



Universidade de Brasília

Instituto de Biologia

Programa de Pós-Graduação em Biologia Molecular

Desenvolvimento de circuitos genéticos mediados por serina-integrases: memória permanente e modulação de genomas procariontes e eucariontes

Marco Antônio de Oliveira

Orientação: Dr. Elibio Rech

**Brasília - DF
Março, 2024**

Desenvolvimento de circuitos genéticos mediados por serina-integrases: memória permanente e modulação de genomas procariontes e eucariontes.

Marco Antônio de Oliveira

Exame de defesa de Tese de Doutorado apresentada ao Programa de Pós-Graduação em Biologia Molecular do Instituto de Ciências Biológicas da Universidade de Brasília como requisito para obtenção do título de Doutor em Biologia Molecular.

Orientação: Dr. Elíbio Rech

Brasília, março de 2024

BANCA EXAMINADORA

Presidente:

Dr. Elibio Leopoldo Rech Filho

Embrapa Recursos Genéticos e Biotecnologia - CENARGEN

Membro 1:

Dr. Marcelo de Macedo Brígido

Universidade de Brasília - UnB

Membro 2:

Dr^a Danielle Biscaro Pedrolli

Universidade Estadual Paulista - UNESP

Membro 3:

Dr^a. Grácia Maria Soares Rosinha

Embrapa Recursos Genéticos e Biotecnologia - CENARGEN

Suplente:

Dr^a. Fabiana Brandão Alves Silva

Universidade de Brasília - UnB

Local: Plataforma virtual

Data: 05/04/2024

Horário: 08h30

AGRADECIMENTOS

Ao dr Elibio, meu orientador, sempre nos estimulando e desafiando na busca de expandir as fronteiras do conhecimento. Obrigado pela oportunidade ímpar de aprendizado e crescimento. O olhar sobre o Fazer Ciência pode ser tão importante quanto a ciência feita.

Aos membros do Prédio de Biologia Sintética da Embrapa Cenargen pelo auxílio nas atividades diárias, com os importantes esforços das doutoras Grácia e Daniela na contribuição para o bom andamento das atividades de pesquisa do grupo.

To dr. John Glass, dr. Yo Suzuki and members of the Department of Synthetic Biology - J. Craig Venter Institute, for the incredible experience, guidance, and opportunity to work with the fascinating Minimal Cell.

Aos membros do clássico Laboratório de Biologia Sintética, a velha guarda do agora PBS, que muito bem me acolheram e enriqueceram os dias com ciência e cafézinhos quando havia. A sempre atenta vigília do disco voador Rayne, Mariana, Thaís, Mayna e Lilian, companheira de PCRs da madrugada. E aos membros honorários Raquel, Raquel e Maria.

Por fim - e definitivamente não menos importante - à minha família por todo o apoio ao longo dessa longa jornada.

À minha mãe, Marisia, que ensinou pelo exemplo, até sem se dar conta, a resolução prática de problemas no melhor estilo “engenharia criativa” que tanto exerço dentro e fora da ciência.

Ao meu pai, Mauro, possivelmente mais ansioso que eu para não precisar perguntar mais “ô, filho, e o doutorado?”, por sempre valoriza e estimular o crescimento através do aprendizado.

Aos meus irmãos Dênia, João Paulo e Fernanda pela rede de apoio mútuo sempre disponível.

A André, meu esposo, por todo carinho, cuidado e perturbação. Saiba que todas as ofertas de ajuda são sempre recebidas com muito amor, a despeito de empecilhos técnicos que porventura existam.

Agradeço ao CNPq, FAPDF e Instituto Nacional de Ciência e Tecnologia - Biologia Sintética pelo financiamento das atividades desenvolvidas.

