



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA E BIODIVERSIDADE

MAYNA DA SILVEIRA GOMIDE

**FUNCIONALIDADE DE SERINA INTEGRASES NO CONTROLE DE
INTERRUPTORES GENÉTICOS EM PLANTAS**

BRASÍLIA – DF
2020



Mayna da Silveira Gomide

**FUNCIONALIDADE DE SERINA INTEGRASES NO CONTROLE DE
INTERRUPTORES GENÉTICOS EM PLANTAS**

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia e Biodiversidade da Universidade de Brasília como parte dos requisitos para obtenção do título de Doutora em Ciências.

Orientador: Dr Elibio Leopoldo Rech Filho
Coorientadora: Profa. Dra Cíntia Marques Coelho

BRASÍLIA – DF
2020



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

FOLHA DE APROVAÇÃO

Mayna da Silveira Gomide

Funcionalidade de serina integrases no controle de interruptores genéticos em plantas

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia e Biodiversidade da Universidade de Brasília como parte dos requisitos para obtenção do título de Doutora em Ciências.

Aprovada em 19 de junho de 2020

Dr Elibio Leopoldo Rech Filho
Presidente - Embrapa Cenargen

Prof. Dr Fernando Araripe Gonçalves Torres
Membro 1 - Universidade de Brasília - UnB

Profa. Dra Danielle Biscaro Pedrolli
Membro 2 - Universidade Estadual Paulista – Unesp

Prof. Dr Robert Neil Gerard Miller
Membro 3 - Universidade de Brasília – UnB

À minha família,
meus pais, Denise e Miguel, meu irmão, Thales,
minhas tias, tio, primos, e, *in memoriam*, meus avós
e tio Sérgio, por compartilharem alegria e
entusiasmo a cada passo do meu caminhar.

AGRADECIMENTOS

Agradeço

Ao Elíbio a abertura à orientação do doutorado, os seus esforços para nos fornecer boas condições para a realização das pesquisas e os incentivos aos estudos sobre biologia sintética.

À Cíntia Coelho estar, desde a minha graduação, sempre presente contribuindo para meu amadurecimento profissional e para uma visão crítica do trabalho científico. Obrigada também pela proposta de realização desse doutorado em Brasília, pelas discussões constantes sobre experimentos e dados e pela preocupação e amizade em toda essa fase.

À Leila ter aberto as portas do seu laboratório e ter abraçado os desafios do presente estudo, contribuindo frequentemente nos ensaios e análises dos resultados. Ainda, obrigada por pensar junto tantas vezes em possíveis soluções para dificuldades encontradas ao longo do caminho.

À Thais, Lílian, Marco, Rayane e Mariana as trocas de experiências no LSB, as sugestões e outras importantes contribuições para esse trabalho, que reuniu no artigo publicado, todos esses esforços, incluindo incansáveis atividades de desenho de vetores, bancada, análise de dados e montagem do manuscrito. Obrigada também pela companhia em divertidos encontros que se estendem além da rotina no laboratório.

Aos também colegas do LSB, de hoje ou que por lá passaram, Raquel, Aline, Hermerson, Luis, Valquíria, Matheus, Tomás, Felipe, Karina, Letícia, Maria Luisa, Tati, Essayra, Saluana, Adriano o harmônico convívio diário e a constante prontidão em ajudar, o que contribuiu para tornar momentos de uma rotina muitas vezes cansativa, mais fáceis e leves.

Ao Cristiano e André a agradável companhia no laboratório, desde que cheguei ao LSB, o auxílio na execução de algumas técnicas, as discussões de possíveis rumos às pesquisas em andamento e o compartilhamento de materiais para vários testes realizados. Também agradeço a avaliação e revisão do manuscrito oriundo do presente estudo.

Ao Eduardo Melo a disponibilização de uso da sala de cultivo *in vitro* do LRA, a cessão de plasmídeos a partir dos quais foram feitas algumas sínteses de vetores usados em ensaios que compuseram o artigo e as revisões do manuscrito.

À Ana Cristina, Regivaldo, Andrielle, Natalia e Margot a disponibilização de equipamentos e importantes auxílios com as técnicas de microscopia e citometria de fluxo, que foram imprescindíveis para a análise dos resultados do presente trabalho.

Ao Martín, Luciana, Cintia Limia e outros membros do laboratório do Inca a colaboração nos ensaios com células humanas, incluindo as discussões acerca da análise dos resultados e estruturação do manuscrito, bem como suas revisões.

Ao Miguel o extenso trabalho na composição gráfica do artigo e a sua paciência por tantas idas e vindas entre acréscimos de dados e ajustes.

À Joseane Padilha o auxílio na análise estatística dos dados e as orientações para uso do programa R.

À UFJF a autorização concedida para me licenciar por quatro anos das atividades como técnica administrativa em educação para a realização desse doutorado. Em especial, agradeço à direção da Faculdade de Medina e à coordenação do NAPE a liberação.

Ao CNPq, CAPES, FAPDF, INCT Biosyn, Embrapa e UnB o financiamento e a estrutura que permitiram a execução dos ensaios necessários para a realização do presente trabalho e resultante publicação.

Aos professores e pesquisadores das disciplinas cursadas ao longo do doutorado e também aqueles que compuseram as bancas de qualificação e defesa de tese, a disponibilidade e o estímulo às trocas de conhecimentos, assim como as sugestões para a construção de um melhor trabalho.

Aos colegas do laboratório de Biologia sintética/Citogenética vegetal, do departamento de Genética e Morfologia da UnB, as enriquecedoras reuniões com discussões de artigos e metodologias científicas.

Aos colegas do LEM os diversos lanches (muitos cuscuzes), churrascos, conversas, cachaças e violões, que, sem dúvida, deixaram os dias pelo Cenargen ainda mais aprazíveis.

Aos colegas do LIMPP, LEG e outros laboratórios do Cenargen as rápidas conversas no café ou corredor para tirar dúvidas sobre técnicas ou simples bate-papos, e aquelas que por algumas vezes puderam se estender a bares e mesmo a viagens à Chapada, trazendo agradáveis momentos de troca e descontração.

À Nara dividir nos últimos quatro anos com muito respeito, zelo, carinho e paciência uma morada que se tornou mais que um espaço compartilhado, se tornou um lar, e com especiais raízes juiz-foranas. Também agradeço à Joana que, por um semestre, trouxe raízes gaúchas para muito bem somar à construção desse ambiente.

Aos amigos que Brasília me permitiu encontrar, Bia, Gustavo, Lara, Rian, Márcio, Lucas, Mônica, Karol a alegria das ocasiões partilhadas, onde a afinidade foi se fazendo presente, permitindo que, aos poucos, a sensação de pertencimento a esse lugar, até então estranho, fosse se estabelecendo.

Aos outros também queridos amigos a torcida e os desejos repletos de positividade, vindos ainda que de longe, nítido em seus olhos e mensagens, que por vezes reanimaram e fortaleceram a busca dos meus objetivos.

Ao Zé a sincera amizade que se perpetua por um amoroso companheirismo diário, que perpassa debates intensos e amenidades. Agradeço também ter sempre palavras que me ajudam a enxergar melhor as decisões compatíveis com as metas e valores que busco.

A Deus a fé que me permite, em muitas vezes, ressignificar diferentes acontecimentos para além das primeiras impressões humanas. Agradeço ter se feito presente entre nós na figura de Jesus para nos ensinar sobre o amor e, depois, ter ainda continuado nos trazendo pessoas para nos ajudar nesse difícil aprendizado sobre amar o próximo. Assim foram, por exemplo, Chiara Lubich e todos aqueles da comunidade dos Focolares, a quem agradeço compartilharem suas dores e vitórias, o que me ajuda a tentar conduzir aquilo que me proponho a fazer dentro dessa dimensão.

À minha família a base formadora. Indo para longe percebi com mais evidência o que isso significa. Entendi melhor que quem sou diante do mundo, e porque assim sou, se relaciona diretamente com o que aprendi durante toda uma vida com vocês (e sempre continuarei aprendendo). Estar nesse ponto hoje é reflexo desse aprendizado. Muito obrigada!

“Requer, de fato, alguma coragem para realizar um trabalho de tão ampla extensão; parece, no entanto, ser o único caminho certo para finalmente chegarmos à solução de uma questão cuja importância não pode ser superestimada em conexão com a história da evolução das formas orgânicas.”

(Gregor Mendel, 1866. Experiments in plant hybridization)

RESUMO

A biologia sintética tem sido caracterizada por abordagens interdisciplinares voltadas para o desenho e/ou reprogramação de sistemas biológicos. Fomenta a síntese de genomas completos ou reduzidos e o desenvolvimento de novas tecnologias de edição e de controle de expressão de genes que permitam o implemento de características desejadas. Desse contexto têm emergido os circuitos genéticos biológicos sintéticos que, baseados na álgebra binária dos circuitos eletrônicos, colocam partes genéticas sob um controle de entrada e saída (*input-output*) para a geração de interruptores (*switches*), portas lógicas e redes genéticas sintéticas. Um elemento chave para tais circuitos são as enzimas recombinases do grupo das serina integrases, que são capazes de reconhecer sítios *attB/attP* e inverter a sequência de DNA inserida entre eles de forma unidirecional. Assim, essas enzimas podem ser utilizadas como ferramentas para ativar e desativar a expressão gênica a partir da inversão de partes biológicas, como promotores, terminadores ou sequências codificadoras. Contudo, o número dessas proteínas completamente caracterizadas para aplicações em plantas é ainda reduzido. É, portanto, de interesse o incremento do número de integrases funcionais para permitir a construção de circuitos genéticos vegetais voltados para a aquisição de múltiplas características vantajosas. Desta forma, o objetivo deste estudo foi avaliar a funcionalidade de seis serina integrases (2, 4, 5, 7, 9 e 13) na ativação de interruptores genéticos unidirecionais em protoplastos de *Arabidopsis thaliana*. A construção dos interruptores envolveu desenho e síntese de um sistema de plasmídeos repórteres e plasmídeos para expressão das integrases. Os repórteres possuem a sequência codificadora do gene *gfp* ou a sequência promotora CaMV 35S invertidas e flanqueadas pelo sítio de reconhecimento das integrases. A saída do sistema é a fluorescência da proteína GFP. Ensaio baseado em cotransformação transiente foram realizados para validar a capacidade das serina integrases ativarem os interruptores pela inversão das partes genéticas. As células fluorescentes positivas resultantes foram avaliadas por microscopia e por citometria de fluxo. Para ativação da sequência codificadora de *gfp*, as integrases 13, 9 e 4 promoveram as maiores proporções de células GFP positivas. As já estabelecidas phiC31 e Bxb1, e a integrase 7 promoveram proporção intermediária e a integrase 2, baixa. Sob ação da integrase 5, não foi observada fluorescência. Para ativação do promotor, as integrases 2 e 4 levaram às maiores porcentagens de células com o repórter GFP. Análise molecular por PCR indicou que todas as integrases rotacionaram as partes genéticas, mesmo a integrase 5, que não apresentou resultado de ativação por medição da fluorescência do GFP. Análises por sequenciamento evidenciaram a formação dos sítios resultantes da recombinação, *attL/attR*, e a correta orientação do promotor e da sequência codificadora do *gfp* para todas as integrases. Ensaio de viabilidade celular mostrou que as integrases não têm atividade citotóxica. Os resultados obtidos demonstraram a funcionalidade das integrases testadas em um sistema de interruptor vegetal e podem embasar o desenvolvimento de circuitos genéticos sintéticos mais finamente regulados para controlar a expressão gênica em plantas.

PALAVRAS-CHAVE: Serina integrases. Interruptores genéticos. Circuitos genéticos sintéticos. Biologia sintética em plantas.

ABSTRACT

Synthetic biology has been characterized by interdisciplinary approaches aiming the design and/or reprogramming of biological systems. It promotes the synthesis of complete or reduced genomes and the development of new gene editing and gene expression control technologies that enable the implementation of desired characteristics. From this context synthetic biological genetic circuits have emerged that, based on the binary algebra of electronic circuits, place genetic parts under an input-output control for the generation of switches, logic gates and synthetic genetic networks. A key element in such circuits are the recombinases serine integrase enzymes, which are capable of recognizing *attB/attP* sites and unidirectionally reversing the DNA sequence inserted between them. Thus, these enzymes can be used as tools to activate and deactivate gene expression by inverting biological parts, such as promoters, terminators or coding sequences. However, the number of these fully characterized proteins for plant applications is still small. Therefore, it is of interest to increase the number of functional integrases to allow the construction of plant genetic circuits for the acquisition of multiple advantageous traits. Thus, the aim of this study was to evaluate the functionality of six serine integrases (2, 4, 5, 7, 9 and 13) in the activation of unidirectional genetic switches in *Arabidopsis thaliana* protoplasts. The construction of the switches involved design and synthesis of a system of reporter plasmids and integrase expression plasmids. Reporters have the *gfp* coding sequence (CDS) or the CaMV 35S promoter sequence inverted and flanked by the integrase recognition sites. The system output is GFP protein fluorescence. Transient cotransformation assays were performed to validate the ability of serine integrases to activate switches by inverting the genetic parts. The resulting positive fluorescent cells were evaluated by microscopy and flow cytometry. For *gfp* CDS activation, integrases 13, 9 and 4 promoted the highest proportions of GFP positive cells. The already established phiC31 and Bxb1, and integrase 7 promoted intermediate proportions, whilst integrase 2 resulted in lowest proportions of GFP positive cells. Under the action of integrase 5, no fluorescence was observed. For promoter activation, integrases 2 and 4 led to the highest percentages of cells with the GFP reporter. Molecular analysis by PCR indicated that all integrases rotated the genetic parts, including even integrase 5, which showed no activation result by measuring GFP fluorescence. Sequence analysis showed the formation of the resulting sites of recombination, *attL/attR*, and the correct orientation of the promoter and the *gfp* CDS for all integrases. Cell viability assays also showed that integrases are not cytotoxic. Taking together, the results obtained demonstrated the functionality of the tested integrases in a plant switch system and may support the development of finely tuned regulated synthetic genetic circuits to control gene expression in plants.

KEYWORDS: Serine integrases. Genetic switches. Synthetic genetic circuits. Plant synthetic biology.

LISTA DE FIGURAS – INTRODUÇÃO

Figura 1 – Quadro negro do físico Richard Feynman.....	16
Figura 2 – Esquema comparativo dos sistemas de edição gênica.....	23
Figura 3 – Portas lógicas com respectivos símbolos e tabelas verdade	29
Figura 4 – Desenho esquemático de uma porta lógica AND hipotética	30
Figura 5 – Reação de clivagem do DNA catalisada por tirosina ou serina recombinase.....	33
Figura 6 – Recombinação mediada por serina integrase	34
Figura 7 – Estrutura de uma serina integrase	35
Figura 8 – Mecanismo de recombinação das serina integrases.....	36
Figura 9 – Integração, excisão, translocação e inversão de sequências mediadas por serina integrase.....	37

LISTA DE FIGURAS – MANUSCRITO

Fig. 1 – Strategy overview of the eukaryotic genetic switch system	62
Fig. 2 – Functional characterization of the genetic switches in human cells.....	65
Fig. 3 – Functional characterization of the genetic switches in bovine cells	66
Fig. 4 – Functional characterization of the genetic switches in plant protoplasts.....	68
Fig. 5 – Promoter as a switchable genetic part in plant protoplasts.....	70
Fig. 6 – Int activity in primary human T lymphocytes from PBMCs isolated from three independent donors.....	72
Fig. 7 – Int activity in stem cells.....	73

LISTA DE FIGURAS – SUPLEMENTARES AO MANUSCRITO

Supplementary Figure 1 – Schematic representation of the vector sets synthesized for the eukaryotic genetic switches and the resulting activated vector	90
---	----

Supplementary Figure 2– <i>In silico</i> prediction of the nuclear localization signal (NLS) for the Int coding sequences	91
Supplementary Figure 3 – Representative EGFP fluorescence images of the three model eukaryotic cell systems cotransfected/cotransformed with integrase expression (pIE) and switch GFP or promoter (pSG or pSP) vectors	92
Supplementary Figure 4 – EGFP fluorescence determined by flow cytometry analysis of HEK 293T cells.....	93
Supplementary Figure 5 – EGFP fluorescence determined by flow cytometry analysis of bovine fibroblasts	95
Supplementary Figure 6 – HEK 293T cells uncropped PCR gel images.....	97
Supplementary Figure 7 – Representative sequence reads showing the <i>attL</i> and <i>attR</i> sites obtained after Int activity in HEK 293T cells compared to the predicted sequences (grey highlighted)	98
Supplementary Figure 8 – Bovine fibroblasts uncropped PCR gel images	99
Supplementary Figure 9 – Representative sequence reads showing the <i>attL</i> and <i>attR</i> sites obtained after Int activity in bovine fibroblasts compared to the predicted sequences (grey highlighted)	100
Supplementary Figure 10 – EGFP fluorescence determined by flow cytometry analysis of <i>A. thaliana</i> protoplasts	101
Supplementary Figure 11 – EGFP-expressing cell percentages of the positive control pCaMV35S-GFP vector and pCaMV35S-GFP-pSG vector.....	103
Supplementary Figure 12 – <i>A. thaliana</i> protoplasts uncropped PCR gel images	104
Supplementary Figure 13 – Representative sequence reads showing the <i>attL</i> and <i>attR</i> sites obtained after Int activity in <i>A. thaliana</i> protoplasts compared to the predicted sequences (grey highlighted).....	105
Supplementary Figure 14 – Int 9 core-site differences compared to the original sequences described.....	106
Supplementary Figure 15 – EGFP fluorescence determined by flow cytometry analysis of <i>A. thaliana</i> protoplasts with the switch promoter system	107
Supplementary Figure 16 – <i>A. thaliana</i> protoplasts uncropped PCR gel images	108

Supplementary Figure 17 – Representative chromatograms showing the recognition site sequences obtained after Int activity in the switch promoter vector (pSP)	109
Supplementary Figure 18 – EGFP fluorescence determined by flow cytometry analysis of PBMCs	110
Supplementary Figure 19 – PBMCs uncropped PCR gel images	111
Supplementary Figure 20 – Representative sequence reads showing the <i>attL</i> and <i>attR</i> sites obtained after Int activity in PBMCs compared to the predicted sequences (grey highlighted).....	112

LISTA DE TABELAS

Tabela 1 – <i>Networks</i> , repositórios e <i>softwares</i> voltados para Biologia Sintética	18
Tabela 2 – Exemplos de funções biológicas das recombinases sítio-específicas.....	32
Supplementary Table 1 – Addgene accession numbers of all plasmids used in this study.....	113
Supplementary Table 2 – Summary of the statistical analysis of the EGFP-positive cell percentages obtained by flow cytometry assays with the switch GFP system in HEK 293T cells, bovine fibroblasts and plant protoplasts.....	114
Supplementary Table 3 – Number of PCR cloned sequences after Int activity on the switch GFP systems and observed covered mutations in HEK 293T, bovine fibroblast and plant protoplast cells.....	115
Supplementary Table 4 – Summary of the statistical analysis of the EGFP-positive cell percentages obtained by flow cytometry assays with the switch promoter system in plant protoplasts	116
Supplementary Table 5 – Number of PCR cloned sequences after Int activity on the switch promoter system and observed covered mutations in protoplasts.....	117
Supplementary Table 6 – Number of PCR cloned sequences after Int activity on the switch GFP system and observed covered mutations in PBMCs.....	118

Supplementary Table 7 – Identification, sequences and target systems of the primers used in this study to amplify *attL* and *attR* sites formed after Int activity in HEK 293T cells, bovine fibroblasts, plant protoplasts and PBMCs 119

SUMÁRIO

1 INTRODUÇÃO	15
1.1 BIOLOGIA SINTÉTICA	15
1.2 PRINCIPAIS ABORDAGENS QUE COMPÕEM A BIOLOGIA SINTÉTICA.....	15
1.2.1 Célula sintética, genomas mínimos e ferramentas de edição gênica	15
1.2.2 Portas lógicas e circuitos genéticos sintéticos	26
1.2.3 Recombinases serina integrases: estrutura, funcionamento e uso como ferramentas de circuitos genéticos sintéticos	31
1.3 BIOLOGIA SINTÉTICA EM PLANTAS.....	40
2 JUSTIFICATIVA	44
3 OBJETIVO.....	45
3.1 OBJETIVO GERAL	45
3.2 OBJETIVOS ESPECÍFICOS	45
4 DESENVOLVIMENTO.....	46
5 REFERÊNCIAS BIBLIOGRÁFICAS	47
6 MANUSCRITO	57
7 MATERIAL SUPLEMENTAR AO MANUSCRITO.....	89
8 CONSIDERAÇÕES FINAIS	131

1 INTRODUÇÃO

1.1 BIOLOGIA SINTÉTICA

A biologia sintética constitui uma área interdisciplinar que engloba ciências biológicas, da computação, biotecnologia, matemática, engenharias, física e química. As interseções acontecem para contribuir no desenho ou redesenho de organismos vivos, que passariam a executar aperfeiçoadas ou novas tarefas. A área está sendo formada tanto por estudos que buscam construir uma célula sintética, recriando todos seus elementos mínimos de funcionamento, como por trabalhos que objetivam sintetizar ou editar genomas e definir novas rotas de atuação ou processos biológicos. Os níveis de alterações podem ser múltiplos, o que acaba por vezes ampliando a denominação de área para biologia de sistemas. As proposições permitem expandir conhecimentos das ciências básicas, como por exemplo, entender essencialidade de genes ou grupos gênicos, e das ciências aplicadas, deletando ou inserindo características indesejadas ou desejadas, respectivamente, ou montando redes com respostas controladas, como nos circuitos eletrônicos. Essa evolução da tecnologia do DNA recombinante foi possível com o avanço técnico e barateamento da execução dos processos, mas muito do que se almeja encontra-se ainda em etapas iniciais.

1.2 PRINCIPAIS ABORDAGENS QUE COMPÕEM A BIOLOGIA SINTÉTICA

1.2.1 Célula sintética, genomas mínimos e ferramentas de edição gênica

O termo biologia sintética desperta a impressão de uma área que busca sintetizar a vida. No entanto, quanto mais o conhecimento biológico dos organismos avança, mais se reconhece sua diversidade e complexidade, dificultando o alcance desse objetivo. Ainda assim, a busca pela compreensão de formas e funcionamentos da vida tem levado a estudos que procuram compreender os processos biológicos essenciais que permitam o desenvolvimento de organismos sintéticos.

A reflexão deixada pelo físico e ganhador do prêmio Nobel, Richard Feynman, costuma ser empregada como justificativa desta busca. Em 1988, quando faleceu, em seu quadro negro na instituição onde lecionava, encontravam-se duas afirmações: "*What I cannot create, I do not understand.*", e "*Know how to solve every problem that has been solved*", ou seja, o que não posso criar, não posso entender; e, saiba como resolver todos os problemas que foram resolvidos (Figura 1). As afirmativas podem suscitar o questionamento: uma vez desvendados o funcionamento de genes e proteínas, de redes metabólicas e comunicações celulares, seria possível, então, recriá-los?

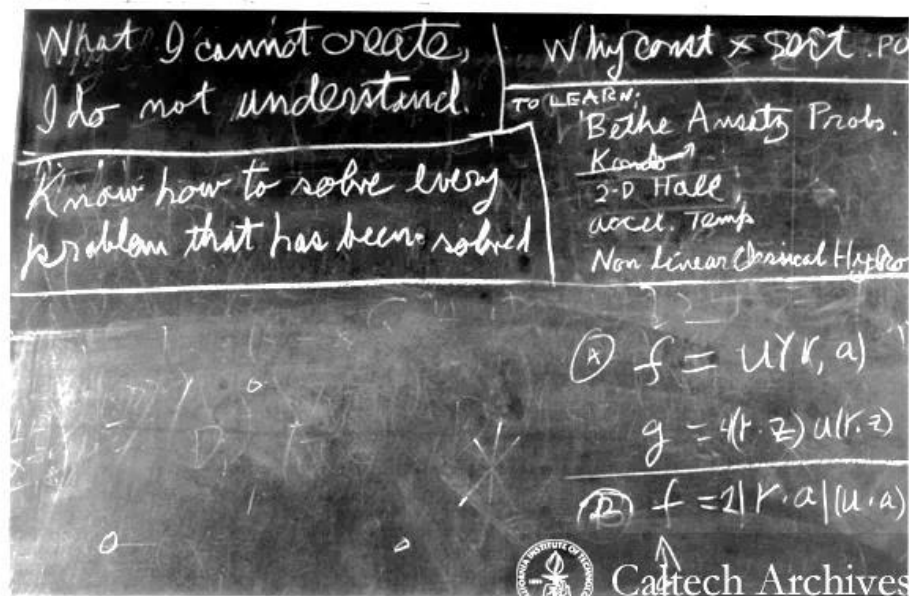


Figura 1 – Quadro negro do físico Richard Feynman. Anotações encontradas no quadro negro da Caltech, instituição onde Richard Feynman era professor, quando faleceu em 1988. As frases "*What I cannot create, I do not understand.*", e "*Know how to solve every problem that has been solved.*" provocam a reflexão se, de fato, para a compreensão da vida é necessário saber como recriá-la. (Fonte: archives-dc.library.caltech.edu/islandora/object/ct1%3A483, consultado em 15 de outubro de 2019).

Em 2010, Craig Venter e colaboradores desenvolveram a primeira célula comandada por um genoma totalmente sintético (GIBSON *et al.*, 2010). Esse estudo apresentou a síntese do genoma completo da bactéria *Mycoplasma mycoides*, com 1,08 Mb. O genoma sintético foi transplantado com sucesso para a bactéria recipiente *M. capricolum*, que, por sua vez, teve seu genoma próprio removido. Não houve modificações que resultassem em diferença fenotípica com relação ao genoma natural de *M. mycoides*. O organismo controlado completamente pelo

genoma transplantado foi denominado *JCVI-syn1.0*. Embora se trate apenas do genoma sintético e não de toda a célula, o trabalho representou importante avanço científico por consolidar progressos significativos já apresentados em estudos prévios desse grupo de pesquisa referentes a técnicas de sequenciamento de genomas, síntese de DNA, montagem de partes genéticas e transferência de grandes sequências de DNA entre organismos (FLEISCHMANN *et al.*, 1995; FRASER *et al.*, 1995; GIBSON *et al.*, 2008; VENTER *et al.*, 2001). Além deste tipo de trabalho, de um modo geral, o avanço das técnicas de biologia molecular desde o DNA recombinante e PCR, adicionalmente à publicação de genomas completos e à expansão e barateamento de sínteses e sequenciamentos de DNA, contribuíram diretamente para o desenvolver progressivo da biologia sintética.

O próximo passo foi dado na direção de aplicar essa estratégia de controle celular por genomas sintéticos a um organismo eucarioto. Um outro grupo formado por cientistas de diversas instituições iniciou um extenso projeto para a síntese dos 16 cromossomos da levedura *Saccharomyces cerevisiae*, formando o consórcio *Synthetic Yeast 2.0* ([Tabela 1](#)). Em 2014 foi feita a primeira publicação com a síntese total de um cromossomo funcional, o cromossomo III, chamado de *synIII* (ANNALURU *et al.*, 2014). Para algumas verificações e possíveis otimizações da técnica, algumas mudanças foram desenhadas na sequência a ser sintetizada, como a substituição dos códons de parada TAG por TAA e deleções de tRNAs, íntrons, regiões subteloméricas, elementos de transposição e dos *loci* sexuais silenciados *HML* e *HMR*. Outra alteração consistiu na inclusão de sítios *loxP*sym da enzima Cre recombinase flanqueando genes considerados não essenciais para a levedura, identificados com base nos dados disponíveis na literatura. A análise, nomeada de SCRaMbLE (do inglês, *synthetic chromosome rearrangement and modification by loxpsym-mediated evolution*), teve como intuito remover esses genes e avaliar possíveis efeitos desta considerada evolução induzida, bem como possibilitar reduções do genoma. A remoção dos genes previamente identificados como não essenciais, no entanto, foi letal para a levedura.

A síntese de outros cinco cromossomos da levedura já foi também concluída e publicada em edição especial sobre o projeto no periódico *Science*, em março de 2017 (MITCHELL *et al.*, 2017; SHEN *et al.*, 2017; WU, Y. *et al.*, 2017; XIE *et al.*, 2017; ZHANG, W. *et al.*, 2017). O projeto Sc2.0 ainda segue em andamento,

Tabela 1. *Networks*, repositórios e *softwares* voltados para Biologia Sintética.

Plataforma	Descrição	Link/Publicação
Synthetic Yeast 2.0	Consórcio internacional de pesquisadores para a síntese dos 16 cromossomos da levedura <i>Saccharomyces cerevisiae</i> .	www.syntheticyeast.org
GP-write	<i>Genome Project-write</i> - projeto formado por pesquisadores de diferentes áreas e instituições com o objetivo de viabilizar formas mais eficientes e de menor custo para a síntese e teste de genomas sintéticos grandes.	engineeringbiologycenter.org
Build-A-Cell	Busca a formação de uma comunidade científica para a colaboração em projetos de construção de células sintéticas.	buildacell.io
BaSyC	Grupo formado por vários laboratórios da Holanda com o intuito de construir uma célula sintética com abordagem <i>bottom-up</i> .	basyc.nl
iGEM	Fundação que promove competição mundial entre estudantes de graduação e pós-graduação para projetos aplicados de biologia sintética. Também compuseram um importante repositório público de partes genéticas dispostas em plasmídeos padronizados.	parts.igem.org
BioBricks	Fundação para benefício público que divulga e conecta iniciativas de biologia sintética, como plataformas e programas. Também tem sido responsável pelas últimas edições da conferência de biologia sintética SBX.0.	biobricks.org
Open WetWare	Plataforma para compartilhamento de informações entre grupos/laboratórios.	openwetware.org
Cello	Programa on-line para desenho de circuitos biológicos sintéticos formados por portas lógicas booleanas.	cellocad.org DOI: 10.1126/science.aac7341
SBOL	Plataforma on-line com símbolos padronizados para construção de circuitos genéticos <i>in silico</i> .	sbolstandard.org
SynBioBeta	Rede que busca integrar pesquisadores, empreendedores e investidores interessados em pesquisas e aplicações biológicas, além de publicar boletins informativos e organizar encontros periódicos.	synbiobeta.com
INCT-BioSyn	Instituto Nacional de Ciência e Tecnologia que reúne pesquisadores de diferentes instituições brasileiras com projetos voltados para o desenvolvimento da Biologia sintética no país.	inctbiosyn.com
EBRC	Grupo formado por membros de iniciativas públicas e privadas que busca roteirizar objetivos de interesse na área de engenharia biológica e articular com programas de pesquisas afins.	ebrc.org

buscando aperfeiçoar formas de verificação e solução de problemas (LIN *et al.*, 2019). A iniciativa estimulou também a formação de um outro grupo de pesquisadores interessados em escrever genomas sintéticos de diferentes organismos eucariotos mais complexos, como linhagens celulares humanas, plantas e outros organismos de interesse. O projeto foi denominado *GP-write* (Tabela 1), com especial denominação para o ramo humano, chamado *HGP-write* (*Human Genome Project-write*).

Em prosseguimento aos estudos de síntese de genomas bacterianos, o grupo de Craig Venter desenvolveu também uma bactéria controlada por um genoma sintético mínimo. Por conhecimentos prévios envolvendo genômica comparativa e testes de mutagênese de transposon, selecionaram genes considerados não essenciais para serem excluídos do genoma de *M. mycoides*. Assim como no estudo de redução de genoma em levedura, a remoção dos genes selecionados foi letal para a bactéria. No entanto, com testes adicionais de mutagênese por transposon, o grupo selecionou um conjunto de genes denominados *quasi-essenciais* para restabelecimento no genoma, considerados não completamente imprescindíveis para a sobrevivência da bactéria, mas necessários para um crescimento robusto. O genoma mínimo passou então por mais três passos de reconstrução até que se chegou à *JCVI-syn3.0*, o organismo controlado pelo menor genoma capaz de se reproduzir de forma autônoma nas condições laboratoriais empregadas, com 531 kb e uma redução de 901 genes (da *JCVI-syn1.0*) para 473 (HUTCHISON *et al.*, 2016).

A síntese de genomas de microrganismos almejava atingir o genoma da bactéria modelo *Escherichia coli*, amplamente utilizada na pesquisa. Seu genoma, muito maior (aproximadamente 4,5 Mb) que o dos micoplasmas, representava um desafio para a síntese utilizando técnicas como a do transplante do genoma completo sintetizado para uma bactéria recipiente. Recentemente, o genoma sintético desse organismo foi desenvolvido com uma técnica de substituição do genoma natural (uma linhagem com 4 Mb) pelo sintético, por partes, até o completo sintético ser formado (FREDENS *et al.*, 2019). Além da síntese, esse trabalho também implementou recodificações no genoma. Os códons de serina TCG e TCA foram substituídos apenas pelos códons AGC e AGT e o códon de parada TAG, apenas por TAA nas sequências gênicas codificadoras. Essas alterações reduziram o número de códons codificadores de aminoácidos de 61 para 59 e reduziu um *stop* códon do genoma. Essa recodificação visou reduzir o número de RNAs

transportadores, bem como facilitar o desenvolvimento de futuras aplicações (FREDENS *et al.*, 2019).

Assim, os estudos de construções de genomas sintéticos possibilitam entendimentos da ciência básica relacionados a funcionalidades e essencialidade de genes, individualmente ou em grupos, podendo melhorar compreensões acerca do comportamento celular. Podem também ser trabalhados como uma ferramenta que permita a otimização de organismos para crescimento controlado e aplicações biotecnológicas.

Além da síntese de genomas, outras iniciativas vêm buscando avançar sobre processos que permitam a montagem de uma célula sintética elementar, como um esboço, constituindo a abordagem *bottom-up*. Esses trabalhos têm essencialmente tratado de avanços voltados para compartimentalização e metabolismo. A capacidade de formar membranas lipídicas estáveis tem avançado com o desenvolvimento de técnicas da microfluídica na produção de vesículas unilamelares pequenas, médias ou grandes, além da integração de proteínas que interferem na conformação de membrana e proteínas de citoesqueleto e transmembrana, incluindo ATP sintase (DESHPANDE *et al.*, 2016; LITSCHER *et al.*, 2018; WEISS *et al.*, 2018). Vislumbra-se ainda a compartimentalização de vias sintéticas, como por exemplo, a que permite a fixação de CO₂ induzida por luz solar e que utiliza membranas tilacoides naturais para formar uma estrutura sintética semelhante a um cloroplasto (MILLER *et al.*, 2020; SCHWANDER *et al.*, 2016). Algumas iniciativas voltadas para a construção de células sintéticas estão apresentadas na [Tabela 1](#).

Assim como desde a proposição das teorias de Oparin e Haldane para a origem da vida, na década de 1920, acompanhadas do posterior experimento de Miller e Urey, em 1953, não se avançou substancialmente no entendimento de como a vida surgiu na Terra, a construção da célula sintética também se encontra apenas na combinação rudimentar de alguns elementos primordiais, estando ainda distante da formação de uma unidade estável e autônoma de controle, desenvolvimento e replicação. No entanto, a área da biologia sintética tem se desenhado não só sobre essa busca, mas também naquela que tem como intenção reprogramar ou editar os organismos. De forma interdisciplinar entre áreas como biologia, ciência da computação e engenharias, essas alterações de alguns genes, produtos ou vias, seja para pesquisas básicas ou para aplicações, constituem outra vertente de trabalho, denominada *top-down*.

Recentemente, inúmeros estudos *top-down* baseiam-se na engenharia de edição gênica de organismos através do uso de nucleases capazes de clivar o DNA e da obtenção de alterações decorrentes dos mecanismos de reparo celular. Após a clivagem bifilamentar da molécula de DNA o reparo pode ocorrer pela via NHEJ (*non homologous end joining*), que une as extremidades cortadas gerando mutações do tipo INDEL (inserção ou deleção), ou pela HDR (*homology directed repair*), que faz uma recombinação utilizando uma sequência homóloga ou uma sequência de interesse flanqueada por braços de homologia como *template*.

O diferencial dessas metodologias está na capacidade de direcionamento dessas enzimas para o *locus* gênico desejado. Em 2001, o grupo de Carroll e Chandrasegaran criou uma proteína quimérica formada por três domínios de ligação ao DNA de nucleases dedo de zinco (*zinc finger nucleases* - ZFN) e um domínio de clivagem não específico de uma enzima de restrição, a FokI. Cada ZFN se liga a nucleotídeos alvo e são necessárias duas proteínas quiméricas posicionadas de forma espelhada, uma em cada fita, para o corte duplo da molécula de DNA ([Figura 2a](#); BIBIKOVA *et al.*, 2001). Apesar de esforços de bioinformática que avançaram sobre a possibilidade de desenhar ZFNs funcionais para alvos gênicos desejados (SANDER *et al.*, 2009; WRIGHT *et al.*, 2006), as proteínas de dedo de zinco ainda configuravam um ferramenta difícil para uso difundido. Além de exigir construção laboriosa, com uma complexa montagem que deveria ser feita para os dois lados da fita próximos ao local de clivagem, nem sempre as proteínas desenhadas *in silico* se mostravam funcionais *in vivo* (ISALAN, 2012).

Em 2009 foi decifrado como ocorre a ligação de proteínas de bactérias patogênicas de plantas do gênero *Xanthomonas* com as sequências de DNA alvo. Elas interferem na expressão de genes do hospedeiro contribuindo para a infecção bacteriana. Por esta ação são chamadas efetores tipo ativadores da transcrição ou, no inglês, *transcription activator-like effectors* (TAL effectors ou TALE). Seu domínio de ligação ao DNA é formado por iguais sequências repetidas de 34 aminoácidos, exceto por dois deles nas posições 12 e 13, cuja variação se relaciona com ligação a um nucleotídeo alvo específico da seguinte forma: NI – A; HD – C; NN – G; NG – T (BOCH *et al.*, 2009; MOSCOU; BOGDANOVE, 2009). A partir de então, seguindo a mesma lógica de construção das ZFNs, essas proteínas TALE tiveram seu domínio de ligação desenhado para alvos específicos do DNA, sendo fusionadas ao domínio de clivagem da nuclease FokI, formando a ferramenta de edição gênica TALEN

(TALE *nucleases*). O funcionamento também acontece com duas construções flanqueando o alvo de modo que os domínios de nucleasse da FokI formem um dímero funcional para que a clivagem da dupla fita ocorra ([Figura 2b](#); CHRISTIAN *et al.*, 2010). O trabalho para a montagem do TALEN era também muito dispendioso, mas técnicas com boa efetividade foram desenvolvidas para a construção de vetores, o que conseguiu estabelecer o TALEN como uma ferramenta eficiente de edição de genes (CERMAK *et al.*, 2011). Aplicações foram desenvolvidas baseadas nessa ferramenta, como por exemplo, a edição em bovinos leiteiros pela inserção de um alelo variante do gene H, alelo PC (observado em gado sem chifre), fazendo com que esses nascessem sem chifres, evitando o sacrificante processo de descorna (CARLSON *et al.*, 2016).

Outros trabalhos utilizando TALEN continuaram sendo desenvolvidos, chegando-se inclusive a kits para inserção de genes via recombinação homóloga. No entanto, a iminente tecnologia de edição CRISPR/Cas9 expande-se rapidamente entre pesquisas e aplicações. Excetuando-se pela clivagem de um DNA alvo, as características de funcionamento diferem agora daquelas descritas para ZFNs e TALEN, sendo oriundas de um sistema de defesa de procariotos. Em 1987 um trabalho registrou o primeiro relato, em *E. coli*, da presença de sequências de 32 nucleotídeos repetidas de forma regular com outras sequências pequenas espaçando essas repetições (ISHINO *et al.*, 1987). Posteriormente, outros trabalhos identificaram clusters com esse mesmo perfil no genoma de diversas bactérias e também em archeas (MOJICA *et al.*, 1995; VAN BELKUM *et al.*, 1998). Esse *locus* foi chamado de CRISPR (do inglês, *clustered regularly interspaced short palindromic repeats*, ou, repetições palindrômicas curtas agrupadas e regularmente espaçadas) e também foram identificados genes codificadores de proteínas associadas a esse cluster (*CRISPR associated genes*, ou genes *cas*) (JANSEN *et al.*, 2002). Com a identificação adicional de que as sequências espaçadoras do cluster CRISPR eram oriundas de plasmídeos ou bacteriófagos, foi sugerida a função de defesa de procariotos contra vírus (BOLOTIN *et al.*, 2005; MOJICA *et al.*, 2005; POURCEL; SALVIGNOL; VERGNAUD, 2005). Assim, nos anos subsequentes compreendeu-se que quando um bacteriófago infecta uma bactéria introduzindo seu genoma, proteínas Cas processam esse material genético e inserem um fragmento na região identificada como espaçadora no cluster CRISPR. Quando a bactéria com essa porção de genoma incorporado é infectada novamente por vírus de mesma espécie,

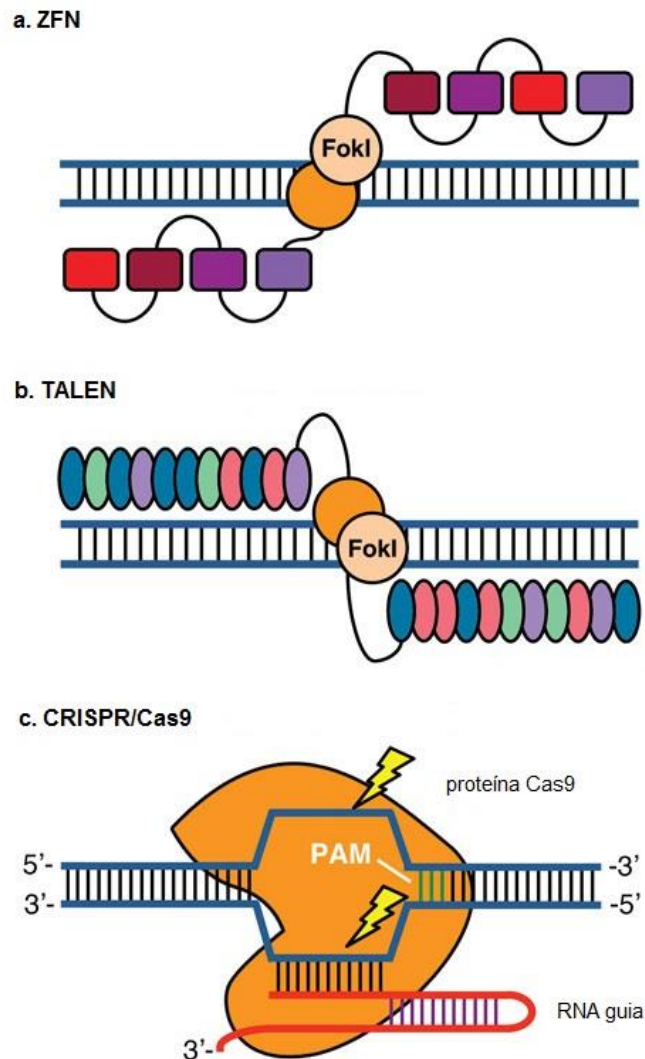


Figura 2 – Esquema comparativo dos sistemas de edição gênica. **a.** Ligação ao DNA por domínios dedo de zinco fusionados à enzima FokI, que após dimerização funcional, cliva o DNA. O híbrido proteico forma as nucleases dedo de zinco (ZFNs). **b.** Efetores TAL fusionados à enzima FokI ligados a alvo gênico específico para clivagem do duplex de DNA. Formam a ferramenta TALEN **c.** Proteína Cas9 ancorada a um RNA guia que por complementariedade de sequência e encontro dos três nucleotídeos do PAM, liga o complexo ao alvo gênico, onde acontece clivagem do DNA. Cas9 e guia compõem o sistema CRISPR/Cas9. (Adaptada de genengnews.com/magazine/219/comparing-genome-editing-technologies/, consultado em 22 de outubro de 2019).

esse *locus* é transcrito e o RNA processado em crRNA, tendo também um tracrRNA (naquelas com sistema de defesa tipo II) a ele complementar na região referente à sequência repetida. Juntos, ambos os RNAs, formam um guia para a proteína Cas9 (sistema tipo II) até a sequência complementar do genoma daquele vírus invasor. A Cas9 ancora em parte da estrutura do RNA guia e quando a região complementar encontra o alvo e o motivo adjacente PAM, ela cliva a dupla fita do DNA (Figura 2c; DELTCHEVA *et al.*, 2011). Desta forma passou-se a vislumbrar o uso desse sistema como ferramenta biotecnológica para clivar alvos específicos do genoma

identificados em um RNA guia, gerando, portanto, uma nova ferramenta de edição de genes.

Em 2012, o trabalho conduzido por Emmanuelle Charpentier e Jennifer Doudna consolida o emprego do mecanismo de defesa bacteriano como uma ferramenta aplicada. Validaram *in vitro* a funcionalidade de um sistema onde um único RNA quimérico (junção do crRNA e tracrRNA) com sequência complementar a um DNA alvo (plasmidial ou oligonucleotídeo duplex) é desenhado e guia uma proteína Cas9, isolada de *Streptococcus pyogenes*. Uma vez localizado o alvo e o motivo PAM por complementariedade, este é clivado pela Cas9 (JINEK *et al.*, 2012). No ano seguinte, os grupos de Feng Zhang e George Church, de forma independente, demonstraram a funcionalidade deste sistema em células de camundongo e humano, com alvos em *loci* endógenos, validando a tecnologia em eucarioto, inclusive com reparo mediado por recombinação homóloga (CONG *et al.*, 2013; MALI *et al.*, 2013).

Por apresentar vantagens como construção mais simples, menor tamanho que ZFNs e TALEN, flexibilidade estrutural (pode-se ter uma Cas9 e desenhar diferentes RNAs guia, tendo-se variados ensaios) e exibir boa eficiência em diversos organismos testados, a ferramenta CRISPR/Cas9 difundiu-se rapidamente por uma ampla diversidade de estudos e aplicações biotecnológicas. Como exemplo, podem ser citados trabalhos com animais (COOPER *et al.*, 2018; KALDS *et al.*, 2019; RICROCH, 2019; WU, G.; BAZER, 2019) e plantas (RAZZAQ *et al.*, 2019; ZHANG, YINGXIAO *et al.*, 2019) editados e estudos relacionados a doenças, como aqueles envolvendo doenças de visão (LEE *et al.*, 2019) ou imunoterapias contra o câncer (LIU, J. *et al.*, 2019). Para plantas agrocliváveis, a ferramenta traz grande potencial para acelerar os processos de seleção convencional e aperfeiçoar o ganho de características relacionadas a resistências e aumento de produtividade. Muitos trabalhos em diferentes culturas já foram realizados, mas, de modo geral, ainda há problemas limitantes envolvendo o *delivery* do sistema CRISPR/Cas9 nas plantas de grande interesse comercial, além dos tempos de regeneração e estabelecimentos de organismos editados. Além disso, outro gargalo ainda em resolução são as legislações que regulamentam esses vegetais editados. Cada país tende a um entendimento legal, havendo, sobretudo, discussão de enquadramento como OGM (organismos geneticamente modificados) ou não. Um avanço universal nesses

marcos regulatórios poderia agilizar a disseminação desses organismos redesenhados com sistemas CRISPR (ZHANG, YI *et al.*, 2020).

