amplification of Bacterial Transgenes. Microorganisms 2022, 10, 236. https://doi.org/10.3390/microorganisms10020236. **(served as the basis of the current PhD dissertation).** IF: 3.84

2. Chauhan IS[#], **Shukla R**[#]**,** Krishna S, Sekhri S, Kaushik U, Baby S, Pal C, Siddiqi MI, Sundar S, Singh N. Recombinant Leishmania Rab6 (rLdRab6) is recognized by sera from visceral leishmaniasis patients. Exp Parasitol. 2016 Nov; 170:135-147. https://doi.org/10.1016/j.exppara.2016.09.010. (**[#]equal first authors**). IF: 2.011

3. Møller-Olsen, C., Ho, S.F.S., **Shukla, R.D**.[#] *et al.* Engineered K1F bacteriophages kill intracellular *Escherichia coli* K1 in human epithelial cells. *Sci Rep* 8, 17559 (2018). https://doi.org/10.1038/s41598-018-35859-6. IF: 4.379

12. Summary

Plasmids are the most common tools used to clone and express a gene of interest in bacteria. Plasmid-based gene expression, however, has multiple drawbacks: as the bacterial cells divide there is variation in the copy number of the plasmid received by the offspring which affects the expression level of the encoded protein. The random inheritance of the plasmids also promotes the segregation of mutants and the loss of the desired function. The maintenance of plasmids always requires selection, which causes antibiotic contamination of the final product and hence puts an economic burden on the industry due to the removal step. Multi-copy integration of a transgene into the bacterial chromosome could be a viable alternative to avoid these problems. Several techniques have been developed with such aims, however, very few projects have succeeded in achieving true gene amplification. In most cases, the prior insertion of a DNA sequence (the landing pad) into the bacterial genome is needed in multiple copies to allow the multi-copy integration of the transgene, which is tedious and time-consuming.

Insertion sequences (IS elements or ISes) are the simplest kind of transposable elements found in bacterial genomes. They are not essential for the survival of the host bacteria and are usually present in multiple types and copies within most strains of most species. Our research focused on developing novel methods for easier engineering of bacterial strains that permit stable and antibiotic-contamination-free production of bioactive compounds for industrial applications. We show that bacterial insertion sequences can be used as robust genome engineering tools for this purpose.

We first aimed to develop a system allowing the one-step integration of a marked IS*3* element into the genome of a bacterium devoid of the element, and then find the locus of integration. We constructed a conditionally replicative plasmid encoding an IS*3* element marked with a spectinomycin resistance gene (SpR), and a plasmid carrying the corresponding transposase gene in an inducible form. We demonstrated the integration of the marked IS*3* at two distinct loci in two different strains of *E. coli* using this system.

In the next step we demonstrated the integration of 1 kbp long selection markers and a 9 kbp long marked operon into IS elements residing on the bacterial genome using one-step λ-Red recombineering. To further extend the possibilities of recombineering, we combined it with concomitant or subsequent cleavage of the targeted IS elements using the CRISPR/Cas9 system. This permitted the integration of resistance genes (SpR, CmR) in two copies and of an unmarked gene (*gfp*) in a single copy into the *E. coli* chromosome. CRISPR/Cas-mediated

recombineering applying concomitant cleavage was also essential to integrate two copies of the marked five-gene operon into the genome in a single round.

Our next goal was to amplify the copy number of the chromosomally inserted resistance genes or the marked *vioABCDE* operon. We developed inducible IS*1* and IS*3* transposase expressing thermosensitive plasmids for this purpose. After transposase expression, we applied selection on antibiotic gradient plates to obtain bacterial colonies having multiple copies of the heterologous cargo. We monitored the copy-amplification process with droplet-digital PCR and verified the copy numbers of the amplified IS elements using Illumina sequencing.

Finally, we set up two different experimental systems to test the stability of the engineered strains carrying the *vioABCDE* operon in single or multiple copies on their chromosomes. In all experiments carried out in the lack of selection, the stability of the genomic co-integrants surpassed that of the strains carrying the same operon on multi-copy plasmids.

Overall we have shown that a selectable gene, an unselectable gene, and a marked multigene operon can be inserted into the bacterial in one or two copies by recombineering or CRISPR/Cas9-assisted recombineering. We show that a resistance gene, as well as a marked multigene operon, can be amplified to up to 12 functional copies in the genome by expressing the corresponding transposase. We call this strategy inPOSE, referring to transgene insertion and amplification by transposition. We recommend using inPOSE for the construction of bacterial strains designed for prolonged-expression of transgenes from their chromosomes.