RESUMO

A capacidade de regular a expressão de genes em resposta a sinais externos é um dos mecanismos centrais envolvidos na manutenção da vida na natureza, sendo também um dos principais objetivos de cientistas no esforço para controlar e reprogramar organismos. Para tal, ferramentas que permitam a manipulação gênica são cruciais na biologia sintética, especialmente na criação de circuitos genéticos complexos e redes regulatórias sintéticas. Serina-integrases (Ints) são importantes candidatos devido a suas propriedades, como a capacidade de executar suas funções sem necessidade de auxiliares endógenos. Além disso, a depender do desenho de seus sítios de reconhecimento (sítios *att*), a edição pode resultar em diferentes tipos de rearranjos, como inserção, excisão ou inversão. Considerando as vantagens advindas de sua plasticidade de uso e robustez, neste trabalho apresentamos a aplicação de Ints na construção de circuitos genéticos integrados aos genomas de dois modelos de graus de complexidade distintos: *Mycoplasma mycoides JCVI-syn3B*, uma bactéria de genoma procariótico minimizado sinteticamente, e *Nicotiana benthamiana*, um importante modelo vegetal com genoma eucariótico complexo. Em *M. mycoides JCVI-syn3B*, Int9 e Int13 foram inicialmente avaliadas como efetores de interruptores genéticos capazes de inverter a sequência codificadora de um gene repórter. Apenas a ativação por Int9 resultou no aumento de expressão do repórter, sendo então selecionada como ativador para um segundo interruptor responsável por controlar a expressão da endonuclease I-CeuI, usada na digestão e consequente deleção do genoma celular para produção de *Simple Cells* (SimCells), uma importante plataforma para aplicações em biologia sintética. O uso do interruptor controlado por Int9 permitiu um fino controle de expressão da nuclease. Por fim, para elaboração de um sistema *kill-switch* que permita a seleção negativa de escapes do processo de remoção do genoma, promotores induzíveis foram testados, com apenas o promotor pA13/AraE-AraR apresentando o comportamento esperado, porém com níveis de expressão muito baixos. Já para *N. benthamiana* um sistema mais complexo foi desenhado. Nomeado Int-Plex@ (*Integrase- Plant Expression*), o sistema consiste no gene repórter *mgfp* flanqueado por sítios *att* das integrases Bxb1, phiC31, Int9 e Int13, podendo ser dividido em dois módulos: inversão e excisão. O primeiro é composto pelos sítios *att* de Int9 e Int13, com sua recombinação resultando em uma inversão do DNA alvo e expressão de mGFP. A ativação com as duas Ints resultará no retorno do gene repórter à sua orientação inicial. Já o módulo de excisão é ativado por Bxb1 ou phiC31. Sua ativação leva à remoção irreversível da sequência de DNA flanqueada por seus sítios *att*. Todas as Ints testadas foram capazes de editar o genoma de *N. benthamiana*. O sistema Int-Plex@ poderá futuramente ser aplicado na modulação de vias metabólicas vegetais de interesse econômico ou ambiental, endereçando questões de biocontenção de transgenes, dada a possibilidade de excisão da construção. Somando aos esforços de avanço em biologia sintética, o trabalho incluiu a implementação de metodologias de produção e uso de sistema de expressão *in vitro* (TxTI) e encapsulamento para geração de sistemas sintéticos, incluindo a comprovação de funcionamento do sistema Int-Plex@ no sistema TxTI.

ABSTRACT

The ability to regulate gene expression in response to external cues is one of the central mechanisms of the differentiation and maintenance of life in nature, as well as one of the main goals of scientists in efforts to control and reprogram organisms. The availability of molecular tools that enable genetic manipulation is crucial for such advances in synthetic biology, especially when creating intricate genetic circuits and activation cascades to work as synthetic regulatory networks. Serine-integrases (Int) are strong candidates for these applications due to their properties, especially their orthogonality and ability to work without any additional cofactors. Moreover, depending on the placement of their attachment (*att*) sites, the recombination can cause different outcomes, including insertion, excision, or inversion of a target sequence. Taking advantage of their plasticity and robustness, here we propose the application of serine-integrases in the assembly of genetic circuits integrated into the genome of two models of opposing complexity: *Mycoplasma mycoides* JCVI-syn3B, a synthetic bacterium hosting a minimized prokaryotic genome, and *Nicotiana benthamiana*, a model plant with its complex eukaryotic genome. In *M. mycoides* JCVI-syn3B, Int9 or Int13 were first tested as effectors in genetic switches capable of inverting a reporter gene sequence, but only Int9 activation resulted in a reporter gene expression increase. Int9 was then selected as an activator for a switch controlling the expression of I-CeuI - an endonuclease used for digestion and removal of the cell's genome for the production of Simple Cells (*SimCells*), a valuable platform for synthetic biology applications. Our Int9-based switch was able to tightly control I-CeuI expression. Finally, envisioning a kill-switch for selection of the *SimCells* generated, new inducible promoters were tested, with only the pA13/AraE+AraR promoter presenting a proper response so far, although yielding considerably low expression levels. As for *N. benthamiana*, a more complex switch was assembled. Named Int-Plex@ (Integrase-Plant Expression), this genetic switch system consists of a reporter *mgfp* gene sequence flanked by a combination of *att* sites of Ints Bxb1, PhiC31, Int13, and Int9 and can be divided into two modules: the inversion module and the excision module. The inversion module is composed of Int9 and Int13 *att* sites and their recombination results in a 180-degree flip of the target DNA, leading to activation of mGFP expression. Concurrent or sequential activation of both Ints will bring the reporter gene to the initial silenced state. The excision module is activated by transient expression of Bxb1 or phiC31. In this case, the DNA sequence flanked by their attachment sites is irreversibly excised from the genome. Integrases were delivered via agroinfiltration, and all four Ints were able to edit the *N. benthamiana* genome successfully. This memory switch system can be used in future genetic circuits to engineer and modulate plant metabolic pathways of economic and environmental importance, while also addressing transgene biocontainment issues given the possibility of cassette excision. Adding to the efforts for advances in synthetic biology, this work includes the implementation of protocols for the production and use of cell-free expression (TxTl) and encapsulation to produce synthetic systems. Int-Plex@ was shown to also work in TxTl reactions.