A aplicação da ferramenta foi ainda expandida com a manipulação da proteína Cas9 para a desativação de seu domínio de clivagem, tornando-a apenas um elemento carreador de outras proteínas fusionadas ao alvo desejado, como ativadores ou repressores da transcrição (LAU; SUH, 2018), editores de base (YANG, B.; YANG; CHEN, 2019), e o sistema *prime editing*, com uma transcriptase reversa que insere pequenas modificações no gene alvo utilizando um molde de RNA (ANZALONE *et al.*, 2019). Já foram também desenvolvidos kits diagnósticos baseados no sistema CRISPR (GOOTENBERG *et al.*, 2018), incluindo o recém-aprovado pela agência reguladora americana FDA (*Food and Drug Administration*) para a detecção do vírus SARS-CoV-2, causador da COVID-19 (SHERLOCK BIOSCIENCES, 2020). Nesse caso a proteína Cas utilizada não é a Cas9 e sim outras como a Cas13 e a Cas12a, que são capazes de clivar moléculas de RNA alvo e permanecerem ativadas, clivando outras moléculas de RNA repórteres.

Apesar de todas essas possibilidades e de muito se vislumbrar com o uso da Cas9, há ainda adicionais pontos negativos a serem enfrentados, como as clivagens fora do gene alvo, os chamados *off-targets*, e possíveis efeitos citotóxicos, como demonstrado para alguns organismos. Esses problemas são pontos críticos para permitir seu uso em humanos, em especial a clivagem de sequência não alvo, tendo por isso gerado grande preocupação na comunidade científica mundial quando, no final de 2018, um pesquisador chinês anunciou ter editado embriões de duas meninas. Sendo assim, o desenvolvimento de ferramentas de controle de expressão gênica, eficientes e mais precisas, ainda é desejável.

As edições de vários genes são obtidas por estratégias multiplex, pelas quais se pretende atingir diferentes *locus* concomitantemente (CAMPA *et al.*, 2019). Além disso, vislumbram-se configurações ainda mais complexas de programação, podendo mesclar edições que levem à perda de função ou adição de genes, regulações da expressão gênica por ativações ou repressões nos diferentes níveis (transcricional, traducional, proteico), ou mesmo interferências em níveis metabólicos secundários (PURNICK; WEISS, 2009). Esse controle de organismos vivos tem sido colocado em paralelo ao controle de sistemas realizado por circuitos lógicos eletrônicos. Esse comparativo implica também em uma busca metodológica que se assemelha àquela empregada nas engenharias e ciências da computação e emerge

mais uma vez com interseções entre tais áreas e as ciências biológicas moleculares, posicionando a biologia sintética também como uma área que almeja essa engenharia de circuitos genéticos sintéticos voltados para a aplicação de novas ou reprogramadas redes genéticas funcionais em organismos vivos.

1.2.2 Portas lógicas e circuitos genéticos sintéticos

Em 1961, Jacob e Monod identificaram a regulação e os genes envolvidos na via de repressão do operon lac em bactéria (JACOB; MONOD, 1961). Com esses achados, traçaram alguns modelos teóricos de mecanismos de funcionamento enzimático, pontuando elementos que poderiam ser combinados em uma variedade do que já chamaram de circuitos, formados por relações de feedbacks ou de ativações e repressões entre regiões gênicas codificadoras e regulatórias (MONOD; JACOB, 1961). A partir dos achados e proposições de Jacob e Monod, outros pesquisadores começaram a pensar na trama de redes genéticas envolvidas nos metabolismos celulares e a encaixá-las em lógicas matemáticas e de circuitos eletrônicos. Essa formalização lógica da biologia teórica começa a ser desenhada baseada em álgebra Booleana (binária) para a compreensão de sistemas genéticos regulatórios. Foram desenvolvidas equações baseadas na presença ou ausência de variáveis (genes, reguladores, promotores) no tempo e em condições simuladas que permitiam projetar, via cálculos e matrizes, combinações necessárias em determinadas situações hipotéticas (KAUFFMAN, 1969; SUGITA, 1975; THOMAS, 1973). Já era explícita a intenção de compactar os sistemas de regulação gênica em uma forma mais objetiva para facilitar possíveis manipulações e eventuais ajustes de modelos após experimentações (THOMAS, 1973).

Novamente, com o avanço técnico da biologia molecular, chegando-se à tecnologia do DNA recombinante e emergência da engenharia genética, os primeiros testes experimentais de redes genéticas sintéticas foram realizados. Essas primeiras publicações, no início dos anos 2000, marcaram a fundação dos circuitos genéticos sintéticos. O *toggle switch*¹, demonstrado em *E. coli*, promove uma regulação gênica biestável e tem desenho semelhante ao *switch* natural do bacteriófago λ , que ocorre

¹ Um *switch*, termo correspondente a interruptor, pode ser considerado uma unidade efetora que detecta um sinal de entrada e promove uma resposta como saída. Pode ser aplicado aos mecanismos biológicos naturais ou desenhado para que, em conjuntos determinados, componham redes genéticas ou circuitos biológicos sintéticos.

para definição de entrada em ciclo lítico ou lisogênico (PTASHANE, 2004). O sistema foi construído com dois promotores repressíveis, de tal modo que o produto gênico de cada um reprime o promotor do outro. Uma sequência codificadora da proteína fluorescente GFP foi posicionada para uma expressão cistrônica após a CDS (do inglês, *coding sequence*) de um dos repressores. A alternância entre o estado de observação e ausência do sinal fluorescente foi alcançado com pulsos de indutor para cada promotor alternadamente (GARDNER; CANTOR; COLLINS, 2000). Seguindo mecanismo semelhante ao supracitado, Elowitz & Leibler (2000) criaram um sistema baseado em *loops* de *feedback*, mas com três repressores, onde o produto gênico de um promotor reprime o seguinte. O mecanismo foi baseado na regulação circadiana que ocorre em cianobactérias (ISHIURA *et al.*, 1998). Um cassete adicional para a expressão de GFP sob controle de um dos promotores repressíveis também compunha o sistema, que foi chamado de *repressilator*. Na presença de um dos indutores o sistema deveria apresentar picos oscilatórios de fluorescência. Apesar de terem sido observadas as variações temporais esperadas, houve ruído no sistema, como diferenças de tempo e intensidades nas manifestações desse parâmetro entre as bactérias testadas, sugerindo efeitos não previstos sobre o modelo planejado (ELOWITZ; LEIBLER, 2000).

Trabalhos posteriores buscaram reduzir esses ruídos aperfeiçoando os *toggle switches* e as redes de regulação oscilatórias com inclusão de *loops* de *feedback* positivos e negativos e maior detalhamento e aperfeiçoamento dos modelos matemáticos (ATKINSON *et al.*, 2003; GOH; KAHNG; CHO, 2008; STRICKER *et al.*, 2008). Apesar das previsões de funcionamento dos circuitos sintéticos envolvendo cálculos matemáticos, quando postos em prática, resultados diferentes do esperado ou com variações levavam à criação de hipóteses para os resultados observados, correções do sistema e novos testes. Essa lógica gerou certo padrão de *workflow* nos trabalhos envolvendo circuitos genéticos sintéticos, compreendendo então: desenho, construção ou síntese, teste, busca de resolução dos problemas observados (*debugging*), redesenho, nova construção e reentrada no processo com novos testes. No entanto, essa lógica metodológica pode ser lenta, além de muito laboriosa. Buscando-se facilitar e ampliar esses trabalhos algumas iniciativas criaram as bibliotecas de partes genéticas, com repositórios públicos compostos por partes como promotores, sequências codificadoras, terminadores, sítios de ligação ribossomal, dentre outras (CAMERON; BASHOR; COLLINS, 2014).

Esse uso de partes genéticas caracterizadas como peças a serem montadas em sistemas programados como jogos de montar do tipo “lego” fomentou a criação, por exemplo, da fundação iGEM (*International Genetically Engineered Machine*). Esse grupo desenvolveu uma competição que hoje envolve estudantes, principalmente de graduação, de todo mundo. Essa iniciativa originou também um importante repositório de partes encontrado na plataforma *on-line* do programa ([Tabela 1](#)). Tais partes registradas devem seguir o padrão de montagem Biobricks ([Tabela 1](#)), outra iniciativa que surge neste contexto e que apoia programas relacionados a desenvolvimento e compartilhamento de informações. Entre eles está a plataforma OpenWetWare, onde pesquisadores podem compartilhar protocolos e conhecimentos sobre seus achados ([Tabela 1](#)).

Os circuitos genéticos sintéticos avançaram em complexidade de desenho, controle e regulação ao serem organizados como portas lógicas, um princípio da álgebra Booleana binária amplamente empregado na engenharia eletrônica. Essas portas lógicas seguem as tabelas verdades onde são mostrados os valores de entrada (*inputs*) e de saída (*output*) do sistema, que serão 0 ou 1. Há padrões de portas lógicas que irão atender a diferentes funções, tais como YES (ou *buffer*), NOT (ou inversora), AND, NAND, OR, NOR, XOR e XNOR, cujas estruturas e símbolos podem ser visualizados na [Figura 3](#).

Para o desenho de um circuito biológico sintético, os componentes de entrada devem ser organizados atendendo à tabela verdade para condicionar exatamente determinada saída (MIYAMOTO *et al.*, 2013). Por exemplo, para uma porta lógica AND, só haverá uma saída positiva (1) se as duas entradas estiverem presentes (1 e 1). Transpondo para um exemplo composto por partes biológicas, pode-se ter a expressão de um gene (*output*) sob um promotor induzível cujo indutor precisa ser composto por duas porções proteicas. Assim, apenas com a expressão das duas partes (*inputs*), há formação do indutor e consequente ativação do gene, o *output* ([Figura 4](#); BROPHY; VOIGT, 2014). Para a composição dos circuitos, mais de uma porta lógica pode ser combinada, tornando o sistema mais complexo, sendo determinado por maior número de elementos modulados.

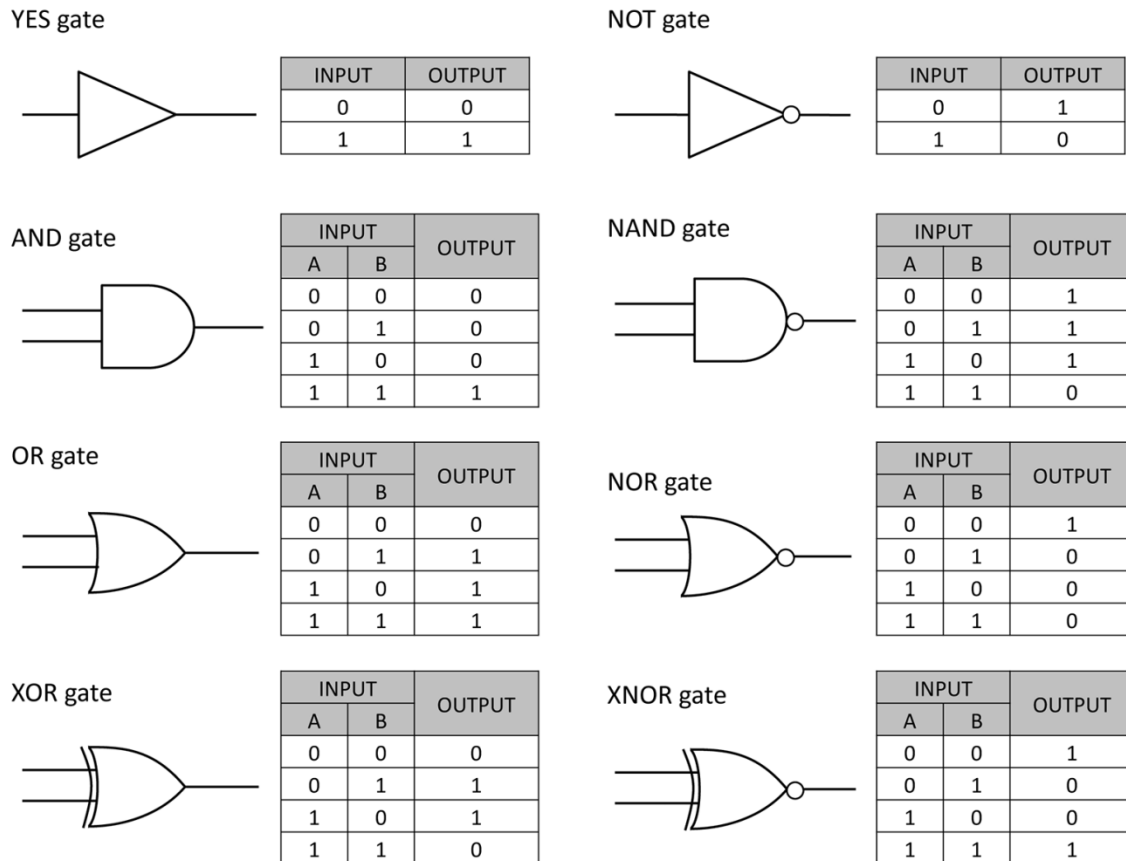


Figura 3 – Portas lógicas com respectivos símbolos e tabelas verdade. As tabelas indicam a ausência (0) ou presença (1) dos sinais de entrada (*input*) e saída (*output*) em cada condição. (Adaptada de Miyamoto *et al.*, 2013).

Uma série de trabalhos passou a utilizar esse sistema de circuitos compostos por portas lógicas, combinando sinais de entrada para a obtenção de determinados resultados. Muitas são as peças biológicas que têm sido combinadas para formar essas interações genéticas sintéticas, como aquelas organizadas nas bibliotecas de partes. Quanto mais partes são definidas e caracterizadas, maiores as possibilidades de combinações. Buscando aumentar o número de peças sensoras, ou seja, capazes de captar um *input*, o grupo de Christopher Voigt fez uma busca por repressores da família TetR em genomas de procariotos e 16 pares de repressores e correspondentes promotores foram selecionados (STANTON *et al.*, 2014). Posteriormente, o grupo expandiu as condições para desenhos de circuitos biológicos desenvolvendo um sistema operacional on-line chamado Cello ([Tabela 1](#)). Utilizando linguagem de descrição para circuitos eletrônicos, é possível informar as partes de composição e o tipo de circuito desejado. O programa monta a sequência de DNA, desenha o circuito baseado nos símbolos padronizados pela plataforma SBOL Visual ([Tabela 1](#)) e simula seu funcionamento. Utilizando o Cello, testaram 60

circuitos em *E. coli*, dos quais 45 responderam corretamente ao *output* esperado (NIELSEN, A. A. K. *et al.*, 2016). Fizeram também trabalhos de montagem de circuitos compostos por elementos do sistema CRISPR/Cas e de controles de vias metabólicas, ambos com interações com o genoma endógeno de *E. coli* (MOSER *et al.*, 2018; NIELSEN, A. A.; VOIGT, 2014).

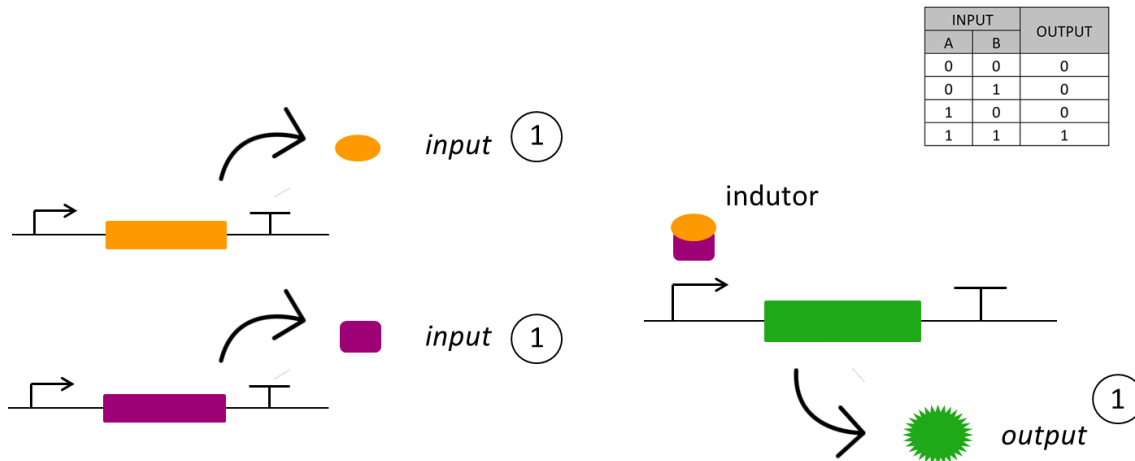


Figura 4 – Desenho esquemático de uma porta lógica AND hipotética. Duas proteínas caracterizam as entradas (*inputs* 1) para a formação de um indutor que ativa o promotor induzível de um terceiro cassete, levando à expressão de uma proteína fluorescente que caracteriza a saída (*output* 1) do sistema. (Adaptada de Brophy e Voigt, 2014).

O desenvolvimento de circuitos sintéticos que recebem sinais e emitem respostas planejadas logo passou a ser alvo também de trabalhos na área de saúde humana. Por exemplo, para a detecção específica de câncer de bexiga, Liu *et al.* (2014) desenvolveram uma porta lógica AND envolvendo CRISPR/Cas9. Os dois módulos de *input* são compostos por um promotor específico de câncer, que promove a expressão do RNA guia, e um promotor específico de tecido de bexiga, para expressão da Cas9. O *output* luciferase só é formado na presença das duas condições indutoras dos promotores das entradas. Também fizeram testes com os genes alvo hBAX (apoptose), p21 (inibição de crescimento) e E-caderina (baixa motilidade celular) como *outputs*, que efetivamente inibiram o crescimento de células tumorais da bexiga, em cultivo *in vitro*.

Além de elementos regulatórios como promotores induzíveis, proteínas recombinases do grupo das serina integrases também surgem como importante peça para ativar ou desativar partes genéticas. Posicionados os sítios de reconhecimento dessas enzimas flanqueando uma sequência de interesse, esta

pode ser rotacionada de modo a ligar ou desligar um sistema, podendo essas enzimas integrarem a composição de portas lógicas e circuitos genéticos sintéticos.

1.2.3 Recombinases serina integrases: estrutura, funcionamento e uso como ferramentas de circuitos genéticos sintéticos

Recombinases sítio-específicas são proteínas que realizam a recombinação de determinada sequência de DNA por um mecanismo de reconhecimento de sítio, que são sequências nucleotídicas curtas e específicas. Esse processo pode desencadear inserção, excisão ou inversão de DNA dependendo do arranjo inicial dos sítios. Foram descritas para funções biológicas em geral relacionadas a vírus, bactérias, cianobactérias e leveduras, tais como inserção ou excisão de genoma de fago ou de elementos de transposição; excisão para redução de dímero plasmidial ou cromossomal, ou para regulação da ativação gênica ligada a desenvolvimento; e inversão para alternância de expressão gênica, como por exemplo, alteração de fase fimbrial em *E.coli* ([Tabela 2](#); GRINDLEY; WHITESON; RICE, 2006).

Essas recombinases podem ser distinguidas em dois grupos ou famílias, das tirosina ou das serina recombinases, que apresentam diferenças evolutivas e estruturais. Uma das diferenças entre elas está relacionada ao resíduo do sítio catalítico, uma tirosina ou uma serina, responsável pela clivagem da molécula de DNA necessária para a recombinação, conforme reação química de transesterificação detalhada na [Figura 5](#) (GRINDLEY; WHITESON; RICE, 2006).

As serina recombinases compõem uma família mais heterogênea. Um conjunto de dados estruturais dessa família pode ser obtido no apanhado realizado por Stark (2014). A $\gamma\delta$ resolvase, envolvida na resolução de integrações de DNA em decorrência de transposições, foi a primeira melhor caracterizada (SANDERSON *et al.*, 1990; YANG, W.; STEITZ, 1995). O resíduo catalítico de serina está localizado na região N-terminal, comumente na posição 10, entretanto, há variação na organização dos domínios, principalmente naqueles da região C-terminal, sendo alguns membros maiores. Algumas recombinases de fagos temperados, ou seja, que inserem seu genoma no da bactéria hospedeira para entrada em ciclo lisogênico, pertencem a este grupo e são chamadas de grandes serina recombinases (*large serine recombinases*), ou apenas, serina integrases (SMITH; THORPE, 2002).

Tabela 2. Exemplos de funções biológicas das recombinases sítio-específicas

	Recombinase	Função biológica
Tirosina recombinases	λ Int e outras integrases de bacteriófagos	Integração e excisão do genoma de bacteriófagos
	Int de Tn916/Tn1545	Integração e excisão para transposição de transposons circulares
	IntI	Integração e excisão de cassetes de genes em integrons
	Cre	Excisão para redução de dímero plasmidial do bacteriófago P1
	XerC/D	Excisão para redução de dímero no cromossomo de <i>E. coli</i> e de muitos outros cromossomos bacterianos e de alguns plasmídeos
	TnpI de Tn4430	Excisão para resolução de cointegrados resultantes da transposição de Tn 4430
	FimB, Fim E	Inversão para alternância de expressão gênica (variação de fase fimbrial) em <i>E. coli</i>
	Rci de R64	Inversão de segmentos <i>shufflon</i> em plasmídeo R64, produzindo várias formas de pili
	XisA, XisC	Excisão para ativação de regulação gênica para desenvolvimento em <i>Anabaena</i>
	Flp	Inversão para amplificação de plasmídeo 2- μ m de levedura
Serina recombinases	TnpR de Tn3/ $\gamma\delta$ e transposons relacionados	Excisão para resolução de cointegrados resultantes de transposição
	Sin de <i>Staphylococcus aureus</i>	Excisão para redução de dímero em plasmídeos de estafilococos
	ParA de RP4	Excisão para redução de dímero em plasmídeo RP4
	Hin	Inversão para alternância de expressão gênica (variação de fase flagelar) em <i>Salmonella</i>
	Gin, Cin	Inversão para alternância de expressão gênica (proteínas de fibra da cauda) em bacteriófagos Mu e P1
	OrfA de IS607/IS1535	Integração e excisão para transposição do elemento IS607 de <i>Helicobacter pylori</i>
	Int de phiC31/Bbv1/phiRv1*	Integração e excisão de bacteriófagos, como os de <i>Streptomyces</i> e micobactérias
	TnpX de Tn4451*	Integração e excisão para transposição de Tn4451 em <i>Clostridium</i>
	SpoIVCA (CisA)*	Excisão para ativação de regulação gênica para desenvolvimento em <i>Bacillus subtilis</i>
XisF*	Excisão para ativação de regulação gênica para desenvolvimento em <i>Anabaena</i>	

* Membros da subfamília grandes serina recombinases

(Adaptada de Grindley, Whiteson e Rice 2006)

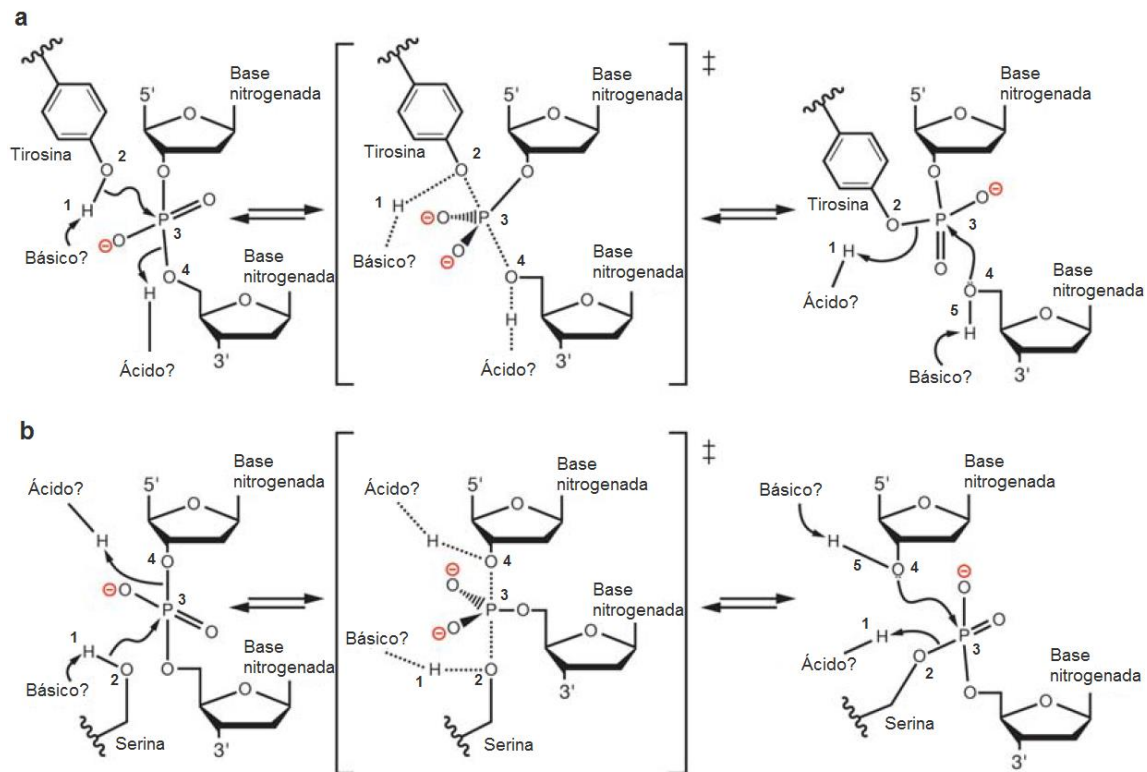


Figura 5 – Reação de clivagem do DNA catalisada por tirosina ou serina recombinase. Quando o DNA é posicionado no sítio catalítico, a enzima cria um microambiente básico nas proximidades do resíduo de tirosina **(a)** ou serina **(b)** que tende a atrair seu hidrogênio (1). O oxigênio (2), buscando uma estabilização, realiza um ataque nucleofílico ao fósforo (3) da cadeia de DNA. Neste ambiente instável outro microambiente, ácido, também propiciado pela enzima, tende a atrair o oxigênio (4) da pentose, que é estabilizado pela formação da hidroxila (5), rompendo assim, a ligação fosfodiéster do DNA. Para a reação contrária de religação da cadeia de DNA, um microambiente ambiente básico induz a ionização da hidroxila (5), e este oxigênio (4) agora comporta-se como o nucleófilo que busca a estabilidade pelo ataque ao fósforo (3). A ligação à enzima é também desestabilizada pelo microambiente ácido, que doa próton (1) ao oxigênio (2) do resíduo catalítico da enzima, rompendo assim, a ligação ao DNA. ? – resíduo(s) específico(s) desconhecido(s). (Adaptada de Grindley, Whiteson e Rice 2006).

As serina integrases fazem distinção entre os processos de inserção e remoção de sequências. Essa distinção está relacionada ao reconhecimento dos sítios (*attachment sites – att*) *attB*, da bactéria, e *attP*, do fago (*phage*) e a formação dos sítios *attL* (*left*) e *attR* (*right*) após a inserção do genoma viral (SMITH, 2014). Os sítios resultantes não são mais clivados e recombinados, a menos que haja interação com uma proteína acessória RDF (*recombination directionality factor*) para mediar o retorno ([Figura 6: OLORUNNIJI et al., 2017](#)).

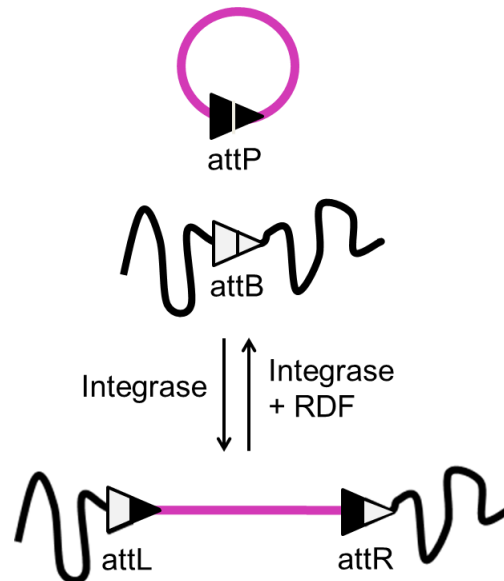


Figura 6 – Recombinação mediada por serina integrase. A inserção do genoma do bacteriófago (com sítio *attP*) no genoma da bactéria (com sítio *attB*) hospedeira é mediada por serina integrase com resultante formação dos sítios *attL* e *attR*. A excisão do genoma viral ocorre apenas com a serina integrase mais a proteína acessória RDF. (Adaptada de Stark, 2017).

As proteínas RDF têm se mostrado de difícil identificação. Entre algumas já determinadas, vê-se poucas sequências homólogas, variação de tamanho e *loci* gênicos muito diversificados (FOGG *et al.*, 2017). Podem ser elencadas as seguintes serina integrases com suas respectivas RDFs conhecidas: phiC31 – gp3 (FOGG *et al.*, 2018); Bxb1 – gp47 (GHOSH; WASIL; HATFULL, 2006); phiRv1 – Rv1584c (BIBB; HATFULL, 2002); A118 – gp44 (MANDALI *et al.*, 2017); TP901-1 – orf7 (BREÜNER; BRØNDSTED; HAMMER, 1999); phiBT1 – gp3^{BT1} (ZHANG, L. *et al.*, 2013); SPBc – SprB (ABE *et al.*, 2014); phiJoe – gp52 (FOGG *et al.*, 2017).

Observações indicam que as serina integrases são encontradas na forma de dímeros compactados ([Figura 7a](#)) e, quando se associam ao sítio de reconhecimento, adotam uma conformação mais estendida, ficando cada integrase posicionada em metade do sítio (GHOSH; PANNUNZIO; HATFULL, 2005). A estrutura C-terminal dessas enzimas foi definida por dois trabalhos que obtiveram cristais de integrase LI, de profago de *Listeria innocua*, ligada ao sítio *attP* (LI *et al.*, 2018; RUTHERFORD *et al.*, 2013), e o domínio N-terminal completado por cristal obtido previamente para $\gamma\delta$ resolvase (YANG, W.; STEITZ, 1995). São compostas por um domínio catalítico N-terminal (cat) e por um multidomínio C-terminal formado por uma alfa hélice αE que parte do cat e chega ao domínio de recombinação (RD), que é ligado por um *linker* a um domínio de fita de zinco (ZD), de onde emerge um

motivo *coiled coil* (CC) protuberante ([Figura 7](#); LI *et al.*, 2018; VAN DUYN; RUTHERFORD, 2013).

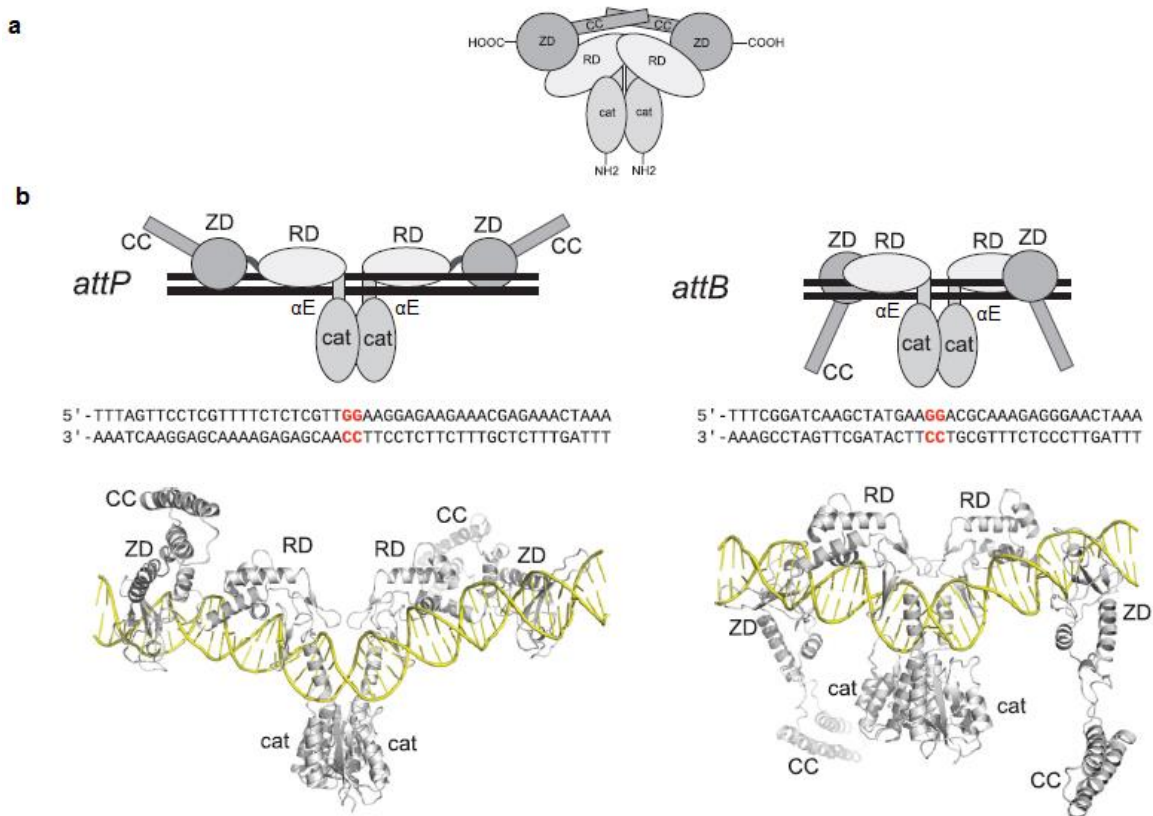


Figura 7 – Estrutura de uma serina integrase. **a.** Dímero de conformação globular observado para serina integrase não ligada a sítio de reconhecimento. **b.** Possíveis estruturas conformacionais do dímero de integrase ligado aos sítios *attP* e *attB* com o domínio catalítico N terminal (cat) e os multidomínios C terminal com as porções αE , domínio de recombinação (RD), domínio fita de zinco (ZD) e o domínio *coiled coil* (CC). A sequência nucleotídica refere-se ao sítio de reconhecimento da integrase A118, com o core marcado em vermelho. Estruturas obtidas por cristalografias da porção C terminal da integrase LI, de *Listeria*, com a porção N terminal completada por cristal obtido para a $\gamma\delta$ resolvasse ligadas aos sítios de reconhecimento no DNA. (Adaptada de Van Duyne e Rutherford, 2013).

Um dímero ligado a um sítio *attP* une-se a um segundo ligado a um sítio *attB*, formando um tetrâmero em uma conformação na qual os sítios posicionam-se de maneira justaposta, chamada de complexo sináptico ([Figura 8](#)). Aproximadamente na metade de um sítio, localiza-se a região de clivagem, o sítio de cruzamento (*crossover site*) ou *core* ([Figura 7b](#)). Cada uma das quatro serina integrases promove a clivagem de uma fita do duplex do DNA nesta posição, gerando uma quebra bifilamentar na molécula e deixando um excedente de dois nucleotídeos. Um mecanismo de rotação de subunidade recombina metades dos sítios *attB* e *attP* e as fitas são religadas, formando os novos sítios *attL* e *attR* ([Figura 8](#); RUTHERFORD;

VAN DUYNE, 2014). Esse mecanismo é definido para as serina recombinases de modo geral e difere do observado nas tirosinas, em que há quebra de uma fita de cada molécula de DNA por vez e as religações formam junções do tipo Holiday, sendo esta mais uma diferença entre as duas famílias de recombinases sítio-específicas (MEINKE *et al.*, 2016).

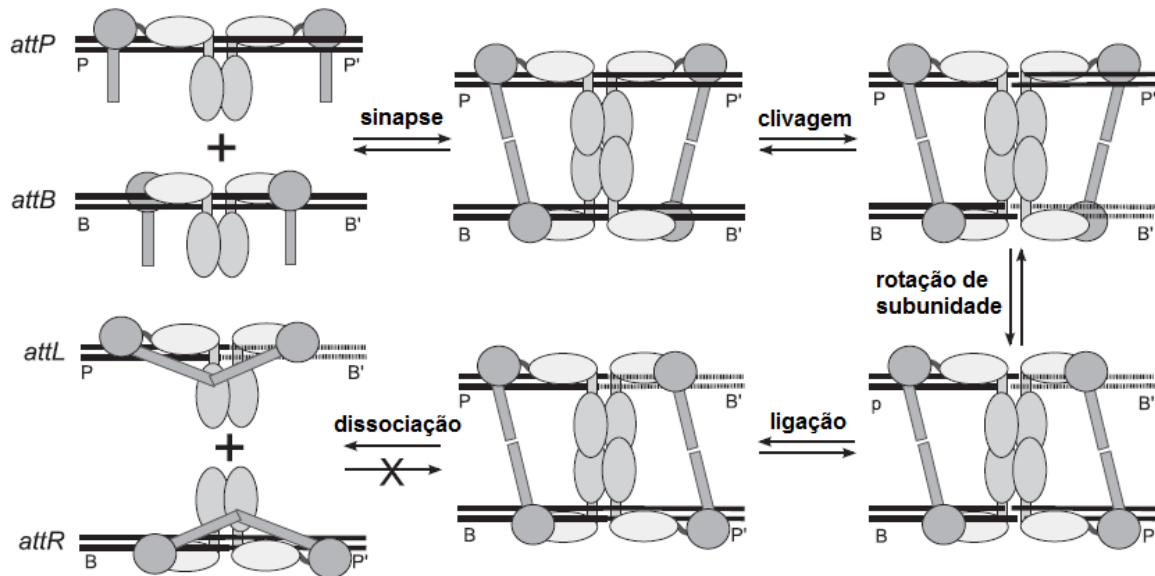


Figura 8 – Mecanismo de recombinação das serina integrases. Um dímero ligado a um sítio *attP* forma um tetrâmero com outro dímero ligado a um *attB* em uma conformação que une os sítios chamada de complexo sináptico. Os resíduos de serina catalisam a quebra de dupla fita em cada sítio, que é seguido por um movimento de rotação de subunidades. Duas integrases ligadas neste caso a P' e B' rotacionam e recombinam a posição dessas duas metades, formando os sítios *attL* e *attR*. É feita a ligação das porções que resulta em recombinação dos sítios e há dissociação do tetrâmero. O movimento é unidirecional na ausência de uma proteína RDF. (Adaptada de Van Duyne e Rutherford, 2013).

Por apresentarem outras características como sítios de reconhecimento pequenos (40 a 60 bp), ausência de necessidade de sítios e proteínas acessórias para dobras do DNA e recombinação ou determinadas conformações específicas do DNA, e por seus resultados de recombinação serem unidirecionais (reação contrária apenas em presença de RDF), as serina integrases despontaram como ferramentas biotecnológicas. Dependendo do direcionamento dos sítios e se estão localizados em uma mesma molécula de DNA ou em moléculas diferentes, lineares ou circulares, podem ser projetadas reações de inserção, deleção ou inversão de uma sequência específica (Figura 9). O emprego dessas proteínas tem envolvido aplicações tais como: remoção de transgene ou inserção de sequência desejada na presença do sítio, denominado, nesse caso, como sítio de ancoragem (*docking site*)

ou plataforma de aterrissagem (*landing pad*); mecanismos de troca de cassetes; montagem de DNA (*assembly*); e ativação ou desativação de expressão gênica por inversão de sequência, possibilitando, nesse caso, a construção de interruptores genéticos (MERRICK; ZHAO; ROSSER, 2018; STARK, 2017).

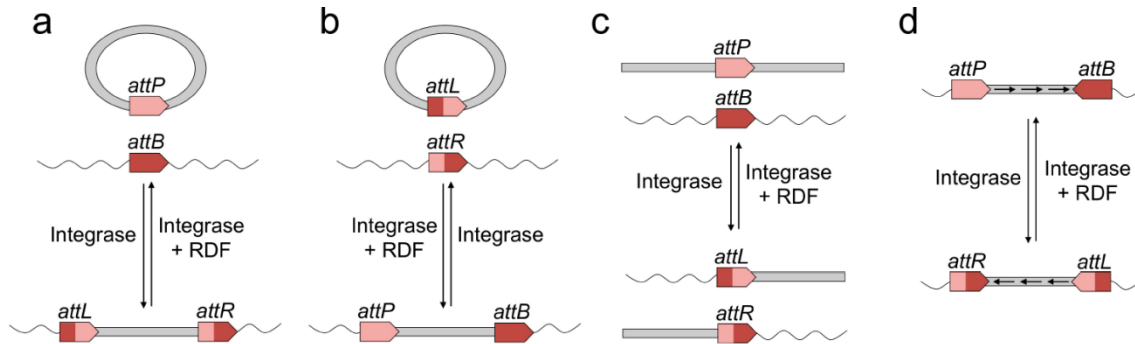


Figura 9 – Integração, excisão, translocação e inversão de sequências mediadas por serina integrase. **a.** Integração de DNA circular em molécula linear. **b.** Excisão de sequência de molécula linear resultando em porção circular. **c.** Translocação entre sequências de moléculas lineares. Nos três casos os sítios de reconhecimento *attB* e *attP* estão posicionados na mesma direção **d.** Inversão de sequência flanqueada pelos sítios *attB* e *attP* posicionados de forma oposta um ao outro. Não há recombinação dos sítios *attL* e *attR* para *attB* e *attP* sem a presença adicional de uma proteína RDF. (Fonte: Merrick, Zhao and Rosser, 2018).

Essa possibilidade de aplicação que permite a ligação ou o desligamento da expressão gênica por meio de rotação de sequências (Figura 9d) ganhou especial interesse para a construção de portas e circuitos lógicos. Apesar do paralelo ainda pouco explorado com as funções computacionais ou eletrônicas, em 1985, foi montado um sistema plasmidial, testado em *E. coli*, com os sítios da tirosina recombinase λ Int posicionados em direcionamentos opostos, de forma que foi possível realizar a inversão do promotor que flanqueavam e consequente ativação da expressão gênica, com já mencionada mudança de estado *off* para *on* (PODHAJSKA; HASAN; SZYBALSKI, 1985). No entanto, tirosina recombinases, tais como as enzimas Cre e Flp, apesar de hoje muito empregadas como ferramentas genéticas, apresentam algumas desvantagens dependendo da intenção de uso, pois têm características como recombinação bidirecional e em muitos casos, necessidade de sítios e proteínas acessórias.

Então, em 2013, Bonnet e colaboradores montaram seis portas lógicas utilizando plasmídeos em *E. coli* com duas entradas ativadas pela expressão induzível das serina integrases TP901-1 e Bxb1, e a saída mensurada pela fluorescência da proteína repórter GFP. O controle das portas foi feito por

mecanismos de inversão ou remoção de promotores ou terminadores flanqueados pelos sítios de reconhecimento das integrases. O fluxo da RNA polimerase para permitir a transcrição do *gfp* era controlado de acordo com os *inputs* dados para atender a tabela verdade de cada porta projetada (BONNET *et al.*, 2013). No mesmo ano, outra publicação independente, de forma muito semelhante também desenvolveu para *E. coli* 16 portas lógicas com as integrases Bxb1 e phiC31 controlando por inversão, terminadores, promotores e também a CDS do gene repórter *gfp* para obtenção da saída de acordo com a tabela verdade de cada porta desenhada (SIUTI; YAZBEK; LU, 2013). Em 2017, um estudo desenvolveu uma série de circuitos utilizando recombinases, e que foram testados em linhagens de células de mamífero. Dos 113 circuitos montados, 109 se apresentaram funcionais em células T Jukart. As construções envolveram múltiplos *inputs*, com 10 recombinases utilizadas, entre tirosinas, para remoção de terminadores, e duas serina integrases, para inversão de promotor e CDS, e os *outputs* mCherry e GFP. Também testaram um circuito utilizando um sistema CRISPR/Cas9 de ativação transcricional, dCas9-VPR, cujos gRNA eram ativados por recombinases (WEINBERG *et al.*, 2017).

Para a expansão de aplicações com maior acurácia, tornam-se necessários estudos amplos de validação de funcionalidade de serina integrases em diferentes organismos, principalmente em eucariotos. Xu *et al.* (2013), objetivando uma avaliação uniforme, testaram 15 serina integrases para atividades de deleção ou inserção por troca de cassete em células de mamíferos. Sete delas foram ativas em sequências integradas no genoma e das demais oito, cinco foram ativas em DNA extracromossomal. Foram verificados alguns erros de recombinação nos sítios. Houve também alguns problemas de crescimento celular que podem ter sido em decorrência de toxicidade. Bxb1 e phiC31 tiveram o melhor desempenho. Posteriormente, também verificaram a atividade das mesmas integrases em *S. cerevisiae* por ensaios de troca de cassete e 10 apresentaram atividade. Apesar de nesses ensaios não terem observado dano de sítios após recombinações, por ensaio de diluição seriada, verificaram toxicidade para sete das enzimas analisadas (XU, Z.; BROWN, 2016).

Além dos circuitos atendendo a portas lógicas Booleanas, um segundo conceito, também vindo da computação, é empregado na construção de redes genéticas sintéticas utilizando recombinases, o de memória. O mecanismo está

ligado a interruptores de ativação ou desativação que podem ser montados por recombinases em que, com uma entrada, uma “memória” (saída conferido por um gene) pode ser ativada ou pode gerar uma cascata de outras ativações e/ou desativações, levando a vários possíveis estados resultantes. Dependendo da construção, essa memória pode ser herdável e reversível ou irreversível. Ham *et al.* (2008), buscando empregar esses conceitos, montaram um sistema de inversão duplo e sobreposto. O desenho tem dois repórteres, GFP e RFP, posicionados em sentido oposto e reverso com uma região entre ambos onde foram incluídos os sítios das recombinases Fim B e Hin (invertases tirosina e serina recombinase, respectivamente) flanqueando dois promotores e um terminador. Dos cinco estados possíveis esperados, três foram alcançados, mas visualizados fracamente, sendo confirmados por PCR. Após o sequenciamento de isolados, não foi verificada inversão feita por Fim B. Apenas com o sequenciamento do PCR foi possível observar sua atividade de inversão prevista, mas a uma taxa de eficiência inferior à de Hin, sendo também observada ação cruzada. A inesperada baixa atividade de Fim B comprometeu a eficiência do sistema e foi hipotetizado pelos autores ser devido a *hairpin* formado pelos sítios posicionados de forma espelhada ou a funcionamentos de Fim B ainda não conhecidos. Isso destaca a importância da identificação de recombinases para compor sistemas de inversão robusto para montagem de interruptores.

Esses resultados reforçam a importância da caracterização funcional das recombinases e a vantagem que as serina integrases podem conferir ao controle de interruptores. Em 2014, buscando expandir o número de serina integrases com o objetivo de uso para construção de circuitos de memória permanente, Yang e colaboradores, por alinhamento de sequências conservadas com grandes serina recombinases, como phiC31 e Bxb1, encontraram mais de 4000 potenciais novas sequências. Com a busca pelos sítios de reconhecimento *attB* e *attP*, chegaram a 34 integrases e seus respectivos sítios. Testaram em *E. coli* 13 integrases, com até 60% de similaridade, para ativação de *switches* e consequente ativação do gene repórter *gfp*. Onze integrases se mostraram funcionais e apenas três delas (Ints 7, 8 e 11) reconheceram os sítios de outra integrase, a Int 10. Também organizaram três integrases (Ints 2, 5 e 7) em uma reação em cascata onde uma ativa a outra, obtendo o esperado resultado final de expressão do repórter (YANG, L. *et al.*, 2014).