Thesis points:

1. Insertion sequence elements IS*1* and IS*3* can both be used as target sequences (landing pads) to integrate an antibiotic resistance gene alone or together with a multi-gene metabolic operon into the *E. coli* chromosome by linear DNA-mediated recombineering.
2. Cleavage of the genomic target using the CRISPR/Cas system was essential for the genomic insertion of the selectable gene or operon in two copies in a single step of recombineering.
3. Cleavage of the genomic target using the CRISPR/Cas was also essential for the insertion of a non-selectable gene (*gfp*) in a single copy using recombineering.

4. The operon inserted into the genome could be copy-amplified *in vivo* by the expression of the transposase enzyme corresponding to the targeted IS element, reaching copy numbers of >10.
5. In the lack of antibiotic selection, the function of the five-gene metabolic operon was maintained for over 40 generations by the genome-engineered cells, as opposed to cells losing the operon within 10-20 generations if encoded on a plasmid.

13. Összefoglaló

A molekuláris biológia hajnala óta a plazmidok a transzgének bakteriális klónozásának legfőbb eszközei. Egyszerűségük és elterjedt voltuk ellenére azonban több ismert hátránnyal is rendelkeznek. Ilyen pl. a kópiaszámuk változó volta, amely a hordozó törzzsel végzett kísérletek reprodukálhatóságát csökkenti. A legtöbb plazmid fenntartása a sejtben állandó szelekciót igényel, és az utódsejteknek való random átadásuk révén a mutáns, funkcióvesztett plazmidok könnyen kizárolagossá válhatnak a sejtben. Ezen hátrányok egyik kikerülésének potenciális módja a transzgének kromoszómába ültetése lehet, ez azonban csak rövid konstrukciók egy-egy kópiában integrálása esetén rutinszerű. Több megoldás is született hosszabb DNS szakaszok beültetésére egy vagy több kópiában, ezek többsége azonban egy specifikus célszekvencia többpéldányos, előzetes beültetését igényelt*E.*

A bakteriális inszerciós szekvenciák (IS elemek) a természetben előforduló legegyszerűbb autonóm mobilis genetikai elemek. A legtöbb baktériumfaj legtöbb törzsében megtalálhatóak különböző típusaik, általában több kópiában is. Jelen munkánk fő célja az volt, hogy az IS elemek felhasználásával olyan bakteriális genommanipulációs eszközt fejlesszünk ki, mely segítségével gyorsan és hatékonyan ültethetünk heterológ géneket vagy operonokat a kromoszómába, azok stabil, szelekciót nem igénylő, hosszútávú expressziója céljából.

Első lépésként egy olyan eszközt alakítottunk ki, mely képes egy IS*3* elemet transzpozíció útján a bakteriális genomba ugratni. Ez egy spektinomycin-rezisztenciával jelölt IS*3*-at hordozó plazmidból, és egy ehhez tartozó transzpozázt kifejező plazmidból állt. Használatukkal demonstráltuk a jelölt IS*3* kétféle *Escherichia coli* törzsbe, két külön lókuszba történő beugrását. Ezzel a lépéssel a teljes genommanipulációs protokollunk olyan törzsekben is használhatóvá tehető, melyek eredetileg nem tartalmazták ezt az IS elemet.

A következő lépésünk a különböző szelektálható génkonstrukciók kromoszómális IS*3* elemekbe történő homológ rekombinációjából állt. Ennek során lineáris donor DNS-en kódolt antibiotikum-rezisztencia géneket, illetve egy ezekhez kapcsolt ötgénes operont (*vioABCDE*) ültettünk a kromoszómába, λ-Red rekombináz enzimek segítségével. A rekombináció hatékonyságát úgy igyekeztünk fokozni, hogy a célszekvenciát a CRISPR/Cas rendszer *in vivo* kifejezésével elvágtuk. Ez nemcsak lehetővé tette a szelektálható gének és operonok két kópiában való beültetését egyetlen lépésben, de a nem szelektálható gének, mint a *gfp* is sikeresen integrálhatóvá vált különböző IS-típusokba, egy másolatban.

Következő célunk a beültetett szelektálható konstrukciók kópiaszám-növelése volt a kromoszómán. Ehhez megszerkesztettük az IS*1*, illetve az IS*3* elem transzpozáz enzimét expresszáló hőérzékeny plazmidokat. A megfelelő transzpozáz enzim kifejezése után a baktérium-populációt antibiotikum-grádiens lemezre szélesztve szelektáltunk a heterológ konstrukciót emelt példányszámban hordozó törzsekr*E*. Az így kiválasztott vonalak genomjában digitális PCR segítségével becsültük meg jelölt IS elemek új kópiaszámát, amit a teljes genomok Illumina-szekvenálásával igazoltunk.