SUMÁRIO

Introdução.....	1
Serina-integrasas como ferramentas de biologia sintética.....	1
A célula mínima <i>Mycoplasma mycoides</i> JCVI-Syn3.....	11

Capítulo I: Construção de interruptores genéticos ativados por serina-integrasas na célula sintética minimizada *Mycoplasma mycoides* JCVI-syn3B: regulação transcricional controlada por Int9 e sua aplicação no controle de expressão da endonuclease I-Ceul para produção de SimCells.....

15

Introdução.....	15
Objetivo Geral.....	17
Objetivos específicos.....	17
Metodologia.....	17
Soluções e Meios de Cultura.....	17
Linhagens celulares, cultivo e manipulação.....	19
Design e construção de plasmídeos.....	21
Transformação de células Syn3b.....	24
Avaliação funcional de interruptores genéticos.....	25
Ensaio de Unidades de Mudança de Cor de Culturas.....	25
Curva de Crescimento e Acúmulo de Conteúdo Genômico.....	26
Caracterização funcional de promotores induzíveis.....	26
Resultados e Discussão.....	26
Conclusões e Perspectivas.....	41

Capítulo II: Desenvolvimento da plataforma Int-Plex@: interruptor de memória genética bimodular controlado pelas serina-integrasas Int9, Int13, phiC31 e Bxb1 capaz de regulação transcricional por meio de inversão ou excisão de sequências integradas ao genoma de *Nicotiana benthamiana*.....

43

Introdução.....	43
Objetivo Geral.....	44
Objetivos específicos.....	44
Metodologia.....	45
Design e construção de plasmídeos.....	45
Geração de Plantas Transgênicas.....	48
Condições de cultivo e manuseio em casa de vegetação.....	49
Agroinfiltração.....	50
Extração de DNA e PCR.....	50
Ensaio de expressão in vitro.....	52

Microscopia Confocal.....	53
Resultados e Discussão.....	53
Conclusões e Perspectivas.....	64
Capítulo III: Implementação de metodologias chave da Biologia Sintética: Sistema de expressão cell-free TxTI (Transcription-Translation), Produção de Lipossomos e Encapsulamento como ponto de partida na construção de Células Sintéticas.....	66
Introdução.....	66
Objetivo.....	67
Objetivos Específicos.....	67
Metodologia.....	67
Protocolo de produção de extrato celular para sistema TxTI.....	67
Protocolo de produção e purificação de RNA polimerase T7 para uso em sistema TxTI.....	69
Protocolo de preparo de mix de aminoácidos para reações TxTI.....	72
Protocolo de preparo de Mix Energético.....	73
Preparo de reações de expressão in vitro com sistema TxTI.....	74
Produção de Lipossomos por Emulsão Inversa.....	75
Vetores de expressão em sistema TxTI.....	77
Western Blotting.....	78
Microscopia Óptica.....	78
Resultados e Discussão.....	78
Conclusões e Perspectivas.....	83
Referências Bibliográficas.....	85
Anexo A. Manuscrito. Protocol for the establishment of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells.....	98
Anexo B. Manuscrito. Development of Int-Plex@ binary memory switch system: plant genome modulation driven by large serine-integrases.....	116