1.3 BIOLOGIA SINTÉTICA EM PLANTAS

Os princípios trazidos pela área da biologia sintética envolvendo biotecnologia, modelos matemáticos e conceitos de engenharia para combinações de elementos atrelados às aplicações experimentais de edições genômicas e execução de redes genéticas interativas com *outputs* controláveis e variados são de grande interesse também na área vegetal. Após a emergência da tecnologia de inserção de transgenes para fornecimento de características de interesse, pretende-se avançar no desenvolvimento e refinamento de plantas agrocliváveis capazes de combinar fenótipos vantajosos. Os interesses são diversos, como aumento de produtividade; resistências a múltiplos patógenos, agentes de herbivoria e a diferentes condições de estresse abiótico; transferência de características como formação de rizóbio para fixação de nitrogênio para plantas não leguminosas ou metabolismo de planta C4 para planta C3; biofábricas para produção de compostos heterólogos; e alteração de rotas metabólicas endógenas para otimização de biocombustíveis como óleos, lignocelulose e açúcares solúveis. No entanto, em comparação com outros organismos, como bactérias, leveduras e modelos animais, a biologia sintética em plantas encontra-se ainda em estágios iniciais, com muitos desafios a serem superados, como por exemplo, aqueles relacionados às técnicas de transformação, ao tempo de estabelecimento estável das alterações genéticas empregadas e ao funcionamento eficiente do sistema implementado para obtenção das características desejadas (BENNING; SWEETLOVE, 2016; DE LANGE; KLAVINS; NEMHAUSER, 2018; SHIH; LIANG; LOQUÉ, 2016; WURTZEL *et al.*, 2019).

Uma abordagem tecnológica que poderia superar algumas dessas dificuldades é o desenvolvimento de locais seguros onde todas as redes projetadas poderiam ser inseridas, sendo transmitidas via ciclo celular e sem comprometer *loci* gênicos da planta. Essa ideia motivou a criação de cromossomos ou minicromossomos sintéticos. Já houve testes com estratégias parecidas com as utilizadas para a geração de cromossomos artificiais de leveduras, com a inclusão de sequência centromérica e de marcas de seleção. Houve também redesenho de cromossomo B de milho com transferência de região telomérica e transgene para transformá-lo em um minicromossomo com sítio de recombinase onde outras sequências poderiam ser integradas. As estratégias, no entanto, enfrentaram

dificuldades como características centroméricas vegetais difíceis de sintetizar e com sequências de funções desconhecidas, a ausência de cromossomo B em diferentes espécies, problemas de segregação meiótica, entre outras (BIRCHLER, 2015; DAWE, 2020). Apesar disso, esta representa uma importante linha de estudo almejada, inclusive, para geração de genomas vegetais sintéticos e mínimos por construções *bottom-up*. Mas há ainda dificuldades adicionais às da construção sintética, como a transferência de sequências grandes, tais como serão essas, para as células vegetais. O nosso grupo apresentou projeto piloto ao GP-write propondo o desenvolvimento de cromossomo sintético de *A. thaliana* com utilização de serina integrases para determinação de essencialidade de genes endógenos e para construção de circuitos genéticos para controles de expressão gênica. Em um segundo momento, a construção seria também realizada em soja, visando uma progressão do sistema para engenharia de plantas cultiváveis.

Como mencionado previamente, para a composição de circuitos sintéticos de interesse é importante caracterizar as partes a serem montadas. Neste sentido, o grupo de Jim Haseloff tem trabalhado no desenvolvimento de um chassi vegetal para tais testes. O modelo é a hepática *Marchantia polymorpha*, que apresenta características vantajosas como pequeno tamanho de genoma (aproximadamente 280 Mb), curto ciclo de vida, rápida propagação através de gema assexuada, baixos níveis de redundância gênica e um método de transformação de esporos (BOEHM *et al.*, 2017). O grupo desenvolveu uma plataforma aberta de partes da planta, chamada MarpoDB. (DELMANS; POLLAK; HASELOFF, 2017).

Buscando também uma via mais rápida para quantificação de partes para circuitos vegetais, Schaumberg *et al.* (2016) fizeram uma série de análises via transformação transiente em protoplastos de *A. thaliana* e de sorgo. Testaram mais de uma centena de combinações de partes para composição de promotores sintéticos repressíveis ativados por dexametasona ou OHT com atividade mensurada por um sistema de duplo *output* de luciferase.

Outra forma de ativação de circuitos que vem sendo aperfeiçoada é aquela com indução por luz, ou via sistema optogenético. No entanto, seu uso para ativação de interruptores em plantas é desafiador devido aos diversos processos de sinalização mediados por luz nas vias metabólicas vegetais, podendo ainda levar a danos colaterais (ANDRES; BLOMEIER; ZURBRIGGEN, 2019). Mas dois sistemas já foram desenvolvidos. O primeiro utiliza luz vermelha para ativação via fitocromo B

fusionado a um ativador transcricional e desativação por luz vermelho-distante (740 nm), que pode manter o sistema desativado mesmo sob luz branca (MÜLLER *et al.*, 2014). Apesar de nos organismos vegetais haver muitos fotorreceptores, eles apresentam reduzida atividade com luzes do espectro do verde. Desta forma, no segundo estudo, Chatelle *et al* (2018) desenvolveram um sistema de controle de expressão responsivo à luz verde, testado tanto em células de mamíferos como em protoplastos de *A. thaliana* e que foi baseado em um sistema transcricional fotoreceptivo de bactérias, CarH.

Frente a uma determinada sinalização, há também como objetivo desenvolver plantas que possam ser utilizadas como biossensores para monitoramento ambiental. Antunes *et al.* (2006) inicialmente desenvolveram um circuito genético sintético para promover a perda de coloração verde de *A. thaliana*. O sistema interfere na expressão de até cinco genes envolvidos em vias de síntese da clorofila. Posteriormente o grupo desenvolveu em tabaco um mecanismo de transdução de sinais formado por um receptor bacteriano quimérico ligado a partes do sistema histidina quinase vegetal ativado por moléculas de TNT. A sinalização ativa um sistema transcricional que induz promotor que controla o circuito genético de perda de clorofila. Futuramente, pretendem que o sensor vegetal seja responsivo, por exemplo, a poluentes ou outros agentes químicos (ANTUNES *et al.*, 2011).

Buscando avançar na compreensão de mecanismos de comunicação entre plantas leguminosas e bactérias da rizosfera, Geddes *et al.* (2019) identificaram um composto denominado rizopina, que é exsudado por bactérias rizóbios nessa região da raiz. Montaram um mecanismo para a síntese de rizopina em plantas transformadas da leguminosa *Medicago truncatula* e do cereal cevada, que por sua vez era exudada pela raiz e ativava biossensor luminescente expressando GFP em bactérias estabelecidas em uma região de rizosfera dessas plantas. Avanços nessa linha podem contribuir em mecanismos de ativação de sistemas bacterianos com sinais oriundos de plantas, incluindo a fixação de nitrogênio em plantas onde esse fenômeno não é observado.

Para a aplicação de engenharia de rotas sintéticas, fica evidente a possibilidade de contribuição da modelagem matemática no trabalho de Foo *et al.* (2018). Esses autores buscaram desenvolver um sistema para impedir a baixa expressão de reguladores positivos de defesa da planta durante a infecção por um patógeno. O sistema foi avaliado em *A. thaliana* sob ataque do fungo *Botrytis*

cinerea. Utilizando resultado de transcriptoma, carregaram algoritmos com os fatores de transcrição diferencialmente expressos após a infecção. Encontraram nove redes regulatórias contendo o fator de transcrição CHE em baixa regulação. Por modelagem e alguns dos dados experimentais identificaram uma rede regulatória entre os nove genes e propuseram um modelo de circuito que por loops de *feedback* controlariam a expressão do fator CHE, que teria expressão normalizada e poderia atuar contra o patógeno infeccioso.

Com relação ao uso de recombinases sítio-específica em plantas, alguns trabalhos prévios foram reportados. Inicialmente, foram testadas recombinações com as tirosina recombinases Cre e Flp. Apesar de resultados positivos, por apresentarem reação reversível, após algumas rodadas de inserção foram observadas recombinações indesejadas (XU, C.; YU, 2011). Então, outras pesquisas introduziram as serina integrases phiC31 e Bxb1 para excisão ou integração de sequências, demonstradas em *A. thaliana*, cevada, trigo, tabaco e arroz, com permanência da recombinação em gerações subsequentes. Essas aplicações podem remover transgenes quando desejado ou marcas de seleção de plantas transformadas, ou, se usadas para inserção, na presença do sítio específico da integrase, tem-se uma plataforma de integração para determinado cassete de expressão (HIRAKAWA; MATSUNAGA, 2016; KAPUSI *et al.*, 2012; KEMPE *et al.*, 2010; THOMSON *et al.*, 2010, 2012). Estratégia de inversão de sequência foi utilizada por Rubtsova *et al.* (2008) para monitorar atividade da phiC31 transformada em trigo. Entretanto, não foram identificados, até agora, trabalhos relatando o uso de serina integrases em plantas para aplicação em sistemas de biologia sintética, como regulação da expressão gênica em circuitos lógicos ou produção de memórias permanente.

2 JUSTIFICATIVA

Considerando os interesses de desenvolvimento de redes de expressão gênica em plantas e as vantagens oferecidas pelas serina integrases como ferramentas de ativação ou desativação de sistemas sintéticos, fazem-se necessários estudos que avancem na compreensão da funcionalidade e acurácia dessas proteínas na ainda pouco explorada área da biologia sintética vegetal. Apesar de trabalhos anteriores caracterizando diferentes serina integrases, os resultados foram muito variados entre os tipos celulares e organismos modelo testados, incluindo ausências de funcionalidade, problemas de toxicidade, acurácia e ortogonalidade, sendo ainda phiC31 e Bxb1 as integrases mais utilizadas. Para construções de sistemas mais amplos, é imprescindível que outras integrases sejam bem caracterizadas, sobretudo para inversão de partes genéticas em plantas. Por isso, as integrases recém-descritas e avaliadas em bactéria por Yang *et al.* (2014) tornam-se boas candidatas para avaliação de sua funcionalidade em eucariotos.

3 OBJETIVO

3.1 OBJETIVO GERAL

Avaliar a funcionalidade de serina integrases para controle da expressão gênica em células vegetais.

3.2 OBJETIVOS ESPECÍFICOS

Desenhar e sintetizar um sistema de interruptor genético formado por dois plasmídeos, sendo um o repórter e o outro o de expressão da serina integrase;

Isolar protoplastos de *A. thaliana* para cotransformação transiente do sistema de dois plasmídeos para avaliação da funcionalidade das serina integrases;

Caracterizar a funcionalidade do sistema de ativação do repórter (GFP) por:

- Observação por microscopia de fluorescência dos protoplastos emitindo fluorescência de GFP para análise fenotípica da funcionalidade do sistema;
- Mensuração da porcentagem populacional de células emitindo fluorescência por citometria de fluxo;
- Avaliação da inversão das sequências alvo — sequência codificadora do gene *gfp* e sequência promotora CaMV 35S — e da formação esperada dos sítios resultantes da recombinação, *attL* e *attR*, por PCR e sequenciamento.

4 DESENVOLVIMENTO

O projeto desenvolvido para a composição desta tese de doutorado está consolidado no manuscrito intitulado “*Genetic switches designed for eukaryotic cells and controlled by serine integrases*”, que segue abaixo, após a seção Referências Bibliográficas. O trabalho da autora da tese diz respeito aos ensaios realizados com protoplastos de *Arabidopsis thaliana*, uma vez que o objetivo do presente estudo é voltado para a investigação da funcionalidade de serina integrases em plantas. A publicação conta ainda com dados gerados por outros colaboradores que fizeram a mesma investigação, porém voltada para células de mamíferos. O artigo foi publicado pelo periódico *Communications Biology* e pode ser acessado pelo seguinte link: doi.org/10.1038/s42003-020-0971-8.

5 REFERÊNCIAS BIBLIOGRÁFICAS

- ABE, K. *et al.* Developmentally-Regulated Excision of the SP β Prophage Reconstitutes a Gene Required for Spore Envelope Maturation in *Bacillus subtilis*. *PLoS Genetics*, v. 10, n. 10, 2014. Disponível em: <<https://dx.plos.org/10.1371/journal.pgen.1004636>>.
- ANDRES, J.; BLOMEIER, T.; ZURBRIGGEN, M. D. Synthetic Switches and Regulatory Circuits in Plants. *Plant Physiology*, v. 179, n. 3, p. 862–884, mar. 2019. Disponível em: <<http://www.plantphysiol.org/lookup/doi/10.1104/pp.18.01362>>.
- ANNALURU, N. *et al.* Total Synthesis of a Functional Designer Eukaryotic Chromosome. *Science*, v. 344, n. 6179, p. 55–58, 4 abr. 2014. Disponível em: <<http://www.sciencemag.org/lookup/doi/10.1126/science.1249252>>.
- ANTUNES, M. S. *et al.* A synthetic de-greening gene circuit provides a reporting system that is remotely detectable and has a re-set capacity. *Plant Biotechnology Journal*, v. 4, n. 6, p. 605–622, nov. 2006. Disponível em: <<http://doi.wiley.com/10.1111/j.1467-7652.2006.00205.x>>.
- ANTUNES, M. S. *et al.* Programmable Ligand Detection System in Plants through a Synthetic Signal Transduction Pathway. *PLoS ONE*, v. 6, n. 1, p. e16292, 25 jan. 2011. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0016292>>.
- ANZALONE, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. [S.l.: s.n.], 2019. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/31634902>>.
- ATKINSON, M. R. *et al.* Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in *Escherichia coli*. *Cell*, v. 113, n. 5, p. 597–607, maio 2003. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0092867403003465>>.
- BENNING, C.; SWEETLOVE, L. Synthetic biology for basic and applied plant research. *The Plant Journal*, v. 87, n. 1, p. 3–4, jul. 2016. Disponível em: <<http://doi.wiley.com/10.1111/tpj.13245>>.
- BIBB, L. A.; HATFULL, G. F. Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, ϕ Rv1. *Molecular Microbiology*, v. 45, n. 6, p. 1515–1526, 2002. Disponível em: <<http://doi.wiley.com/10.1046/j.1365-2958.2002.03130.x>>.
- BIBIKOVA, M. *et al.* Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases. *Molecular and Cellular Biology*, v. 21, n. 1, p. 289–297, 1 jan. 2001. Disponível em: <<http://mcb.asm.org/cgi/doi/10.1128/MCB.21.1.289-297.2001>>.
- BIRCHLER, J. A. Promises and pitfalls of synthetic chromosomes in plants. *Trends in Biotechnology*, v. 33, n. 3, p. 189–194, mar. 2015. Disponível em: <<http://dx.doi.org/10.1016/j.tibtech.2014.12.010>>.
- BOCH, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science*, v. 326, n. 5959, p. 1509–1512, 11 dez. 2009. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.1178811>>.
- BOEHM, C. R. *et al.* Synthetic Botany. *Cold Spring Harbor Perspectives in Biology*, v. 9, n. 7, p. a023887, jul. 2017. Disponível em: <<http://cshperspectives.cshlp.org/lookup/doi/10.1101/cshperspect.a023887>>.
- BOLOTIN, A. *et al.* Clustered regularly interspaced short palindrome repeats (CRISPRs)

have spacers of extrachromosomal origin. *Microbiology*, v. 151, n. 8, p. 2551–2561, 1 ago. 2005. Disponível em:

<<https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.28048-0>>.

BONNET, J. *et al.* Amplifying Genetic Logic Gates. *Science*, v. 340, n. 6132, p. 599–603, 3 maio 2013. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.1232758>>.

BREÜNER, A.; BRØNDSTED, L.; HAMMER, K. Novel organization of genes involved in prophage excision identified in the temperate lactococcal bacteriophage TP901-1. *Journal of Bacteriology*, v. 181, n. 23, p. 7291–7297, 1999. Disponível em:

<<https://jlb.asm.org/content/181/23/7291>>

BROPHY, J. A. N.; VOIGT, C. A. Principles of genetic circuit design. *Nature Methods*, v. 11, n. 5, p. 508–520, 29 maio 2014. Disponível em:

<<http://www.nature.com/articles/nmeth.2926>>.

CAMERON, D. E.; BASHOR, C. J.; COLLINS, J. J. A brief history of synthetic biology. *Nature Reviews Microbiology*, v. 12, n. 5, p. 381–390, 1 maio 2014. Disponível em:

<<http://dx.doi.org/10.1038/nrmicro3239>>.

CAMPA, C. C. *et al.* Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nature Methods*, v. 16, n. 9, p. 887–893, 12 set. 2019.

Disponível em: <<http://www.nature.com/articles/s41592-019-0508-6>>.

CARLSON, D. F. *et al.* Production of hornless dairy cattle from genome-edited cell lines.

Nature Biotechnology, v. 34, n. 5, p. 479–481, 6 maio 2016. Disponível em:

<<http://www.nature.com/articles/nbt.3560>>.

CERMAK, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research*, v. 39, n. 12, p. e82–e82, jul. 2011. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr218>>.

CHATELLE, C. *et al.* A Green-Light-Responsive System for the Control of Transgene Expression in Mammalian and Plant Cells. *ACS Synthetic Biology*, v. 7, n. 5, p. 1349–1358, 18 maio 2018. Disponível em: <<https://pubs.acs.org/doi/10.1021/acssynbio.7b00450>>.

CHRISTIAN, M. *et al.* Targeting DNA Double-Strand Breaks with TAL Effector Nucleases.

Genetics, v. 186, n. 2, p. 757–761, out. 2010. Disponível em:

<<http://www.genetics.org/lookup/doi/10.1534/genetics.110.120717>>.

CONG, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, v.

339, n. 6121, p. 819–823, 15 fev. 2013. Disponível em:

<<http://www.sciencemag.org/cgi/doi/10.1126/science.1231143>>.

COOPER, C. A. *et al.* Innovative approaches to genome editing in avian species. *Journal of Animal Science and Biotechnology*, v. 9, n. 1, p. 15, 9 dez. 2018. Disponível em:

<<https://jasbsci.biomedcentral.com/articles/10.1186/s40104-018-0231-7>>.

DAWE, R. K. Charting the path to fully synthetic plant chromosomes. *Experimental Cell Research*, v. 390, n. 1, p. 111951, maio 2020. Disponível em:

<<https://doi.org/10.1016/j.yexcr.2020.111951>>.

DE LANGE, O.; KLAVINS, E.; NEMHAUSER, J. Synthetic genetic circuits in crop plants.

Current Opinion in Biotechnology, v. 49, p. 16–22, fev. 2018. Disponível em:

<<http://dx.doi.org/10.1016/j.copbio.2017.07.003>>.

DELMANS, M.; POLLAK, B.; HASELOFF, J. MarpoDB: An open registry for *Marchantia polymorpha* genetic parts. *Plant and Cell Physiology*, v. 58, n. 1, p. pcw201, 18 jan. 2017.

Disponível em: <<https://academic.oup.com/pcp/article-lookup/doi/10.1093/pcp/pcw201>>.

DELTCHEVA, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, v. 471, n. 7340, p. 602–607, 30 mar. 2011. Disponível em: <<http://www.nature.com/articles/nature09886>>.

DESHPANDE, S. *et al.* Octanol-assisted liposome assembly on chip. *Nature Communications*, v. 7, n. 1, p. 10447, 22 abr. 2016. Disponível em: <<http://www.nature.com/articles/ncomms10447>>.

ELOWITZ, M. B.; LEIBLER, S. A synthetic oscillatory network of transcriptional regulators. *Nature*, v. 403, n. 6767, p. 335–338, jan. 2000. Disponível em: <<http://www.nature.com/articles/35002125>>.

FLEISCHMANN, R. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, v. 269, n. 5223, p. 496–512, 28 jul. 1995. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.7542800>>.

FOGG, P. C. M. *et al.* Genome Integration and Excision by a New *Streptomyces* Bacteriophage, ϕ Joe. *Applied and Environmental Microbiology*, v. 83, n. 5, p. 1–16, 1 mar. 2017. Disponível em: <<http://aem.asm.org/lookup/doi/10.1128/AEM.02767-16>>.

FOGG, P. C. M. *et al.* Recombination directionality factor gp3 binds Φ C31 integrase via the zinc domain, potentially affecting the trajectory of the coiled-coil motif. *Nucleic Acids Research*, v. 46, n. 3, p. 1308–1320, 2018. Disponível em: <<https://academic.oup.com/nar/article/46/3/1308/4708263>>.

FOO, M. *et al.* A Framework for Engineering Stress Resilient Plants Using Genetic Feedback Control and Regulatory Network Rewiring. *ACS Synthetic Biology*, v. 7, n. 6, p. 1553–1564, 15 jun. 2018. Disponível em: <<https://pubs.acs.org/doi/10.1021/acssynbio.8b00037>>.

FRASER, C. M. *et al.* The Minimal Gene Complement of *Mycoplasma genitalium*. *Science*, v. 270, n. 5235, p. 397–404, 20 out. 1995. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.270.5235.397>>.

FREDENS, J. *et al.* Total synthesis of *Escherichia coli* with a recoded genome. *Nature*, v. 569, n. 7757, p. 514–518, 15 maio 2019. Disponível em: <<http://dx.doi.org/10.1038/s41586-019-1192-5>>.

GARDNER, T. S.; CANTOR, C. R.; COLLINS, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, v. 403, n. 6767, p. 339–342, 2000. Disponível em: <[10.1038/35002131](http://dx.doi.org/10.1038/35002131)>.

GEDDES, B. A. *et al.* Engineering transkingdom signalling in plants to control gene expression in rhizosphere bacteria. *Nature Communications*, v. 10, n. 1, p. 3430, 31 dez. 2019. Disponível em: <<http://dx.doi.org/10.1038/s41467-019-10882-x>>.

GHOSH, P.; PANNUNZIO, N. R.; HATFULL, G. F. Synapsis in Phage Bxb1 Integration: Selection Mechanism for the Correct Pair of Recombination Sites. *Journal of Molecular Biology*, v. 349, n. 2, p. 331–348, jun. 2005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0022283605003244>>.

GHOSH, P.; WASIL, L. R.; HATFULL, G. F. Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biology*, v. 4, n. 6, p. 0964–0974, 2006. Disponível em: <<https://dx.plos.org/10.1371/journal.pbio.0040186>>.

GIBSON, D. G. *et al.* Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome. *Science*, v. 319, n. 5867, p. 1215–1220, 29 fev. 2008.

Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.1151721>>.

GIBSON, D. G. *et al.* Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science*, v. 329, n. 5987, p. 52–56, 2 jul. 2010. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.1190719>>.

GOH, K.-I.; KAHNG, B.; CHO, K.-H. Sustained Oscillations in Extended Genetic Oscillatory Systems. *Biophysical Journal*, v. 94, n. 11, p. 4270–4276, jun. 2008. Disponível em: <<http://dx.doi.org/10.1529/biophysj.107.128017>>.

GOOTENBERG, J. S. *et al.* Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science*, v. 360, n. 6387, p. 439–444, 27 abr. 2018. Disponível em: <<https://www.sciencemag.org/lookup/doi/10.1126/science.aaq0179>>.

GRINDLEY, N. D. F.; WHITESON, K. L.; RICE, P. A. Mechanisms of Site-Specific Recombination. *Annual Review of Biochemistry*, v. 75, n. 1, p. 567–605, jun. 2006. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev.biochem.73.011303.073908>>.

HAM, T. S. *et al.* Design and Construction of a Double Inversion Recombination Switch for Heritable Sequential Genetic Memory. *PLoS ONE*, v. 3, n. 7, p. e2815, 30 jul. 2008. Disponível em: <<http://dx.plos.org/10.1371/journal.pone.0002815>>.

HIRAKAWA, T.; MATSUNAGA, S. *Chromosome and Genomic Engineering in Plants*. New York, NY: Springer New York, 2016. v. 1469. Disponível em: <<http://link.springer.com/10.1007/978-1-4939-4931-1>>. (Methods in Molecular Biology).

HUTCHISON, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science*, v. 351, n. 6280, p. aad6253–aad6253, 25 mar. 2016. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.aad6253>>.

ISALAN, M. Zinc-finger nucleases: how to play two good hands. *Nature Methods*, v. 9, n. 1, p. 32–34, 28 jan. 2012. Disponível em: <<http://dx.doi.org/10.1038/nmeth.1805>>.

ISHINO, Y. *et al.* Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, v. 169, n. 12, p. 5429–5433, dez. 1987. Disponível em: <<http://jlb.asm.org/lookup/doi/10.1128/jb.169.12.5429-5433.1987>>.

ISHIURA, M. *et al.* Expression of a Gene Cluster *kaiABC* as a Circadian Feedback Process in Cyanobacteria. *Science*, v. 281, n. 5382, p. 1519–1523, 4 set. 1998. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.281.5382.1519>>.

JACOB, F.; MONOD, J. On the Regulation of Gene Activity. *Cold Spring Harbor Symposia on Quantitative Biology*, v. 26, n. 0, p. 193–211, 1 jan. 1961. Disponível em: <<http://symposium.cshlp.org/cgi/doi/10.1101/SQB.1961.026.01.024>>.

JANSEN, R. *et al.* Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, v. 43, n. 6, p. 1565–1575, mar. 2002. Disponível em: <<http://doi.wiley.com/10.1046/j.1365-2958.2002.02839.x>>.

JINEK, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, v. 337, n. 6096, p. 816–821, 17 ago. 2012. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.1225829>>.

KALDS, P. *et al.* Sheep and Goat Genome Engineering: From Random Transgenesis to the CRISPR Era. *Frontiers in Genetics*, v. 10, n. September, p. 1–27, 3 set. 2019. Disponível em: <<https://www.frontiersin.org/article/10.3389/fgene.2019.00750/full>>.

- KAPUSI, E. *et al.* phiC31 Integrase-Mediated Site-Specific Recombination in Barley. *PLoS ONE*, v. 7, n. 9, p. e45353, 14 set. 2012. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0045353>>.
- KAUFFMAN, S. A. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, v. 22, n. 3, p. 437–467, mar. 1969. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0022519369900150>>.
- KEMPE, K. *et al.* Transgene excision from wheat chromosomes by phage phiC31 integrase. *Plant Molecular Biology*, v. 72, n. 6, p. 673–687, 2 abr. 2010. Disponível em: <<http://link.springer.com/10.1007/s11103-010-9606-7>>.
- LAU, C.-H.; SUH, Y. In vivo epigenome editing and transcriptional modulation using CRISPR technology. *Transgenic Research*, v. 27, n. 6, p. 489–509, 4 dez. 2018. Disponível em: <<https://doi.org/10.1007/s11248-018-0096-8>>.
- LEE, J. H. *et al.* Gene therapy for visual loss: Opportunities and concerns. *Progress in Retinal and Eye Research*, v. 68, n. Janeiro de 2019, p. 31–53, jan. 2019. Disponível em: <<https://doi.org/10.1016/j.preteyeres.2018.08.003>>.
- LI, H. *et al.* Serine Integrase attP Binding and Specificity. *Journal of Molecular Biology*, v. 430, n. 21, p. 4401–4418, out. 2018. Disponível em: <<https://doi.org/10.1016/j.jmb.2018.09.007>>.
- LIN, Y. *et al.* Improving Chromosome Synthesis with a Semiquantitative Phenotypic Assay and Refined Assembly Strategy. *ACS Synthetic Biology*, v. 8, n. 10, p. 2203–2211, 18 out. 2019. Disponível em: <<https://pubs.acs.org/doi/10.1021/acssynbio.8b00505>>.
- LITSCHER, T. *et al.* Beating Vesicles: Encapsulated Protein Oscillations Cause Dynamic Membrane Deformations. *Angewandte Chemie International Edition*, v. 57, n. 50, p. 16286–16290, 10 dez. 2018. Disponível em: <<http://doi.wiley.com/10.1002/anie.201808750>>.
- LIU, J. *et al.* Building Potent Chimeric Antigen Receptor T Cells With CRISPR Genome Editing. *Frontiers in Immunology*, v. 10, n. MAR, p. 1–8, 19 mar. 2019. Disponível em: <<https://www.frontiersin.org/article/10.3389/fimmu.2019.00456/full>>.
- LIU, Y. *et al.* Synthesizing AND gate genetic circuits based on CRISPR-Cas9 for identification of bladder cancer cells. *Nature Communications*, v. 5, n. 1, p. 5393, 6 dez. 2014. Disponível em: <<http://dx.doi.org/10.1038/ncomms6393>>.
- MALI, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science*, v. 339, n. 6121, p. 823–826, 15 fev. 2013. Disponível em: <[http://www.sciencemag.org.cyber.usask.ca/content/339/6121/823.full.pdf%5Cnfile:///Articles/2013/Mali/Science 2013 Mali.pdf](http://www.sciencemag.org.cyber.usask.ca/content/339/6121/823.full.pdf%5Cnfile:///Articles/2013/Mali/Science%2013%20Mali.pdf)>.
- MANDALI, S. *et al.* Control of recombination directionality by the Listeria phage A118 protein Gp44 and the coiled-coil motif of its serine integrase. *Journal of Bacteriology*, v. 199, n. 11, p. 1–17, 2017. Disponível em: <<https://jb.asm.org/lookup/doi/10.1128/JB.00019-17>>.
- MEINKE, G. *et al.* Cre Recombinase and Other Tyrosine Recombinases. *Chemical Reviews*, v. 116, n. 20, p. 12785–12820, 26 out. 2016. Disponível em: <<https://pubs.acs.org/doi/10.1021/acs.chemrev.6b00077>>.
- MERRICK, C. A.; ZHAO, J.; ROSSER, S. J. Serine Integrase: Advancing Synthetic Biology. *ACS Synthetic Biology*, v. 7, n. 2, p. 299–310, 16 fev. 2018. Disponível em: <<https://pubs.acs.org/doi/10.1021/acssynbio.7b00308>>.
- MILLER, T. E. *et al.* Light-powered CO₂ fixation in a chloroplast mimic with natural and

synthetic parts. *Science*, v. 368, n. 6491, p. 649–654, 8 maio 2020. Disponível em: <<https://www.sciencemag.org/lookup/doi/10.1126/science.aaz6802>>.

MITCHELL, L. A. *et al.* Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science*, v. 355, n. 6329, p. eaaf4831, 10 mar. 2017. Disponível em: <<http://www.sciencemag.org/lookup/doi/10.1126/science.aaf4831>>.

MIYAMOTO, T. *et al.* Synthesizing Biomolecule-Based Boolean Logic Gates. *ACS Synthetic Biology*, v. 2, n. 2, p. 72–82, 15 fev. 2013. Disponível em: <<https://pubs.acs.org/doi/10.1021/sb3001112>>.

MOJICA, F.J.M. *et al.* Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Molecular Microbiology*, v. 17, n. 1, p. 85–93, jul. 1995. Disponível em: <http://doi.wiley.com/10.1111/j.1365-2958.1995.mmi_17010085.x>.

MOJICA, FRANCISCO J.M. *et al.* Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution*, v. 60, n. 2, p. 174–182, fev. 2005. Disponível em: <<http://link.springer.com/10.1007/s00239-004-0046-3>>.

MONOD, J.; JACOB, F. General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. *Cold Spring Harbor Symposia on Quantitative Biology*, v. 26, p. 389–401, 1 jan. 1961. Disponível em: <<http://symposium.cshlp.org/cgi/doi/10.1101/SQB.1961.026.01.048>>.

MOSCOU, M. J.; BOGDANOVA, A. J. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science*, v. 326, n. 5959, p. 1501–1501, 11 dez. 2009. Disponível em: <<http://www.sciencemag.org/content/326/5959/1501.full>>.

MOSER, F. *et al.* Dynamic control of endogenous metabolism with combinatorial logic circuits. *Molecular Systems Biology*, v. 14, n. 11, p. 1–18, 29 nov. 2018. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.15252/msb.20188605>>.

MÜLLER, K. *et al.* A red light-controlled synthetic gene expression switch for plant systems. *Mol. BioSyst.*, v. 10, n. 7, p. 1679–1688, 2014. Disponível em: <<http://xlink.rsc.org/?DOI=C3MB70579J>>.

NIELSEN, A. A. K. *et al.* Genetic circuit design automation. *Science*, v. 352, n. 6281, p. aac7341–aac7341, 1 abr. 2016. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.aac7341>>.

NIELSEN, A. A.; VOIGT, C. A. Multi-input <sc>CRISPR</sc> / <sc>C</sc> as genetic circuits that interface host regulatory networks. *Molecular Systems Biology*, v. 10, n. 11, p. 763, 24 nov. 2014. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.15252/msb.20145735>>.

OLORUNNIJI, F. J. *et al.* Control of serine integrase recombination directionality by fusion with the directionality factor. *Nucleic Acids Research*, v. 45, n. 14, p. 8635–8645, 21 ago. 2017. Disponível em: <<https://academic.oup.com/nar/article/45/14/8635/3896347>>.

PODHAJSKA, A. J.; HASAN, N.; SZYBALSKI, W. Control of cloned gene expression by promoter inversion in vivo: construction of the heat-pulseactivated att-nutL-p-att-N module. *Gene*, v. 40, n. 1, p. 163–168, jan. 1985. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/037811985900381>>.

POURCEL, C.; SALVIGNOL, G.; VERGNAUD, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional

tools for evolutionary studies. *Microbiology*, v. 151, n. 3, p. 653–663, 1 mar. 2005. Disponível em: <<https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.27437-0>>.

PTASHANE, M. *A genetic switch: phage lambda revisited*. [S.l.: s.n.], 2004.

PURNICK, P. E. M.; WEISS, R. The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology*, v. 10, n. 6, p. 410–422, jun. 2009. Disponível em: <<http://dx.doi.org/10.1038/nrm2698>>.

RAZZAQ, A. *et al.* Modern Trends in Plant Genome Editing: An Inclusive Review of the CRISPR/Cas9 Toolbox. *International Journal of Molecular Sciences*, v. 20, n. 16, p. 4045, 19 ago. 2019. Disponível em: <<https://www.mdpi.com/1422-0067/20/16/4045>>.

RICROCH, A. Global developments of genome editing in agriculture. *Transgenic Research*, v. 28, n. S2, p. 45–52, 18 ago. 2019. Disponível em: <<http://link.springer.com/10.1007/s11248-019-00133-6>>.

RUBTSOVA, M. *et al.* Expression of active Streptomyces phage phiC31 integrase in transgenic wheat plants. *Plant Cell Reports*, v. 27, n. 12, p. 1821–1831, 17 dez. 2008. Disponível em: <<http://link.springer.com/10.1007/s00299-008-0604-z>>.

RUTHERFORD, K. *et al.* Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Research*, v. 41, n. 17, p. 8341–8356, set. 2013. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt580>>.

RUTHERFORD, K.; VAN DUYN, G. D. The ins and outs of serine integrase site-specific recombination. *Current Opinion in Structural Biology*, v. 24, n. 1, p. 125–131, fev. 2014. Disponível em: <<http://dx.doi.org/10.1016/j.sbi.2014.01.003>>.

SANDER, J. D. *et al.* An affinity-based scoring scheme for predicting DNA-binding activities of modularly assembled zinc-finger proteins. *Nucleic Acids Research*, v. 37, n. 2, p. 506–515, fev. 2009. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkn962>>.

SANDERSON, M. R. *et al.* The crystal structure of the catalytic domain of the site-specific recombination enzyme $\gamma\delta$ resolvase at 2.7 Å resolution. *Cell*, v. 63, n. 6, p. 1323–1329, dez. 1990. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/009286749090427G>>.

SCHAUMBERG, K. A. *et al.* Quantitative characterization of genetic parts and circuits for plant synthetic biology. *Nature Methods*, v. 13, n. 1, p. 94–100, 16 jan. 2016. Disponível em: <<http://www.nature.com/articles/nmeth.3659>>.

SCHWANDER, T. *et al.* A synthetic pathway for the fixation of carbon dioxide in vitro. *Science*, v. 354, n. 6314, p. 900–904, 18 nov. 2016. Disponível em: <<http://www.sciencemag.org/lookup/doi/10.1126/science.aah5237>>.

SHEN, Y. *et al.* Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science*, v. 355, n. 6329, p. eaaf4791, 10 mar. 2017. Disponível em: <<http://www.sciencemag.org/lookup/doi/10.1126/science.aaf4791>>.

SHERLOCK BIOSCIENCES. *Sherlock Biosciences Receives FDA Emergency Use Authorization for CRISPR SARS-CoV-2 Rapid Diagnostic*. Disponível em: <<https://sherlock.bio/sherlock-biosciences-receives-fda-emergency-use-authorization-for-crispr-sars-cov-2-rapid-diagnostic/>>. Acesso em: 12 maio 2020.

SHIH, P. M.; LIANG, Y.; LOQUÉ, D. Biotechnology and synthetic biology approaches for metabolic engineering of bioenergy crops. *The Plant Journal*, v. 87, n. 1, p. 103–117, jul. 2016. Disponível em: <<http://doi.wiley.com/10.1111/tpj.13176>>.

SIUTI, P.; YAZBEK, J.; LU, T. K. Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology*, v. 31, n. 5, p. 448–452, 10 maio 2013. Disponível em: <<http://dx.doi.org/10.1038/nbt.2510>>.

SMITH, M. C. M. Phage-encoded Serine Integrases and Other Large Serine Recombinases. *Mobile DNA III*. [S.l.]: American Society of Microbiology, 2014. p. 253–272. Disponível em: <<http://www.asmscience.org/content/book/10.1128/9781555819217.chap11>>.

SMITH, M. C. M.; THORPE, H. M. Diversity in the serine recombinases. *Molecular Microbiology*, v. 44, n. 2, p. 299–307, 25 abr. 2002. Disponível em: <<http://doi.wiley.com/10.1046/j.1365-2958.2002.02891.x>>.

STANTON, B. C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nature Chemical Biology*, v. 10, n. 2, p. 99–105, 8 fev. 2014. Disponível em: <<http://dx.doi.org/10.1038/nchembio.1411>>.

STARK, W. M. Making serine integrases work for us. *Current Opinion in Microbiology*, v. 38, p. 130–136, ago. 2017. Disponível em: <<http://dx.doi.org/10.1016/j.mib.2017.04.006>>.

STARK, W. M. The Serine Recombinases. *Microbiology Spectrum*, v. 2, n. 6, p. MDNA3-0046–2014, 2014. Disponível em: <<http://www.asmscience.org/content/book/10.1128/9781555819217.chap3>>.

STRICKER, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature*, v. 456, n. 7221, p. 516–519, 29 nov. 2008. Disponível em: <<http://www.nature.com/articles/nature07389>>.

SUGITA, M. Functional analysis of chemical systems in vivo using a logical circuit equivalent. *Journal of Theoretical Biology*, v. 53, n. 1, p. 223–237, set. 1975. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0022519375901137>>.

THOMAS, R. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, v. 42, n. 3, p. 563–585, dez. 1973. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0022519373902476>>.

THOMSON, J. G. *et al.* PhiC31 recombination system demonstrates heritable germinal transmission of site-specific excision from the Arabidopsis genome. *BMC Biotechnology*, v. 10, n. 1, p. 17, 2010. Disponível em: <<http://bmcbiotechnol.biomedcentral.com/articles/10.1186/1472-6750-10-17>>.

THOMSON, J. G. *et al.* The Bxb1 recombination system demonstrates heritable transmission of site-specific excision in Arabidopsis. *BMC Biotechnology*, v. 12, n. 1, p. 9, 21 dez. 2012. Disponível em: <<https://bmcbiotechnol.biomedcentral.com/articles/10.1186/1472-6750-12-9>>.

VAN BELKUM, A. *et al.* Short-sequence DNA repeats in prokaryotic genomes. *Microbiology and molecular biology reviews : MMBR*, v. 62, n. 2, p. 275–93, jun. 1998. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/9618442>>.

VAN DUYNE, G. D.; RUTHERFORD, K. Large serine recombinase domain structure and attachment site binding. *Critical Reviews in Biochemistry and Molecular Biology*, v. 48, n. 5, p. 476–491, 28 set. 2013. Disponível em: <<http://www.tandfonline.com/doi/full/10.3109/10409238.2013.831807>>.

VENTER, J. C. *et al.* The Sequence of the Human Genome. *Science*, v. 291, n. 5507, p. 1304–1351, 16 fev. 2001. Disponível em: <<http://www.sciencemag.org/lookup/doi/10.1126/science.1058040>>.

WEINBERG, B. H. *et al.* Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nature Biotechnology*, v. 35, n. 5, p. 453–462, 27 maio 2017. Disponível em: <<http://dx.doi.org/10.1038/nbt.3805>>.

WEISS, M. *et al.* Sequential bottom-up assembly of mechanically stabilized synthetic cells by microfluidics. *Nature Materials*, v. 17, n. 1, p. 89–96, 16 jan. 2018. Disponível em: <<http://www.nature.com/articles/nmat5005>>.

WRIGHT, D. A. *et al.* Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nature Protocols*, v. 1, n. 3, p. 1637–1652, 16 ago. 2006. Disponível em: <<http://www.nature.com/articles/nprot.2006.259>>.

WU, G.; BAZER, F. W. Application of new biotechnologies for improvements in swine nutrition and pork production. *Journal of Animal Science and Biotechnology*, v. 10, n. 1, p. 28, 19 dez. 2019. Disponível em: <<https://jasbsci.biomedcentral.com/articles/10.1186/s40104-019-0337-6>>.

WU, Y. *et al.* Bug mapping and fitness testing of chemically synthesized chromosome X. *Science*, v. 355, n. 6329, p. eaaf4706, 10 mar. 2017. Disponível em: <<http://www.sciencemag.org/lookup/doi/10.1126/science.aaf4706>>.

WURTZEL, E. T. *et al.* Revolutionizing agriculture with synthetic biology. *Nature Plants*, p. 1–4, 18 nov. 2019. Disponível em: <<http://www.nature.com/articles/s41477-019-0539-0>>.

XIE, Z.-X. *et al.* “Perfect” designer chromosome V and behavior of a ring derivative. *Science*, v. 355, n. 6329, p. eaaf4704, 10 mar. 2017. Disponível em: <<http://www.sciencemag.org/lookup/doi/10.1126/science.aaf4704>>.

XU, C.; YU, W. *Plant Chromosome Engineering*. Totowa, NJ: Humana Press, 2011. v. 701. Disponível em: <<http://link.springer.com/10.1007/978-1-61737-957-4>>. (Methods in Molecular Biology).

XU, Z. *et al.* Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnology*, v. 13, n. 1, p. 87, 2013. Disponível em: <<http://bmcbiotechnol.biomedcentral.com/articles/10.1186/1472-6750-13-87>>.

XU, Z.; BROWN, W. R. A. Comparison and optimization of ten phage encoded serine integrases for genome engineering in *Saccharomyces cerevisiae*. *BMC Biotechnology*, v. 16, n. 1, p. 13, 9 dez. 2016. Disponível em: <<http://dx.doi.org/10.1186/s12896-016-0241-5>>.

YANG, B.; YANG, L.; CHEN, J. Development and Application of Base Editors. *The CRISPR Journal*, v. 2, n. 2, p. 91–104, abr. 2019. Disponível em: <<https://www.liebertpub.com/doi/10.1089/crispr.2019.0001>>.

YANG, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nature Methods*, v. 11, n. 12, p. 1261–1266, 26 dez. 2014. Disponível em: <<http://www.nature.com/articles/nmeth.3147>>.

YANG, W.; STEITZ, T. A. Crystal structure of the site-specific recombinase $\gamma\delta$ resolvase complexed with a 34 by cleavage site. *Cell*, v. 82, n. 2, p. 193–207, jul. 1995. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0092867495903070>>.

ZHANG, L. *et al.* Control of directionality in *Streptomyces* phage ϕ BT1 integrase-mediated site-specific recombination. *PLoS ONE*, v. 8, n. 11, p. 1–8, 2013.

ZHANG, W. *et al.* Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science*, v. 355, n. 6329, p. eaaf3981, 10 mar. 2017. Disponível em:

<<http://www.sciencemag.org/lookup/doi/10.1126/science.aaf3981>>.

ZHANG, YI *et al.* A CRISPR way for accelerating improvement of food crops. *Nature Food*, v. 1, n. 4, p. 200–205, 30 abr. 2020. Disponível em: <<http://dx.doi.org/10.1038/s43016-020-0051-8>>.

ZHANG, YINGXIAO *et al.* The emerging and uncultivated potential of CRISPR technology in plant science. *Nature Plants*, v. 5, n. 8, p. 778–794, 15 ago. 2019. Disponível em: <<http://dx.doi.org/10.1038/s41477-019-0461-5>>.

6 MANUSCRITO

Genetic switches designed for eukaryotic cells and controlled by serine integrases

Mayna S. Gomide^{1,2,3}, Thais T. Sales^{1,2}, Luciana R.C. Barros⁴, Cintia G. Limia⁴, Marco A. de Oliveira^{1,2}, Lilian H. Florentino¹, Leila M.G. Barros¹, Maria L. Robledo⁴, Gustavo P.C. José¹, Mariana S.M. Almeida¹, Rayane N. Lima¹, Stevens K. Rehen^{5,6}, Cristiano Lacorte¹, Eduardo O. Melo^{1,7}, André M. Murad¹, Martín H. Bonamino^{4,8*}, Cintia M. Coelho^{9*} and Elibio Rech^{1*}.

1 Brazilian Agriculture Research Corporation – Embrapa – Genetic Resources and Biotechnology – CENARGEN, Brasília, 70770917, DF, Brazil;

2 Department of Cell Biology, Institute of Biological Science, University of Brasília, Brasília, 70910900, DF, Brazil;

3 School of Medicine, Federal University of Juiz de Fora, Juiz de Fora, 36036900, MG, Brazil;

4 Molecular Carcinogenesis Program, Research Coordination, National Cancer Institute (INCA), Rio de Janeiro, 20231050, RJ, Brazil;

5 D'Or Institute for Research and Education (IDOR), Rio de Janeiro, 22281100, RJ, Brazil;

6 Institute of Biomedical Sciences, Federal University of Rio de Janeiro, Rio de Janeiro, 21941902, RJ, Brazil;

7 Graduation Program in Biotechnology, Federal University of Tocantins, Gurupi, 77402970, TO, Brazil;

8 Vice-Presidency of Research and Biological Collections (VPPCB), FIOCRUZ – Oswaldo Cruz Foundation Institute, Rio de Janeiro, 21040900, RJ, Brazil;

9 Department of Genetic and Morphology, Institute of Biological Science, University of Brasília, Brasília, 70910900, DF, Brazil;

* Corresponding authors (emails: elibio.rech@embrapa.br; cintiacoelhom@unb.br; and mbonamino@inca.gov.br).

Abstract

Recently, new serine integrases have been identified, increasing the possibility of scaling up genomic modulation tools. Here, we describe the use of unidirectional genetic switches to evaluate the functionality of six serine integrases in different eukaryotic systems: the HEK 293T cell lineage, bovine fibroblasts and plant protoplasts. Moreover, integrase activity was also tested in human cell types of therapeutic interest: peripheral blood mononuclear cells (PBMCs), neural stem cells (NSCs) and undifferentiated embryonic stem (ES) cells. The switches were composed of plasmids designed to flip two different genetic parts driven by serine integrases. Cell-based assays were evaluated by measurement of EGFP fluorescence and by molecular analysis of *attL/attR* sites formation after integrase functionality. Our results demonstrate that all the integrases were capable of inverting the targeted DNA sequences, exhibiting distinct performances based on the cell type or the switchable genetic sequence. These results should support the development of tunable genetic circuits to regulate eukaryotic gene expression.