Végül két különböző kísérleti rendszert állítottunk fel annak céljára, hogy a genomban több kópiában hordozott *vioABCDE* operon expressziójának stabilitását vizsgáljuk. Mindkét rendszer használata során azt láttuk, hogy szelekció nélkül a genomi lókuszokról történő expresszió 40 generáció után is változatlanul fennál. Ezzel szemben szelekció hiányában a plazmidról történő kifejeződés szignifikánsan csökkent mértékét, vagy teljesen hiányzó voltát lehetett megfigyelni már 10-20 generáció után.

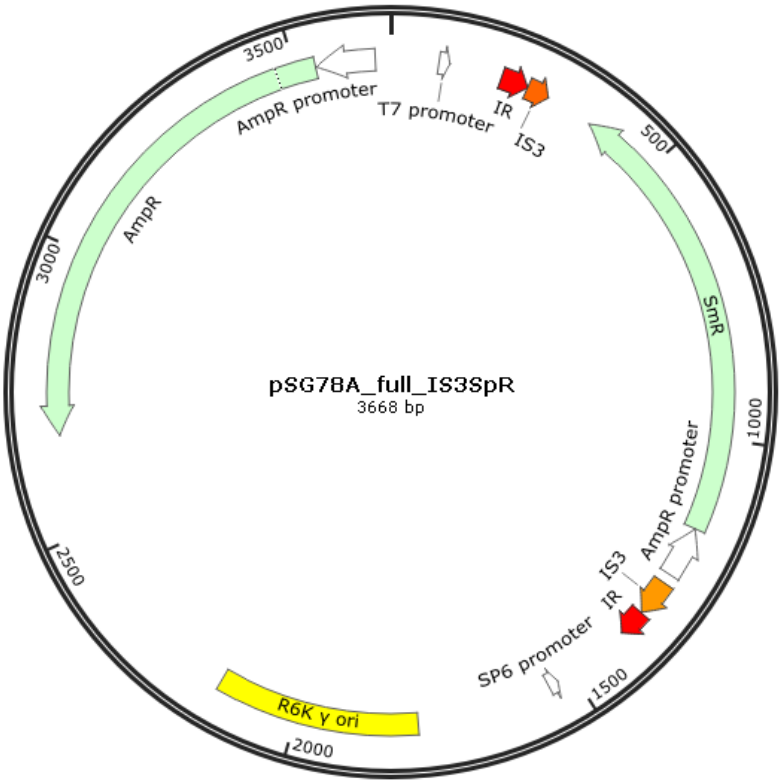
Összességében megmutattuk, hogy lineáris donor DNS felhasználásaval, λ-Red rekombinázok segítségével szelektálható és nem szelektálható gének ültethetők kromoszómális IS elemekbe. CRISPR/Cas-vágás segítségével már az integráció is történhet két kópiában, a megcélzott IS elemnek megfelelő transzpozáz kifejezésével pedig a transzgén kópiaszáma akár 12-re is növelhető. E stratégiát inPOSE-nak nevezzük, utalva az integrációs lépésre, és az utána következő, replikatív transzpozíció útján történő kópiaszám növelésre. Tapasztalataink alapján az inPOSE eszközrendszer használata megfontolandó minden olyan munkában, amely heterológ gének vagy operonok bakteriális kromoszómáról történő tartós kifejezését célozza meg.

Tézispontok:

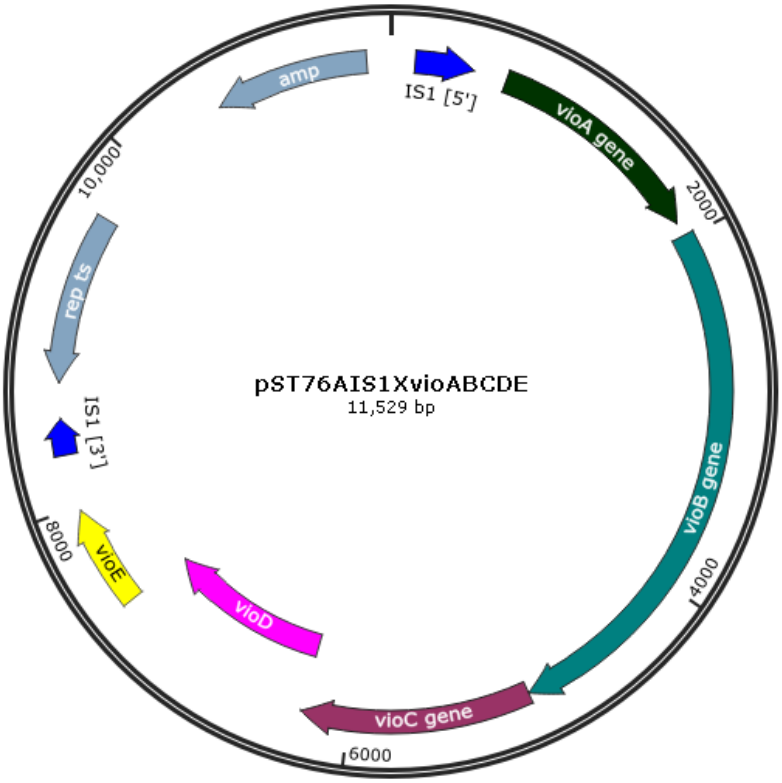
1. Az *Escherichia coli* IS*1* és IS*3* típusú inszerciós szekvenciája is rekombinációs célpontként használható különböző antibiotikum-rezisztencia gének önállóan, vagy egy ötgénes operonnal együtt való kromoszómába ültetésére, lineáris donor DNS felhasználásával.
2. A rekombinációs célpontok *in vivo* vágása CRISPR/Cas-al nélkülözhetetlen volt a szelektálható gén vagy operon két kópiában történő genomi-integrációjához.
3. A rekombinációs célpontok CRISPR/Cas-al történő *in vivo* vágása nélkülözhetetlen volt az egyébként nem szelektálható gének, mint a *gfp* egy példányban történő kromoszómába-illesztéséhez is.
4. A rezisztenica-génnel kapcsolt operon kópiaszáma akár 10 fölé is növelhető volt az integrációs célpontnak megfelelő transzpozáz enzim sejten belüli kifejezésével.
5. Antibiotikum-szelekció hiányában a kromoszómába ültetett ötgénes operon 40 generáció után is megőrizte funkcióját, szemben a plazmidon hordozott operonnal, amely 10-20 generáció között elveszett.

14. Appendix

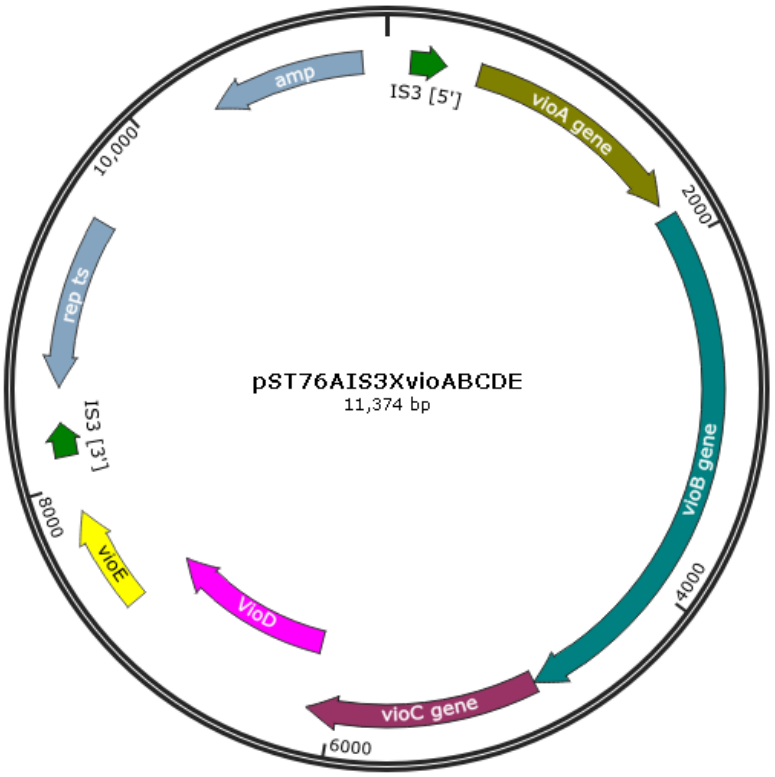
S1.



S2.



S3.


amp
IS3 [5']
vioA gene
2000
vioB gene
4000
vioC gene
6000
VioD
vioE
8000
IS3 [3']
rep ts
10,000
pST76AIS3XvioABCDE
11,374 bp