LISTA DE FIGURAS

- FIGURA 1.** A e B) Diagramas de formação dos homo-tetrâmeros de serina-integrases após ligação ao DNA, clivagem e recombinação dos sítios att. C) Estruturas cristalográficas dos tetrâmeros indicando alterações conformacionais após giro de 90°. A- adaptação de Merrick et al.11 ; B e C - Adaptação de Yuan et al.10..... **2**
- FIGURA 2.** Esquematização de possíveis resultados de recombinação por serina-integrases. A) Integração, Excisão, Inversão e Troca de Cassete Mediada por Recombinase (RMCE - Recombinase Mediated Cassette Exchange). B) Inversão demonstrando em detalhes o funcionamento de um interruptor baseado em integrases. A inversão da sequência codificadora resulta em alinhamento de sua orientação com relação ao seu promotor Pconst, ativando a expressão do GFP. B- adaptação de Yang et al.29..... **4**
- FIGURA 3.** Mapeamento de ocorrências de inserção de Tn5 no genoma de Syn1.0 por meio de ensaios de bombardeamento de transposon para identificação de genes essenciais. Setas pretas representam inserções detectadas na primeira passagem. Setas rosa representam inserções detectadas após a quarta passagem. Adaptado de Hutchinson et al. (2016)76..... **13**
- FIGURA 4.** Representação dos genomas da célula sintética *M. mycoides* JCVI-syn1.0 e da célula minimizada *M. mycoides* JCVI-syn3.0 Regiões em vermelho representam genes mantidos durante o processo de minimização. Adaptado de Hutchinson et al. (2016)76..... **14**
- FIGURA 5.** Mapas esquemáticos das construções realizadas..... **22**
- FIGURA 6.** Modelo de construção de plasmídeos pelo método de montagem in vivo de fragmentos com extremidades homólogas..... **23**
- FIGURA 7.** Mapa das construções utilizadas no estudo. As setas vermelhas representam alinhamento de sequências obtidas para confirmação das construções. Mutações corrigidas pela sobreposição de sequências corretas foram consideradas erros de sequenciamento..... **28**
- FIGURA 8.** Mapeamento de integração dos interruptores genéticos controlados por Int9 ou Int13 ao genoma de Syn3B. A) Sequenciamento genômico da linhagem Syn3B_Int9iBP. B) Sequenciamento genômico da linhagem Syn3B_Int13iBP..... **29**
- FIGURA 9.** Ensaios de ativação do interruptor controlado por integrase por meio da indução do efector por tetraciclina..... **31**
- FIGURA 10.** Sequenciamento e detecção da formação de sítios attL e attR após indução de expressão da respectiva integrase efectora. A) attR Int9; B) attL Int9; C) attR Int13 e D) attL Int13..... **33**
- FIGURA 11.** Ensaios de indução sequencial de culturas previamente expostas à tetraciclina..... **34**
- FIGURA 12.** Curvas de crescimento com marcação de DNA por DAPI de linhagens contendo a endonuclease I-CeuI com expressão controlada por Int9..... **36**
- FIGURA 13.** A) Ensaios CCU de linhagens carregando a endonuclease I-CeuI com expressão controlada por Int9 após indução por tetraciclina. B) Plaqueamento dos mesmos tratamentos para acompanhamento de formação de colônias..... **38**
- FIGURA 14.** Quantificação de expressão do repórter mCherry sobre regulação de diferentes promotores induzíveis na presença de variadas concentrações dos respectivos indutores..... **40**
- FIGURA 15.** A) Mapa da construção repórter pART27_mgfp(rc) inserida no genoma de *N. benthamiana*. B) Portas lógicas constituídas pelos módulos de inversão e excisão do sistema