Introduction

Recombinant DNA technology was a landmark for the development of cells that are able to perform several tasks in different areas of research, such as therapy¹, diagnosis², biosensing³, bioremediation⁴, and plant and animal genetic improvement⁵⁻⁷. Over the last decade, the desirable traits were mostly monogenic, with transgene expression being mainly transcriptionally controlled by tissue/development-specific or chemically/physically inducible promoters, acting as activators or repressors separately or in loops^{8,9}. Currently, novel approaches in synthetic biology are being used for the design of new molecular entities that would lead to the creation and/or improvement of pathways, reprogramming organisms for new tasks^{10,11}. This scenario demands the development of tools to finely tune polygenic expression.

Large serine-type phage integrases have been described as useful genomic tools for gene manipulation¹²⁻¹⁷. These proteins belong to a superfamily of site-specific serine recombinases, which differ in many aspects from their counterparts, the tyrosine recombinases. The latter are bidirectional proteins, which is a major drawback to their use to control gene expression due to the instability of the on/off end product¹⁸. Instead, the phage-encoded serine integrases (Ints) are capable of unidirectional recombination that leads to permanent DNA fragment inversion, which ultimately could be used to modulate gene expression¹⁹. To perform this function, Ints recognize specific attachment sites, named *attB* and *attP*²⁰. When both sites are placed in opposite directions flanking a genetic part (such as a promoter, coding sequence and/or terminator), Ints perform a 180° flip of these genetic parts, which can turn gene expression on/off. As a result of this recombination event, the attachment sites are converted into different sequences, called *attL* and *attR*²¹. This recombination process can only be reversed in the presence of a cognate protein, called recombination directionality factor (RDF)¹³.

Ints have been previously used to build logic gates based on Boolean algebra to regulate gene expression in prokaryotic organisms^{22,23}. In these reports, the authors utilized Ints to rotate promoters or terminators, hence controlling RNA polymerase flow in a way resembling an electronic transistor, now called a transcriptor²². Most importantly, GFP expression for every Int used was observed as predicted by the truth table of each designed logic gate that was evaluated^{22,23}. Another strategy used

to regulate prokaryotic gene expression was recombinase-based state machines, a system in which several combinations of integrase inputs that inverted or excised genetic parts produced different outputs²⁴. However, in eukaryotic organisms, there is a scarcity of effective tools that allow broad and precise gene regulation, which is an essential requirement for multiplex gene control. To date, only a limited number of Ints have been successfully tested in eukaryotic cells, showing a restricted cell-type-dependent functionality^{25–28}. Nevertheless, most of these studies aimed to integrate or excise a DNA fragment from the genome of a particular organism, rather than using the Ints as regulators of gene expression^{28–35}. In a different approach, Weinberg et al.²⁶ used a combination of four tyrosine and two serine recombinases in mammalian cells to develop a six-input/one-output Boolean logic lookup table. However, the relatively small number of functional Ints characterized in eukaryotic cells restricts the scale-up of the use of these proteins to build genetic circuits.

Recently, Yang et al.¹⁸ identified 34 putative Ints, showing functionality for 11 of these Ints in prokaryotic cells. In an attempt to fill this existing gap in eukaryotic cells, we built unidirectional genetic switches to evaluate the functionality of six out of these 11 Ints that have already been tested in prokaryotic cells. To this end, we chose three different model systems: the human embryonic kidney cell lineage (HEK 293T), bovine primary fibroblasts, and *Arabidopsis thaliana* protoplasts. In addition, as a medically important proof of concept, unidirectional genetic switches using these Ints were also evaluated in peripheral blood mononuclear cells (PBMCs), neural stem cells (NSCs) differentiated from induced pluripotent stem cells, and undifferentiated human embryonic stem cells (hES, BR-1 cell line).

Here we report that all tested Ints are functional genetic switch controllers, activating the coding sequence or the promoter switches designed to be turned on in the eukaryotic cells. The frequency of cells emitting the reporter fluorescence varies among the tested integrases and cell types, and, in some cases, the switch activation is proven by molecular tests. Additionally, Ints show accuracy in their site recognition and recombination process, and are not cytotoxic for the cell models assayed. These data put the evaluated Ints as suitable candidates to regulate gene expression in wide synthetic genetic networks that now can be built for several eukaryotic organisms.

Results

Unidirectional genetic switches design

In our system, the HEK 293T lineage, bovine primary fibroblasts, and *A. thaliana* protoplasts were chosen as human, nonhuman mammal and plant models, respectively. Ints 2, 4, 5, 7, 9, and 13 evaluated in *Escherichia coli* by Yang et al.¹⁸ were selected to be tested in eukaryotic switch systems. The phiC31 and Bxb1 integrases were also evaluated because they were shown to be functional in genetic switches in human cells²⁶ and plants³² (only phiC31). We designed unidirectional genetic switches composed of two sets of synthesized plasmids. The first set contains either a human or plant-optimized (*A. thaliana*) coding sequence of an Int under the control of a strong species-specific constitutive promoter, named integrase expression vectors (set 1, pIE). The second set contains the reporter gene *egfp* in the reverse complement orientation flanked by the recognition sites *attB* and *attP* of the corresponding Int under distinct constitutive promoters, for either plant or animal systems. The resulting plasmids were called switch GFP vectors (set 2, pSG) ([Fig. 1](#); [Supplementary Fig. 1](#); [Supplementary Methods](#)). Therefore, eight plasmids were generated for each vector set of the mammalian systems, and the same number was generated for the plant system ([Supplementary Table 1](#)). One plasmid of set 1 and one plasmid of set 2 were used to transiently cotransfect mammalian cells and cotransform plant protoplasts (test condition). The negative control cells were cotransfected/cotransformed with only one of the two plasmids plus a mock plasmid to keep the DNA concentration constant. As a positive control (pGFP), the cells were cotransfected/cotransformed with a plasmid containing *egfp* in forward orientation under the same constitutive promoter as that in the plasmids from set 2 plus a mock plasmid. HEK 293T control cells were not cotransfected with the mock plasmid. In these transient assays many copies of both plasmids for each test or control conditions are inserted in the cells, according to the concentrations described in the [Methods](#) section.

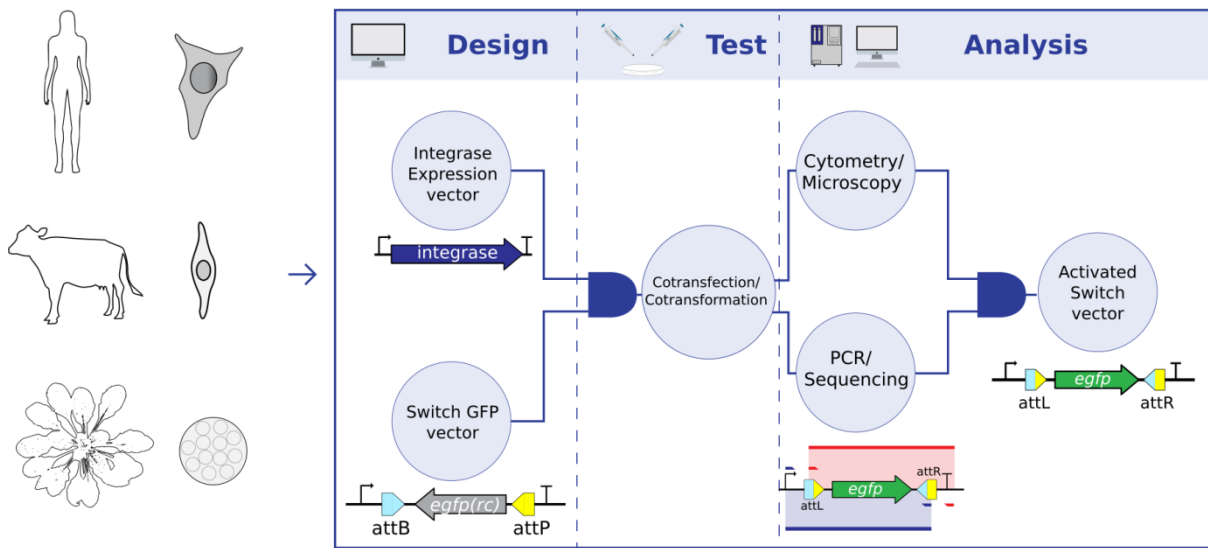


Fig. 1 Strategy overview of the eukaryotic genetic switch system. The human cell lineage HEK 293T, bovine fibroblasts and *A. thaliana* protoplasts were the selected model systems. The first step involved the design of two plasmid sets: the integrase expression vectors to express Ints 2, 4, 5, 7, 9, 13, phiC31, or Bxb1 and the switch GFP vectors with the *attB* and *attP* recognition sites of the respective Int flanking an *egfp* coding sequence in a reverse complement (*rc*) orientation. Acting as a schematic AND gate, combination of the corresponding plasmids of each of the vector sets results in the second step in the process, the test, accomplished by cotransfection or cotransformation assays of mammalian and plant cells, respectively. The third and last step led to the development of analytical methods that include the inputs of an additional schematic AND gate. Microscopy/flow cytometry analyses were used to detect EGFP fluorescence in cells resulting from the flipping action of the integrase. PCR/sequencing was used in the analysis of the *egfp* coding sequence rotated to the correct forward orientation flanked by the formed *attL* and *attR* sites. Both analytical inputs provide evidence of the activated switch vector output. The PCRs used one primer pair to amplify the complete *attL* site and the entire *egfp* coding sequence, now in the forward orientation (blue), and a second primer pair to amplify the complete *attR* site and the *egfp* sequence (red).

The rationale for this system was that if a particular Int is functional, it would switch the *egfp* coding sequence to the forward orientation, leading to *egfp* expression and the formation of the *attL* and *attR* sites, referred to as the activated switch vector (Fig. 1; Supplementary Fig. 1). The six Ints were chosen because their *attL* sites formed after flipping did not lead to the formation of any ATG upstream of the *egfp* ORF. Although there was not ATG formation in its *attL*, Int 10 was not selected to be evaluated because its recognition sites were recognized by other integrases when evaluated in bacterial cells¹⁸. All assays were evaluated at 48 h for mammalian cells and 24 h for *A. thaliana* protoplasts, corresponding to the maximum EGFP-positive cell frequency.

No nuclear localization signal (NLS) was added to the Ints since previous studies showed contradictory results regarding the addition of an NLS at the N- and/or C-termini of these proteins^{17,25}. Additionally, *in silico* analysis demonstrated

the existence of potential cryptic NLSs in the Ints evaluated in this study with various scores, except for Int 5, for which no NLS was predicted (Supplementary Fig. 2).

Functional characterization of Ints as switch controllers

In the mammalian groups, EGFP-expressing cells were detected among HEK 293T cells cotransfected with pIE + pSG vectors corresponding to Int 13, phiC31, or Bxb1 (Fig. 2a, b; Supplementary Figs. 3 and 4) and in bovine fibroblasts cotransfected with pIE + pSG vectors corresponding to Int 9, 13, or Bxb1 (Fig. 3a, b; Supplementary Figs. 3 and 5). In addition, low levels of cells emitting EGFP fluorescence could also be observed in tests activated by Ints 2, 4, and 5 in HEK 293T cells and by Ints 2 and phiC31 in bovine fibroblasts (Fig. 2b; Fig. 3b; Supplementary Figs. 4 and Fig. 5). In general, flow cytometry analysis showed that the abundance of EGFP-positive cells ranged from 0.03 to 16.02% for the tested switches in these cell groups, indicating different levels of Int functionality (Fig. 2b; Fig. 3b). In HEK 293T cells, Kruskal-Wallis statistics corroborated that the percentages of EGFP-positive cells cotransfected with pIE + pSG corresponding to Ints 13, phiC31 and Bxb1 resulted in significant differences compared with the negative controls ($p=1.71 \times 10^{-5}$, 1.10×10^{-4} , and 1.50×10^{-4} , respectively; Fig. 2b; Supplementary Table 2). In bovine fibroblasts, in addition to Ints 9, 13, and Bxb1 ($p=1.10 \times 10^{-6}$, 2.04×10^{-6} , and 9.18×10^{-7} , respectively), cells cotransfected with pIE + pSG corresponding to Ints 2, 5 and phiC31 were also significantly different from the negative controls ($p=1.87 \times 10^{-6}$, 3.08×10^{-5} , and 1.61×10^{-6} , respectively; Fig. 3b; Supplementary Table 2). Analyzing only the herein tested integrases, Int 13 led to the highest number of EGFP-positive cells in both models evaluated. The positive cell frequencies in the Int 13 tests were 33% and 16% compared to the positive controls (pGFP) for HEK 293T cells and bovine fibroblasts, respectively.

To confirm Int functionality by targeted *egfp* sequence rotation and correct *attL/attR* sites formation, PCR and sequencing analysis were performed. To this end, two pairs of primers were used. One pair annealed to the promoter sequence (forward) and to the *attR* site (reverse) (Fig. 1, blue color), and the other pair annealed to the *attL* site (forward) and to a region next to the terminator sequence (reverse) (Fig. 1, red color). Thus, PCR amplifications are expected in only activated pSG vectors. Interestingly, although EGFP fluorescence was not detected through

microscopy and flow cytometry analysis of cells tested with some of the Ints, the PCR and sequencing analysis of DNA extracted from tested cells confirmed the predicted *attL/attR* sites for Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1 and that the *egfp* had flipped to the forward orientation ([Fig. 2c](#); [Fig. 3c](#); Supplementary Figs. [6](#), [7](#), [8](#) and [9](#)). These results confirm that all the evaluated Ints were functional, however, their efficiency may set a threshold for detection of EGFP fluorescence by flow cytometry or microscopy analysis.

An important limiting factor to the use of Ints in genetic switches is their potential cytotoxicity. Thus we measured the viability of mammalian cells after 48h of the integrase activity using MTT assay. Neither HEK 293T cells nor fibroblasts showed marked viability impairment comparing with control groups ([Fig. 2d](#); [Fig. 3d](#)).

For *A. thaliana* protoplasts, EGFP-expressing cells were detected after cotransformation with the pIE + pSG vectors corresponding to Int 2, 4, 7, 9, 13, phiC31, or Bxb1. ([Fig. 4a, b](#); Supplementary Figs. [3](#) and [10](#)). All Int tests led to a statistically significant frequency of EGFP-positive cells compared to the respective negative controls, although Int 5 led to a very low percentage of EGFP-expressing cells (p-value for Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1 groups were 2.57×10^{-11} , 2.22×10^{-7} , 2.28×10^{-6} , 2.45×10^{-9} , 7.24×10^{-7} , 2.01×10^{-8} , 1.28×10^{-6} , and 1.07×10^{-6} , respectively; [Fig. 4b](#); Supplementary Table [2](#)). As observed for HEK 293T cells and bovine fibroblasts, the flow cytometry data showed a variable number of cells expressing EGFP. The frequency of EGFP-positive cells ranged from 0.1% for Int 5 to 24.2% for Int 13 (Supplementary Table [2](#)). This highest value showed Int 13 to be superior to phiC31 and Bxb1 in the induction of EGFP activation. Notably, overall, plant protoplasts resulted in higher numbers of EGFP-accumulating cells than the mammalian systems tested, indicating robustness in the vegetal system ([Fig. 2b](#), [3b](#), and [4b](#); Supplementary Table [2](#)). In this system, the pCaMV35S-GFP plasmid, with known EGFP expression efficiency, was used as a positive control³⁶. The cauliflower mosaic virus (CaMV) 35S promoter of this plasmid (alignment with GenBank V00140.1) has some single nucleotide polymorphisms (SNPs) compared to the CaMV 35S promoter chosen for the plant switch vector syntheses (iGEM registry BBa_K1547006 and alignment with GenBank V00141.1). These SNPs, however, did not result in statistically significant differences in the frequency of EGFP-expressing cells ($p=0.1277$; Supplementary Fig. [11](#)).

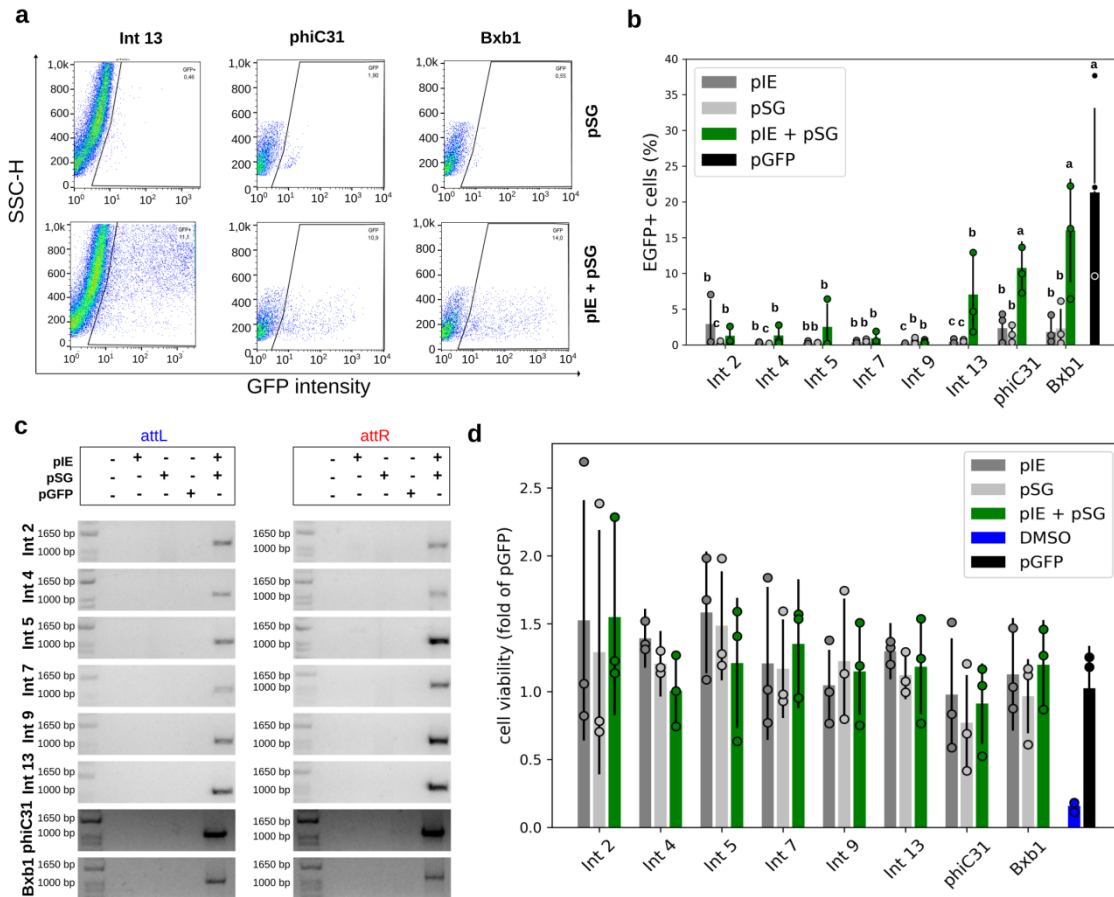


Fig. 2 Functional characterization of the genetic switches in human cells. **a** Flow cytometry distribution of HEK 293T cells at 48 h post transfection for Ints 13, phiC31 and Bxb1, the integrases that led to the highest EGFP-expressing cell frequencies. The heat map indicates the scattering of high cell concentrations (warm colors) to low cell concentrations (cool colors). The gate encompasses the EGFP-expressing cell population. **b** Bar graph plots showing the total average percentage and standard deviation of a cell population expressing EGFP in biological repeat assays ($n=3$) and circles showing the technical duplicate or triplicate average of each assay on the y axis. The x axis contains the different conditions. For each Int data group, different letters indicate significant differences ($p<0.05$). **c** Amplicons obtained through PCR analysis using two specific primer sets, the first set to verify *attL* formation and the second set to verify *attR* (highlighted in Fig. 1). The expected amplicon sizes in the Int test groups varied from 1021 to 1104 bp for *attL* and from 1058 to 1084 bp for *attR*. **d** Bar graph plots showing the viable cells average (circles corresponding to technical replicates averages) and standard deviation normalized with pGFP of OD measurements obtained after MTT assays ($n=3$). DMSO corresponds to the impairment negative control. Negative control cells were transfected with only one of the two vector sets, that is, integrase expression (pIE) or switch GFP (pSG) vectors. Positive control cells (pGFP) have an *egfp* sequence in the forward orientation under the control of the EF1 alpha promoter. All the data are representative of two or three technical and three biological replicates.

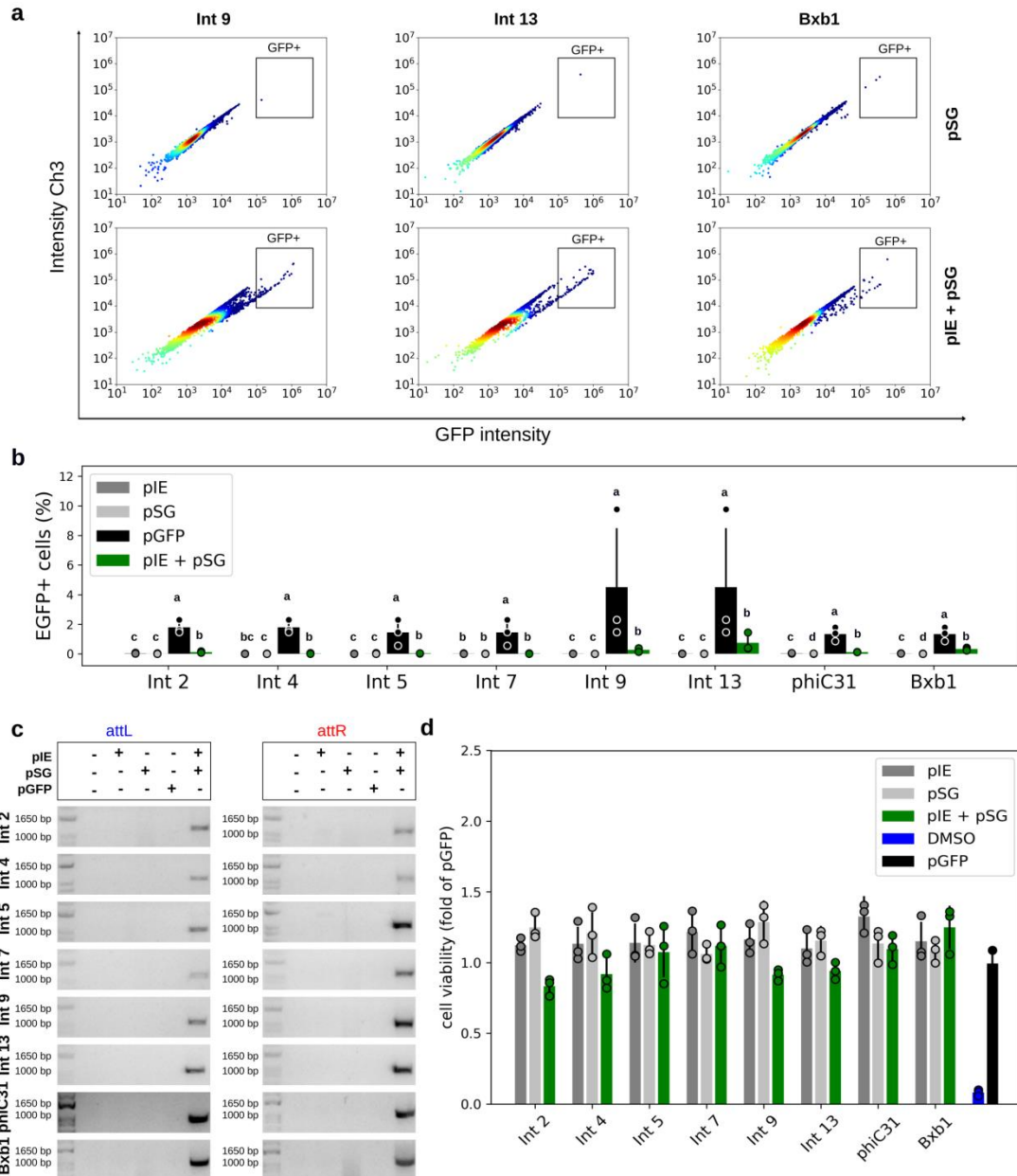


Fig. 3 Functional characterization of the genetic switches in bovine cells. **a** Flow cytometry distribution of bovine fibroblasts at 48 h post cotransfection for Ints 9, 13, and Bxb1, the integrases that led to the highest EGFP-expressing cell frequencies. The heat map indicates the scattering of high cell concentrations (warm colors) to low cell concentrations (cool colors). The gate encompasses the EGFP-expressing cell population. **b** Bar graph plots showing the total average percentage and standard deviation of a cell population expressing EGFP in biological repeat assays ($n=3$) and circles showing the technical triplicate average of each assay on the y axis. The x axis contains the different conditions. For each Int data group, different letters indicate significant differences ($p<0.05$). **c** Amplicons obtained through PCR analysis using two specific primer sets, the first set to verify *attL* formation and the second set to verify *attR* (highlighted in Fig. 1). The expected amplicon sizes in the Int test groups varied from 1021 to 1104 bp for *attL* and from 1058 to 1084 bp for *attR*. **d** Bar graph plots showing the viable cells average (circles corresponding to technical replicates averages) and standard deviation normalized with pGFP of OD measurements obtained after MTT assays ($n=3$). DMSO corresponds to the impairment negative control. Negative control cells were transfected with only one of the two vector sets, that is, integrase expression (pIE) or switch GFP (pSG) vectors, plus a mock plasmid. Positive control cells (pGFP) have an *egfp* sequence in the forward orientation under the control of the EF1 alpha promoter. All the data are representative of three technical and three biological replicates.

Additionally, PCR and sequencing analysis, following the same strategy previously described for the mammalian cells, showed *attL/attR* sites formation and *egfp* flipping to the forward orientation for all Ints evaluated ([Fig. 4c](#); Supplementary Figs. [12](#) and [13](#)), even for the Int 5 test condition, in which EGFP-positive cells were detected at a very low level by flow cytometry analysis.

The cell viability in protoplasts after Ints activity was measured using fluorescein diacetate (FDA) hydrolysis assay. Also, as observed in mammalian cells, in protoplasts no marked cell impairment was observed after 24 h of the integrases transformation ([Fig. 4d](#)).

Another important concern related to the use of Ints is the accuracy with which the Ints mediate the site-specific recombination. Therefore, the individual sequence clone reads, obtained from PCR amplicons, were aligned with the expected flipping plasmid sequence, trimmed, and analyzed for all three cell model systems evaluated. In addition to confirming Int functionality, DNA sequencing demonstrated minimal errors in the recombined *attL/attR* sites, indicating that these proteins are not error prone. For only Int 4 in the bovine fibroblast system and Int 7 in the protoplast system, one covered SNP was observed at the *attR* sites. However, mutations were also observed in the *egfp* sequence for most clones sequenced, suggesting that the observed mutations may be due to PCR errors instead of imprecise Int recombination (Supplementary Table [3](#); Supplementary Data [1, 2, 3](#)).

Sequencing analyses also evidenced a one-base difference for the Int 9 *attL/attR* attachment sites compared to the original described by Yang et al.¹⁸ (Supplementary Fig. [14](#)). These data indicate that the adenine after the previously marked *attB* core and a thymine after the *attP* should be part of the Int 9 crossover site. However, this arrangement would lead to a one-base difference between the *attB/attP* cores, making it impossible for dinucleotide overhang recombination to occur after integrase subunit rotation³⁷. Thus, we found a possible arrangement with the next CT dinucleotide as the core for the Int 9 attachment sites.

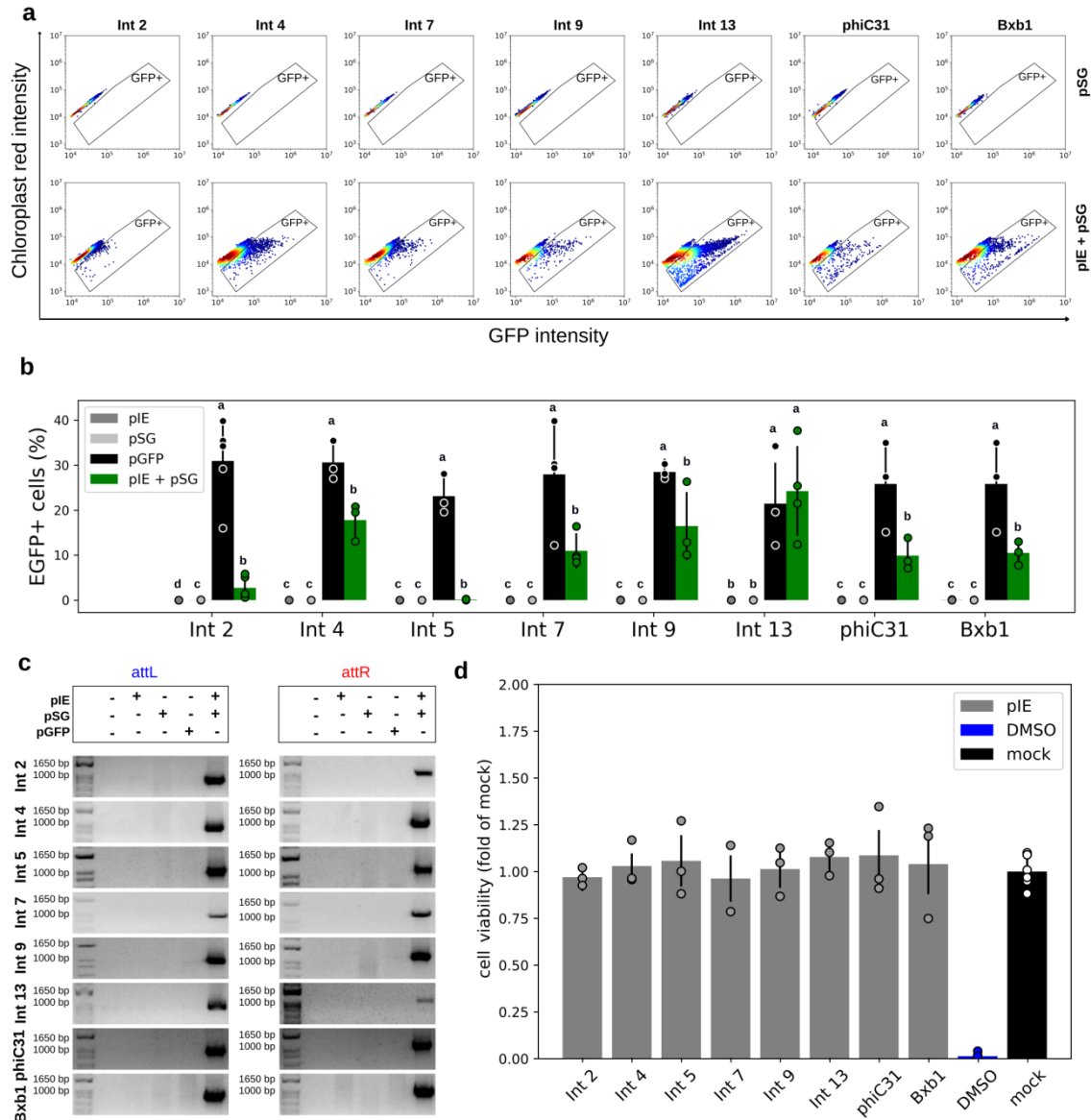


Fig. 4 Functional characterization of the genetic switches in plant protoplasts. **a** Flow cytometry distribution of protoplasts at 24 h post cotransformation for Ints 2, 4, 7, 9, 13, phiC31 and Bxb1, the integrases that led to the highest EGFP-expressing cell frequencies. The heat map indicates the scattering of high cell concentrations (warm colors) to low cell concentrations (cool colors). The gate encompasses the EGFP-expressing cell population. **b** Bar graph plots showing the total average percentage and standard deviation of a cell population expressing EGFP in biological repeat assays (n in Supplementary Table 2) and circles showing the technical triplicate average of each assay on the y axis. The x axis contains the different conditions. For each Int data group, different letters indicate significant differences ($p < 0.05$). **c** Amplicons obtained through PCR analysis using two specific primer sets, the first set to verify *attL* formation and the second set to verify *attR* (highlighted in Fig. 1). The expected amplicon sizes in the Int test groups varied from 948 to 983 bp for *attL* and from 1136 to 1132 bp for *attR*. Negative control cells were transfected with only one of the two vector sets, that is, integrase expression (pIE) or switch GFP (pSG) vectors, plus a mock plasmid. Positive control cells (pGFP) have an *egfp* sequence in the forward orientation under the control of the CaMV 35S promoter. All the data are representative of three technical and three or more biological replicates. **d** Bar graph plots showing the viable cells average (circles corresponding to technical replicates averages) and standard deviation normalized with mock plasmid positive cells obtained after FDA assays. DMSO corresponds to the impairment negative control. The cells were transfected with pIE vectors in technical triplicates and three biological replicates (except for Int 7, for which there was three technical and two biological replicates).

Promoter as switchable genetic part and Ints orthogonality

We wanted to evaluate whether a different genetic component could be used as a switchable part and if these proteins were orthogonal. To this end, *A. thaliana* protoplast was chosen as a model organism, since we observed a high number of cells expressing EGFP in this system, indicating robust Int activity. Three Ints with variable levels of efficiency were selected: Ints 2, 4, and 5. Although Int 13 led to the highest *egfp* activation in protoplasts, it was not selected because there is an ATG start codon in the *attR* site formed after recombination, what would remove the *egfp* gene sequence frame; and because Yang et al.¹⁸ reported a constitutive promoter activity for Int 13 *attP* site.

In this proposal, the switch GFP construct (set 2 plasmid) was redesigned, placing the constitutive promoter, CaMV 35S, in the reverse complement orientation flanked by the *attB* and *attP* sites of Ints 2, 4, and 5 in tandem, followed by the *egfp* sequence in the forward orientation; this construct was named the switch promoter vector (pSP) (Fig. 5a). Flow cytometry data showed 32.2, 38.6, and 12.5% EGFP-positive cells in the protoplast population cotransformed with the pIE Int vectors 2, 4, and 5 + pSP, respectively (Fig. 5b, c; Supplementary Figs. 3 and 15). Residual EGFP accumulation on the pSP negative control cells was observed; however, all three Int tests exhibited statistically significant differences compared to the negative control groups ($p=1.51 \times 10^{-11}$; Supplementary Table 4). Notably, a higher EGFP-positive cell frequency was observed for the Int 2 test than for the flipping strategy with the switch GFP vector.

PCR and sequencing analysis corroborated these results, showing *attL/attR* sites formation and CaMV 35S flipping, driving the RNA polymerase to transcribe the *egfp* gene (Fig. 5d; Supplementary Fig. 16). Because we used primer sets flanking the recognition sites of Ints 2, 4, and 5 (Fig. 5a, activated switch promoter vector), we could observe that each Int only recognized its own *attB/attP* sites (Supplementary Fig. 17), with just one covered deletion observed at the formed *attL* site of Int 5 (Supplementary Table 5, Supplementary Data 4). These results indicate that Ints 2, 4, and 5 are accurate and orthogonal. In conclusion, the promoter sequence can also be used to build genetic switches.

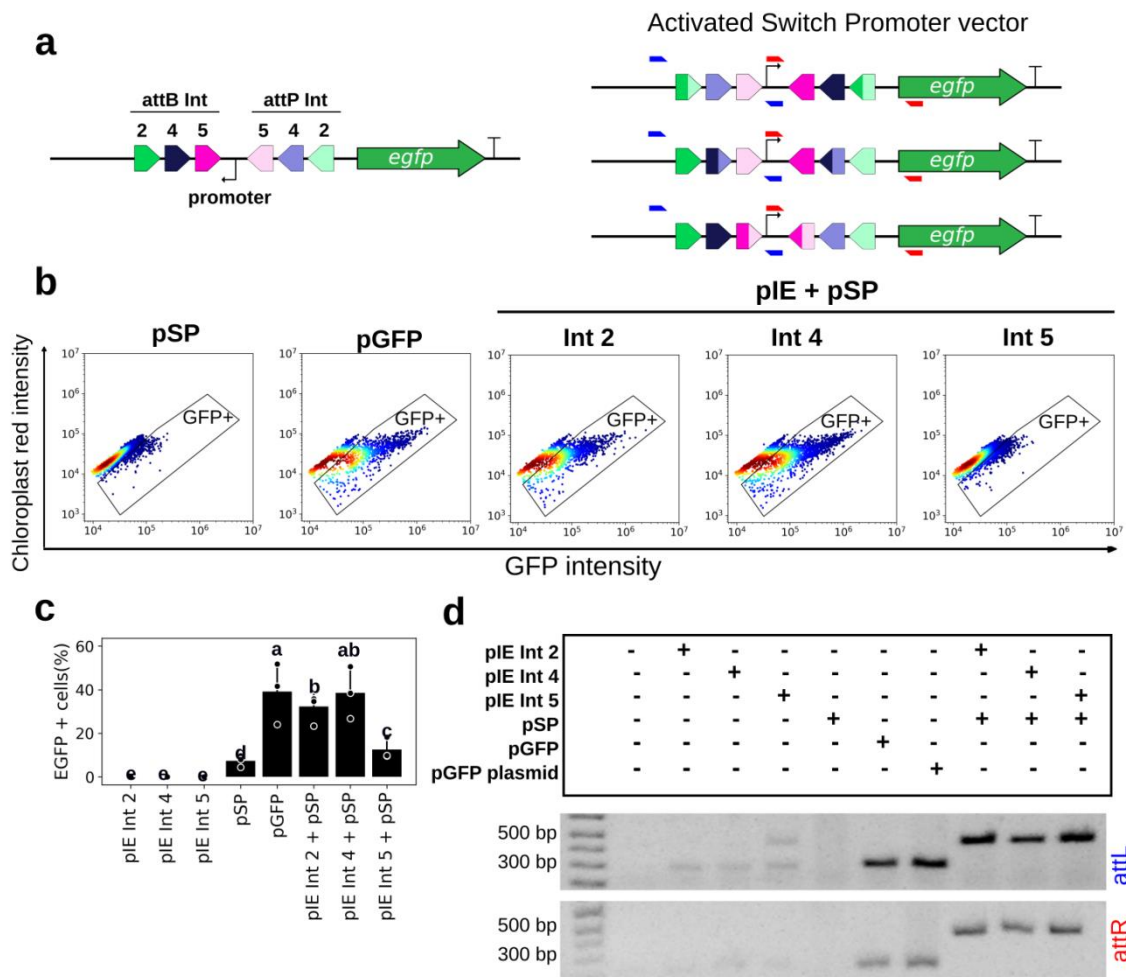


Fig. 5 Promoter as a switchable genetic part in plant protoplasts. **a** Schematic representation of the CaMV 35S promoter as a switchable genetic part. In this case, the switch vector was redesigned to contain the promoter in the reverse complement sequence orientation flanked by the *attB* and *attP* sites of three different Ints (2, 4 and 5) in tandem; this construct was named the switch promoter vector (pSP). The integrase expression vectors (pIE) for each Int were cotransformed separately with the pSP, and each resulting activated promoter vector is shown. **b** Flow cytometry distribution of protoplasts at 24 h post cotransformation. The heat map indicates the scattering of high cell concentrations (warm colors) to low cell concentrations (cool colors). The gate encompasses the EGFP-expressing cell population. **c** Bar graph plots showing the total average percentage and standard deviation of a cell population expressing EGFP in biological repeat assays ($n=3$) and circles showing the technical triplicate average of each assay on the y axis. The x axis contains the different conditions. Different letters indicate significant differences ($p < 0.05$). **d** Amplicons obtained through PCR analysis using two specific primer sets. The first primer set (blue) was complementary to the pSP vector backbone sequence (forward) and promoter sequence (reverse). Expected amplicon size: 429 bp. The second primer set (red) was complementary to the promoter sequence (forward) and *egfp* coding sequence (reverse). Expected amplicon size: 438 bp (primers are colored and marked in letter a). Negative control cells were cotransformed with one of the integrase expression vectors (pIE) plus a mock plasmid or with the switch promoter vector (pSP) plus a mock plasmid. Positive control cells (pGFP) have a plasmid containing the *egfp* sequence under the control of the CaMV 35S promoter in the forward orientation plus a mock plasmid. Expected amplicon size for pGFP: *attL* gel: 285 bp; *attR* gel: 227 bp. These amplicons were smaller than those obtained under the test conditions due to the absence of the Int attachment sites. All the data are representative of three technical and three biological replicates.

Ints activity in primary human T lymphocytes and stem cells

As a proof of concept, unidirectional genetic switches using integrase expression (pIE) and switch GFP (pSG) vectors were tested in human peripheral blood mononuclear cells (PBMCs), the main source of T lymphocytes. These cells were chosen due to their significance in several different areas of health research, such as vaccine development, hematological malignancies, high-throughput screening, and immunology. Recently, T lymphocytes have gained importance in cancer immunotherapies^{38,39} through their clinical use with or without genetic manipulation. Considering the results obtained for HEK 293T cells, we selected Int 13, an integrase with highly detectable functionality, and Int 4, which showed a low frequency of EGFP-expressing cells, to be tested with PBMCs; phiC31 and Bxb1 were also evaluated. Flow cytometry analysis showed that the Int 13 and Bxb1 tests resulted in the same EGFP-expressing cell frequency (7%) in PBMCs extracted from three healthy independent donors ([Fig. 6a, b](#); [Supplementary Fig. 18](#)). These results indicate that Int 13 and Bxb1 were able to promote the inversion of the *egfp* coding sequence to the forward orientation, leading to that high EGFP-expressing cells frequency, equivalent to 87.5% of the result observed in the positive control population (pGFP).

PCR and sequencing analysis performed using the same strategy previously described for the other switch GFP assays corroborated these data and expanded the data, showing that Ints 4 and phiC31 were also capable of flipping the *egfp* coding sequence ([Fig. 6c](#); [Supplementary Figs. 19](#) and [20](#)). The DNA sequence reads obtained were also aligned with the activated plasmid expected sequence, and the covered mutations were counted. Only a few SNPs were observed at the formed *attL/attR* sites ([Supplementary Table 6](#), [Supplementary Data 5](#)).

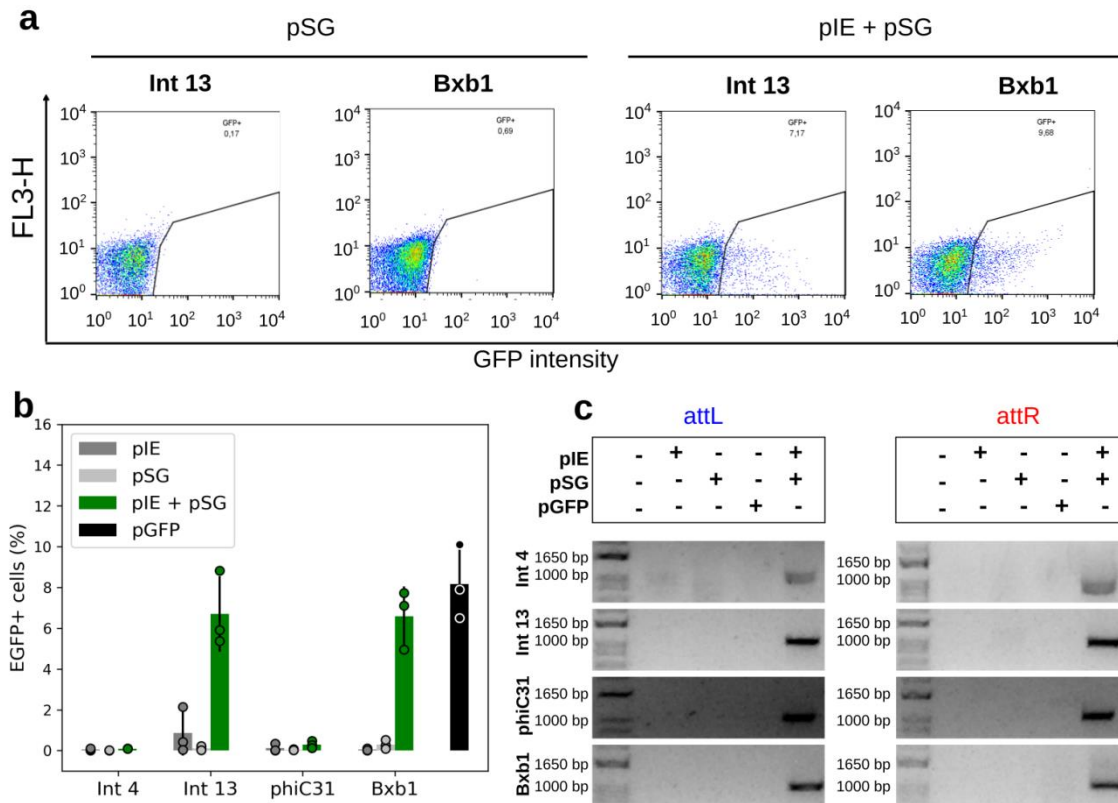


Fig. 6 Int activity in primary human T lymphocytes from PBMCs isolated from three independent donors. **a** Flow cytometry distribution of PBMCs at 48 h post electroporation for GFP switch tests with Int 13 and Bxb1, the integrases that led to the highest EGFP-expressing cell frequencies. The heat map indicates the scattering of high cell concentrations (warm colors) to low cell concentrations (cool colors). The gate encompasses the EGFP-expressing cell population. **b** Bar graph plots showing the total average percentage and standard deviation of a cell population expressing EGFP in biological repeat assays ($n=3$) and circles corresponding to single data points from each donor material. The x axis contains the different conditions. In the PBMCs assays, Ints 2, 13, phiC31 and Bxb1 were evaluated. **c** Amplicons obtained through PCR analysis using two specific primer sets, the first set to verify *attL* formation and the second set to verify *attR* (highlighted in Fig. 1). The expected amplicon sizes in the Int test groups varied from 1021 to 1104 bp for *attL* and from 1058 to 1084 bp for *attR*. Negative control cells were electroporated with only one of the two vector sets, that is, integrase expression (pIE) or switch GFP (pSG) vectors. Positive control cells (pGFP) have an *egfp* sequence in the forward orientation under the control of the EF1 alpha promoter. All the data were representative of three donors, corresponding to biological triplicates with single measurements.

Additionally, this study aimed to investigate Int functions in another human cell type used as an efficient *in vitro* model in studies on several diseases and embryogenesis. Switch systems using integrase expression vectors (pIE Ints 2, 9, 13, phiC31 and Bxb1) and the respective switch GFP (pSG) vectors were thus evaluated in neural stem cells (NSCs), and undifferentiated human embryonic stem (hES) cells. The plasticity of these cells has made them the focus of basic developmental research and also of the challenging regenerative medicine field. As an example, NSCs exhibit promise in the treatment of neurodegenerative diseases⁴⁰. The hES-BR-1 cell line was the first hES cell line established from the Brazilian population and

is a relevant model for stem cell assays⁴¹. The flow cytometry measurements were within a narrow range, with 2.0 to 3.9% EGFP-positive cells observed in the Int 2, 9 and 13 tests (Fig. 7), indicating that these Ints are functional in these relevant disease modeling systems. PhiC31 and Bxb1 tests exhibited the highest values, showing activity in between 9.6 and 24.8% of EGFP-positive cells (Fig. 7).

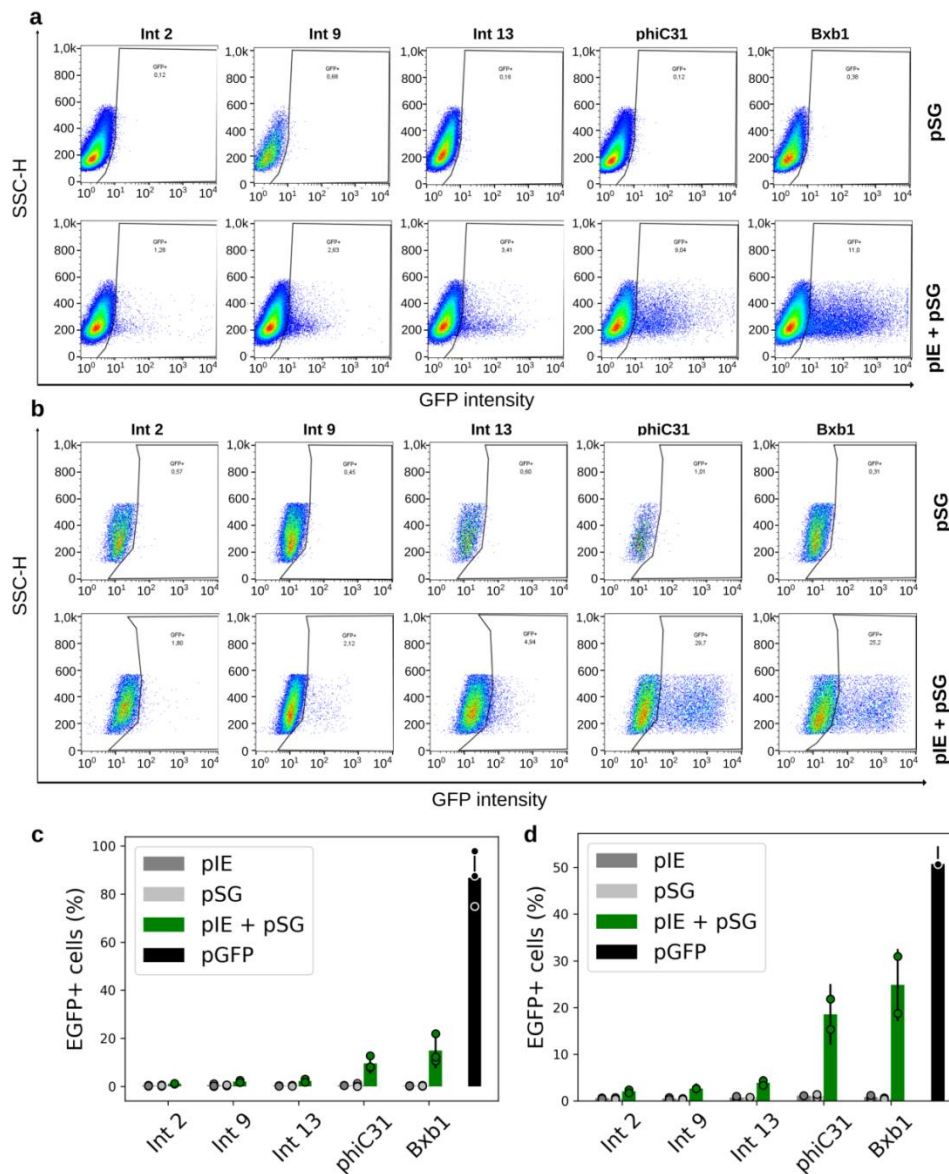


Fig. 7 Int activity in stem cells. Flow cytometry distribution of NSCs (a) and hES cells (b) at 48 h post electroporation for GFP switch tests with Ints 2, 9, 13, phiC31, and Bxb1. The heat map indicates the scattering of high cell concentrations (warm colors) to low cell concentrations (cool colors). The gate encompasses the EGFP-expressing cell population. Bar graph plots showing the total average percentage and standard deviation of NSCs (c) and hES cell (d) populations expressing EGFP in biological repeat assays (n=3 for NSCs; n=2 for hES cells) and circles showing the technical triplicate average of each assay on the y axis. The x axis contains the different conditions. In the stem cell assays, Ints 2, 9, 13, phiC31, and Bxb1 were evaluated. Negative control cells were transfected with only one of the two vector sets, that is, integrase expression (pIE) or switch GFP (pSG) vectors. Positive control cells (pGFP) have an *egfp* sequence in the forward orientation. All the data are representative of three technical and three (NSCs) or two (hES cells) biological replicates.

Discussion

Although tools based on activators and repressors, including recently modified CRISPR-Cas9 systems⁴², are used to regulate eukaryotic gene expression, there remains a need for new technologies that are scalable and precise for multiplex gene regulation control. In our study, through eukaryotic cell-based assays, we demonstrated the widespread use of six Ints 2, 4, 5, 7, 9, and 13 as genetic switchers in mammalian and plant cells. All the Ints tested were able to i) recognize their respective predicted *attB/attP* sites, ii) generate the predicted *attL/attR* sites, and iii) invert the *egfp* or promoter sequences. Furthermore, our results demonstrated that these six Ints have different degrees of functionality depending on the eukaryotic cell type or the genetic part to be flipped, potentially leading to a tunable system. Of the six Ints evaluated, the flow cytometry analysis showed that Int 13 led to the highest proportion of EGFP-expressing cells in all three eukaryotic model systems. Int 13 (NCBI No. WP_012095429.1) is a protein with 55.5 KDa in a range of 54.0 to 67.1 KDa among the eight integrases evaluated, and it was identified from a prophage inside *Bacillus cytotoxicus* NVH 391-98^{18,43}. Int 9 for bovine fibroblasts and Int 4 and Int 9 for plant protoplasts were set at the second position. PhiC31 and Bxb1, characterized elsewhere, yielded the highest EGFP percentage values only for HEK 293T cells, slightly overcoming Int 13 activity. On the other hand, for bovine fibroblasts, Bxb1 exhibited activity close to that of Int 9, and phiC31 showed low activity. In protoplasts, these integrases promoted an intermediate effect, similar to that of Int 7. Ints 9, 4, and 13 yielded higher percentages of EGFP-positive cells than phiC31 and Bxb1 in this plant model. Moreover, interestingly, in the study conducted by Yang et al.¹⁸ in bacteria, Ints 2, 4, 5, 7, and 13 yielded 100% GFP-positive cell populations and Int 9 yielded approximately 80%. In eukaryotic model systems, different inherent factors of these highly complex organisms can be hypothesized to interfere with integrase functionality, leading to different overall results. However, this variability of Ints potency in eukaryotic systems can be used to design genetic circuits with distinct functionalities and modulation proprieties.

Despite none of these Ints were previously evaluated as regulators of gene expression in eukaryotic cells, two studies evaluated different Ints performing other functions in yeast (*Saccharomyces cerevisiae*) and in two mammalian cell types^{25,29}. These authors showed that ten out of 14 integrases were active in *S. cerevisiae*

through recombinase exchange reactions and that seven out of 15 Ints promoted site-specific deletions in both lineages of mammalian cells. They also showed that seven Ints were cytotoxic in yeast and that all 15 were cytotoxic in mammalian cells to some degree, and some of the Ints were error prone in the mammalian systems. Here, the *attL/attR* sites were amplified and sequenced, and no mutations in these sites were found for Ints 2 and 9 in any of the models tested. Regarding the other Ints, only a few mutations were observed in some models. Additionally, our cell viability assays indicated no limiting toxicities related to the integrases activity, suggesting that the systems reported herein can be used in stable expression-based experiments.

Taking into account the CaMV 35S promoter-based switch construction for protoplasts, despite a residual leak observed in the switch promoter vector negative control, Int 4 and Int 5 led to a result that was compatible with the switch GFP system. Int 2, however, exhibited a much higher number of EGFP-positive cells when the promoter was flipped than when the *egfp* sequence was. Moreover, this system suggested that these three Ints are orthogonal, as previously demonstrated in bacteria¹⁸.

Therefore, these proteins can be multiplexed in different combinations of genetic switches or logic gates based on Boolean algebra to build genetic circuits. Such applications would facilitate differential control of metabolic routes or synthetic transgenic systems in livestock animals or crop plants, allowing precise and efficient regulation of gene expression regarding, for example, abiotic/biotic conditions or growth timing responses to activate or deactivate resistance, defense, or nutritional improvement genes.

An important challenge to be considered in studies that aim to develop biotechnological tools with wide applicability is proof of concept. Here, in addition to the evaluation of Ints in model systems, unidirectional genetic switches were evaluated in cells of great clinical relevance, such as primary T lymphocytes, and stem cells^{39,44,45}. The therapies based on T lymphocytes and stem cells represent a field of research with a demand for transcription systems with fine-tuning capabilities⁴⁶, such as the Int-based systems described in this study. As for the model systems HEK 293T cells, bovine fibroblasts, and plant protoplasts, our results in PBMCs, NSCs and hES cells showed that the evaluated Ints were able to recognize the *attB/attP* sites, precisely forming the *attL/attR* sites and performing the 180°

rotation (flip) of the *egfp* coding sequence. Accordingly, not all the Ints tested led to high frequencies of EGFP detection by flow cytometry and microscopy analysis, indicating differential Int activity in these cells. These results are very relevant because, first, we were able to show that these phage proteins are active in eukaryotic cells used in a variety of biologically important studies, reinforcing the robustness of the Int platform⁴⁷. Second, we can foresee the construction of genetic circuits to improve the already successful cancer immunotherapy strategies and for a wide range of potential applications of Ints in disease modeling *in vitro* and in therapeutic-based approaches in human stem cells. One important contribution in this context can be using Ints to refine specific temporal/space gene activation/deactivation, minimizing potential undesired side effects.

The use of Ints can also be expanded to the study of essential genes in eukaryotic organisms. Recently, Cre/Lox tyrosine integrases were utilized in a SCRaMBLE analysis to identify nonessential genes from chromosome III of *S. cerevisiae*⁴⁸. However, due to the use of only one integrase, all genes called nonessential were extracted from the genome of this organism at the same time, resulting in immediate loss of cell viability due to unknown genome redundancy⁴⁸. Our work led to a considerable increase in the number of functionally characterized Ints that were available for use in eukaryotic cells. Now, these Ints can be multiplexed to flank several endogenous genes in random combinations, making it possible to knockout separate groups of genes. Ultimately, this multiplex strategy can allow the investigation of functional gene redundancy on the genome-scale. Furthermore, as different Ints have different degrees of functionality, these proteins can be used in studies of vital multifamily genes, allowing modulation of gene expression, or even to evaluate extrachromosomal toxic protein-coding genes. Due to their accuracy, Ints can also be used to investigate the roles of specific domains of selected genes or gene families, flanking predicted functional domains with their *attB/attP* sites and triggering the flipping of these sequences to the nonfunctional reverse orientation in a specific tissue or developmental stage. Finally, these proteins can also be used as DNA barcodes to identify cell lineages in developmental and evolutionary studies⁴⁹.

Importantly, in this study, six Ints were evaluated, so the number of functional Ints in eukaryotic cells can be substantially increased by taking into account the pool

of more than 4,000 integrases observed by Yang et al.¹⁸ in sequence databases with predicted recognition sites for 34 of these proteins.

Furthermore, studies regarding nuclear localization and protein accumulation could be performed, leading to improvement in Int functionality with the addition of NLSs or removal/modification of putative degradation signals. Last, the results presented in this work indicate that Ints 2, 4, 5, 7, 9, and 13 can further be used for a myriad of biotechnological applications.

Methods

Integrases and plasmids. The serine integrases 2, 4, 5, 7, 9, 13¹⁸, phiC31⁵⁰ and Bxb1²² were codon optimized by an online codon optimization tool (IDT software) for expression compatibility in the respective eukaryotic system: for mammalian cells, the proteins were codon optimized for expression in *Homo sapiens*, and for protoplasts, the proteins were codon optimized for expression in *A. thaliana*. Two plasmid sets were synthetically constructed (Epoch Life Science Inc.) to compose the genetic switch systems. The first plasmid set was constructed to express the Int gene (integrase expression vectors (pIE)). The second plasmid set, the switch vectors, carried the reporter *egfp* gene (switch GFP vectors (pSG)) or its promoter (switch promoter vector (pSP)) in reverse complement orientation, flanked by the *attB* site and the reverse complement sequence of the *attP* site of the Int expressed by the first plasmid. Addgene accession numbers of all vectors used in this study are described in Supplementary Table [1](#). Vector's part sequences are also provided in Supplementary [Methods](#).

Mammalian system plasmids. For the integrase expression vectors, Ints 2, 4, 5, 7, 9, 13, phiC31 and Bxb1 were placed under the control of the ubiquitin promoter and β -globin poly(A) signal terminator. The coding sequences of these integrases were cloned into the pUB-GFP plasmid (Addgene, 11155), replacing the GFP-coding sequence, which resulted in a set of integrase-expressing vectors called pUB-HspINTX (X=2, 4, 5, 7, 9, 13, phiC31 or Bxb1). For the switch GFP vectors, the *egfp* coding sequence (Addgene, 11154) was cloned in reverse complement orientation, flanked by the *attB* site and the reverse complement sequence of the *attP* site of the individual Ints 2, 4, 5, 7, 9, 13, phiC31 and Bxb1. These cassettes were cloned into

the pEF-GFP plasmid (Addgene, 11154), replacing the original *egfp* coding sequence under the regulation of the EF1 alpha promoter and β -globin poly(A) signal terminator. The resulting switch GFP vectors were named pEF-GFP(rc)X (X=2, 4, 5, 7, 9, 13, phiC31 or Bxb1) (Supplementary Fig. 1a). The GFP expression-positive control pT3-Neo-EF1 α -GFP (Addgene, 69134) was used for HEK 293T and PBMCs assays; the pEF-GFP plasmid (Addgene, 11154) was used for bovine fibroblast assays; and the pT2-GFP (previously kindly provided⁵¹) was used for NSCs and hES cell assays.

Plant system plasmids. In this system, the Ints 2, 4, 5, 7, 9, 13, phiC31 and Bxb1 CDS, under the control of the actin2 gene promoter⁵² and NOS terminator (pBI426 plasmid⁵³), were cloned into the pUC57, pSB3K3 or pBluescript II SK(-) vectors by Epoch Life Science Inc. These plasmids resulted in a set of integrase expression vectors, individually called pAct-AtINTX (X=2, 4, 5, 7, 9, 13, phiC31 or Bxb1). The *egfp*³⁶ coding sequence in a reverse complement orientation, flanked by *attB* and the reverse complement sequence of the *attP* attachment site of the Ints 2, 4, 5, 7, 9, 13, phiC31 and Bxb1, was placed under the control of the CaMV 35S promoter (iGEM registry BBa_K1547006) and the NOS terminator, constituting a set of switch GFP vectors for plants. These cassettes were inserted into the plasmids pUC18 or pBluescript SK(-) (Epoch Life Science Inc.), resulting in p35S-GFP(rc)X (X=2, 4, 5, 7, 9, 13, phiC31 or Bxb1) (Supplementary Fig. 1b). For the plant system, a second switch vector was constructed. This vector consisted of the reverse complement of the CaMV 35S promoter flanked by the *attB* and the reverse complement sequence of the *attP* sites of the Ints 2, 4 and 5, sequentially positioned together (Fig. 5a). This promoter cassette was synthesized (Epoch Life Science Inc.) and cloned, replacing the CaMV 35S sequence in the positive control vector pCaMV35S-GFP³⁶. The final vector was called p35S(rc)2_4_5-GFP and was used as the switch promoter vector (pSP).

HEK 293T maintenance and PBMCs isolation. The human embryonic kidney cell lineage HEK 293T (a gift from Dr. Elio Vanin of St. Jude Children's Research Hospital) was cultivated in 75-cm² tissue culture flasks with 15 ml of Dulbecco's modified Eagle's medium (DMEM; Gibco), 10% fetal bovine serum (FBS; HyClone) and penicillin-streptomycin (10 U/ml; Gibco). The cells were detached and seeded

every 2-3 days at below 80% confluence. For PBMCs, white blood cells from healthy blood donors were collected using a leukocyte reduction filter (RS; Haemonetics) and washed with phosphate-buffered saline (PBS). To isolate PBMCs, a density gradient centrifugation was performed using Ficoll-Hypaque®-1077 (GE HealthCare) (deceleration off; centrifugation for 20 min at 800 x g) followed by three washes with PBS. The use of PBMCs from healthy donors was approved by an Institutional Review Board (the Brazilian National Cancer Institute (INCA) Ethics Committee), and donors signed review board-approved informed consent forms.

Human stem cell culture. Neural stem cells (NSCs) differentiated from induced pluripotent stem cells (iPSCs)⁵⁴ and pluripotent human embryonic stem cells (hES, BR-1 cell line)⁴¹ were used. All human stem cell experiments were approved by the ethics committee of Copa D'Or Hospital (CAAE number 60944916.5.0000.5249, approval number 1.791.182). The cells were cultured/maintained in neural advanced DMEM/F12 and neurobasal medium (50% v/v) plus neural induction supplement (NIS) medium (all from Thermo Fisher Scientific), called NEM (neural expansion medium), over Geltrex (Thermo Fisher Scientific) at 37°C in 5% CO₂ as previously described^{54,55}. hES cells were cultured in mTeSR medium (Stem Cell Technologies) over Geltrex.

HEK 293T transfection. A total of 4×10^6 HEK 293T cells were plated in 75-cm² flasks with 15 ml of DMEM (Gibco), 10% FBS (HyClone) and penicillin-streptomycin (10 U/ml; Gibco). After 24 h, the medium was removed, and 10 ml of fresh DMEM/FBS was added. In the next step, 5 µg of each integrase expression vector pUB-IntX (X= 2, 4, 5, 7, 9, 13, phiC31 or Bxb1) and each switch GFP vector pEF-GFP(rc)X carrying the respective integrase site were added to 500 µl of 2X CaCl₂ at 250 nM. Next, 500 µl of HBS (pH 7.1) was slowly added while the solution was vortexed at 10,000 rpm. Bubbles were produced in the solution with a Pasteur pipette and mixed. The solution was incubated for 10 min at room temperature and then dripped with the Pasteur pipette throughout the flask. The medium was changed after 16 h. The results were analyzed 48 h after the transient transfection.

PBMC, NSC, and hES electroporation. Cells (1×10^7 PBMCs, 1×10^6 NSCs or hES cells) were transferred to a sterile 0.2-cm cuvette (Mirus Biotech®) and

electroporated as previously described⁵¹. Briefly, PBMCs and hES cells were resuspended in 100 µl of 1SM buffer, and NSCs were resuspended in 100 µl of 1S buffer. PBMCs were electroporated with 5 µg of each integrase expression (pIE) and switch GFP (pSG) plasmids using the U-14 program of the Lonza® Nucleofector® II electroporation system. NSCs and hES cells were electroporated with 12 µg of pIE plasmids and 8 µg of pSG plasmids using A-33 and A-23, respectively, from the Lonza® Nucleofector® II electroporation system. The mock control was electroporated with 100 µl of 1SM (PBMCs and hES cells) or 1S (NSCs) buffer without plasmid (used to set flow cytometry gates). After transfection, PBMCs were gently resuspended in 1 ml of prewarmed RPMI medium supplemented with 2 mM L-Glu and 20% fetal calf serum (FCS; Gibco). NEM was used for NSCs, and mTeSR was used for hES cells. All cells were cultured for 48 h after electroporation for transient transfection, and then, the analyses were performed.

Bovine fibroblast isolation. Fibroblasts were isolated according to the protocol described by Freshney⁵⁶ with some modifications. The cells were removed from 14-month-old Nelore (*Bos indicus*) bull oxtail by biopsies and washed three times in 0.05% trypsin (Gibco). The cells were then transferred to 25-cm² cell culture flasks and incubated in DMEM (Gibco) supplemented with 10% FBS (Gibco) and penicillin-streptomycin at 37°C in a 5% CO₂ atmosphere. After three passages, or when the fibroblast cultures showed homogeneity, the cells were ready for transfection. Cell cultures with 60% to 70% confluence were picked. The use of the bovine cells was approved by the Ethics Committee on the Use of Animals (CEUA) of Embrapa Genetic Resources and Biotechnology in March 2013 under the reference number 001/2013.

Bovine fibroblast cell transfection. After growth, the cells were enzymatically dissociated with a trypsin solution (0.5% trypsin, 0.2% EDTA), and after 10 min, the reaction was inactivated using DMEM (Gibco). The cells were counted in a Neubauer chamber, transferred (10⁵ cells/well) to 24-well culture dishes and grown for 24 h, when they reached >70% confluence. Primary bovine fibroblasts were cotransfected with 350 ng of each of the two plasmid sets utilizing Lipofectamine LTX & Plus Reagent (Invitrogen) and cultured in Opti-MEM (Invitrogen) according to the

manufacturer's instructions. The results were analyzed 48 h after transient transfection.

Protoplast isolation. The protoplasts were obtained following the protocol described by Yoo et al.⁵⁷ with some modifications. *A. thaliana* ecotype Columbia was grown under a 12 h light/12 h dark cycle at 22°C. Four to 5 weeks after seeding, approximately 20 young leaves were collected, transferred to a plate with W5 solution (154 mM NaCl, 125 mM CaCl₂, 5 mM KCl, 2 mM MES, pH 5.7), and scalped by using a blade. Then, the leaves were placed on a digestion plate containing 5 ml of enzyme solution [0.5 M mannitol, 20 mM KCl, 20 mM MES (pH 5.7), 0.2% pectolyase (Sigma-Aldrich), 0.5% driselase (Kyowa Hakko Bio Co., Ltd.), 1.5% cellulase (Sigma-Aldrich), 10 mM CaCl₂, 1 mg/ml BSA]. A 15-20 pol Hg vacuum was applied three times for approximately 5 seconds, and the plate was incubated at room temperature in a platform shaker at 40 rpm for 3 h. The digested sample was filtered through a 44- μ m mesh, and the W5 solution was added to wash the obtained protoplasts, followed by centrifugation at 100 x g for 2 min. After two additional washing steps followed by centrifugation with 10 ml of W5, the protoplasts were resuspended in 1 ml of MMg solution (0.4 M mannitol, 15 mM MgCl₂, 4 mM MES, pH 5.7), and the concentration was adjusted to 4-5 x 10⁵ protoplasts/ml.

Protoplast transformation. Cotransformation was performed in a 15-ml Corex tube using 100 μ l of 4-5 x 10⁵ protoplasts/ml, 10 μ g of each desired plasmid DNA of the two sets of vectors and 110 μ l of 40% PEG solution [PEG 4000 (Sigma), 0.2 M mannitol, 100 mM CaCl₂] for each reaction (scaling up to 6 reactions per tube). After 15 min, the reaction was stopped with 2 volumes of W5 solution, centrifuged at 100 x g for 2 min, resuspended in 500 μ l of W1 solution (0.5 M mannitol, 20 mM KCl, 4 mM MES, pH 5.7) (each reaction) and plated on a 12-well plate. The plates were incubated at room temperature in a platform shaker at 40 rpm, and the results were analyzed after 24 h transient transformation.

Flow cytometry. HEK 293T cells, NSCs, hES cells, bovine fibroblasts, and *A. thaliana* protoplasts were analyzed by flow cytometry in technical triplicates. PBMCs were analyzed in one sample for each of the three donors. HEK 293T cells, PBMCs, NSCs and hES cells were detached/resuspended from the culture flasks with cold

PBS and washed again with PBS. Signal acquisition was performed with a FACS Calibur (BD Biosciences), and the cells with FSC/SSC patterns compatible with viable cells were gated. Then, a second selection was done for 7AAD negative population. Analyses were performed using FlowJo software, version 10 (Tree Star). Bovine fibroblasts were trypsinized from 24-well plates, where 200 μ l of DMEM was added per well. The contents were transferred to a microtube and centrifuged at 687 x g for 5 min, and the supernatant was removed, leaving approximately 50 μ l for analysis. Protoplasts from each well of 12-well plates were transferred to a microtube and centrifuged at 100 x g for 1 min, and the supernatant was removed, leaving a volume of approximately 50 μ l. The entire contents of each tube were analyzed. Bovine and protoplast cells were analyzed on an imaging flow cytometer (Amnis FlowSight) under 488-nm laser excitation and a power of 60 and 10 mW, respectively. Signal acquisition in bovine cells was determined by channel 3 intensity (filters at 566-635) versus channel 2 intensity (green EGFP reporter emission, 532-555 nm). Signal acquisition in the protoplast population was determined by channel 4 intensity (chloroplast autofluorescence red emission, 610/30 nm) versus channel 2 intensity. Before the first reading, a no-gating population was acquired under intensity channels and the imaging system from the Amnis FlowSight cytometer allowed visually gating the viable cells population. Once the gate was set up, the samples were acquired using this gate. The results from bovine and protoplast cells were processed and visualized using IDEAS software. For mammalian systems, at least 10,000 single cells were analyzed, and for protoplasts, at least 1,000 single cells were analyzed.

DNA extraction, amplification, and sequencing. For HEK 293T cells, PBMCs, and bovine fibroblasts, total DNA was extracted from the pool of technical triplicates using DNeasy Blood & Tissue Kits (Qiagen) after a 48-h assay. Plant protoplasts were also pooled after a 24-h assay, and the DNA was extracted using the DNeasy Plant Mini Kit (Qiagen). For all samples, the PCRs were carried out using appropriate primer pairs (Supplementary Table [7](#)) to amplify the *attL* and *attR* site-containing regions from the reporter switch vectors (pSG and pSP) using Platinum Taq DNA polymerase (Invitrogen). The Wizard® SV Gel and PCR Clean-Up System (Promega) was used to clean the expected amplicon from the agarose gel. The PCR products were cloned into the pGEM-T Easy vector (Promega) according to the manufacturer's protocol

and transformed into *E. coli* DH10B or XL1-blue chemically competent cells by heat shock. The plasmid DNA was extracted by the Wizard® Plus SV Miniprep DNA Purification System (Promega) and sequenced by Macrogen, always using the M13F and M13R or SP6 and T7 primer pair for coverage sequencing. The obtained sequences were aligned with the expected activated plasmid sequences and trimmed using Geneious software (version 7.0.6).

Cell viability assays. HEK 293T cells and bovine fibroblasts were plated in 96 well plates at a density of 1×10^5 cells/well in triplicates and grown for 24 h at 37°C in a 5% CO₂ atmosphere. The cells were cotransfected with pIE and pSG vectors or only with one of these vectors plus a mock plasmid as previously described proportionally to a 96 well plate assay. For the impairment negative control were used 20 µL dimethyl sulfoxide (DMSO; Sigma-Aldrich) for a final volume of 200 µL per well. After 48 h transient cotransfection, the cells were incubated with 15 µL of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT; Thermo Fisher Scientific) (5 mg/mL) for 4 h at 37°C. Subsequently, the medium was removed and 150 µL DMSO (Sigma-Aldrich) was added to each well to dissolve formazan crystals. A microplate reader (Sunrise reader, Tecan) calibrated to read absorbance at 595 nm was used to quantify the formazan product. Protoplasts were transformed with pIE vector of each integrase or with a mock plasmid in triplicates as previously described. The impairment negative control was obtained with 50 µL DMSO (Sigma-Aldrich) added per well to complete the final incubation volume of 500 µL. The fluorescein diacetate (FDA; Sigma-Aldrich) assay was performed following Lin et al.⁵⁸ protocol with some modifications. After 24 h transient transformation, each sample was incubated with 3 µL FDA work solution [10 µL stock solution (5 mg/mL) in 2.5 mL W1 solution] by 8 min and analyzed by flow cytometry as previously described. The work solution was remade every 2 h. Channel 2 intensity was used to acquire FDA positive cells population in an adequate gate corresponding to viable cells.

Statistics and Reproducibility

The dataset for the genetically activated cell proportion (EGFP+) obtained by flow cytometry was analyzed using R software (version 3.6.0). The nonparametric Kruskal-Wallis test was used to determine significant differences between controls and test conditions of each Int group at the 5% statistical probability level.

Data Availability

All plasmids constructed for this study are available in [Addgene](#) repository and the accession numbers are listed in Supplementary Table 1. Also, the genetic sequence parts used are listed in Supplementary Information. The complete sequence alignment data set is available in Supplementary Data. The full dataset of positive GFP cell proportion obtained by flow cytometry analysis in all cell experiments and the OD acquisitions of MTT assays were deposited in Dryad Digital Repository ([doi:10.5061/dryad.dr7sqv9tv](https://doi.org/10.5061/dryad.dr7sqv9tv)⁵⁹). Any other data are available from the corresponding authors upon request.

References

1. Wirth, T., Parker, N. & Ylä-herttua, S. History of gene therapy. *Gene* **525**, 162–169 (2013).
2. Ueki, H., Watanabe, M., Kaku, H., Huang, P. & Li, S. A novel gene expression system for detecting viable bladder cancer cells. *International J. Oncol.* **41**, 135–140 (2012).
3. Meer, J. R. Van Der & Belkin, S. Where microbiology meets microengineering : design and applications of reporter bacteria. *Nat. Rev. Microbiol.* **8**, 511–522 (2010).
4. Fasani, E., Manara, A., Martini, F., Furini, A. & Dalcorso, G. The potential of genetic engineering of plants for the remediation of soils contaminated with heavy metals. *Plant. Cell Environ.* **41**, 1201–1232 (2018).
5. Christou, P. Plant genetic engineering and agricultural biotechnology 1983 – 2013. *Trends Biotechnol.* **31**, 125–127 (2013).
6. Ahmad, N. & Mukhtar, Z. Genetic manipulations in crops : Challenges and opportunities. *Genomics* **109**, 494–505 (2017).
7. Eenennaam, A. L. Van. Genetic modification of food animals. *Curr. Opin. Biotechnol.* **44**, 27–34 (2017).
8. Lewandoski, M. Conditional control of gene expression in the mouse. *Nat. Rev. Genet.* **2**, 743–755 (2001).
9. Hernandez-garcia, C. M. & Finer, J. J. Identification and validation of promoters and cis -acting regulatory elements. *Plant Sci.* **217–218**, 109–119 (2014).
10. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410–422 (2009).
11. Khalil, A. S. & Collins, J. J. Synthetic biology : applications come of age. *Nat. Publ. Gr.* **11**, 367–379 (2010).
12. Arber, W. Host-Controlled Modification of Bacteriophage. *Annu. Rev. Microbiol.* **19**, 365–378 (1965).
13. Merrick, C. A., Zhao, J. & Rosser, S. J. Serine Integrases: Advancing Synthetic Biology. *ACS Synth. Biol.* **7**, 299–310 (2018).
14. Stark, W. M. Making serine integrases work for us. *Curr. Opin. Microbiol.* **38**, 130–136 (2017).
15. Smith, M. C. M. Phage-encoded Serine Integrases and Other Large Serine

- Recombinases. in *Mobile DNA III* 253–272 (American Society of Microbiology, 2014). doi:10.1128/microbiolspec.MDNA3-0059-2014
16. Rutherford, K. & Van Duyne, G. D. The ins and outs of serine integrase site-specific recombination. *Curr. Opin. Struct. Biol.* **24**, 125–131 (2014).
 17. Brown, W. R. A., Lee, N. C. O., Xu, Z. & Smith, M. C. M. Serine recombinases as tools for genome engineering. *Methods* **53**, 372–379 (2011).
 18. Yang, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nat. Methods* **11**, 1261–1266 (2014).
 19. Olorunniji, F. J., Rosser, S. J. & Stark, W. M. Site-specific recombinases : molecular machines for the Genetic Revolution. *Biochem. J.* **473**, 673–684 (2016).
 20. Nash, H. A. Integration and Excision of Bacteriophage λ : The Mechanism of Conservative Site Specific Recombination. *Annu. Rev. Genet.* **15**, 143–167 (1981).
 21. Rutherford, K., Yuan, P., Perry, K., Sharp, R. & Van Duyne, G. D. Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res.* **41**, 8341–8356 (2013).
 22. Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P. & Endy, D. Amplifying Genetic Logic Gates. *Science*. **340**, 599–603 (2013).
 23. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31**, 448–452 (2013).
 24. Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S. & Lu, T. K. Synthetic recombinase-based state machines in living cells. *Science*. **353**, aad8559 (2016).
 25. Xu, Z. & Brown, W. R. A. Comparison and optimization of ten phage encoded serine integrases for genome engineering in *Saccharomyces cerevisiae*. *BMC Biotechnol.* **16**, 13 (2016).
 26. Weinberg, B. H. *et al.* Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat. Biotechnol.* **35**, 453–462 (2017).
 27. Wang, X. *et al.* Bxb1 integrase serves as a highly efficient DNA recombinase in rapid metabolite pathway assembly. *Acta Biochim. Biophys. Sin. (Shanghai)*. **49**, 44–50 (2017).
 28. Kapusi, E., Kempe, K., Rubtsova, M., Kumlehn, J. & Gils, M. phiC31 Integrase-Mediated Site-Specific Recombination in Barley. *PLoS One* **7**, e45353 (2012).
 29. Xu, Z. *et al.* Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol.* **13**, 87 (2013).
 30. Thomson, J. G. *et al.* The Bxb1 recombination system demonstrates heritable transmission of site-specific excision in Arabidopsis. *BMC Biotechnol.* **12**, 9 (2012).
 31. Fogg, P. C. M., Haley, J. A., Stark, W. M. & Smith, M. C. M. Genome Integration and Excision by a New Streptomyces Bacteriophage, ϕ Joe. *Appl. Environ. Microbiol.* **83**, 1–16 (2017).
 32. Rubtsova, M. *et al.* Expression of active Streptomyces phage phiC31 integrase in transgenic wheat plants. *Plant Cell Rep.* **27**, 1821–1831 (2008).
 33. Kempe, K. *et al.* Transgene excision from wheat chromosomes by phage phiC31 integrase. *Plant Mol. Biol.* **72**, 673–687 (2010).
 34. Yu, Y. *et al.* Improved site-specific recombinase-based method to produce selectable marker- and vector-backbone-free transgenic cells. *Sci. Rep.* **4**, 4240 (2014).

35. Thomson, J. G., Chan, R., Thilmony, R., Yau, Y.-Y. & Ow, D. W. PhiC31 recombination system demonstrates heritable germinal transmission of site-specific excision from the Arabidopsis genome. *BMC Biotechnol.* **10**, 17 (2010).
36. Chiu, W. *et al.* Engineered GFP as a vital reporter in plants. *Curr. Biol.* **6**, 325–330 (1996).
37. Van Duyne, G. D. & Rutherford, K. Large serine recombinase domain structure and attachment site binding. *Crit. Rev. Biochem. Mol. Biol.* **48**, 476–491 (2013).
38. Barros, L. *et al.* Immunological-based approaches for cancer therapy. *Clinics* **73**, 1–11 (2018).
39. Chicaybam, L., Laino Sodr e, A. & Bonamino, M. Chimeric antigen receptors in cancer immuno-gene therapy: Current status and future directions. *Int. Rev. Immunol.* **30**, 294–311 (2011).
40. Tang, Y., Yu, P. & Cheng, L. Current progress in the derivation and therapeutic application of neural stem cells. *Nat. Publ. Gr.* **8**, e3108 (2017).
41. Fraga, A. M. *et al.* Establishment of a Brazilian Line of Human Embryonic Stem Cells in Defined Medium : Implications for Cell Therapy in an Ethnically Diverse Population. *Cell Transplant.* **20**, 431–440 (2011).
42. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nat. Publ. Gr.* **17**, 5–15 (2015).
43. Guinebretiere, M.-H. *et al.* *Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus cereus* Group occasionally associated with food poisoning. *Int. J. Syst. Evol. Microbiol.* **63**, 31–40 (2013).
44. Vargas, J. E. *et al.* Retroviral vectors and transposons for stable gene therapy : advances , current challenges and perspectives. *J. Transl. Med.* **14**, 1–15 (2016).
45. Dakic, V. *et al.* Harmine stimulates proliferation of human neural progenitors. *PeerJ.* **4**, e2727 (2016).
46. Chicaybam, L. & Bonamino, M. H. Moving receptor redirected adoptive cell therapy toward fine tuning of antitumor responses. *Int. Rev. Immunol.* **33**, 402–416 (2014).
47. Zhu, F. *et al.* DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. *Nucleic Acids Res.* **42**, e34 (2014).
48. Annaluru, N. *et al.* Total Synthesis of a Functional Designer Eukaryotic Chromosome. *Science.* **344**, 55–58 (2014).
49. Kalhor, R. *et al.* Developmental barcoding of whole mouse via homing CRISPR. *Science.* **361**, eaat9804 (2018).
50. Grandchamp, N. *et al.* Hybrid lentivirus-phiC31-int-NLS vector allows site-specific recombination in murine and human cells but induces DNA damage. *PLoS One* **9**, e99649 (2014).
51. Chicaybam, L. *et al.* An Efficient Electroporation Protocol for the Genetic Modification of Mammalian Cells. *Front. Bioeng. Biotechnol.* **4**, 1–13 (2017).
52. Arag o, F. J. L., Vianna, G. R., Carvalheira, S. B. R. C. & Rech, E. L. Germ line genetic transformation in cotton (*Gossypium hirsutum* L .) by selection of transgenic meristematic cells with a herbicide molecule. *Plant Sci.* **168**, 1227–1233 (2005).
53. Datta, R. S. S., Hammerlindl, J. K., Pelcher, L. E., Crosby, W. L. & Selvaraj, G. A bifunctional fusion between β -glucuronidase and neomycin phosphotransferase: a broad-spectrum marker enzyme for plants. *Gene* **101**, 239–246 (1991).

54. Casas, B. S. *et al.* hiPSC-derived neural stem cells from patients with schizophrenia induce an impaired angiogenesis. *Transl. Psychiatry* **8**, 48 (2018).
55. Garcez, P. P. *et al.* Zika virus impairs growth in human neurospheres and brain organoids. *Science*. **352**, 816–818 (2016).
56. Freshney, R. I. *Culture of Animal Cells: a manual of basic technique*. (Wiley-Liss, Inc: New York, 1994).
57. Yoo, S. D., Cho, Y. H. & Sheen, J. Arabidopsis mesophyll protoplasts: A versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572 (2007).
58. Lin, H.-Y., Chen, J.-C. & Fang, S.-C. A Protoplast Transient Expression System to Enable Molecular, Cellular, and Functional Studies in Phalaenopsis orchids. *Front. Plant Sci.* **9**, 1–13 (2018).
59. Gomide, M. *et al.* Genetic switches designed for eukaryotic cells and controlled by serine integrases. Dryad data sets. doi:<https://doi.org/10.5061/dryad.dr7sqv9tv> (2020).

Acknowledgments

We thank Joseane Padilha da Silva, MSc, from Embrapa Genetic Resources & Biotechnology for providing assistance with the statistical analyses; Dr Miguel Andrade, for figures graphic design; Adriano Caiado A. F. de Almeida, MSc, for contributions to the p35S(rc)2_4_5-GFP (pSP) cloning process; Lara Rodrigues Nesralla, MSc, who took part in the adaptation of the protocols for protoplast extraction and transformation; Dr Andrielle Thainar Mendes Cunha, for assistance with the flow cytometry methodology; and INCA's blood bank and hemotherapy service for the preparation of donor samples. This research was funded by INCT BioSyn (National Institute of Science and Technology in Synthetic Biology), CNPq (National Council for Scientific and Technological Development), CAPES (Coordination for the Improvement of Higher Education Personnel), Brazilian Ministry of Health, and FAPDF (Research Support Foundation of the Federal District), Brazil.

Author contributions

Conceived and designed the experiments: E.R., C.M.C., M.H.B., and A.M.M. Performed vector design and cloning: C.M.C., M.S.G., G.P.C.J., M.S.M.A., L.M.G.B., and R.N.L. Performed the *Arabidopsis thaliana* protoplast experiments: M.S.G., L.M.G.B., L.H.F., and C.L. Performed the bovine fibroblast experiments: T.T.S.,

M.A.O., E.O.M., G.P.C.J., and R.N.L. Performed the HEK 293T cell, PBMCs, NSCs and hES cell experiments: L.R.C.B., C.G.L., M.L.R., L.H.F., M.A.O., and S.K.R. Wrote the manuscript: E.R., C.M.C., M.S.G., T.T.S., L.M.G.B., R.N.L. and M.H.B. All the authors analyzed, interpreted and discussed the results and contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

7 MATERIAL SUPLEMENTAR AO MANUSCRITO

Supplementary Information

Genetic switches designed for eukaryotic cells and controlled by serine integrases

Mayna S. Gomide^{1,2,3}, Thais T. Sales^{1,2}, Luciana R.C. Barros⁴, Cintia G. Limia⁴, Marco A. de Oliveira^{1,2}, Lilian H. Florentino¹, Leila M.G. Barros¹, Maria L. Robledo⁴, Gustavo P.C. José¹, Mariana S.M. Almeida¹, Rayane N. Lima¹, Stevens K. Rehen^{5,6}, Cristiano Lacorte¹, Eduardo O. Melo^{1,7}, André M. Murad¹, Martín H. Bonamino^{4,8*}, Cintia M. Coelho^{9*} and Elibio Rech^{1*}.

1 Brazilian Agriculture Research Corporation – Embrapa – Genetic Resources and Biotechnology – CENARGEN, Brasília, 70770917, DF, Brazil;

2 Department of Cell Biology, Institute of Biological Science, University of Brasília, Brasília, 70910900, DF, Brazil;

3 School of Medicine, Federal University of Juiz de Fora, Juiz de Fora, 36036900, MG, Brazil;

4 Molecular Carcinogenesis Program, Research Coordination, National Cancer Institute (INCA), Rio de Janeiro, 20231050, RJ, Brazil;

5 D'Or Institute for Research and Education (IDOR), Rio de Janeiro, 22281100, RJ, Brazil;

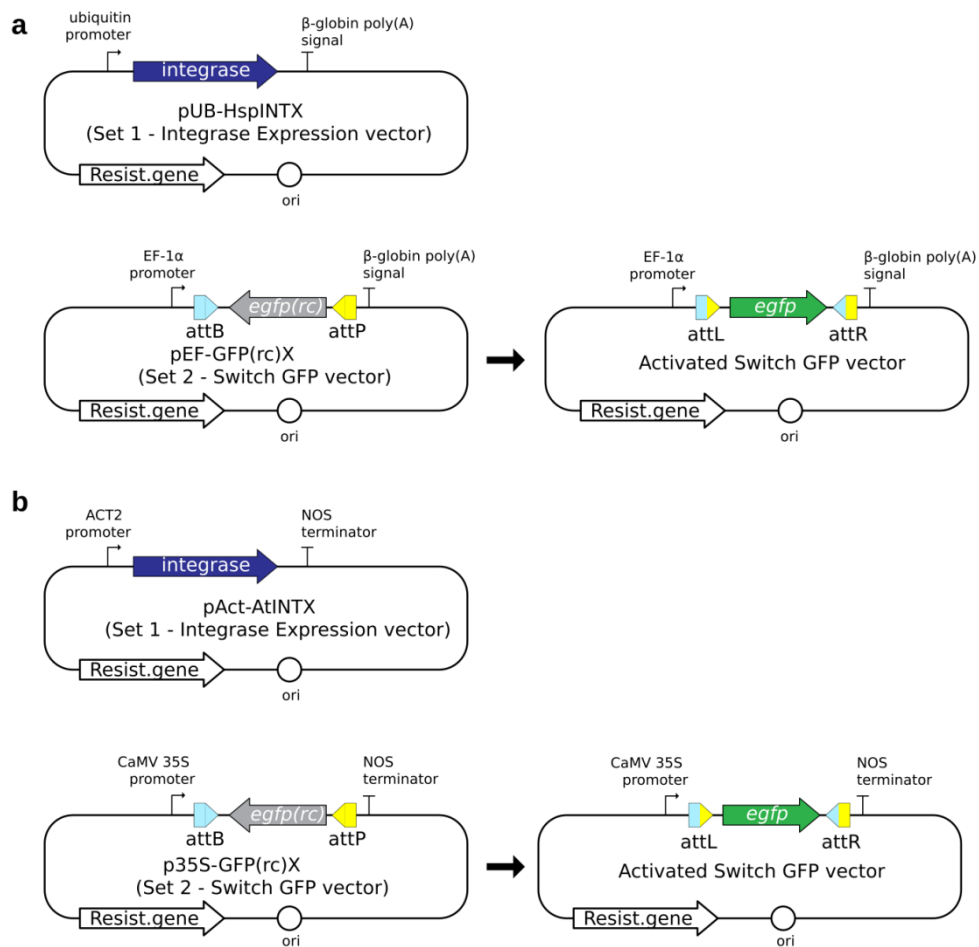
6 Institute of Biomedical Sciences, Federal University of Rio de Janeiro, Rio de Janeiro, 21941902, RJ, Brazil;

7 Graduation Program in Biotechnology, Federal University of Tocantins, Gurupi, 77402970, TO, Brazil;

8 Vice-Presidency of Research and Biological Collections (VPPCB), FIOCRUZ – Oswaldo Cruz Foundation Institute, Rio de Janeiro, 21040900, RJ, Brazil;

9 Department of Genetic and Morphology, Institute of Biological Science, University of Brasília, Brasília, 70910900, DF, Brazil;

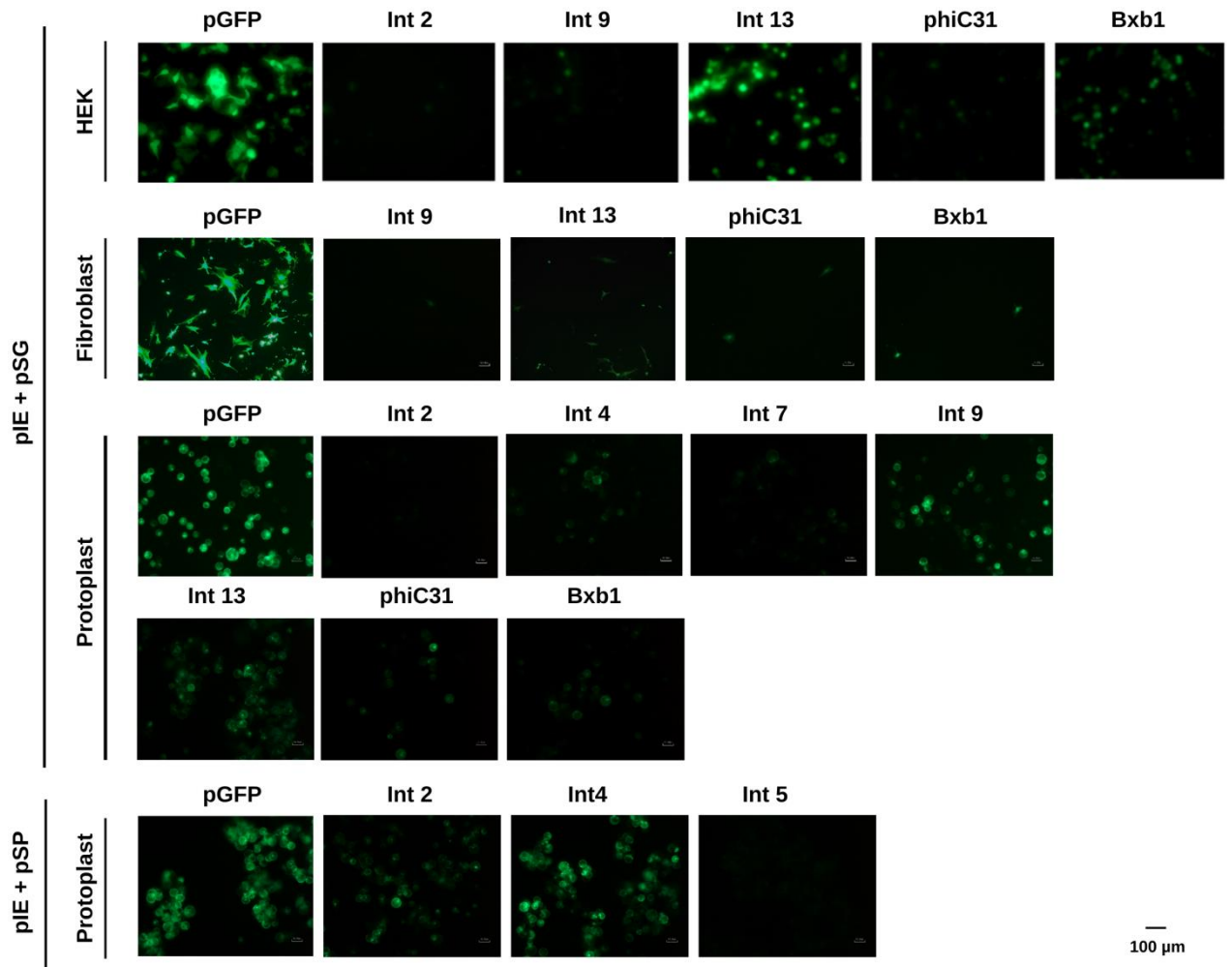
* Corresponding authors (emails: elibio.rech@embrapa.br; cintiacoelhom@unb.br; and mbonamino@inca.gov.br).