Int-Plex@.....	46
FIGURA 16. Esquema de edições esperadas do cassete integrado pART27_mgfp(rc) após ativação de integrases. A) Módulo de inversão editado por Int13. B) Módulo de excisão editado por phiC31. C) Módulo de inversão após edição sequencial por Int9 e Int13. A dupla inversão retorna o gene mgfp a seu estado inicial silenciado.....	47
FIGURA 17. Teste em sistema de expressão in vitro TxTl do módulo de inversão de Int-Plex@. A) Detecção da expressão de deGFP em transluminador de luz Azul. B) Confirmação por sequenciamento de ocorrência de dupla edição do material submetido à sistema TxTl na presença de vetores efetores de Int9 e Int13. C) Amplificações para detecção de todos os eventos de rearranjo observados no sistema TxTl. Marcador de peso molecular: 1kb Plus (invitrogen).....	55
FIGURA 18. Localização dos sítios de integração do sistema Int-Plex@ identificados no genoma de <i>N. benthamiana</i> para os quatro eventos independentes avaliados. Os círculos concêntricos coloridos representam o genoma de cada um dos eventos alinhado ao genoma selvagem Niben2.6.1 utilizado como referência. Picos representam a integração. Do mais externo para o mais interno: eventos E1 > E3 > E10 > E15.....	57
FIGURA 19. Resultados representativos obtidos para um dos eventos avaliados após infiltração com Int9. A) Sequenciamento de fragmentos contendo sítio attL Int9 ou sítio attR Int9 amplificados após flipagem de mgfp. B) Microscopia de fluorescência a laser confocal realizada em região infiltrada mostrando ativação da expressão de mGFP.....	59
FIGURA 20. Resultados representativos obtidos para um dos eventos avaliados após infiltração com Int13. A) Sequenciamento de fragmentos contendo sítio attL Int13 ou sítio attR Int13 amplificados após flipagem de mgfp. B) Microscopia de fluorescência a laser confocal realizada em região infiltrada mostrando ativação da expressão de mGFP.....	59
FIGURA 21. Identificação de sintomas foliares indicativos de disseminação sistêmica do vetor viral PVX_GW contendo integrase efetora.....	60
FIGURA 22. Ativação do módulo de excisão do sistema Int-Plex@. A) Fragmentos esperados em amplificação após excisão do gene mgfp em decorrência de infiltração com integrases phiC31 ou Bxb1. B) Amplificações obtidas após tratamentos. Obtenção de fragmentos dos dois tamanhos indicam heterogeneidade da amostra de DNA, com redução do tempo e extensão da polimerase selecionando preferencialmente a detecção do fragmento editado. Amostras indicadas por triângulos foram tratadas com a versão da integrase contendo a sequência NLS. NT, evento positivo não transfectado com integrase.....	62
FIGURA 23. Alinhamento de sequências obtidas a partir das amplificações do DNA de amostras tratadas com integrases Bxb1.....	63
FIGURA 24. Detecção de molécula circular contendo gene mgfp excisado do genoma após tratamento com integrases phiC31.....	64
FIGURA 25. Mapa da unidade transcricional principal do plasmídeo 2041p para uso em sistema TxTl, contendo promotor T7Max, sequência RBS e terminador T500.....	78
FIGURA 26. Levantamento de custos de produção de sistema TxTl e comparação com sistemas comerciais. Preços levantados em novembro/2021.....	79
FIGURA 27. Western Blotting de expressão de proteínas diversas em sistema TxTl. A) Proteínas de soja, CVN, GRFT e GAGE. B) Proteína de teia de aranha Masp2 16x, com uso de T7pol purificada adicionada à reação.....	80
FIGURA 28. Produção de lipossomos. A) Sistema de extrusão manual com membrana de 0.1	

um. B) Microscopia de campo claro dos lipossomos gerados. C) Microscopia de campo claro após extrusão dos lipossomos.....	81
FIGURA 29. Microscopia de campo claro para medição de lipossomos produzidos por emulsão invertida.....	82
FIGURA 30. Western blotting para detecção de expressão de proteínas de soja em sistema TxTl e encapsulamento. A revelação foi realizada com uso de anticorpo anti-His conjugado a fosfatase alcalina e substrato cromogênico NBT/BCIP.....	83

LISTA DE TABELAS

TABELA 01. Aplicações de serina-integrases na edição de DNA em sistemas eucariontes.....	6
TABELA 2. Oligonucleotídeos utilizados na genotipagem de linhagens <i>M. mycoides</i> JCVI-syn... 20	20
TABELA 3. Quantificação de reads contendo sequências dos sítios att de Int9 obtidas por meio de sequenciamento nanopore. As sequências mCh pertencem ao gene repórter mcherry.....	33
TABELA 4. Oligonucleotídeos utilizados na detecção das edições realizadas no cassete repórter pART27_mgfp(rc) ou pART27_degfp(rc).....	51
TABELA 5. Dados descritivos do sequenciamento nanopore realizado para mapeamento de sítios de integração do sistema Int-Plex@.....	56
TABELA 6. Quantidades de cada aminoácido para preparo de solução estoque 20 mM do Mix de aminoácidos para reações TxTl.....	72
TABELA 7. Reagentes e concentrações para preparo de estoque 10X do Mix Energético para reações TxTl.....	74
TABELA 8. Composição e ordem de preparo de reação TxTl completa.....	74
TABELA 9. Proteínas de soja selecionadas para processo de expressão em sistema TxTl e encapsulamento em lipossomos.....	81

Introdução

A Biologia Sintética começa a se estabelecer como campo de pesquisa no início dos anos 2000 como consequência do desenvolvimento e amadurecimento da biologia molecular trazidos com o advento da tecnologia do DNA Recombinante¹⁻³. Campo de caráter altamente multidisciplinar, que abrange áreas bastante variadas dentre as quais a modelagem matemática, engenharia eletrônica, computação e biologia molecular, um dos pontos focais da Biologia Sintética é a reprogramação de sistemas biológicos e construção de organismos sintéticos para aplicação no avanço de áreas como agricultura, energética e medicina⁴. Suas linhas de pesquisa podem ser classificadas em dois grandes grupos de acordo com a abordagem aplicada, denominadas *Top-Down* e *Bottom-Up*. De modo geral, na abordagem *Top-Down* sistemas complexos vivos já existentes são utilizados como ponto de partida para aquisição de conhecimento a seu respeito, em uma espécie de engenharia reversa de componentes moleculares, e modificações são inseridas visando otimizar e agregar novas funções e aplicações a esses sistemas. Já a abordagem *Bottom-Up* implica na criação de componentes isolados necessários para a montagem de um novo organismo, tendo como ponto de chegada idealizado a criação de uma célula completamente sintética⁵. Apesar de poder ser entendida como um área aplicada com objetivos delineados, os esforços em biologia sintética possuem amplo impacto no desenvolvimento da ciência e compreensão da vida e seus mecanismos, uma vez que não só aglutina e utiliza conhecimentos de áreas diversas, mas ao longo do processo também gera valiosas informações e ferramentas.

Serina-integrases como ferramentas de biologia sintética

A capacidade de regular a expressão de genes em resposta a estímulos externos é um dos mecanismos centrais de diferenciação e manutenção da vida na natureza, e assim se apresenta também como um valioso recurso a ser dominado por cientistas na busca de reprogramar e controlar organismos. Assim, faz-se crucial o desenvolvimento de ferramentas moleculares diversas que possibilitem a manipulação gênica e metabólica das células, em especial, nesse contexto, aquelas que permitam a construção de circuitos genéticos complexos capazes de emular sinteticamente vias regulatórias. Dentro do grupo de ferramentas moleculares com tal aplicação podemos destacar as serina-integrases, enzimas capazes de promover rearranjos de DNA de maneira direcionada⁶⁻⁹. Originalmente presentes na natureza como um mecanismo utilizado por bacteriófagos para sua integração ao genoma bacteriano hospedeiro, as serina-integrases formam, juntamente com tirosina-integrases, as principais superfamílias das chamadas Recombinases Sítio-Específicas (*Site Specific Recombinases* - SSR)^{6,8,9}. Seu

mecanismo de ação primário se dá por meio da recombinação entre dois sítios alvo específicos a cada integrase denominados sítio *attP* (originalmente encontrado no genoma do fago- *Phage*) e sítio *attB* (presente no genoma Bacteriano). Ao reconhecer seus sítios cognatos, homodímeros das integrases são formados em cada um dos sítios *att*, culminando na formação de um tetrâmero ao ocorrer a aproximação dos dois complexos¹⁰. Após a ligação as integrases catalisam quebras de dupla fita na região central de seus sítios, conhecida por *core*, seguida de rearranjos conformacionais responsáveis pela junção cruzada das extremidades formadas em cada sítio, sendo a sua ligação realizada pela própria integrase. Os novos sítios recombinados são chamados de *attL* (Left - esquerda) e *attR* (Right - direita) por se encontrarem agora flanqueando o material genético do fago integrado ao genoma do hospedeiro (**Figura 1**)¹¹.