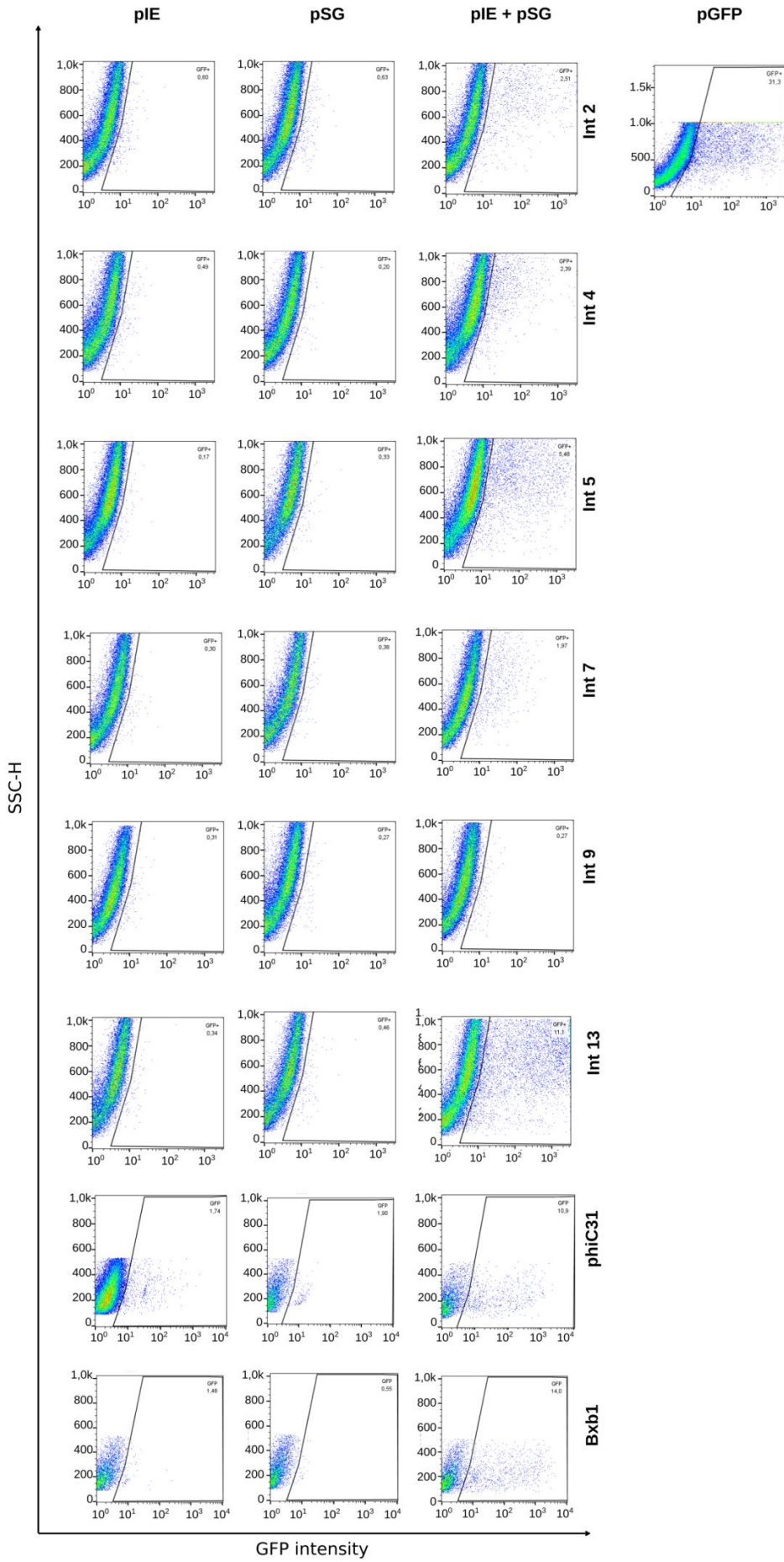
Supplementary Figure 1. Schematic representation of the vector sets synthesized for the eukaryotic genetic switches and the resulting activated vector. For human, bovine, and plant cells, two sets of plasmids were built to evaluate Int functionality. **a** For mammalian systems, one plasmid contains the human codon optimized sequence of Int X (X= 2, 4, 5, 7, 9, 13, phiC31 or Bxb1) under the control of the ubiquitin promoter and β-globin poly(A) signal terminator; this set was named the integrase expression vector (pIE) set, composed of plasmids named pUB-HspINTX. The other plasmid is composed of the reporter *egfp* gene in the reverse complement (rc) orientation flanked by the recognition sites *attB/attP* of that particular Int, under the control of a different strong constitutive promoter, namely, the EF1alpha promoter, and the same terminator; this set was named the switch GFP vector (pSG), with plasmids named pEF-GFP(rc)X. **b** For plant protoplasts, the pIE vectors set contains the *A. thaliana* codon optimized sequence of Int X under the actin2 promoter and the NOS terminator, plasmids named pAct-AtINTX. The pSG vectors set has the *egfp* sequence in the reverse complement orientation flanked by the recognition sites *attB/attP* of that particular Int, under the control of the CaMV 35S promoter and the same terminator, plasmids named p35S-GFP(rc)X (additional information on the plasmids is provided in Supplementary Table 1). Both plasmids of each set were used to cotransfect/cotransform the mammalian and plant systems. It was hypothesized that if a particular Int was functional, it would switch the *egfp* coding sequence to the forward orientation, leading to EGFP expression and formation of the *attL/attR* sites (activated switch GFP vector).

INTEGRASE	NLS
INT_02	¹⁵⁸ RTTRLNAKRGGAHGPVDPGYKRRYPD ¹⁸⁴ [6.0]
INT_04	²⁶⁰ RERQRRRLGIEENHYTIPFQAKYMLSKFLRC ²⁹⁰ [5.9]
INT_05	
INT_07	¹³¹ RENLAERVKFGIEQMIDEGKKPGGHSPYGYKFDKD ¹⁶⁵ [5.4]
INT_09	¹³ EQKEKGHSIEEQERKLRAYS DINDWKIHKVY ⁴³ [5.3]
INT_13	²⁸³ VNRFIKKRKDGTEYC ²⁹⁷ [8.5]
Bxb1	⁵² PFDRKRRPNL ⁶¹ [5.0]
PhiC31	²²¹ REIKTHKHLPFKPGSQAAIHPGSITGLCKRMDAD ²⁵⁴ [4.8]

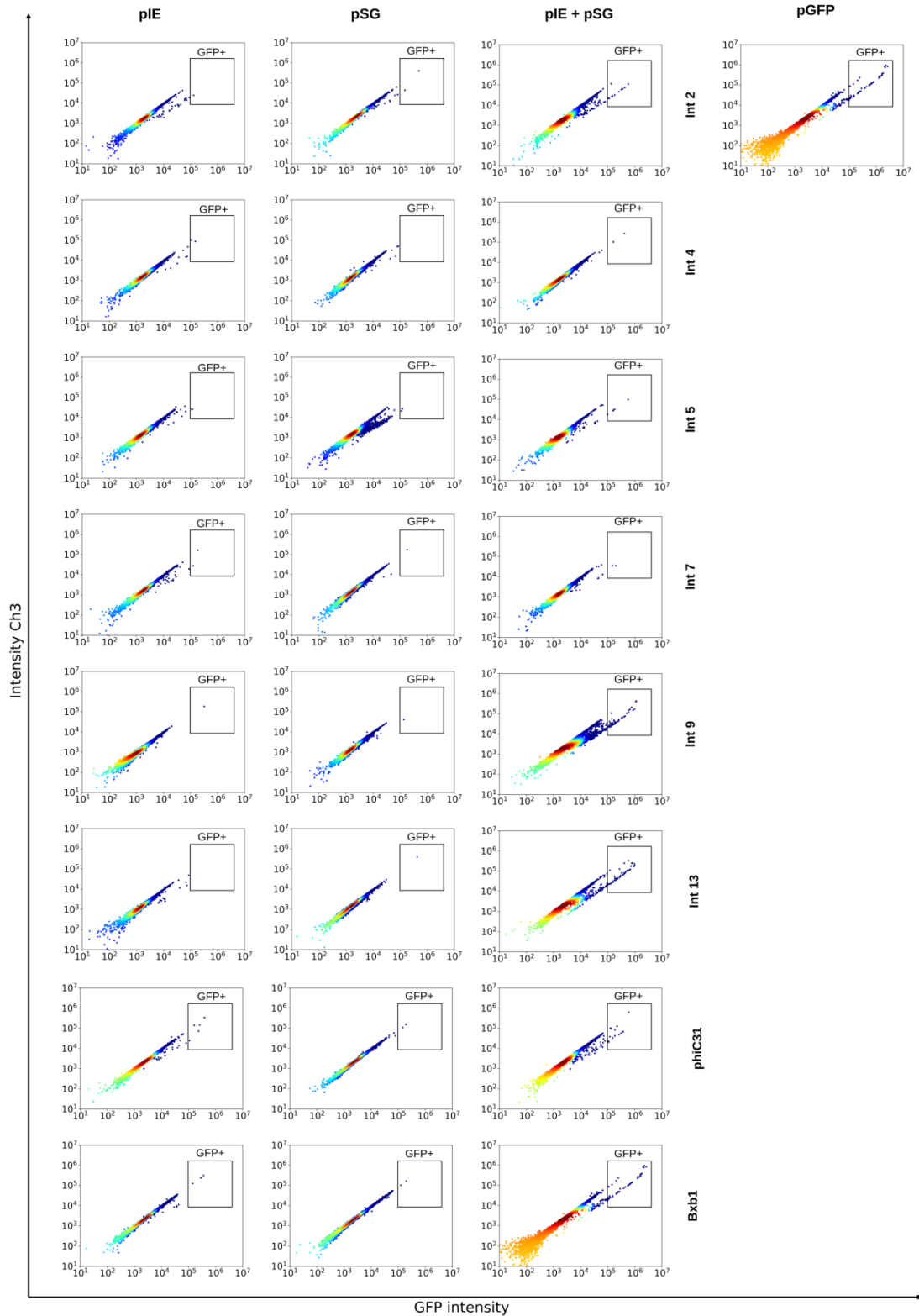
Supplementary Figure 2. *In silico* prediction of the nuclear localization signal (NLS) for the Int coding sequences. The NLS was predicted using NLS Mapper (available at http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi, accessed in 10/08/2018) with an intermediate cut-off score (5.0). The positions of the first and last amino acids of the predicted NLS sequence are denoted as superscripted numbers. The basic amino acids arginine (R) and lysine (K) are highlighted in cyan. Scores are indicated in brackets.



Supplementary Figure 3. Representative EGFP fluorescence images of the three model eukaryotic cell systems cotransfected/cotransformed with integrase expression (pIE) and switch GFP or promoter (pSG or pSP) vectors. HEK 293T cells, bovine fibroblasts and plant protoplasts were observed using an Axiovert 135M (Carl Zeiss) fluorescence microscope. The images were acquired using an attached DS-Ri1 digital camera (Nikon) and the capture software Nikon Digital Sight DS-L3 (Nikon) under a UV light with filter set 15 (Carl Zeiss). Excitation: BP 546; beam splitter: FT 580; emission: LP 590. For each cell model, all images were acquired with the same acquisition setting.

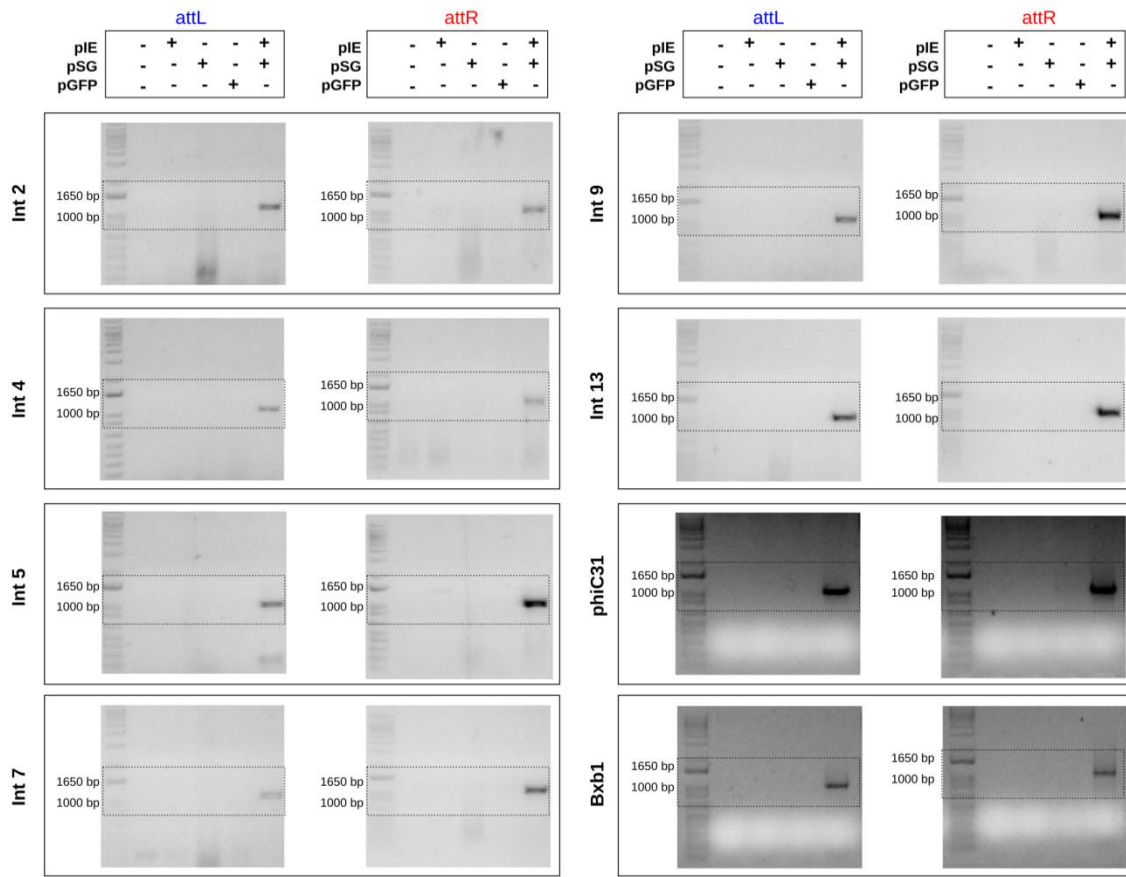


Supplementary Figure 4. EGFP fluorescence determined by flow cytometry analysis of HEK 293T cells. Representative scatter plots are shown indicating the EGFP fluorescence-emitting population in the gate. Experimental groups were analyzed after 48 h. The integrase expression vectors (pIE) of the Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1 or the switch GFP vectors (pSG) containing the *egfp* gene in reverse complement orientation flanked by *attB/attP* sites of the Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1 indicate negative controls. pIE + pSG indicates the cells cotransfected with integrase expression and switch GFP vectors (test condition). Positive control cells were transfected with the pT3-Neo-EF1 α -GFP plasmid containing the *egfp* sequence in the forward orientation (pGFP).

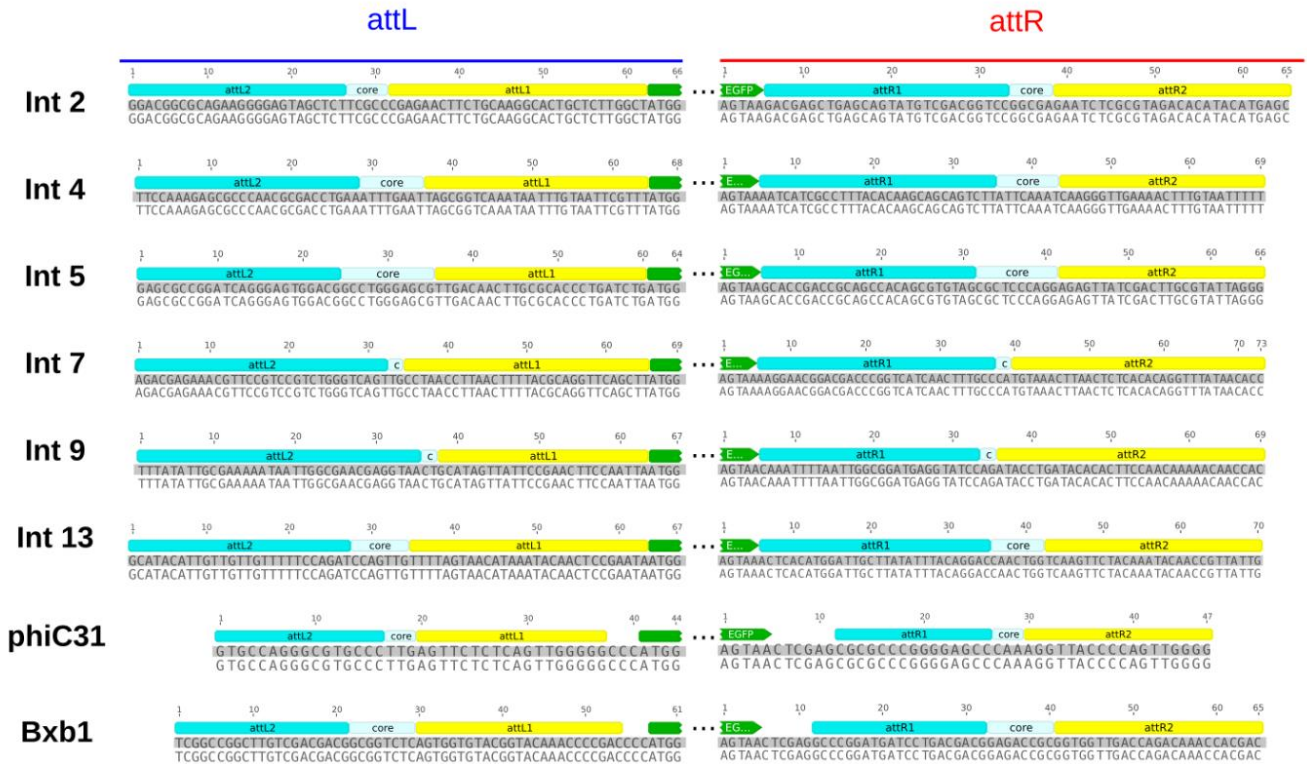


Supplementary Figure 5. EGFP fluorescence determined by flow cytometry analysis of bovine fibroblasts. Representative scatter plots are shown indicating the EGFP fluorescence-emitting population in the gate. Experimental groups were analyzed after 48 h. Negative control cells were cotransfected with one of the plasmids from the integrase expression vectors (pIE) of the Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1 or one of the switch GFP vectors

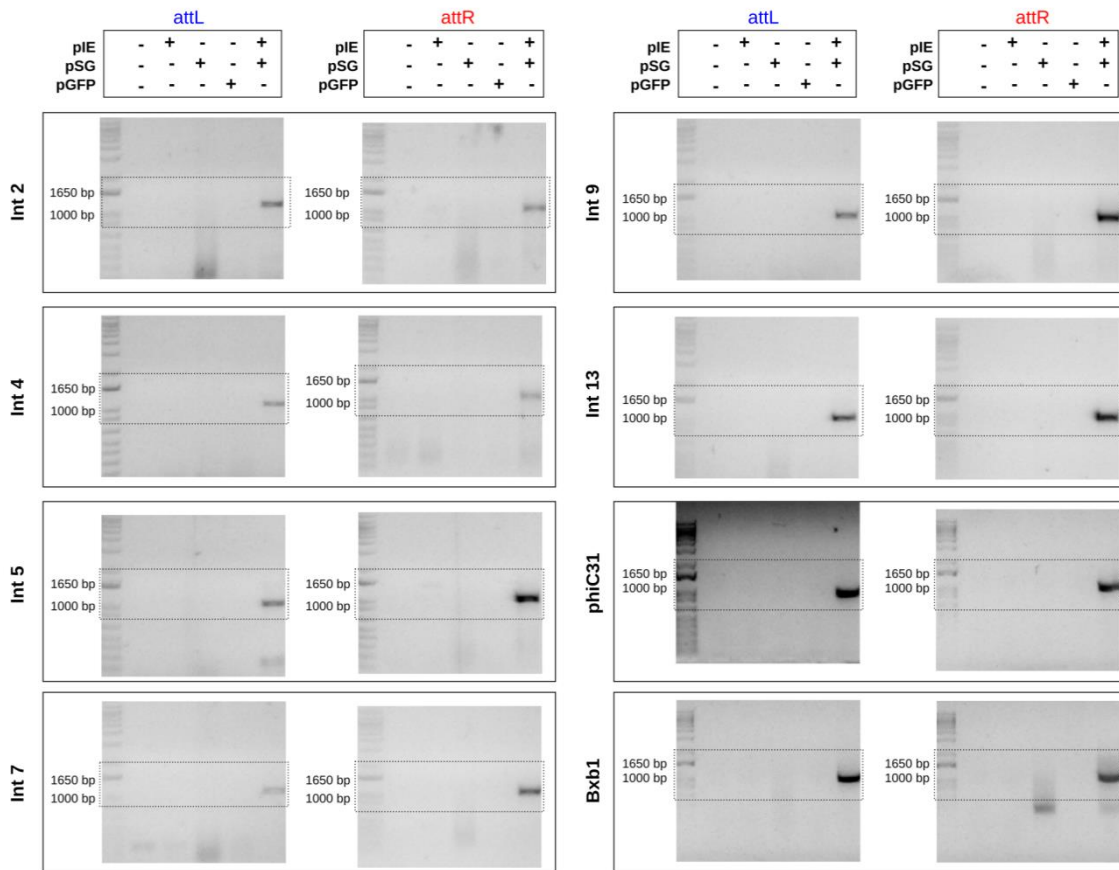
(pSG) containing the *egfp* gene in reverse complement orientation flanked by *attB/attP* sites of the Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1 plus a mock plasmid. pIE + pSG indicates the cells cotransfected with integrase expression and switch GFP vectors (test condition). Positive control cells were cotransfected with the pEF-GFP plasmid containing the *egfp* sequence in the forward orientation (pGFP) plus a mock plasmid.



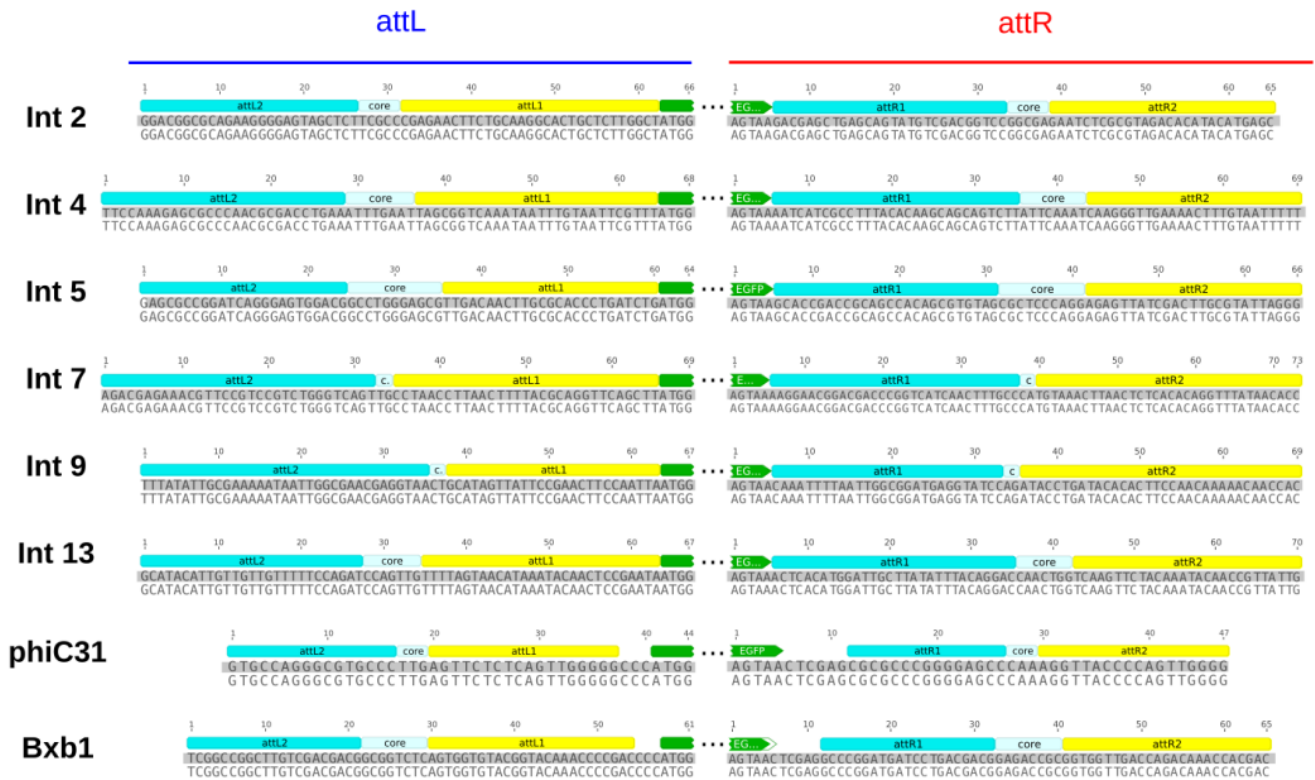
Supplementary Figure 6. HEK 293T cells uncropped PCR gel images. Amplicons obtained using two specific primer sets to verify *attL* (blue) and *attR* (red) formation after *egfp* flipping as shown in the main text Fig. 2c.



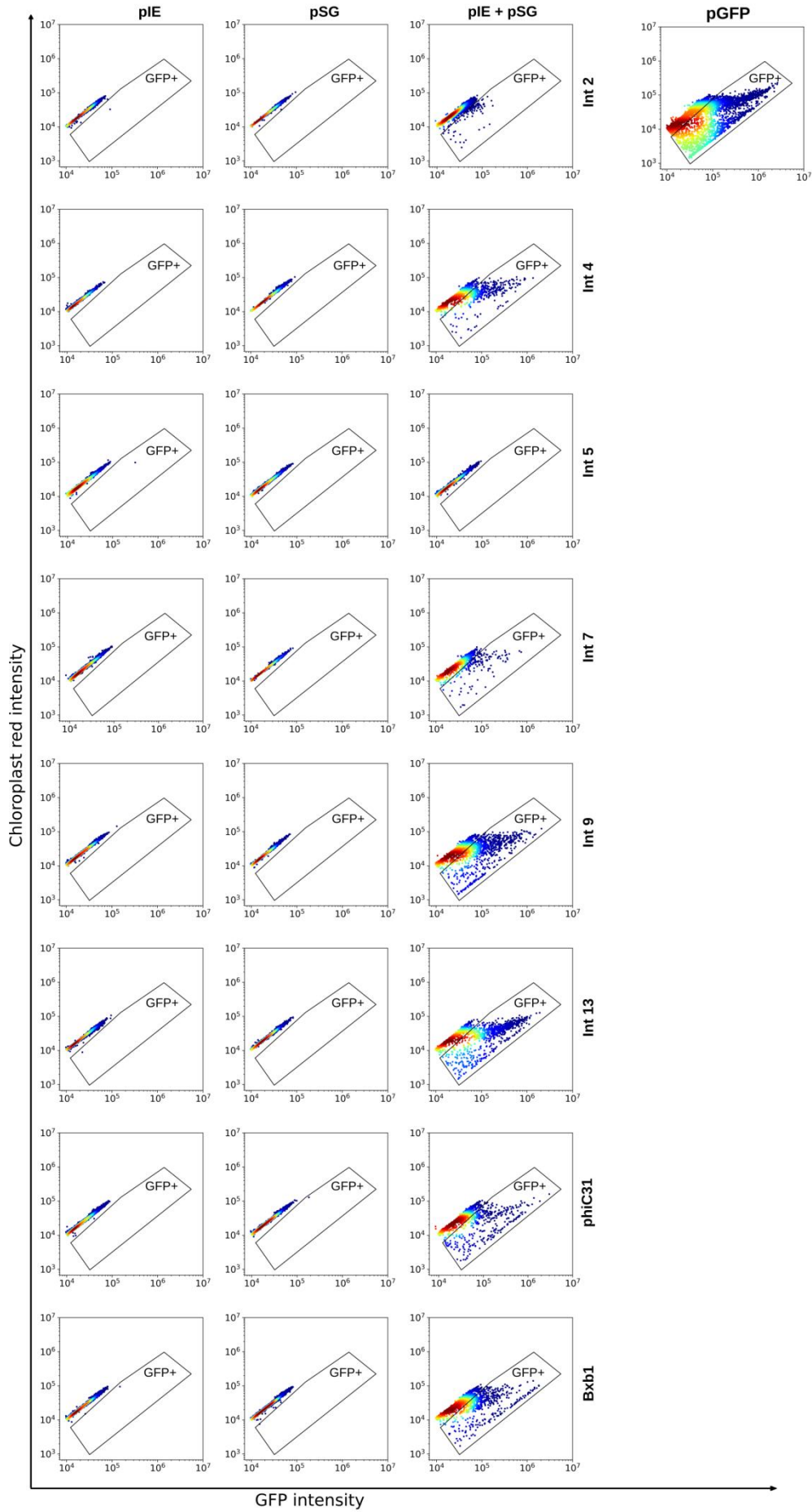
Supplementary Figure 7. Representative sequence reads showing the *attL* and *attR* sites obtained after Int activity in HEK 293T cells compared to the predicted sequences (grey highlighted). *attL1* and *attR1* correspond to the flipped *attP* and *attB* parts, respectively. *attL2* and *attR2* correspond to the previous *attB* and *attP* parts, respectively. Additional information in Supplementary Table 3 and Supplementary Data 1.



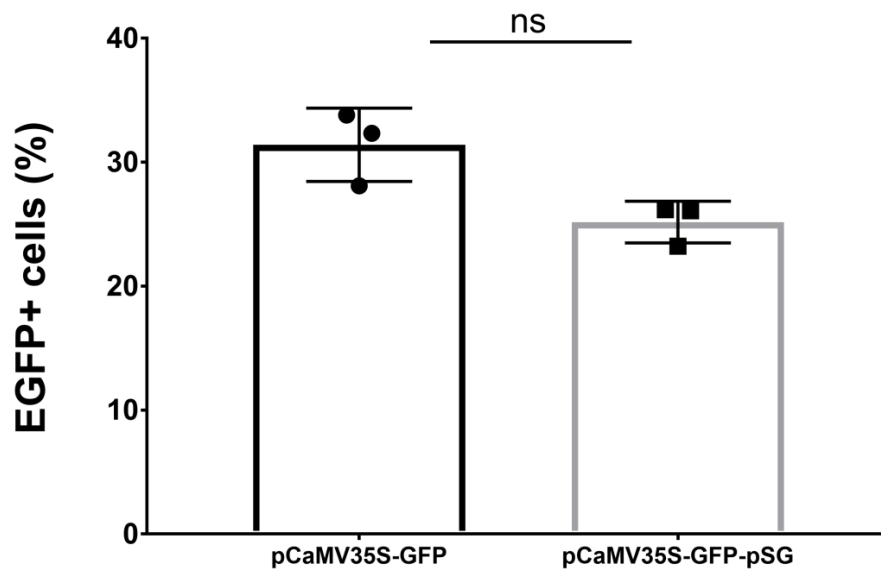
Supplementary Figure 8. Bovine fibroblasts uncropped PCR gel images. Amplicons obtained using two specific primer sets to verify *attL* (blue) and *attR* (red) formation after *egfp* flipping as shown in the main text Fig. 3c.



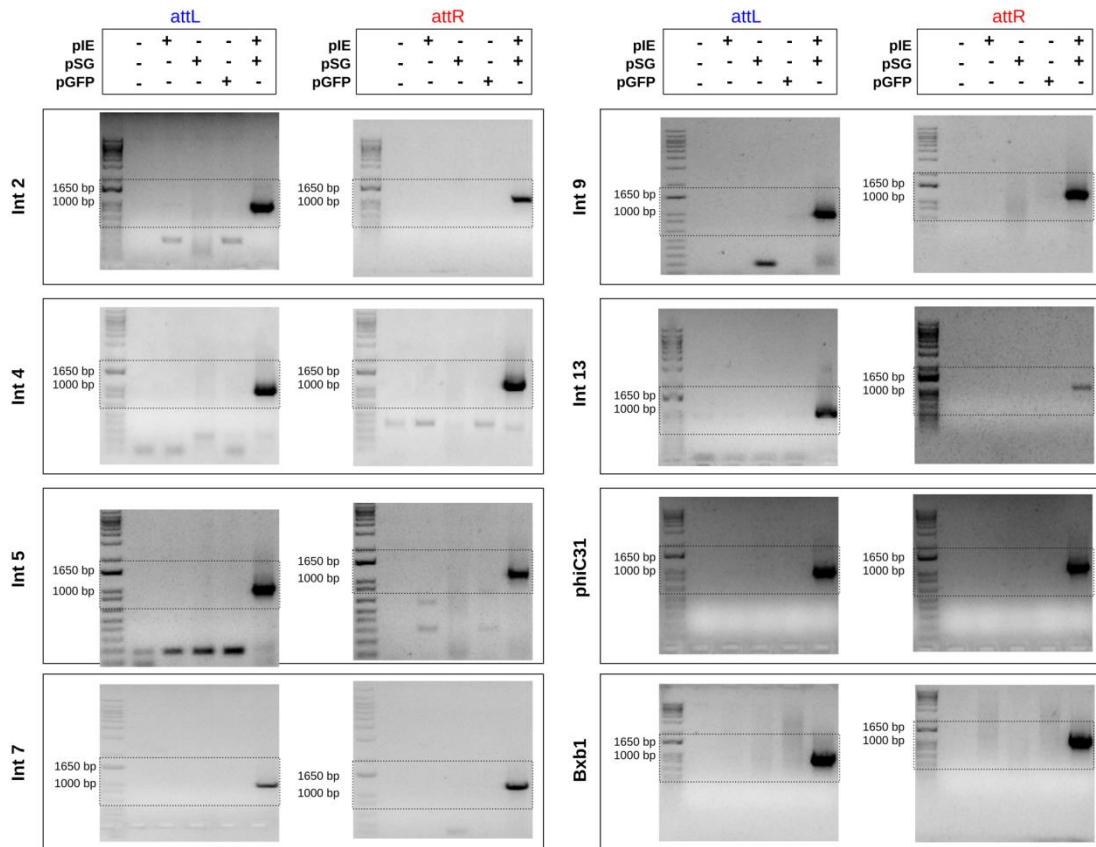
Supplementary Figure 9. Representative sequence reads showing the *attL* and *attR* sites obtained after Int activity in bovine fibroblasts compared to the predicted sequences (grey highlighted). *attL1* and *attR1* correspond to the flipped *attP* and *attB* parts, respectively. *attL2* and *attR2* correspond to the previous *attB* and *attP* parts, respectively. Additional information in Supplementary Table 3 and Supplementary Data 2.



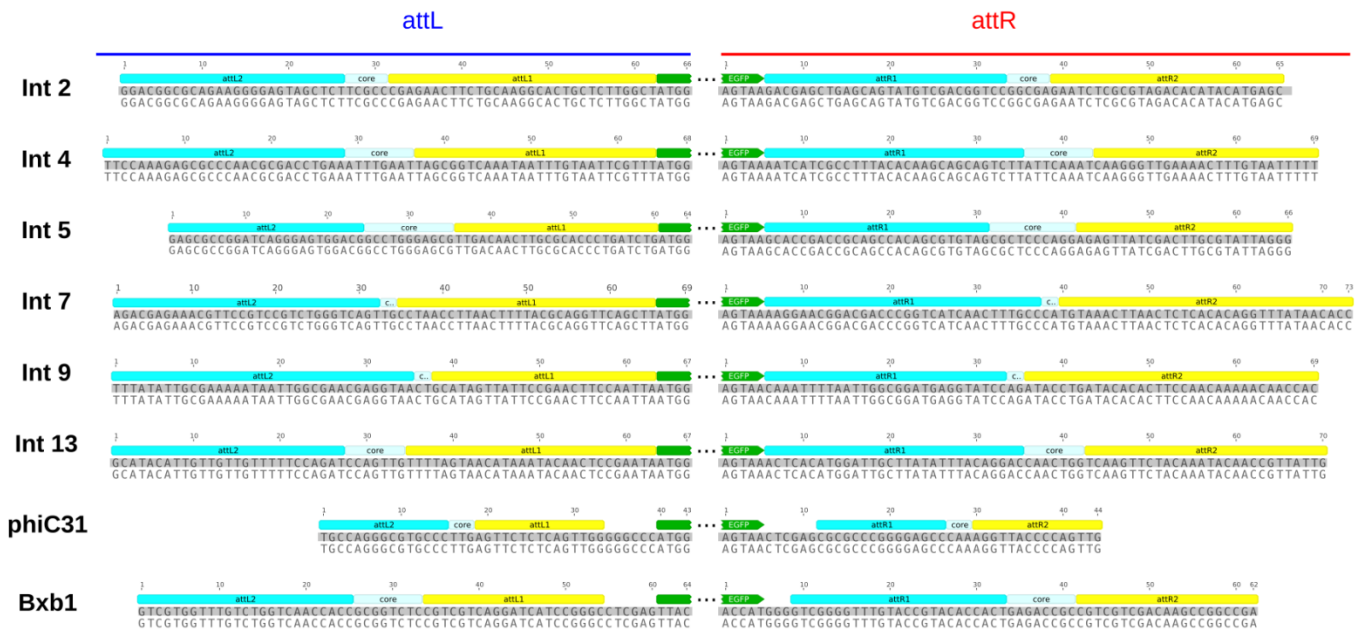
Supplementary Figure 10. EGFP fluorescence determined by flow cytometry analysis of *A. thaliana* protoplasts. Representative scatter plots are shown indicating the EGFP fluorescence-emitting population in the gate. Experimental groups were analyzed after 24 h. Negative control cells were cotransformed with one of the plasmids from the integrase expression vectors (pIE) of the Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1 or the switch GFP vectors (pSG) containing the *egfp* gene in reverse complement orientation flanked by *attB/attP* sites of the Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1 plus a mock plasmid. pIE+ pSG indicates the cells cotransformed with integrase expression and switch GFP vectors (test condition). Positive control cells were cotransformed with the pCaMV35S-GFP plasmid containing the *egfp* sequence in the forward orientation (pGFP) plus a mock plasmid.



Supplementary Figure 11. EGFP-expressing cell percentages of the positive control pCaMV35S-GFP vector and pCaMV35S-GFP-pSG vector. pCaMV35S-GFP (pGFP) has a CaMV 35S promoter with some SNPs compared with the CaMV 35S promoter used in the switch GFP vectors (pSG). Then, the CaMV 35S promoter from pSG was cloned, replacing the pCaMV35S-GFP promoter, resulting in the pCaMV35S-GFP-pSG plasmid. Protoplasts were transformed with both plasmids separately, and flow cytometry analysis showed that the percentage of EGFP-expressing cell populations obtained with the two constructs did not result in statistically significant differences. The statistical analysis was performed in GraphPad Prism 7, applying a paired T test. p value=0.1277. Assays were performed in five or six technical replicates and in three biologically independent experiments.



Supplementary Figure 12. *A. thaliana* protoplasts uncropped PCR gel images. Amplicons obtained using two specific primer sets to verify *attL* (blue) and *attR* (red) formation after *egfp* flipping as shown in the main text Fig. 4c.



Supplementary Figure 13. Representative sequence reads showing the *attL* and *attR* sites obtained after Int activity in *A. thaliana* protoplasts compared to the predicted sequences (grey highlighted). *attL1* and *attR1* correspond to the flipped *attP* and *attB* parts, respectively. *attL2* and *attR2* correspond to the previous *attB* and *attP* parts, respectively. Additional information in Supplementary Table 3 and Supplementary Data 3.

Original described

```

attB  TTTATATTGCGAAAAATAATTGGCGAACAGGTACTGGATACCTCATCGCCAATTAAATTTG ... .. TAATTGGAAGTTCGGAATAACTATGCAGATACCTGATACACACTTCCAACAAAAACAACCAC attP
AAATATAACGCTTTTTATTAACCGCTTGGTCCATTGACCTATGGAGTAGCGGTTAATTTTAAAC ... .. ATTAACCTTCAAGCCTTATTGATACGCTATGGAATATGTGTGAAGGTTGTTTTGTGGTG
attL  TTTATATTGCGAAAAATAATTGGCGAACAGGTACTGCATAGTTATTCCGAACTTCCAATTA ... .. CAAATTTTAAATTGGCGGATGAGGTATCCAGATACCTGATACACACTTCCAACAAAAACAACCAC attR
AAATATAACGCTTTTTATTAACCGCTTGGTCCATTGACCTATGGAGTAGCGGTTAATTTTAAAC ... .. GTTTAAATTAACCGCCTACTCCATAGGTCATGGAATATGTGTGAAGGTTGTTTTGTGGTG
                                     core                                     core

```

Identified after sequencing

```

attB  TTTATATTGCGAAAAATAATTGGCGAACAGGTACTGGATACCTCATCGCCAATTAAATTTG ... .. TAATTGGAAGTTCGGAATAACTATGCAGATACCTGATACACACTTCCAACAAAAACAACCAC attP
AAATATAACGCTTTTTATTAACCGCTTGGTCCATTGACCTATGGAGTAGCGGTTAATTTTAAAC ... .. ATTAACCTTCAAGCCTTATTGATACGCTATGGAATATGTGTGAAGGTTGTTTTGTGGTG
attL  TTTATATTGCGAAAAATAATTGGCGAACAGGTACTGCATAGTTATTCCGAACTTCCAATTA ... .. CAAATTTTAAATTGGCGGATGAGGTATCCAGATACCTGATACACACTTCCAACAAAAACAACCAC attR
AAATATAACGCTTTTTATTAACCGCTTGGTCCATTGACCTATGGAGTAGCGGTTAATTTTAAAC ... .. GTTTAAATTAACCGCCTACTCCATAGGTCATGGAATATGTGTGAAGGTTGTTTTGTGGTG
                                     core                                     core

```

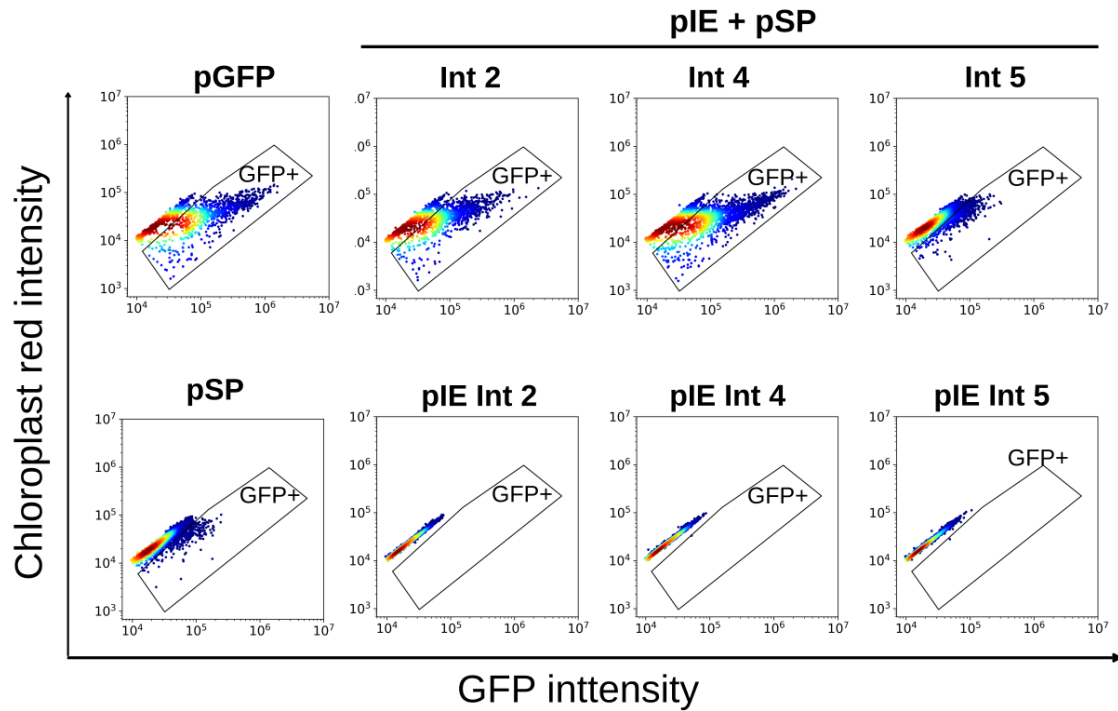
possible core-site dinucleotide identified

```

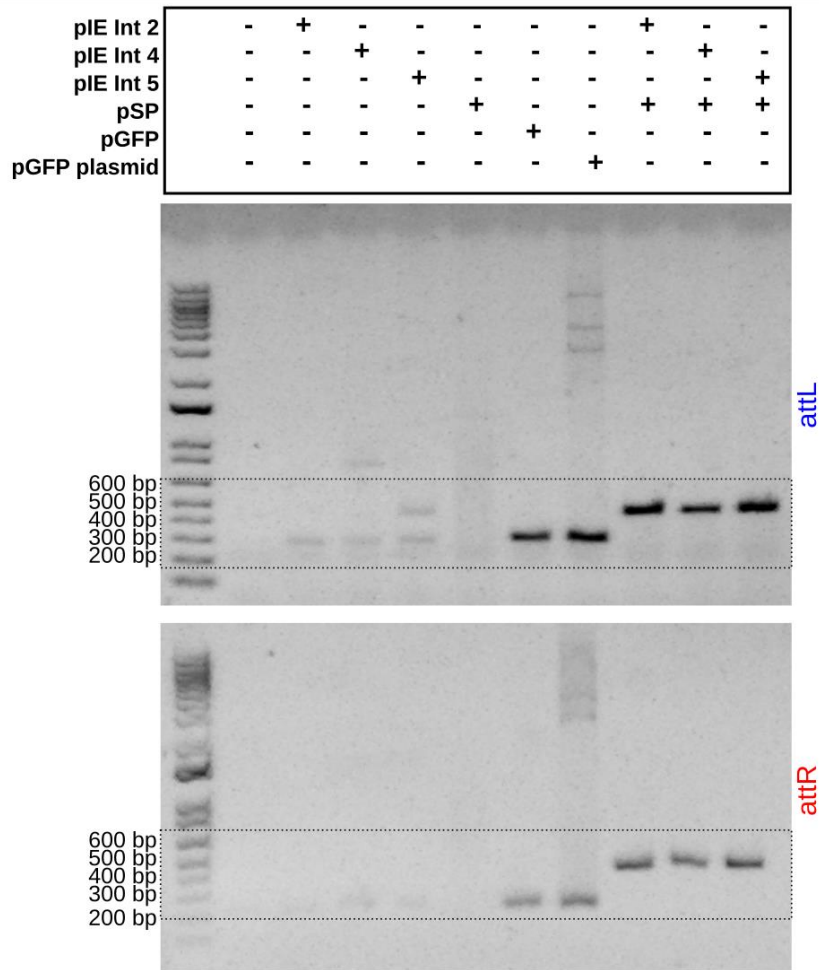
attB  TTTATATTGCGAAAAATAATTGGCGAACGAGGTACTGGATACCTCATCGCCAATTAAATTTG ... .. TAATTGGAAGTTCGGAATAACTATGCAGATACCTGATACACACTTCCAACAAAAACAACCAC attP
AAATATAACGCTTTTTATTAACCGCTTGGTCCATTGACCTATGGAGTAGCGGTTAATTTTAAAC ... .. ATTAACCTTCAAGCCTTATTGATACGCTATGGAATATGTGTGAAGGTTGTTTTGTGGTG
attL  TTTATATTGCGAAAAATAATTGGCGAACGAGGTACTGCATAGTTATTCCGAACTTCCAATTA ... .. CAAATTTTAAATTGGCGGATGAGGTATCCAGATACCTGATACACACTTCCAACAAAAACAACCAC attR
AAATATAACGCTTTTTATTAACCGCTTGGTCCATTGACCTATGGAGTAGCGGTTAATTTTAAAC ... .. GTTTAAATTAACCGCCTACTCCATAGGTCATGGAATATGTGTGAAGGTTGTTTTGTGGTG
                                     core                                     core

```

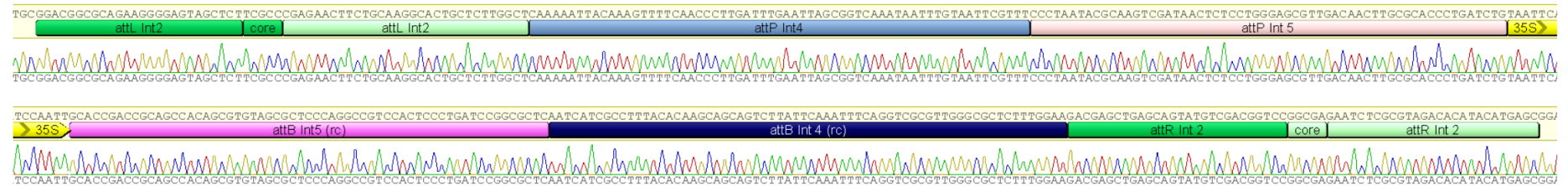
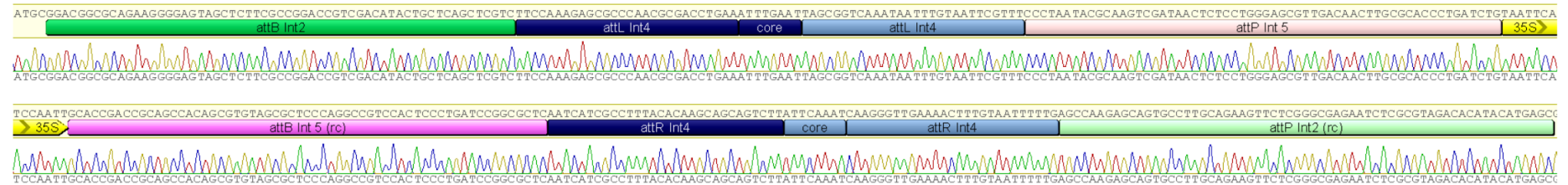
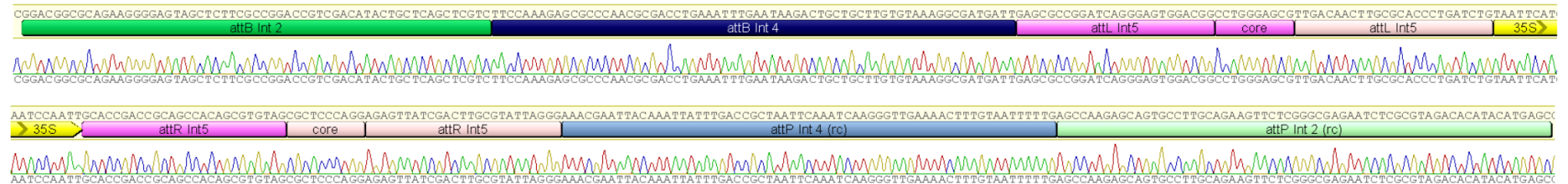
Supplementary Figure 14. Int 9 core-site differences compared to the original sequences described¹. The Int 9 sequence alignments indicated one additional nucleotide near to core-site (black arrow top). Considering the integrase functional mechanism that leads to a rotational and religation of half part of the recognition *attB/P* sites forming the *attL/attR* sites (green and red parts), a possible solution was to consider CT nucleotide as Int9 core-site.



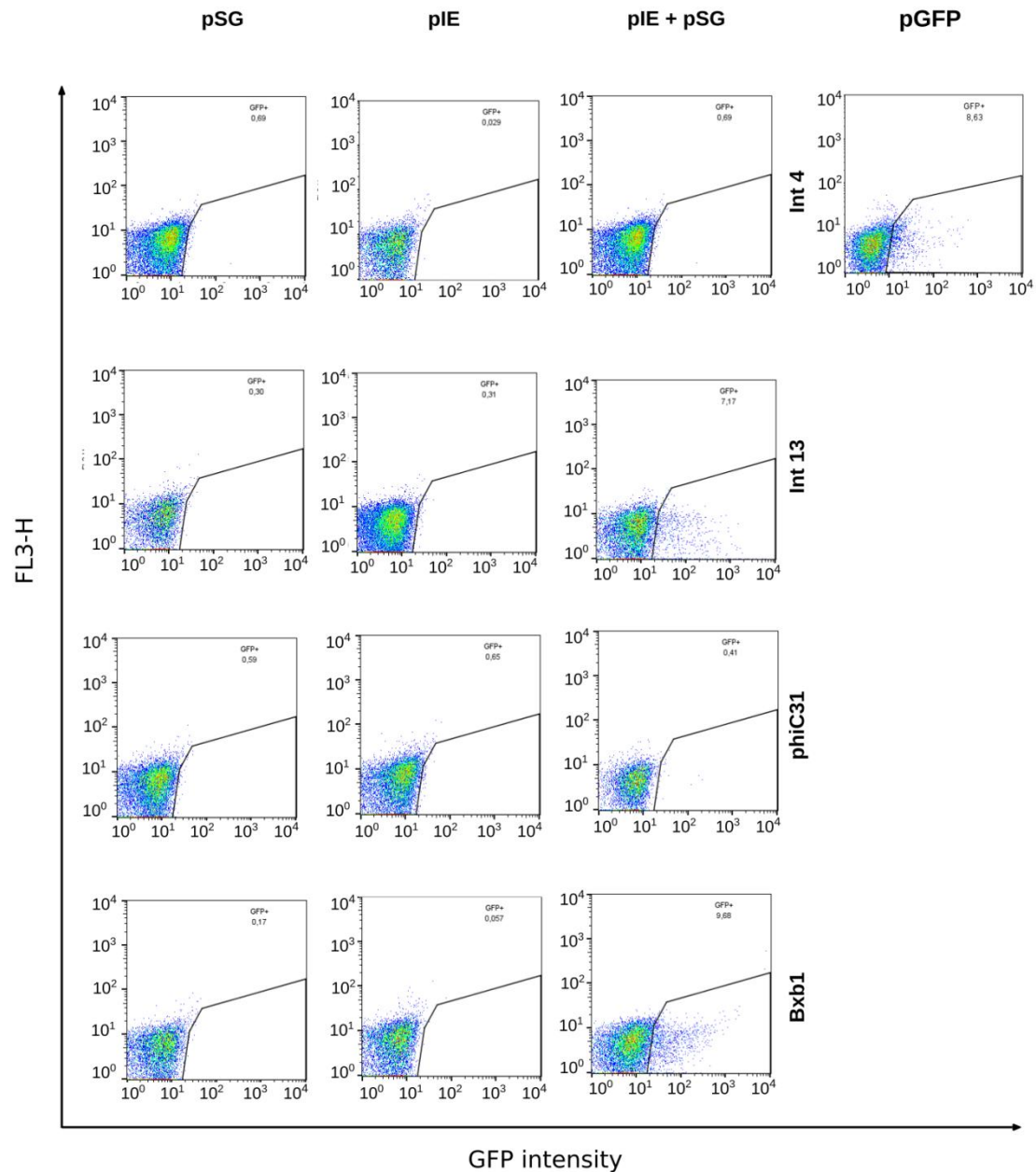
Supplementary Figure 15. EGFP fluorescence determined by flow cytometry analysis of *A. thaliana* protoplasts with the switch promoter system. Representative scatter plots are shown indicating the EGFP fluorescence-emitting population in the gate. Experimental groups were analyzed after 24 h. Negative control cells were cotransformed with one of the plasmids from the integrase expression vectors (pIE) of Ints 2, 4 and 5 or with the switch Promoter vector (pSP) containing the CaMV 35S promoter in reverse complement orientation flanked by *attB/attP* sites of Ints 2, 4 and 5 in tandem plus a mock plasmid. pIE+ pSP indicates the cells cotransformed with integrase expression vectors and the switch promoter vector (test condition). Positive control cells were cotransformed with the pCaMV35S-GFP plasmid containing the CaMV 35S promoter in the forward orientation (pGFP) plus a mock plasmid.



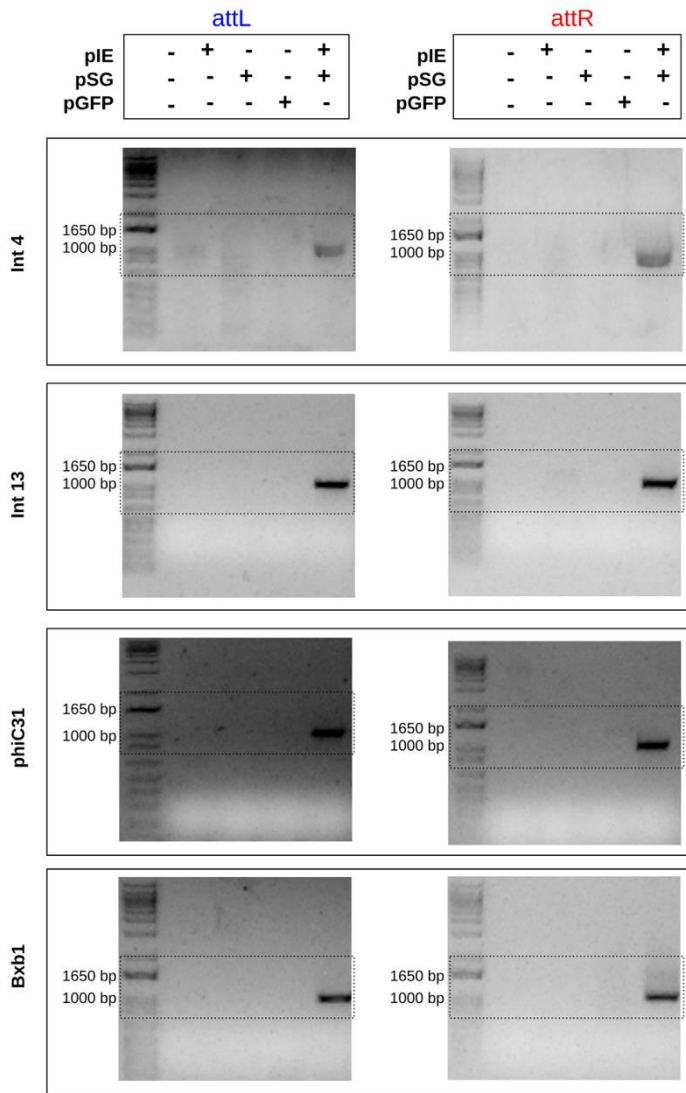
Supplementary Figure 16. *A. thaliana* protoplasts uncropped PCR gel images. Amplicons obtained using two specific primer sets to verify *attL* (blue) and *attR* (red) formation after CaMV 35S promoter flipping as shown in the main text Fig. 5d.

a**b****c**

Supplementary Figure 17. Representative chromatograms showing the recognition site sequences obtained after Int activity in the switch promoter vector (pSP). **a** Recognition sites after Int 2 activity, resulting in *attL/attR* Int 2 sites and *attB/attP* Ints 4 and 5 sites in the expected positions. **b** Recognition sites after Int 4 activity, resulting in *attL/attR* Int 4 sites and *attB/attP* Ints 2 and 5 sites in the expected positions. **c** Recognition sites after Int 5 activity, resulting in *attL/attR* Int 5 sites and *attB/attP* Ints 2 and 4 sites in the expected positions. The core represents the region where cleavage occurred (crossover site). Additional information is provided in Supplementary Data 4.



Supplementary Figure 18. EGFP fluorescence determined by flow cytometry analysis of PBMCs. Representative scatter plots are shown, indicating the EGFP fluorescence-emitting population in the gate. Experimental groups were analyzed after 48 h. Cells electroporated with the integrase expression vectors (pIE) of the Ints 4, 13, phiC3,1 and Bxb1 or the switch GFP vectors (pSG) containing the *egfp* sequence in reverse complement orientation flanked by *attB/attP* sites of the Ints 4, 13, phiC31, and Bxb1 indicate negative controls. pIE + pSG indicates the cells coelectroporated with integrase expression and switch GFP vectors (test condition). Positive control cells were transfected with the pT3-Neo-EF1 α -GFP plasmid containing the *egfp* sequence in the forward orientation (pGFP).



Supplementary Figure 19. PBMCs uncropped PCR gel images. Amplicons obtained using two specific primer sets to verify *attL* (blue) and *attR* (red) formation after *egfp* flipping as shown in the main text Fig. 6c.



Supplementary Figure 20. Representative sequence reads showing the *attL* and *attR* sites obtained after Int activity in PBMCs compared to the predicted sequences (grey highlighted). *attL1* and *attR1* correspond to the flipped *attP* and *attB* parts, respectively. *attL2* and *attR2* correspond to the previous *attB* and *attP* parts, respectively. Additional information in Supplementary Table 6 and Supplementary Data 5.

Supplementary Table 1. Addgene accession numbers of all plasmids used in this study.

Vector set	Plasmid name	Addgene accession number
pSG	INCTbiosyn-pEF-GFP(rc)2	#127504
	INCTbiosyn-pEF-GFP(rc)4	#127505
	INCTbiosyn-pEF-GFP(rc)5	#127506
	INCTbiosyn-pEF-GFP(rc)7	#127507
	INCTbiosyn-pEF-GFP(rc)9	#127508
	INCTbiosyn-pEF-GFP(rc)13	#127509
	INCTbiosyn-pEF-GFP(rc)phiC31	#127510
	INCTbiosyn-pEF-GFP(rc)Bxb1	#127511
pIE	INCTbiosyn-pUB-HspINT2	#127512
	INCTbiosyn-pUB-HspINT4	#127513
	INCTbiosyn-pUB-HspINT5	#127514
	INCTbiosyn-pUB-HspINT7	#127515
	INCTbiosyn-pUB-HspINT9	#127516
	INCTbiosyn-pUB-HspINT13	#127517
	INCTbiosyn-pUB-HspINTphiC31	#127518
	INCTbiosyn-pUB-HspINTBxb1	#127519
pSP	INCTbiosyn-p35S(rc)2_4_5-GFP	#127520
pSG	INCTbiosyn-p35S-GFP(rc)2	#127521
	INCTbiosyn-p35S-GFP(rc)4	#127522
	INCTbiosyn-p35S-GFP(rc)5	#127523
	INCTbiosyn-p35S-GFP(rc)7	#127524
	INCTbiosyn-p35S-GFP(rc)9	#127525
	INCTbiosyn-p35S-GFP(rc)13	#127526
	INCTbiosyn-p35S-GFP(rc)phiC31	#127527
	INCTbiosyn-p35S-GFP(rc)Bxb1	#127528
pIE	INCTbiosyn-pAct-AtINT2	#127529
	INCTbiosyn-pAct-AtINT4	#127530
	INCTbiosyn-pAct-AtINT5	#127531
	INCTbiosyn-pAct-AtINT7	#127532
	INCTbiosyn-pAct-AtINT9	#127533
	INCTbiosyn-pAct-AtINT13	#127534
	INCTbiosyn-pAct-AtINTphiC31	#127535
	INCTbiosyn-pAct-AtINTBxb1	#127536

Supplementary Table 2. Summary of the statistical analysis of the EGFP-positive cell percentages obtained by flow cytometry assays with the switch GFP system in HEK 293T cells, bovine fibroblasts and plant protoplasts.

	Cell type	Int	Mean \pm sd				Assays Repetition (n)	Kruskal-Wallis statistics (χ^2)	df	p-value
			pGFP	pIE + pSG	pIE	pSG				
Switch GFP system	HEK 293T	Int 2	21.32 \pm 11.84 ^a	1.29 \pm 1.09 ^b	2.93 \pm 3.44 ^b	0.33 \pm 0.22 ^c	3	20.30	3	0.00015
		Int 4	21.32 \pm 11.84 ^a	1.33 \pm 1.27 ^b	0.36 \pm 0.10 ^b	0.17 \pm 0.08 ^c	3	24.94	3	1.59 $\times 10^{-5}$
		Int 5	21.32 \pm 11.84 ^a	2.52 \pm 3.29 ^b	0.35 \pm 0.19 ^b	0.30 \pm 0.14 ^b	3	17.51	3	0.00055
		Int 7	21.32 \pm 11.84 ^a	0.93 \pm 0.83 ^b	0.36 \pm 0.18 ^b	0.47 \pm 0.27 ^b	3	18.31	3	0.00038
		Int 9	21.32 \pm 11.84 ^a	0.48 \pm 0.24 ^b	0.21 \pm 0.08 ^c	0.58 \pm 0.41 ^b	3	21.82	3	7.13 $\times 10^{-5}$
		Int 13	21.32 \pm 11.84 ^a	7.07 \pm 5.31 ^b	0.44 \pm 0.30 ^c	0.39 \pm 0.32 ^c	3	24.79	3	1.71 $\times 10^{-5}$
		phiC31	21.32 \pm 11.84 ^a	10.79 \pm 3.72 ^a	2.36 \pm 2.00 ^b	1.37 \pm 1.22 ^b	3	20.96	3	0.00011
		Bxb1	21.32 \pm 11.84 ^a	16.02 \pm 7.25 ^a	1.82 \pm 1.98 ^b	2.30 \pm 2.74 ^b	3	20.29	3	0.00015
	Bovine Fibroblast	Int 2	1.80 \pm 0.49 ^a	0.12 \pm 0.04 ^b	0.04 \pm 0.03 ^c	0.02 \pm 0.03 ^c	3	29.37	3	1.87 $\times 10^{-6}$
		Int 4	1.80 \pm 0.49 ^a	0.03 \pm 0.03 ^b	0.02 \pm 0.02 ^{bc}	0.01 \pm 0.01 ^c	3	23.04	3	3.97 $\times 10^{-5}$
		Int 5	1.44 \pm 0.79 ^a	0.04 \pm 0.02 ^b	0.01 \pm 0.01 ^c	0.02 \pm 0.03 ^c	3	23.57	3	3.08 $\times 10^{-5}$
		Int 7	1.44 \pm 0.79 ^a	0.03 \pm 0.02 ^b	0.02 \pm 0.02 ^b	0.02 \pm 0.01 ^b	3	20.95	3	0.00011
		Int 9	4.51 \pm 3.99 ^a	0.28 \pm 0.12 ^b	0.02 \pm 0.02 ^c	0.01 \pm 0.01 ^c	3	30.46	3	1.10 $\times 10^{-6}$
		Int 13	4.51 \pm 3.99 ^a	0.74 \pm 0.57 ^b	0.02 \pm 0.02 ^c	0.01 \pm 0.01 ^c	3	29.19	3	2.04 $\times 10^{-6}$
		phiC31	1.34 \pm 0.48 ^a	0.13 \pm 0.03 ^b	0.04 \pm 0.04 ^c	0.02 \pm 0.01 ^d	3	29.68	3	1.61 $\times 10^{-6}$
		Bxb1	1.34 \pm 0.48 ^a	0.33 \pm 0.12 ^b	0.03 \pm 0.02 ^c	0.01 \pm 0.01 ^d	3	30.84	3	9.18 $\times 10^{-7}$
	Protoplast	Int 2	30.95 \pm 8.79 ^a	2.71 \pm 2.29 ^b	0.00 \pm 0.01 ^d	0.02 \pm 0.03 ^c	5	52.31	3	2.57 $\times 10^{-11}$
		Int 4	30.57 \pm 4.31 ^a	17.77 \pm 3.70 ^b	0.00 \pm 0.00 ^c	0.00 \pm 0.00 ^c	3	33.76	3	2.22 $\times 10^{-7}$
		Int 5	23.08 \pm 5.52 ^a	0.14 \pm 0.13 ^b	0.00 \pm 0.00 ^c	0.01 \pm 0.02 ^c	3	28.96	3	2.28 $\times 10^{-6}$
		Int 7	27.91 \pm 10.99 ^a	10.96 \pm 3.92 ^b	0.01 \pm 0.03 ^c	0.00 \pm 0.00 ^c	4	43.01	3	2.45 $\times 10^{-9}$
		Int 9	28.44 \pm 3.03 ^a	16.44 \pm 7.59 ^b	0.01 \pm 0.02 ^c	0.00 \pm 0.01 ^c	3	31.33	3	7.24 $\times 10^{-7}$
		Int 13	21.40 \pm 9.12 ^a	24.24 \pm 9.98 ^a	0.01 \pm 0.02 ^b	0.01 \pm 0.02 ^b	4	38.70	3	2.01 $\times 10^{-8}$
		phiC31	25.84 \pm 8.74 ^a	9.87 \pm 3.14 ^b	0.01 \pm 0.02 ^c	0.01 \pm 0.02 ^c	3	30.15	3	1.28 $\times 10^{-6}$
		Bxb1	25.84 \pm 8.74 ^a	10.47 \pm 2.57 ^b	0.04 \pm 0.10 ^c	0.01 \pm 0.02 ^c	3	30.53	3	1.07 $\times 10^{-6}$

Different letters indicate significant differences among the data in each line obtained by the Kruskal-Wallis test at the 5% statistical probability level.

Supplementary Table 3. Number of PCR cloned sequences after Int activity on the switch GFP systems and observed covered mutations in HEK 293T, bovine fibroblast and plant protoplast cells.

Cell type	Sequence description	Int2	Int 4	Int 5	Int 7	Int 9	Int 13	phiC31	Bxb1	
Switch GFP system	HEK 293T	Total sequenced clones	7	9	16	16	16	17	10	9
		Selected high quality sequences	11	16	31	31	31	31	20	18
		<i>attL</i> coverage mutations	0	0	0	0	0	0	0	0
		<i>egfp</i> coverage mutations	2 SNPs	3 SNPs	6 SNPs	4 SNPs	7 SNPs 2 deletions	0	4 SNPs	5 SNPs
		<i>attR</i> coverage mutations	0	0	0	0	0	0	0	0
	Bovine Fibroblast	Total sequenced clones	11	18	19	24	16	24	10	7
		Selected high quality sequences	16	32	33	48	32	44	20	14
		<i>attL</i> coverage mutations	0	0	0	0	0	0	0	0
	Protoplast	Total sequenced clones	29	37	18	36	12	40	23	15
		Selected high quality sequences	50	66	32	72	23	74	33	28
		<i>attL</i> coverage mutations	0	0	0	0	0	0	0	0
	Protoplast	<i>egfp</i> coverage mutations	2 SNPs 1 deletion	2 SNPs	3 SNPs	10 SNPs	0	5 SNPs	4 SNPs	8 SNPs
<i>attR</i> coverage mutations		0	0	0	1 SNP	0	0	0	0	

Supplementary Table 4. Summary of the statistical analysis of the EGFP-positive cell percentages obtained by flow cytometry assays with the switch promoter system in plant protoplasts.

Switch Promoter system - Protoplast	Mean \pm sd	pGFP	39.12 \pm 12.30 ^a
		pIE Int2 + pSP	32.21 \pm 5.88 ^b
		pIE Int4 + pSP	38.57 \pm 10.60 ^{ab}
		pIE Int5 + pSP	12.55 \pm 4.57 ^c
		pSP	7.35 \pm 2.46 ^d
		pIE Int2	0.00 \pm 0.00 ^e
		pIE Int4	0.01 \pm 0.02 ^e
		pIE Int5	0.01 \pm 0.02 ^e
	Assays Repetition (n)		3
	Kruskal-Wallis statistics (χ^2)		64.99
	df		7
	p-value		1.51 x 10 ⁻¹¹

Different letters indicate significant differences among the data obtained by the Kruskal-Wallis test at the 5% statistical probability level.

Supplementary Table 5. Number of PCR cloned sequences after Int activity on the switch promoter system and observed covered mutations in protoplasts.

	Int2		Int 4		Int 5	
	Switch Promoter system - Protoplast	Total sequenced clones	32	Total sequenced clones	28	Total sequenced clones
	Selected high quality sequences	56	Selected high quality sequences	39	Selected high quality sequences	54
	attL Int 2 coverage mutations	0	attB Int 2 coverage mutations	0	attB Int 2 coverage mutations	0
	attP Int 4 coverage mutations	0	attL Int 4 coverage mutations	0	attB Int 4 coverage mutations	2 SNPs
	attP Int 5 coverage mutations	0	attP Int 5 coverage mutations	0	attL Int 5 coverage mutations	1 deletion
	5' end CaMV 35S promoter coverage mutations	0	5' end CaMV 35S promoter coverage mutations	0	5' end CaMV 35S promoter coverage mutations	1 SNP
	3' end CaMV 35S promoter coverage mutations	0	3' end CaMV 35S promoter coverage mutations	0	3' end CaMV 35S promoter coverage mutations	0
	attB Int 5 rc coverage mutations	0	attB Int 5 rc coverage mutations	0	attR Int 5 coverage mutations	0
	attB Int 4 rc coverage mutations	3 SNPs	attR Int 4 coverage mutations	0	attP rc Int 4 coverage mutations	0
	attR Int 2 coverage mutations	0	attP rc Int 2 coverage mutations	1 SNP	attP rc Int 2 coverage mutations	0
	5' end <i>egfp</i> CDS coverage mutations	1 SNP	5' end <i>egfp</i> CDS coverage mutations	0	5' end <i>egfp</i> CDS coverage mutations	0

rc: reverse complement orientation

Supplementary Table 6. Number of PCR cloned sequences after Int activity on the switch GFP system and observed covered mutations in PBMCs.

Switch GFP system	Cell type	Sequence description	Int 4			Int 13			phiC31			Bxb1		
			Donor 1	Donor 2	Donor 3	Donor 1	Donor 2	Donor 3	Donor 1	Donor 2	Donor 3	Donor 1	Donor 2	Donor 3
			PBMC	Total sequenced clones	10	8	8	11	9	11	9	9	11	9
Selected high quality sequences	17	16		14	22	17	20	17	16	19	17	20	22	
<i>attL</i> coverage mutations	0	1 SNP		0	0	1 SNP	0	0	0	0	1 SNP	0	1 SNP	
<i>egfp</i> coverage mutations	3 SNPs	1 SNP		6 SNPs	10 SNPs	5 SNPs	7 SNPs	1 SNP	2 SNPs	3 SNPs	5 SNPs	3 SNPs	0	
<i>attR</i> coverage mutations	0	0		0	0	0	1 SNP	0	0	0	2 SNPs	0	0	

Supplementary Table 7. Identification, sequences and target systems of the primers used in this study to amplify *attL* and *attR* sites formed after Int activity in HEK 293T cells, bovine fibroblasts, plant protoplasts and PBMCs.

Switch GFP system	Primers to amplify <i>attL</i>		
	Promoter	Forward primer (5' > 3')	Target System
	EFa_966F	TTCTCGAGCTTTTGGAGTACGTCGTCTTTAGGTTG	Mammalian
	35S_282F	ATTGATGTGATATCTCCACTGACGTAAGGGATGACGCAC	Plant
	<i>attR</i>	Reverse primer (5' > 3')	Target System
	attR_Int2_R	GTGTCTACGCGAGATTCTCGCCGGACCGTCGACATACTGC	General
	attR_Int4_R	AGTTTTCAACCCTTGATTTGAATAAGACTGCTGCTTGTGT	General
	attR_Int5_R	ATAACTCTCCTGGGAGCGCTACACGCTGTGGCTG	General
	attR_Int7_R	CTGTGTGAGAGTTAAGTTTACATGGGCAAAGTTGATGAC	General
	attR_Int9_R	TGGAAGTGTGTATCAGGTAACCTGGATACCTCATC	General
	attR_Int13_R	GTAGAACTTGACCAGTTGGTCTGTAAATATAAGCAATCC	General
	attR_phiC_R2	CCAACTGGGGTAACCTTTGGGCTCC	General
	attR_Bxb1_R2	CTGGTCAACCACCGCGGTCTCCGTCGTCAGGATC	General
	Primers to amplify <i>attR</i>		
	<i>attL</i>	Forward primer (5' > 3')	Target System
	attL_Int2_F	GGAGTAGCTCTTCGCCCGAGAACTTCTGCAAG	General
	attL_Int4_F	CGACCTGAAATTTGAATTAGCGGTCAAATAATTTGTA	General
	attL_Int5_F	GACGGCCTGGGAGCGTTGACAACCTGCGCACC	General
	attL_Int7_F	GTCCGTCTGGGTCAAGTTGCCTAACCTTAACCTTTAC	General
	attL_Int9_F	ATAATTGGCGAACGAGGTATCTGCATAGTTATTCCGAAC	General
attL_Int13_F	TCCAGATCCAGTTGTTTTAGTAACATAAATACA	General	
attL_phiC_F	TGCCAGGGCGTGCCCTTGAGTTCTCTCAGT	General	
attL_Bxb1_F	TGTCGACGACGGCGGTCTCAGTGGGTACGGT	General	
Backbone	Reverse primer (5' > 3')	Target System	
TermiAni_205R	AATGATTTGCCCTCCCATATGTCCTTCCGAGTG	Mammalian	
NOS _t _283R	ATAACAATTTACACAGGAAACAGCTATGACATGATTACG	Plant	
BB_Termi_R*	GTAAACGACGGCCAGTGAATTGTAATACGACTC	Plant	
Switch Promoter system	Primers to amplify <i>attL</i>		
	Backbone and Promoter	(5' > 3')	Target System
	Pré_Ints_sitesAt_312F	GCGAAAGGGGATGTGCT	Plant
	35S_125R	TAGGAGCCACCTTCCTTTTCC	Plant
	Primers to amplify <i>attR</i>		
	Promoter and <i>egfp</i>	(5' > 3')	Target System
	35S_64F	ATCCTTCGCAAGACCCTTCC	Plant
SGFP_150R	TGGTGCAGATGAACTTCAGG	Plant	

*used only with attL_phiC_F and attL_Bxb1_F

Supplementary Methods

Genetic part sequences used to build the vector sets (5' > 3')

Integrase recognition sites (**orange: core**)

attB Int 2

ggacggcgcagaaggggagtagctct**cgccg**gaccgtcgacatactgctcagctcgtc

attP Int 2 reverse complement sequence

agccaagagcagtgccctgcagaagttctcg**ggcga**gaatctcgcgtagacacatacatgagc

attB Int 4

ttccaaagagcgcaccaacgcgacctgaa**attgaa**taagactgctgcttgtaaaggcgatgatt

attP Int 4 reverse complement sequence

aaacgaattacaaattatttgaccgcta**attcaaat**caaggggtgaaaactttgtaatttt

attB Int 5

gagcgcgggatcagggagtgacgg**cctgggagc**gctacacgctgtggctgcggtcgggtgc

attP Int 5 reverse complement sequence

cagatcaggggtgcgaagttgtca**cgctcccagg**agagttatcgactgctgattaggg

attB Int 7

agacgagaaacgttccgtccgtctgggtcagtt**ggc**aaagttgatgaccgggtcgtccgttcctt

attP Int 7 reverse complement sequence

aagctgaacctgcgtaaaagttaaggttagg**cat**gtaaacttaactctcacacaggtttataacacc

attB Int 9

tttatattgcgaaaaataattggcgaacgaggtaa**ct**ggatacctcatccgccaattaaaatttg

attP Int 9 reverse complement sequence

taattggaagttcgggaataactatg**ag**atacctgatacacactccaacaaaaacaaccac

attB Int 13

gcatacattgttgttttccagat**ccagttg**gtcctgtaaataagcaatccatgtgagt

attP Int 13 reverse complement sequence

tattcggagttgtattatgtactaaaa**caactgg**tcaagttctacaaatacaaccgttattg

attB phiC31

tgccagggcgtgccct**tg**ggctccccgggcgcg

attP phiC31 reverse complement sequence

ccaactgagagaact**caa**aggttaccaccagttg

attB Bxb1

tcggccggctgtcgacgacg**gggtctc**gctcgtcaggatcatccgggc

attP Bxb1 reverse complement sequence

gtcgggggtttgtaccgtacaccact**gagaccgc**gggtggttgaccagacaaaccacgac

Expected *attL* and *attR* sites resulting from inversion recombination
(**orange: core**)

attL Int 2

ggacggcgcagaaggggagtagctct**cgccc**gagaacttctgcaaggcactgctcttggtc

attR Int 2

gacgagctgagcagtatgtcgacgggtcc**ggcga**gaatctcgcgtagacacatacatgagc

attL Int 4

ttccaaagagcgcccaacgcgacctgaaattggaattagcgggtcaaataattgtaattcgttt

attR Int 4

aatcatcgcctttacacaagcagcagctctattcaaatcaaggggtgaaaactttgtaattttt

attL Int 5

gagcgccggatcagggagtggacggcctgggagcgttgacaactgcgaccctgatctg

attR Int 5

gcaccgaccgcagccacagcgtgtagcgctcccaggagagttatcgactgctgattagg

attL Int 7

agacgagaaaacttccgtccgtctgggtcagttgctaacttaacttttacgcaggtcagctt

attR Int 7

aaggaacggacgacccgggtcatcaacttggccatgtaaacttaactctcacacaggtttataacacc

attL Int 9

tttatattcgaaaaataattggcgaacgaggtaaatgcatagttattccgaacttccaatta

attR Int 9

caaatttaattggcggatgaggtatccagatacctgatacacacttccaacaaaaacaaccac

attL Int 13

gcatacattgttgttttccagatccagttgttttagtaacataaatacaactccgaata

attR Int 13

actcacatggattgcttatattacaggaccaactggcaagttctacaatacaaccgttattg

attL phiC31

tgccagggcgtgcccctgagttctctcagttgg

attR phiC31

cgcgcccggggagccaaaggttaccacagttg

attL Bxb1

tcggccgctgtcgcagcagcgcggtctcagtggtgtacggtacaaaccccgac

attR Bxb1

gcccggatgatcctgacgacggagaccggtggtgaccagacaaaccacgac

Mammalian plasmid parts

Ubiquitin C promoter

ggcctccgcccgggttttggcgctcccggggcggcccccctcctcacggcgagcgtgccacgtcagacgaagggcgag
cgagcgtctgatcctccgcccggacgctcaggacagcggcccgtgctcataagactcggccttagaaccacagtatcagca
gaaggacatttaggacgggacttgggtgactctagggcactggtttcttccagagagcggaacagggcaggaaaagtagtcc
cttctcggcgattctcggagggatcctcgtggggcgggtgaacgccgatgattatataaggacgcgcccgggtgtggcacagctag
ttccgtcgcagccgggatttgggtcgcggttctgtttgtggatcgctgtgatcgtcacttggtagtagcgggctgctgggctggccg
gggcttctgtggccgcccgggcccgtcgggtgggacggaagcgtgtggagagaccgccaagggctgtagtctgggtccgagc
aaggttgcctgaactgggggtgggggagcgcagcaaatggcggctgtcccagcttgaatggaagacgctgtgaggc
gggctgtgaggtcgttgaacaaggtgggggcatggtgggcccgaagaaccaaggtctgaggccttcgtaatgcccggaa
agctcttattcgggtgagatgggctggggcaccatctggggaccctgacgtgaagttgtcactgactggagaactcggtttgcgtc
tgttgcgggggcccagttatggcgggtccgttgggcagtgacccgtaccttgggagcgcgcgccctcgtcgtcgtgacgct
accgttctgttggcttataatgcaggtggggccacctcgggtagggtgtcggtaggcttttctccgtcgcaggacgcaggggtc
gggctagggttaggctctcctgaatcgacaggcggccggacctcgttggaggaggataagtgaggcgtcagtttcttggctg
gtttatgtacctatcttctaagtagctgaagctccggtttgaactatgcgctcgggggtggcagtggttttgaagtttttaggcac
ctttgaaatgtaataatttgggtcaatatgtaatttctagtgtagactagtaaattgtccgctaaattctggccgtttttgcttttttag
ac

EF1alpha promoter

gctccggtgcccgtcagtgggcagagcgcacatcgcccacagtccccgagaagtggggggagggtcgccaattgaaccgg
 tgcctagagaaggtggcgcgggtaaaactgggaaagtgatgtcgtgactggctccgcttttcccagggtgggggagaacc
 gtatataagtgcagtagtcgccgtgaacgttcttttcgcaacgggttgcgcccagaacacaggttaagtccgtgtgtggtcccgc
 gggcctggcctcttacgggtatggccttgcgtgccttgaattacttccacgcccctggctgcagtagctgattcttgatcccagct
 tgggttgaagtgggtggagagttcgaggccttgcgcttaaggagccccttgcctcgtgcttgagttagggcctggcctggg
 gctggggccgcccgtgcgaatcgtgtggcaccttgcgcctgtctcgtgctttcgataagtctctagccattaaaattttgatgac
 ctgctgcgacgctttttctggcaagatagcttgaatgcgggccaagatctgcacactggatttccggttttggggccgcccggg
 gcgacggggcccgtgctcccagcgcacatgctcggcgaggcggggcctgagcgcggccaccgagaatcggacggggg
 tagtctcaagctggccggcctgctctggtgctgctcgcgcccggctgtatcgccccgctggggcggcaaggctggcccgg
 cggcaccagttgctgagcggaaagtggccgctcccggcctgctgcagggagctcaaatggaggacgcggcgtcggg
 agagcgggcccgtgagtcacccacacaaaggaaaaggccttccgtcctcagccgtcgttcatgtactccacggagtagc
 gggcgcctcaggcacctcgattagttctcagcctttggagtagctgctttaggttggggggagggtttatcgatggagttt
 cccacactgagtggtggagactgaagtagccagctggcactgatgtaattctccttgaatttccctttttagttggatctt
 ggtcattctcaagcctcagacagtggtcaaagtttttctccatttcagggtgctgga

egfp reverse complement sequence

ttactgtacagctcgtccatgccgagagtatcccggcggcgggtcacgaactccagcaggaccatgtgatcgcgcttctcgttgg
 ggtcttctcagggcggactgggtgctcaggtagtggttgcgggcagcagcagcggggccgtcggcagtggggtgtctgctgg
 tagtggcggcagctgcacgctcctcctgatgtgtggcgatcttgaagttcaccttgatccgttcttctgcttgcggccatg
 atatagacgttggctgtttagttgactccagcttgcggcaggatgttccgtcctccttgaagtcgatccctcagctcgtatg
 gggtcaccagggtgctgcctcgaactcactcggcgggctctgtagttgcccgtcgtccttgaagaagtgggtcgtcctgga
 cgtagcctcgggcatggcggactgaagaagtcgtgcttcatgtgttggggtagcggctgaagcactgcacgcccaggt
 cagggtgtcagaggggtggccaggggcagggcagcttccgggtgggtcagatgaactcagggtcagcttccgtagggtg
 catcggcctcggcctcggcggacacgctgaactgtggcgtttacgtcggcctcagctcagaccaggtgggaccaccccgg
 gaacagctcctcggccttgcaccat

 β -globin poly(A) signal terminator

agcggccgactcctcaggtgcaggtgcctatcagaaggtgggtggtggtggccaatgccctggctcacaataaccactga
 gatcttttccctcgcacaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaggaaatttttcatg
 caatagtggttgaatttttgtgtctcactcggaaaggacatattggaggcacaatcatttaaaccatcagaatgagtttggtt
 agagttggcaacatagccatagctggtgcatgaacaaagggtgataaaagggtcatcagatatgaaacagcccctg
 ctgtcattccttattccatagaaaagcctgactgaggttagattttttatattttgttttattttttttaaaccctaaaatttcc
 ttacatgtttactagccagattttcctcctcctgactactccagctatagctgtcccttctcttatgaagatccctcagcctgag
 cccaagcttggcgtaatcatg

Integrase 2 *H. sapiens* codon optimized sequence

atgccatcgccccggagttcctgagttggcctatccgggacaagagtttctgcttactctacgggagggctccaggatccg
 aagagaaaaggtcgaagcgtccaaagccaattggacgaagggagggcaacctgtcttgatgctggatggcctattgtgggga
 attaaagatgtagacagatcagcatctgcctatgctaggaggacgcgagatgaatttgaagagatgatcgcaggaatccaggc
 gggggagtgcggattctgtggttccgaagcgtctcgatattacagagatttggaggcctatgtcggctcggcagatcgcgg
 gaagcaggcgttcttctgtataacggtcagggtgacgatcttagtaagagtgcagatagaaaggcagcggctcaggatgctg
 gaacgcagaaggcgaagccgacgatataagagaaagaaacctccgaccaccagactaatgccaagcaggggggagct
 catggcctgtcccagatggatataagagacgatatgatccagattccggagatctgtcgaccaaatcccgcaccccgcaga
 gccggttgataactgagatttttagcgggacgcccggcggcagccactggccgctatctgcccagatcctaatgagaggggc
 gaaactaccaccaggcgaagcatggcagaggcaccactgcatgcatcctcgaatccagcttatatcggctatagaag
 acatctcgtgtgacacaggtaaaggaatgtgggcacctatttgcgatgataggactttgcggaaaccttccaggctgcccagg
 aatcctcagcttccagggagacagttgagcccggccctgaagcgaacatttgcagacaggatcgcactttgcggagaa
 cccccgacgaaccgcccgttaggagtgactgtcaggccgcacaaatacaattgtccacaagatagatgtggcgtatg
 cgagaagaccgatggatgcttgcgaggagtcctcatcacgtgctggcctccgacgaagcagttgagcttttgaagata
 atactgacgatgaaccgacaagaaaagcaggatcagactcaaagttctgaagagcagctcagggctgcacaaaaacagg
 ctagaacactccgcccggacggaatggcatgctccttagtatagattccttggccggccttggaaagccgaacttactccacagata
 gacaaggcagggcaggagagtcggtcttgcacgtacccgtctcttgagagacctgtgggcaaccgagctgatgtgac

cgcgcatggaacgaagcactcacgttgccccagcgccgaatgatactgcgcatggtgtcactatccgcctctcaaagcaggat
ctcgaggtgtacggcgatagaccaggtcgaattaccttgagttatgttgagaaccaggatttaagcccgtggcggaatcg
agcaaaacagtaa

Integrase 4 *H. sapiens* codon optimized sequence

atgatcaccacacgcaaggtggccatatacgtccgctatctacgactaatcaagccgaagagggatattctatacagggccaa
atcgattcccttatacaagtactgagggctatgggtggatcatatatgaggaatatactgacgcagggtttcagggcgcaaaattg
atcgcccgccatgagtaagcttactgatgccaacacaagagattcgataaccattctggtgtacaaattggatcggtgtcaa
gaagcgtgccccgatacacttacttgtaaggatgtatttaatacaaaacaacatccactctctcccttcaggagaatatcgacac
atctccgcgatgggaaatctttttgacctctcttctgcatagcggagtttgagcgagagcagatcacggagcgcgatgacctg
gggaaaataggtcgggcaaatcaggggaagacaatggcgtggacctatacacggctcggctacgattataataaggagaagg
gtgaacttatactgatcctgccaaggcaccatcgtgaaaatgatatacaccgattactgaagggtatgtcaattcaaaagatag
tgataaactcaataaaatggattacaacggcaaggattgcacatggttcccacggcgtgaaacatctctggataatccgggtg
tactacggatgacaagatatacaataagctttcccggcaatcatcagccgataattacaaggaggtgtttgataagacca
gcgcgagagacaacggagcgcactcggcatcgaggaaaaccattatacaataccgtttcaggcaaaatatactcagtaagt
tcctcaggtgtccagtgccgctcccgaatgggcttgattgggtcgcggcgcaagaagggaaggcaagcgcagtaagaa
atactactgctcaatagccgaccgaaacggacagcttcatgtgacactcccctgtatgacgccgaaacctggaagattatgtg
ctgcacgaaatagctaaaatacagaaaagacctcaatagcaagtagacagaagcatattgaagatcacgaactcaaatata
agcgagaacgcgatagaggcaaatataaataaaacctgaaatcaacttagcaagctgaacaacctttaccttaacgacctcatta
ctcttgaggacctcaagacacagactaacacactcatcgcaaaaagcggctgttgagaatgaactgacaagacatgcgat
aatgacgcgaactgacagacaagagacaatagcggacttctggcattgccagacggttgactatggactatgaaggtaa
aagtatccgtggagcttctcgttcagcgcgtcaaagttgacagagacaacatagacatacattggacgtttaa

Integrase 5 *H. sapiens* codon optimized sequence

atgccgggaatgaccacggaaaccggaccagatccggcaggttgatagacctgtttgccgcaaatctaaggcagtgaaaag
ccgggctaattggagctgggcaacggagaaaacaggaaatctctattgtgctcaggagacgctggggcgcaaggctcgtgcc
ctgctcggcatgcaagtcgccatgtgtggaaggaagtaggtagcgcgagtcgattcagaaaaggaaaggcaagggatgatc
agagtaaagccctaaagcacttgagagtggggaggtcggggcgctctggtgctatcggctgaccgatgggataggggagggc
gcgggagcaattctaaaattatagagcccgaggatggatgccgcaagattgctttcggctgggatgaggatacaggtcgac
ccgtattggattctacaaaataaagggaccggggtagctgatcagacgagcggaggaagcccgagaagaagcagaaaaa
ctgtcagagcgcgtccgcgataactaaggcacatcagcgcgagaacggggaatgggtaaacgcgcgcgaccatattggcctcc
gagtggtctgtgactgttagtgatgaggagggcgacgagatgatgaaaggaaattggctgcagacgatgaagatgctggtg
gaccgatggtctcaaaaagctgaagcggcagcggctggtcttactctcccgttaccgaccgacttctcagcaggtaccgc
ccatgcaatgaacactcgggaaattccatcaccgacagggcggctcatgtagctgtgacagttaggacatgatccagaacc
ccgcttatgccccgtggcagaccacagggcagggcaggtggaaaacaaaggcgttgacattctacaacggagaggggaaa
aggttctctgtcatgcacggcccaccttgtaacagatgaagaacaggaagcggccaaggcggcagtcaaaaggcgaaga
cggagtaggtgtccgctgacggctcagatcatgacacacgccgcaaacatctcctcagtggttaggtgcatgtccgggatgt
ggtggctcatgctcactcaggtaacggttaccggtgttgagatctagtgttaaaggggctgtccagcaccgacgtacgtggc
cagaaagtcagtcgaggagtatgtgcttttaggtgggcccgtaaactggcggcctccgagcctgatgacctttgtaatagctgt
ggcagatcgtggcagctttgacacaccacaggttccgaagatgaaaaatacgttaaagctgcggttcgagaagcagag
aagaatctggccggtgctgcgagaccggcagaacggcgtctatgatggcccggagcagttctttgcgccggcttatcaa
gaagcccttcaacgctgcagggcggtaaggatgctgtctcagaaagctccgcacccgacagtggtatgaaagttggatagtg
atagttccgattacgaggagctgtggctcggcgacccccgacaatgaggaatgcgattatgatcttgcacgacgagattgg
gtggcaagggaacaacgagggagaccgtttgatggagatgaacgggtcaagattaaatgggcagctcggacgtaa