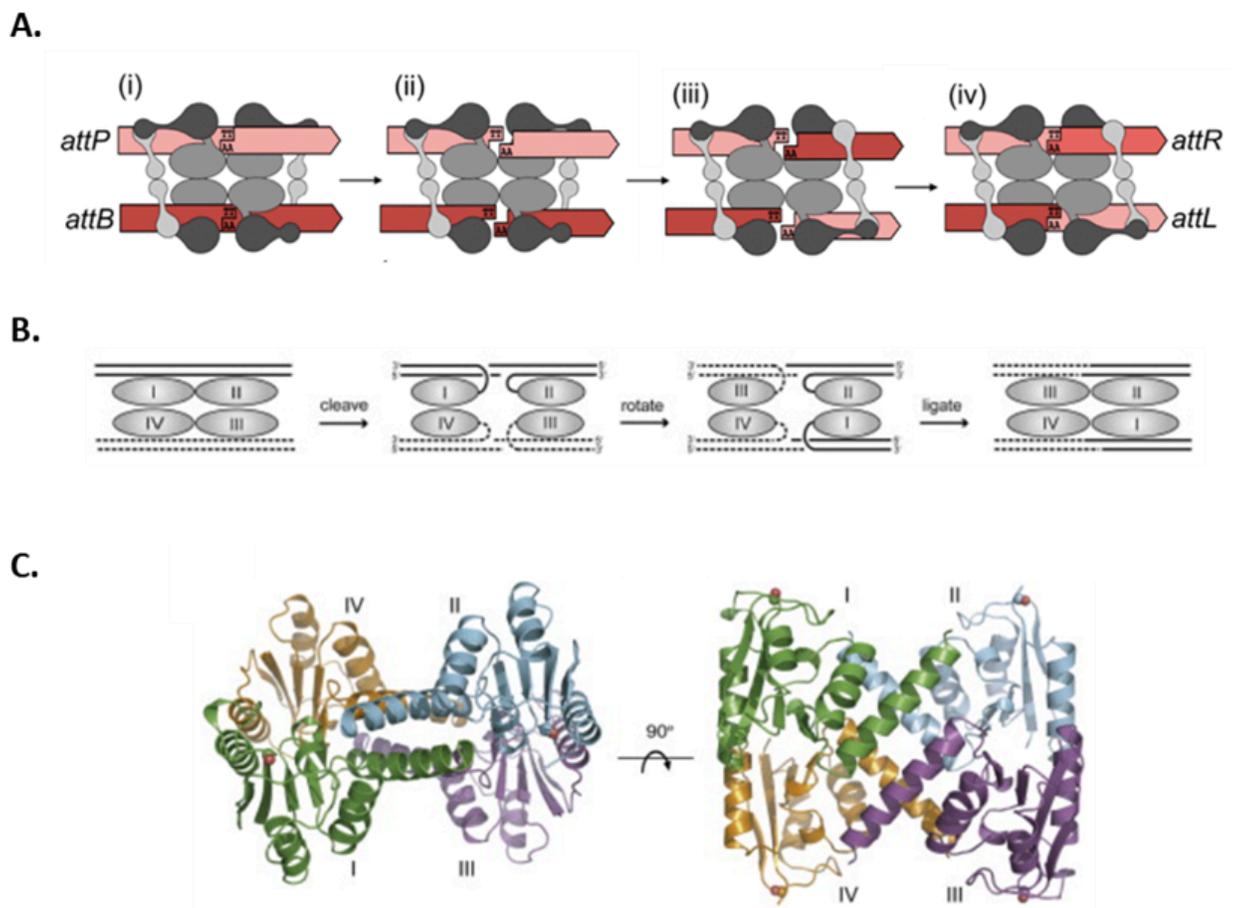


FIGURA 1. A e B) Diagramas de formação dos homo-tetrâmeros de serina-integrases após ligação ao DNA, clivagem e recombinação dos sítios *att*. C) Estruturas cristalográficas dos tetrâmeros indicando alterações conformacionais após giro de 90°. A- adaptação de Merrick et al.¹¹ ; B e C - Adaptação de Yuan et al.¹⁰