Integrase 7 *H. sapiens* codon optimized sequence

atgaaggtagctatatacgtccgctcaacagatgaacaagctaaggagggattctccattccggcgcagagagagaggctc
agggctttgtccagtcagggtgggagatcgtgcaagaatatacgaagagggctggagcgaaggacctgatcgcca
cagatgcagcagtgctgaaggacataaaaaaagggaacatagacatagtgctggtctataggctgaccgactaccaggtca
gtcctcgatctctacctgttctcaaacattcgaaaagtataacgtagccttccgatccgctaccgaggtttacgacacaagcaca
gctatggggcgacttttattaccttggttggcgtttggcgcaatgggaacgcgaaaatctcgcagagagagtaaaagttcggcata
gaacagatgatcgatgaaggcaaaaagcctgtgggcatcaccatattggctataagttcgataaagacttcaactgtacgattat
tgaagaggaggcagacgtttaggtatgataaccgcatgactcgcgatggatggctataggagattgccgatcgcactgaatg
agcttatggtaaaaccaagaatagctaaagagtggaatcataacagtggtcgcgacatcctactaactacgacatataatcgccac
gtaccgatggggagacaaggtggttccaaacaaccaccacctatcattagtgaactttgttaagaaagcgaaaaaagaaa

aagaaaagcgaggagtggatcgaaagagagttggaagtctctggtcttccagtggttaattgtggaggacacaag
 atgcaggggacttcgacaaacgagcaaaagacttactacagatgcacgaaggtcacaggattaccaatgagaaaaata
 tcttgaacctgttgatgagatacagttgctgataacatccaaggaatattttatgtctaaatttagtgaccgatatgatcaacag
 gaagtagtagacgtttctgctcttacgaaagaactggagaagatcaagcgccaaaaggagaagtggatgatctctatattggac
 gatcgcaatcccattcctaaggaagagttgttgcgaaaatcaatgaactgaacaagaaagaagaagagatatattccaagttg
 agcgaggtagaggaggacaaagagcccgtcgaagagaaatgcaatagactgtccaagatgattgatttaagcagcaattcg
 aacaggcgaacgacttcacgaaaaaggaattgttttagcataattgaaaaaatcgtaatttatagggaaaaaggttaagctgaa
 gaagataactctggactacactctcaagtaa

Integrase 9 *H. sapiens* codon optimized sequence

atgaagtgccattacacgagtgagtagtcttggaaacagaaggagaaggggcatagtagggaacaggaacgcaagc
 tgcgggctacagtgacatcaacgattgaaaaatccataaagtatacactgacgcaggatacagcggggcgaaaaaggatcg
 cccagcacttcaggaaatgttgatgagattgataatttgacctcgtctcaagttggataggctacaaggagtgtgaaa
 gacctcttgagattctcgagctgttgaacaagaacgtcctgtttgatcagccactgaggtctacgatactacgtcagctatgg
 gtcggctcttctactctggtggcgctatggctgagtggaacgcacgacaattcaggagagaactcgcgatggggcgacgagc
 gagtgcggaaggacttgccaagactgtccaccgtttactacgacaggggtgaatgataagttgtcccaacgaatataaaa
 aggtcctgcgctcgcggtcgaagaggcaaaaaagggacaagcttgcgagagatcaccattaaactgaataactccaagtat
 aaagctccattgggcaagaattggcatcgaagtgtgataggaaatgcactgacgagcccgttggcagggggcacttggctttg
 gcgacattttgtcgagaacacccatgaggcgattataagtgaagaagagtagcaggagataaaaactcaggatctctgagaaa
 acgaatagcacaatagttaagcacaacgccatattcagatcaaaactgctgtgtccaattgcaaccagaagctgacactgaat
 accgtgaaacatacccaaaagaacaaggaggttggattctaaactgtactctgcagtaattgcaagaacacaaagaacaaa
 aatgcatgcaacattgacgagggagaggtgctgaagcaatttataaactgaagcaattgacctgacgtcctacaaaataga
 aaaccagccgaaagaaattgaggatgtgggtagcagatcgaaaagcttgcgaaggagcgggcacgatgccagacactttc
 attgagggcatgatggacaaagatgaagctttccatcattcaagaatagataaggagatccatgagtagagaagcggag
 gataacgacaaggggaagacgttcaactacgagaagataaagaattcaagtagctcccttctaatgggtgggaactcatggaa
 gacgagctgaaaacggaattcatcaaaatggctatcaaaaacatccacttcgagtagtgaaggggattaaaggtaaaagaca
 aatagctcaagataaccgggtatagatttactaa

Integrase 13 *H. sapiens* codon optimized sequence

atggcagtcggaatttatattcgagtagtaccgaagaacagggcgtccgaaggccactctatagaatccccaaaaaagaagctg
 gccagctactgcgagattcagggatgggacgattatcgctctacatcgaggaaggatctccgggaaaaacacaaatcgcca
 aagttgaaactgctcatggaacatattgaaaagggaagattaatattctcctcgatatacgctcgcagcactcacgctctgtg
 atcgattgcacaagctgctaacttttgcaggagcatgggtgtgcatttaatcagcgacagagacctatgatacgcgacagc
 gaatggtcgcagtgccatgggtattgtcagctcttggcccaatgggaaacagagaatatgtccgaacggataaaaacttaaccttg
 aacacaaggtgctcgtcgaaggggagagagtcggagcgatcccctacggattcgacctcgcgacgagaaaagcttgaag
 aacgaaaaatctgcatattgtgacatgggtgaaagagtcgagaacgggtggtcagtagcaatcggattgtcaactattgaaactt
 actaataatgatcggaactgggtcccgaacgggtgttctgcgctgtcggaaacccggccctgtacggcgccacacgggtggaac
 gacaagatcgtgaaacacgcacgaaggcattatctctaaagaagatttaacggctcagcagatattggcagaccgctcc
 atacaccaccgaaggacgtgaaaggtacttatattcaaggggtcttgcgagtagtcccgtctgcgaccaaacctgagtgtaaa
 ccgctttataaaaaacggaaggatgggacggagtagtggcgctctgtatagatgccaacatgcataaaacagaataaata
 caatcttgccatcggagaagctcgtatcctgaaggctcctcaatgaatataatgtccacagtagagttcagacgggtgaggatgaagt
 aatacccaaaaaaagtgagcgagaaatgctcagtagcacagctcaacagatcgctagaaagcgagagaagtacaaaagg
 cgtgggctcagacctatgagtgacgatgagttgaaaagctgatgtagaaacgagagaaacgatgatgagtgtaacaaa
 agcttgagctgtgaaagaccaataaagatcgacgagacttacctaaagagatagttatatttccaccagacgttcaacgatc
 tggagctgagaagcaaaaaagaatttattctcaaagtttattcgcagcattcgatacactgtaaggagcaacaaccaattcggcc
 ggataaatctaagacggggaagggaagcagaaagttatcataacggaggttagtctatcagtaa

phiC31 *H. sapiens* codon optimized sequence

atggacacatacgcaggagcgtatgatcggcagtcggcgagcagagaaacagcagtgccgcaagcccggcgaccagcg
 atccgcaatgaagataaagcggcagacctccagcgggaagtagaaagagatggaggtagggttagattgtaggacattttcc
 gaggctccaggaacctccgcattcggcagcggcagaacgacctgagttgagcgcacctgaatgaatgtaggccggcaggct
 caacatgatcatagtatatgacgtgtcagcttctccaggctcaaggtgatggacgctattccaatcgtctcagagcttctcgtctcg
 gagtgacctcgtcagcaccgaagggtgtctcaggcaggggaacgtaattggatcttatacatcttattatgcgctggagcctt
 cccacaaggaatcatccctcaagtccgcaagatactcgatactaaagaacctgcagcgcgagttgggtgggtattgttgggaaa
 ggccccatacggttcgaattggtgagcgaacgaaggagattaccaggaatggcgcatgggtaattgtgtaataaacaactg

gctcacagcacaacccccactgaccggccccttcgagtttgaacccgacgtgattagatgggtggggggagatcaagacgcat
 aagcaccttccgtttagccgggtcacaggccgcaattcatcctggttctatcacaggtttgtgcaaaccggatggatgctgatgcg
 gtgcctaccgaggggaaacgatcggcaaaaagacggcttcatcagcctgggatccggccacggttatgcgatacttagaga
 cccacgcatcgcgggcttgcagccgaggttatataaaaaaaagccagatggcacacctaccacaaaatagaagggtac
 cggattcagcagatccgattactctcaggcctgttgagctggattgctggtccgataattgagcccgtgaatggatgaactccag
 gcgtggctcagcggtagggggcgcggaagggcctgagccgagggcaggctatactgtccgccatggataaactgtattgca
 gtgtggcggtaatgacctcaaaaagaggagaagaaagcatcaaggattcatalcagatgcagaaggcgcaaagttgtggacc
 cctctgcacctgtcaacatgaggggacatgcaatgtatcaatggctgcgctcgacaagttttagccgagcggatattcaataag
 attagacacgcccaggggagacgaggagactctgcactgctctgggagggcagcccggcgtttgggaaactgacggaggccc
 ctgaaaaaagtgtgaacgcgctaacctgtagcggagagagcggatgcaactcaacgcgttgaagaactttatgaggaccgg
 gcagcaggcgcctatgacgggtccagtcggggcgaagcacttaggaaacaacaggccgcacttactcttccagcagggggg
 cggaggaacgactcgcgtgagttggaagcggcagagggccccaagtggccttgaccagtggttcccgaagatgccgatgca
 gatccaacagggcctaagagctgggtggggcgggcgtccgtcgatgataagagagtgttgttggttctgtggacaagattgt
 ggtaaccaagtcaactactggctcgggtcaaggtaccccaatagagaagcggggccagcataacatgggcaaagcccccgac
 ggacgacgatgaggacgatgctcaagacggaaccgaagatgttgctgcttag

Bxb1 *H. sapiens* codon optimized sequence

atgccccactgggtggaattaggctctccagagttactgacgctacgaccagccctgagaggcaactgaaagctgccagcag
 ctctgtgcgcaacggggctgggacgtcgttgagtgccggaagacctcgatggttccggtgcagtagatcccttcgataggaagc
 ggcgcccgaattggctcgatggctcgttccgaagagcaaccattcgatgtgatcgttcctatcgagtcgacaggtgacaaga
 agtattcggcatctcaacaattggttcattgggcagaagatcacaataatggtgtctccgctacggaagcacactttgatacaa
 ctacaccgttgcagcagttgtcatagctctgatgggtaccgttgcacaatggaattggaagctattaagagcgaatcggagc
 gccgagcactttaacatcagggcaggcaagtagccgagcagctctcccaccctgggctacctgccgactagagtggtggcga
 atggcgggtggtgcccgaaccggtccaacgggaaagaatcctgaagtgtaccatcgagttgtggacaatcacgagcccctcac
 ctggtagctcatgattgaatcgcggggagttttagccgaaagattatttgcgccagctgcaagggagggaaccacagggtc
 gggagtgagtgctactgctctgaaaagatccatgattagcgaagccatgctgggatacgcaacccttaacggcaaaaccgta
 gggacgatgacggagcggcccctgggtcggggccgaaccaattcttactcgcgagcaactggaagcactgcccggctgagctggtc
 aaaaacatcacgggctaaaccggcgggtgagcacaccaagttgttctcagatattgtttgcgcccgtttgtggagagccagcatac
 aagtttccgggggtgggagaaaacatccccgctatagatgccgagcagatgggctcccgaagcattgtggaaaacggtaacagta
 gcgatggctgagtggtgagtcattctgcaagaacaagtccttacttctgctgacgcagaaacgcctggaaaaggtatgggta
 gctgggtccgactccgctgtagaactgtctgaagtcaatgcagaactggtcgatctcacatccctattggctctccagcgtaccga
 gctggctcccgcgaacgcgaggcacttgatccagaatagcggcactcggcgagacaagaggaattggaggggttgaa
 gctagaccttggatgggaatggagggaaacaggccagcgggtcggggattgtggcgagcaagatactcggcggaaga
 atacgtggcttaggagcatgaacgtccggctgacgttcgacgttagaggggggcttaccaggactatcgattcggagattgacg
 gagtatgaacaacacttgaggcttgtagcgtagtgaggagagactgcatacagggatgtcttag

Plant plasmid parts

Actin2 promoter

aaaattagaacgaacttaattatgatctcaatacattgatacatatctcatctagatctaggttatcattatgaagaaagtttgacg
 aatatggcagcaaaaatggctagactcgatgtaattggatctcaactcaacattatacttataccaaacattagttagacaaaatt
 taaacaactatTTTTATGTATGCAAGAGTCAGCATATGTATAATTGATTCAGAATCGTTTTGACGAGTTCGGATGTAGTAGCCATTAT
 TTAATGTACATACTAATCGTGAATAGTGAATATGATGAAACATTGTATCTTATTGTATAAATATCCATAAACACATCATGAAAGACACT
 TCTTTCACGGTCTGAATTAATTATGATCAATCTAATAGAAAACGAATTAATTACGTTGAATTGTGAAATCAATTGAACAAGCCA
 ACCACGACGACGACTAACGTTGCTGGATTGACTCGTTAAGTTAACCCTAAAAAACGGAGCTGTATGTAACACGGGATC
 GAGCAGGTACAGTCATGAAGCCATCAAAGCAAAAAGAACTAATCCAAGGGCTGAGATGATTAATGTTAAAAATTGTTAAC
 ACGAGGGGAAAAGGCTGTCTGACAGCCAGGTACGTTATCTTACCTGTGGTCGAAATGATTCGTGTCTGCGATTTAATTATTTTTGAA
 AGGCCGAAAATAAGTTGTAAGAGATAAACCCGCCTATATAAATTCATATTTTTCTCTCCGCTTGAATTGTCTGTTCTCTCTCAC
 TTTCATCAGCCGTTTGAATCTCCGGCAGCTGACAGAGAAGAACAAGGAAGAAGACTAAGAGAGAAAAGTAAGAGATAATCCA
 GGAGATTCATCTCCGTTTGAATCTCTCAATCTCATCTTCTCCGCTCTTCTTCCAAGGTAATAGGAACCTTCTGGATCTACTTATTGCT
 GGATCTCGATCTGTTTCTCAATTCCTTGAGATCTGGAATTCGTTAATTTGGATCTGTGAACCTCCACTAAATCTTTGGTTTACTAGAATC
 GATCTAAGTTGACCGATCAGTTAGCTCGATTATAGCTACCAGAATTTGGCTGACCTGTAGGAGAGATCCATGTTTACTGTTACCTGGG
 AATGATTTGTATGTGAATTGAATCTGAACCTGTTGAAGTTAGATTGAATCTGAACACTGTCAATGTTAGATTGAATCTGAACACTGTT
 AAGGTTAGATGAAGTTGTATAGATTCTCGAAACTTAGGATTTGATGTCTGACGTTGAACAGAAAGCTATTTCTGATCAATCAG
 GGTATTGACTGTATTGAACCTTTTTGTGTTGTCAGCTATAAAAA

CaMV 35S promoter

taattcatcaaatataactatagaccctaatttcatatgagacttttcaacaagggtaatatccggaaacctcctcggattccatt
 gccagctatctgtcactttattgtgaagatagtggaaggaagggtggctcctacaaatgccatcattgcgataaaggaaaggc
 catcgttgaagatgcctctgccgacagtgggtcccaaagatggacccccaccacgaggagcatcgtggaaaaagaagacgttc
 caaccacgtcttcaaagcaagtggtgattgatgatatctcactgacgtaaggatgacgcacaatcccactatccttcgcaagac
 ccttctctataataaggaagttcatttcatttggagaggactaactgtgccaaattatcttccaaactcctaattccaatt

egfp reverse complement sequence

ttactgtacagctcgtccatgccgagagtgtcccggcggcgggtcacgaaactccagcaggaccatgtgatcgcgcttctcgttgg
 ggtcttctcagggcggactgggtgctcaggtagtggttgcgggcagcagcagcggggccgtcgcgcatgggggttctcgtg
 tagtggcggcagctgcacgctgccctcctcgatgttggcggatcttgaagttcaccttgatgccgttctcgttgcggccatg
 atatagacgttggctgtttagttgactccagcttgtccccaggatgttgcgctcctcctgaagtcgatcccttcagctcgtatg
 gggtcaccaggggtgctgcctcgaactcacctcggcgggcttctgtagttgccgtcgtcctgaagaagatgggtgctcctgga
 cgtagccttcgggcatggcggactgaagaagtcgtgcttcatgttgcgggtagcggctgaagcactgcacgccgtagggt
 gaaggtggtcacgaggggtggccagggcagggcagcttgcgggtggtgcagatgaacttcagggtcagcttgcgtaggtgg
 catcgcctcgcctcgcggacacgctgaactgtggcggttacgtcgcgctcagctcgaccagatgggcaccaccccggt
 gaacagctcctcgccttctcaccat

NOS terminator

gaattccccgatcgttcaaacatttggcaataaagtttctaagattgaatcctgttgcgggtctgcgatgattatcatataatttctgtt
 gaattacgttaagcatgtaataatttaacatgtaatgcatgacgttatttatgagatgggttttatgatttagagctccgcaattatacattta
 atacgcgatagaaaaaaaatcgcgcgcaaaacttgataaattatcgcgcggtgtcatctatgaggactagatcg

Integrase 2 *A. thaliana* codon optimized sequence

atgccaatagcgcagaaatttctcacttgctatccagggtcaagagtccccgcgtacctttacggacgagcttctagggatccga
 aacgtaaggaagatctgttcaatctcaattagacgaagggagagctacgtgtcctgatgcaggggtggccaattgctggtgagttc
 aaagatgtcgtatagatcggcctccgcatatgctagacgtacacgtgacgaatttgaggaaatgatagcaggaattcaggccggc
 gaatgtagaattctcgtggccttgaagcatccagatattatcgagatttagaggcgtatgtcaggcttagacgagatgtcgtgaag
 ccggagtctgttgttacaacgggtcaggtttacgatttaagcaaatccgccgaccgaaaagcgaccgcacaggacgcccgttaa
 tgcagaggggtgaggcagatgatattagagaacgtaattgagaacaactagattgaatgcaaagagaggaggcgtcatggcc
 ccgtgcctgatgggtataaaaggagatacagaccctgattcaggtgatttagttgatcaaatcccacatccagatagagctggtctg
 ataacagaaatattccgtagggcagcagcggccgagcctctcgtgctatattgctgagactaaacgagagaggagaaacaac
 tcatagaggtaaagcatggcaaaggcaccatctgcacgctatcctgagaaatccagcgtacatcggtcataggagacatctgg
 cgtggatacaggtaaaggaatgtgggaccaaatagtgatgaagacttcgctgaaacattccaagctgtccaggaaatactc
 tttaccaggaaggcaactatcggccgaccagaagcccaacacctaagactggcattgctttatcgggtgagcatccagac
 gagccaccactacgttcggtcactgtaagaggtagaactaattaactgttctacacgatgatgttgcgatgagggaggatcg
 tatggatgctttgtggaagaatcagttacttggcttgcgagtgatgaagcgggtggctgcttccaagataatacagatgatgaga
 gaacgagaaaagccagaattcgaactaaagtctagaagagcaattagaagccgctcaaaaacaggctagaacactccgtcct
 gatgggatgggatgctcttctatcgactcttctgctgctcagggccgaacttacgccacagatagacaaggcaaggcaag
 agagcagatccctgatgtacccgcacttttgcgtgacctttaggcaagccagcagacggtgatcgtgatggaacgaagc
 ttaaccctcctcagagacgtatgatacttaggtggtttagcattagactattcaaggcgggttctcgtggtgtcagggcaata
 gagcctggtcgaattacgctcagctacgttggagaacctggatttaagccagtggtggtaacagagcaaaaacagtga

Integrase 4 *A. thaliana* codon optimized sequence

atgattacgaccagaaaggtggctatatacgttcgtgacgactactaatcaggctgaggaaggggtatagtattcaaggtcaaat
 cgattcactaataaagctcgaagctatgggctggatcatctataggaataactgatgccgggtcagtgagggaagattg
 ataggcccgtatgagtaaatcaacggatgcaaagcacaagagatttgatacgttctgtctataaactgacagattaagta
 gatccgtcagggcacggtatacctgggtgaaggatgtcttcaatcagaacaatatacacttcgtgagcctacaggaaaacatcgat
 acttctcagcgtatgggtaattctattctgacctcttatcagcgtatagctgatttgagcagagaacaaattcaggagcgaatgacg
 atgggcaagattgggagcgaatctggtaagacaatggctggacatatactccttcggatacactacaacaaggagaag
 ggagagctcatcctgaccagccaaagctcctatcgtgaagatgatttacaccgactattaaaagggtatgccatacagaaaat
 tgtgacaagctgaataagatggattacaatggtaagattgcacttgggtcccgcgtgagcctcctagacaatccag
 tctactatggaatgacaaggtataacaacaattatccttggaaatcatcagcctattatccaaggaactattcgacaagacc

cagagagagagacagaggcgacgattgggcattgaagagaatcattacacaattccattccaggctaagtacatgctgtctaaa
 ttcttgcgatgcaggcagtgcggtcaaggatggggctcgaacttgggagacctcgaagaaggaaggtaaaaggtcaaagaa
 atactattgtctgaattctagacctaaaggagcggtctctcgcgacactcccctatacgtgctgagacactggaagattacgtgct
 gcatgaaattgccccaaatccaaaaggatccgtcaatcgcgagtcgacaaaaacatatcgaggatcacgagtaaagtacaag
 cgtgaacgtattgaagccaatataaataaaaaccgtaaatcaactgtccaagttgaacaatttga cctaaacgacttgatcaccttg
 gaggacctcaagactcagacgaacacacttatagcgaagaacgtctctcgaaaacgaacttgataagacatgacgacaacg
 atgacgaactgataggcaggaaactattgcccacttctggtctacctgatgtgtggacgatggactatgagggccagaagat
 gctgtgagttgtagtccagagagttaaagtgacaggggataacatagatatacattggacctttaa

Integrase 5 *A. thaliana* codon optimized sequence

atgccaggcatgacgacagagaccggacccgacccggcagggtaatagacctgttctgtagaaaaagcaaagctgtcaa
 ctagggtaatggggcggtcaaggagaaaacaggaaattccatagcagcgcaggagacgctcgggaggaaggtagcc
 gcgtgttaggaatgcaagtcaggcatgtatggaaggaggtaggatctgccagcagattcagaaagggaaagggcaggggatg
 accagtccaaagctctaaaggcccttgagtctggagaggtggcgcactgtggtgctatcgtctagatagatgggatagagggg
 gtgcagggtctatccttaagataatcgaaccggaggacggtatgcctcgtcattgtgtttggctgggacgaagatacagggcagg
 ccggtgctggacagtactaacaacgagaccgtggtgaaactaacgacgagcagaagaggctcgtgaggaggcggaaaa
 gcttagtgagagggtaagagataactaaagcgcataaaagagaaaaacggggaatgggtaaagtctaggcgcctacggact
 caggggtgacttgcacagatccgatgaagaaggcgtatgactcagatgaacgaaagcttcagcagacgatgaagacgctg
 gaggcccgatggtctaacgaaagcggaggccgcacgactcgtgtcacactcccgtcactgatcactatcctacgcgggaa
 cggctctatgctatgaatacccgagaaataccctcctcactggggggccatggattgcccgaacggctcagggatgatacaga
 accccgcctacgcgggctggcagacaaccgggacagggcgtaaacaacgaagacttacctttataatggtgaaggaaa
 gcgtgttagcgtgatgcacggaccaccacttgtacggatgaggaacaggaggcagcgaaggcagcggttaaggagaggat
 ggggtcggagttcccttagacgggtcagaccatgacacgagggcgaagcactactctctggcgaatgcatgctcgtgctgcg
 ggggtcctgttcatacagtggaatggctatagatgttgagatcttcagtcgaaaggcgttccccgccccacgatgctcgcg
 aggaagtccgtagaagagatgtgtcttttaggtgggctgtaagctcgtcgcctcagagccggatgatcctttgtatagcgggtgc
 agataggtggcggttaaccatctcagggcagcgaagatgagaaatagcgaagctgcccgttcgagaggcagagaa
 aatctaggccgttctcgtgatgcgcaaaatggagtctacgacgggcccggcggagcaattttcgacccgcatacaagag
 gcattaagcacacttcaggcagccaaggatgccgttagtgaagctcagcagtgctgctgtagacgtgagctggatcgtagac
 agtagcgattacgaggagctatggtaagagctacgccgacaatgaggaatgccataatagacacgtgtatagatgagattgg
 gttgcaagggcagaggggaaggccgttcgatggggacgagcagagtaaaagattaaatgggcagcgcgtacataa

Integrase 7 *A. thaliana* codon optimized sequence

atgaaggtagcgtatgtcctgtgtcaactgatgagcaggccaaggaaggatttccatcccagcgcagagagaaagatta
 agggcattttgtcctcacaaggctgggagatcgtacaggagtatatcgaagaggggtggtctgcaaggatttgacagacctc
 agatgcagagactgcttaaggatataaaaaaggtaaacatcgacattgtcctctatacaggtggataggctaactaggctcgtt
 ctggatctttatctgctcctgcaaacatttgagaagtacaacgtggctttctgtagtctacggaggatatacagatacgtcaactgcaat
 gggccgtttattcattaccttagtgccgctcggcacaatgggagcgtgaaaaacctggcagagagggtaaaattggaattgaa
 agatgattgacgagggaagaagcccggagggtcacagctccctacgggtataaattgataaggactcaactgactataatcga
 ggaggaagctgatgctgctgatgattacaggatgtactcgtatggatggctaccgtatgctcggaccgacttaattgatt
 gatggtaaacccaggatagccaaggaatggaatcataattcgttcgagatattcttaaaaacgatatttataggtacatacag
 gtggggtgataaagtagtccccaaataatcatccgcaataatctccgaaactttgtcaaaaaagcacagaaggaaaaagaaa
 agcgtggcgtgaccgaaagcagtcggaagtttcttttacaggactgttgcaatgcgggaattgccccggacacaagatgca
 agggcacttcgataagcgtgagcaaaagacttattaccgttgactaaatgccaccgtatcacaacgagaagaacatactaga
 gccacttctgacgagatacagttgtaattacgagtaaaagactttatgagtaatttagcgcagatacgtcaacaggaggt
 ttagacgtgagcgcgtgactaaagaactggagaagataaagaggcagaaggagaagtggtacgatttgatataatgga cgtc
 gaaatccgatccccagggaagagctatttgcgaataaacgaattgaacaagaaagaaggagatataatagtaagctaag
 tgaagtgaagaagataaagaaccggtagaagagaagtacaaccgtctctccaaaatgattgattttaaacaacagttcgaac
 aagcaaatgatttcacaaaaaggaaactttgtctctatcttcgaaaagattgtaatctatagagagaaagggaaattaagaag
 atcacactagattatactttgaagtaa

Integrase 9 *A. thaliana* codon optimized sequence

atgaaggtcgtatatacaccagagtaagcactctggagcaaaaaggagaaaggtcattctatagaagaacaggagaggaag
 ctgagagcactaccgatataaacgattggaaaattcataaagtgtatacagacgcccgtattccggggcgaagaaagaccg
 accagcactacaagagatgttgaacgaaatagataacttcgattgggtcgtctgttataagctcgaccgactcactcgtctgcaaa
 gacctactgaaattctcagagctatttgagaataaaaatgactcttctgatctcgcaccgaagtctatgataccacatcagccatgg

gacgacttttcgtgacgttggtggcgcgatggcagagtgggaaagaacgactattcaagaaaggaccgcaatgggaagacgt
 gccagtgccagaaagggcttgcgaaaacggtaacctcttctattatgacagggtaatgacaagttcgtgcctaagtagtaca
 gaaagtgtgcttccgctggaagaagcgaagaaagggactagccttagagaaataacaatcaaacgaataacagcaag
 tacaaggcgccactcggaaaaaattggcacaggtccgtaacggaacgcactcacctcacctgttctagagggccatctagat
 ttggggacatttctgaggagaacactcacgagggcagataatcagaagaggagtagcagggaaattaagcttagaatctctgaga
 agacgaaactcaacaatcgtaaacataatgcatctcaggagtaaaactactatgtccaaactgcaatcaaaaagcttacctgaat
 accgftaaacatacacccaagaataaggaagtggtgattcaagctatatttctgcagtaactgtaagaatacgaaaaaataaga
 atgcttgaacattgacgaaggtgaggttttaagcaattctataattatctaaagcaattgatctgacttctataaaatcgagaacc
 agccaaaggaaatcgaagatgtgggtatcgacattgagaagttgaggaaaagcgcgagcgcagatgccagacgctctcattga
 aggtatgatggacaaagatgaggcttttctattatcacgaatcgacaaggagatacacgagatgagaaaaaggaaagataa
 cgacaaagggaaaaacattcaattatgaaaaaatcaagaacttaagtacagcttctcaatgggtgggaactcatggaggatga
 acttaagaccgagttcataaagatggcaatcaaaaatacattttgaatatgtgaaagggataaaggggaagaggcaaaattct
 tgaagatcacgggcatagaattttattaa

Integrase 13 *A. thaliana* codon optimized sequence

atggctgtcgaatttacatacgagtgagtactcaggagcaggccagtgaggggcattccatagaaagccaaaagaaaaaatt
 ggcgttactcgcgagattcaggggtgggatgactataggtttatattgaggagggaatatccgggaagaatacaaatcgtccaa
 agcttaaaccttattggagcatattgaaaagggtaaaattaataactactcgtgtaccgattagaccgtctaacgaggtcagta
 cgattgcacaagtattgaattttctgcaggaacacggttgcgcctcaagtccgcaactgaaactacgcacacaaccaccgcaa
 acggaaggatgagtagtgggattgtactcctgccagtgaggagacggaataatgagcgaacgtataaaactgaatctcg
 aacacaaagtactagtagaggggaaaggggtgggctatcccctacggattgatctaagtgatgacgaaaagctagtga
 aatgaaaaaagtcaattctgctggatattggtggagcagtagaaaaatggatggcctcaatcgaatcgtgaattacctgaatct
 tacgaacaatgaccgaaattggagtcggaatggcgtctcgtctactcgaacccggcgtctacggggcgaccctgtggaa
 cgataagatagctgagaacacacacgaaggaattatcagtaaagagaggtcaatagattgcagcaaatcctagcagatcgctc
 tattcatcacaggcgtgatgtaagggtacctacattttcagggcgtattaaagatgtccggtttgtgatcagacattgagcgtaacag
 attcatcaagaagcgaaggatggcacggagtactgtggcgtttgtatagatgccagccctgtattaaacaaaataatacaatc
 ttgcaattggcgcgagcacgatttctgaaggctctcaacgagtagatgtccacgggtggaatttcaaacgttgaggacgaggtcatcc
 ctaagaagagtgagcgcgagagatgctggaaagtcaactcaacagatagctagaaagagagaaaagtatcaaaaagcgtggg
 ctctgattgatgtctgacgcgagtttgaaaagcttattggttgagacccgagagacttacgatgaatgcaagcagaaaactcgagt
 ctgagaggaccctatcaaaatcgacgaaactacttaaggagatagttatattgtccatcaaacctcaacgacctggagtcg
 agaaacagaaggagttcatctcaagtttattcgtacaatacgttacaccgtaaggaacagcaaccaataaggccagacaag
 agcaagactgggaaaggcaagcagaaagtataactgaggtagaattctatcaataa

phiC31 *A. thaliana* codon optimized sequence

atggacacatacgcgggtgcctatgacagggcagtcctcgtgaacgtgaaaatagtagtgcggtctccccgccacgcagcgttcc
 gccaatgaggacaaagcagcagacctccaacgagaggttgagcgcgagatgggggcaggttcaggttcgtaggccacttctctga
 agcgcgccgactccgctgctggcaccgaggagacccgagttcgcgctataactaatgagtgccgagcgggtcgactca
 acatgataaattgatagcgttagtaggttctccagactgaaagtaatggatgcaataccaatagatcagagctctagcctggg
 cgttacgatagtgacacccaggagggcgtattccgacaagggaacggtatggatcttatacatctaatatgagactgatcca
 gtcacaaggaatcctcctcaagagcgcgaagattctgataactaagaacctacagagggagctcgggggttacgtcggggga
 aaagcccgtacggtttgaactgtcagtgaaaccaaggagattacgaggaacgggaggatgtgcaatgtgtcataaacaagt
 tggcgcactcaactactccattgactggtcgttcgagtttgaccgagcgaatcagatgggtggcgtgagatcaaaactcata
 agcatttaccctttaaaccgggagccaggcggcaatacaccgggaagcatcacaggcctgtgcaaacgtatggacgcccagc
 gctgtgccgaccagagggcgcgagactattgggaaaaagactgcctcctcagcatgggatcccgcctacagttatgagaatttactg
 accccagaatcgtcggcttcgcagcggaaagtattataaaaagaagccagatggcacaccacaacaaaattgaggggta
 caggatccagagagatcctatcacttaaggccagttgagttgattgtggaccaataattgaaccgcccagtggtatgagctc
 aggcctggttgatggaagaggtcgtggaaaaggcctcagccgtggacaggccatcctatctgctatggacaaattgtattgca
 atgctggcgcggtgatgacaagcaagcagggcaggaatctatcaaggacagctacaggtgtcgtagacgtaaaagttgtgacc
 ctacagcactggccagcatgagggtagctgcaatgtctcatggcggcgttagacaagttcgtggctgaacgtatattaacaaa
 attcgacatgacaggggcgacgaggagactctagcgttactatgggaagccgcgctgattgggaagttaacggaggcgc
 tgaaaaaagcggagagagagctaatctctgtagccgagcgtgctgatgactgaatgcgctagaggagctgtatgaggaccgtg
 ccgcccgggcatagatgggcccgttgaaggaagcatttccgtaaacagcaagcggctctgacactgcgtcagcaaggagc
 cgaagagcgcgattagcggagctgaagccgaggcgcctaaactccactgatcagtggttccctgaagacgcggacgccc
 gatccgactgggcccattctgtgggtgagcattccgtgatgacaagcgtgtatttgggtctatttgcgacaagatcgtag

tgactaaatcaactacaggccgtggccagggcacgccaatcgagaaacgtgagcagataacatgggctaaccacctacgg
acgacgatgaagacgatgcacaagatggaactgaggacgtgcagcatag

Bxb1 *A. thaliana* codon optimized sequence

atgagggccttagtctgatccgactatccagggcactgatgccaccacatcaccagaaagacaactagatcatgccagcaa
ttgtgcgcacaaagaggatgggatggttaggcgttgcggaggacctagatgtatcaggtgccgtggacccctttgatagaaaac
gaagacaaacctagcgcgttggctcgcctttgaagagcaaccgtttagtgatcgtggcatacaggggtgatagactaactag
gagtatacgacacctacagcagtttagtaccattgggctgaggatcataagaagctagtggtctctgcaactgaggccatttcgaca
cgaccacccctttgcccgggttgcattgcgttaatgggcaccgtagcacaatggaactcgaggcaatcaaggaaagaaaacc
gaagcgcagcacacttaacattcgagccgggaagtaccgaggctctctcccccttggggctatctcccgcagaggggtggatg
gtgaatggagattagtagcagacccccgtccagaggggaaaggattctagaagtgtaccaccgagtggtggataatcacgaaccg
ctccacctcgtggcgcagcacttaaacctgcgtgggggtgctctctccaaaggactatcttgcgcagcttcagggtcgtgaacctcag
ggcgagagtggtccgctactgcttgaacgtagtagtagtagcagggcaatgtaggatagctacgctaaacggcaaaaaccg
tccgagatgatgacggcgcgcccgtagtagcagctgaacctatactgacaagagaacagcttgaggcttacgtgcggagttagt
taagactagcagggcaaaaaccagcagtttagtaccgcttactcagggctctgtttgcgcagctctgtggtgagcccgcgta
caaattcgtggtggggggcgtaagcaccgcgttaccgttgcgtatgtagggctcccaaagcattgcgggaatggaaccgtg
gcaatggcggagtgggacgcttctgcgaggagcaagtattggatctcctcggggatgctgaacgactggagaaagtgtgggta
gctgggtctgattctgctgtagcagctagcgggaagtaaatgctgaactagtggaactaacgtccctcattgggtctccggcatatagag
ctggcagccctcaaagggaagccctgatgcgcgtatcggccctgcagcagagacaggaagaattggaggggctggagg
ctaggccatcaggctgggagtgccgagaaaacggggcagcgatttggtgattggtggagagagcaggatacggccgctaaaaa
tacgtggcttaggagtagaacgtgctgctgacgtttagtctgctggtggtctcacacgaactattgacttgggtgacttgaagaata
cgagcagcacttacgattgggaagtgtgtgaacgtctacatacgggcatgagttag

References

1. Yang, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nat. Methods* **11**, 1261–1266 (2014).

8 CONSIDERAÇÕES FINAIS

Pode-se destacar como um objetivo da biologia sintética a introdução de novas funções em um organismo ou a interação com atividades endógenas modificando alguns dos seus processos biológicos para execução de funções desejadas e inibição de indesejadas. Dentro desse foco, as serina integrases, empregadas como ferramentas de ativação ou desativação da expressão gênica a partir da inversão de partes genéticas, podem contribuir em diversos tipos de engenharias projetadas para diferentes organismos. No presente estudo, o interesse foi voltado para o avanço de possibilidades no âmbito da biologia sintética de plantas. Para esses organismos o trabalho representou um avanço ao adaptar um chassi de testes vegetais, os protoplastos de *A. thaliana*, para a caracterização da funcionalidade de oito integrases na inversão de partes genéticas. Além de seis dessas integrases terem sido pela primeira vez caracterizadas em eucariotos, a expansão do número dessas proteínas validadas em plantas é importante para permitir construções mais complexas, sobretudo com algumas delas tendo exibido desempenhos nas células vegetais até mais satisfatórios que as já bem estabelecidas phiC31 e Bxb1, com destaque para a Int 13.

Os resultados obtidos em protoplastos demonstraram que todas as integrases foram funcionais para a ativação do interruptor genético, havendo, no entanto, variações nas proporcionalidades de células fluorescendo pela presença da proteína repórter EGFP. A partir daí é possível que sejam traçadas novas questões buscando se entender os motivos que levaram a essas variações. São diversas as perguntas que podem ser levantadas, tais como: uma vez transcritos, podem os mRNAs de algumas integrases estarem sendo identificados por alguma sinalização para degradação? Há variação na produção e acúmulo das diferentes proteínas integrase? Todas têm a mesma capacidade de entrar no núcleo, onde irão agir sobre o plasmídeo repórter? Alguns dos sítios de reconhecimento *attB* e *attP* flanqueando a parte genética alvo de rotação, podem estar comprometendo a transcrição do repórter? A complexidade do núcleo eucariótico pode estar interferindo de forma diferencial na eficiência das diferentes integrases? As limitações nas técnicas de transformação poderiam também estar afetando de forma diferencial os resultados de ativação das integrases? As hipóteses para essas

questões suscitam uma série de ensaios que poderiam melhorar o conhecimento dessas integrases e, assim, da mesma forma, seu uso como ferramenta.

Entretanto, o avanço trazido com o presente trabalho, pela demonstração de funcionalidade das Ints 2, 4, 5, 7, 9 e 13, e ainda sem efeitos citotóxicos e com ortogonalidade verificada para 2, 4 e 5, já permite a progressão para uma linha mais aplicada da ferramenta. Os próximos passos implicariam na testagem de suas atividades em plantas com uma inserção estável das integrases e/ou do cassete formado pela parte genética alvo. Para esta etapa pode-se esperar inicialmente desafios de transformação estável, de métodos de *delivery* para as integrases, caso apenas os alvos sejam integrados no genoma da planta hospedeira, ou de método de ativação da expressão das integrases, como por promotores induzíveis, por exemplo. Esse último ponto relaciona-se com a intenção de montagem de um sistema mais controlável que aquele de expressão de uma proteína sob um promotor constitutivo, como costuma ser feito para a expressão de transgenes. Avançando em desenho e proposição, tais mecanismos poderiam compor um circuito genético sintético, sendo formado, por exemplo, por genes que confirmam resistências a estresse abiótico, como seca ou salinidade, ou resistência a um patógeno. Para tais genes de interesse não serem expressos constantemente os mesmos poderiam estar posicionados de forma invertida, serem flanqueados por sítios de integrases, e ainda estarem sob o comando de um promotor forte e constitutivo, mas, portanto, não haveria formação da proteína de interesse. Apenas as sequências codificadoras das integrases poderiam ficar sob regulação de determinado promotor induzível, que seria ativado pela presença de fator abiótico, praga ou mesmo por um indutor químico ou físico aplicado. Desta forma, apenas os grupos de genes flanqueados pelo sítio de determinada integrase ativada naquela condição seriam invertidos e conseqüentemente ativados. Além disso, se o sítio de uma segunda integrase for posicionado *in tandem* ao da primeira, flanqueando o gene de interesse, esta segunda poderia ser utilizada para desligar o sistema em um momento posterior e sob outra sinalização. Para integrases como phiC31 ou Bxb1, a proteína RDF correlata junto da integrase poderia também ser aplicada para promover essa movimentação contrária, sem a necessidade de posicionamento de um segundo par de sítio. Para as Ints aqui caracterizadas, as respectivas RDFs, no entanto, até onde verificou-se pela literatura, ainda não foram identificadas. Este é

um exemplo de um modelo mais complexo e refinado de controle para construções de redes genéticas ou circuitos genéticos sintéticos que pode ser feito por integrases em plantas. Autores já chamaram essa possibilidade de construção de *smart plants*. Esse desenho de circuitos genéticos empregando integrases pode ser feito também voltado para aplicações de biossensores ou de construção de rotas metabólicas sintéticas para um melhor controle da expressão de partes genéticas envolvidas, podendo inclusive o desenho atender ao modelo de portas lógicas booleanas.

Além disso, esse controle de expressão gênica mediado por integrases pode ser feito utilizando as outras ações que as integrases são capazes de fazer, dependendo do posicionamento do sítio, que são os mecanismos de inserção ou deleção. Um dos primeiros usos destinados a essas enzimas em plantas foi voltado para a remoção de transgenes ou marcas de seleção, ficando apenas a pequena sequência do sítio como “cicatriz” no genoma, aplicação essa que pode ainda ser utilizada. O mecanismo de inserção depende da presença de um sítio no local alvo e abre a possibilidade de diferentes sequências serem inseridas naquele organismo que possuir o sítio de “aterrissagem” no material genético.

As integrases podem também serem utilizadas como ferramentas em outras estratégias de biologia sintética, como nas avaliações de função e essencialidade de genes, grupos gênicos ou sequências gênicas relacionadas a determinados domínios proteicos. Sítios de integrases flanqueando tais partes genéticas e invertidas de acordo com o projetado pode trazer diversas elucidaciones, como por exemplo, aquelas voltadas para o desenvolvimento de um genoma mínimo vegetal, (estratégia *top-down* da biologia sintética), compreensão de redundâncias gênicas, muito presente no genoma de plantas, ou mesmo proporcionar um melhor entendimento a respeito dos domínios de genes NLR, que estão ligados a conferência de resistência vegetal a efetores de patógenos, trazendo assim, possíveis avanços no sistema de imunidade de plantas.

Analisando a ferramenta integrase em relação ao sistema CRISPR/Cas9, apesar das diferenças nos mecanismos de funcionamento, ambas ferramentas biotecnológicas podem desativar um gene ou promover um ganho de função. Uma das maiores preocupações do sistema CRISPR/Cas9, os *off-targets*, entretanto, apesar de potencialmente também poder acontecer com as integrases, devido a sua

especificidade fortemente atrelada à sequência do sítio, parece menos provável. No entanto, há alguns relatos de estudos que buscaram explorar um possível reconhecimento de pseudo-sítios para o aproveitamento de sequências endógenas como potenciais sítios, evidenciando outra dificuldade do uso das integrases, que é a necessidade de presença dessa sequência específica no material genético. Então, se por um lado o reconhecimento de pseudo-sítio pode ser vantajoso, por outro traz a mesma preocupação de possíveis *off-targets*, levantando também a necessidade desse tipo de investigação para trabalhos com integrases. A possibilidade de alternância de estado controlado que pode ser oferecida pelo sistema integrase com um acréscimo de proteína RDF pode ser também considerada uma vantagem para essa ferramenta em relação ao sistema CRISPR.

Hoje, alguns possíveis caminhos podem ser considerados para a resolução do gargalo da necessidade do sítio específico no genoma alvo. Com o desenvolvimento da ferramenta *prime editing*, que consiste em uma Cas9 desativada fusionada a uma transcriptase reversa e um RNA guia fusionado a uma sequência de RNA complementar àquela que se pretende inserir no genoma, os sítios das integrases poderiam ser inseridos em locais específicos direcionados pelo guia. Entretanto, no estudo que desenvolveu a estratégia, o tamanho máximo de inserção foi mostrado para um sítio de recombinase de 44 bp, menor, portanto, que os das Ints aqui avaliadas, além da eficiência do ensaio não ter sido alta. Uma outra possibilidade seria a melhor elucidação das especificidades da proteína com a sequência de DNA onde a mesma se liga no sítio para que integrases possam ser projetadas para *loci* específicos no genoma e, possivelmente, reconhecê-los como sítios. Porém, essa relação de especificidade proteína-sítio ainda não foi elucidada. Ainda nessa intenção, há um trabalho que buscou fusionar uma integrase com uma Cas9 para que, ao invés de um sítio ser reconhecido, a proteína híbrida faça a função de integrase, mas com a especificidade de ligação ao DNA feita pelo RNA guia, substituindo, desta forma, a necessidade da presença do sítio. Contudo, a eficiência de funcionamento desse sistema foi muito baixa.

Muitos ainda são os desafios para as aplicações desse sistema em organismos vegetais, mas igualmente diversas são as possibilidades de aplicação em diferentes projetos de biologia sintética para plantas. A partir das funcionalidades demonstradas em células vegetais de diferentes integrases, esse trabalho permite

que outros pesquisadores possam avançar no planejamento e aplicações de seus projetos conhecendo melhor a ferramenta que pretendem utilizar, minimizando, assim, possíveis erros ou efeitos inesperados advindos de um desconhecimento de peças do sistema.