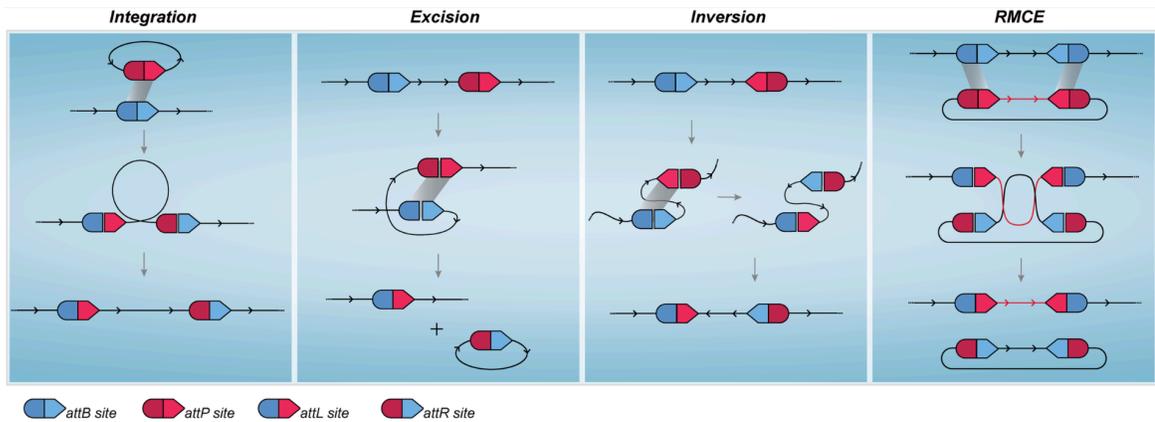
Apesar de o mecanismo geral de ação ser compartilhado entre as duas principais famílias de SSR, há importantes diferenças entre seus membros. Serina-Integrases e Tirosina-Integrases

recebem esse nome devido à presença dos respectivos aminoácidos em seu sítio catalítico e diferem funcionalmente em dois aspectos principais: direcionalidade e efetores auxiliares. Enquanto as recombinações realizadas por tirosina-integrases podem ser revertidas pelo mesmo sistema, a ação de serina-integrases se dá de maneira unidirecional, isto é, elas não são capazes de promover a recombinação entre sítios *attL* e *attR* uma vez formados, apenas entre o par de sítios *attB* e *attP* - sendo a reversão da unidirecionalidade possível apenas na presença de uma outra proteína conhecida como RDF (*Recombination Directionality Factor*, ou Fator de Direcionalidade de Recombinação em português)⁷. As proteínas RDF possuem origem também no genoma do profago e são específicas para cada integrase, com poucos pares Int-RDF identificados e caracterizados até o momento, sendo alguns deles phiC31 – gp3¹²; Bxb1 – gp47¹³; TP901-1 – orf7¹⁴; phiRv1 – Rv1584c¹⁵; SPBc – SprB¹⁶; phiJoe – gp52¹⁷; A118 – gp44¹⁸; phiBT1 – gp3BT1¹⁹. Outra forma de possibilitar a reversibilidade da recombinação se dá por meio do uso de isoformas contendo mutações ativadoras de bidirecionalidade^{20,21}. Tais mutações, como as identificadas na região C-terminal de phiC31 resultam na geração de isoformas com potencial de auto-ativação da recombinação reversa dos sítios *attL/attR* mesmo na ausência do RDF gp3. O segundo principal aspecto de diferenciação entre as duas famílias é a capacidade de serina-integrases realizar todo o processo de recombinação, incluindo clivagem e ligação da fita dupla de DNA, de maneira independente, enquanto tirosina-integrases usualmente necessitam de cofatores codificados pelo hospedeiro para agirem - os chamados Fatores de Integração do Hospedeiro (IHF) normalmente responsáveis por alterações topológicas do DNA necessárias para a ação destas integrases.

Apesar de tirosina-integrases como os sistemas Cre/Lox²²⁻²⁴ e Flp/FRT^{25,26} serem historicamente mais amplamente utilizadas em estudos de biologia molecular e metodologias bem estabelecidas como clonagem Gateway^{27,28}, as serina-integrases têm ganhado espaço na biologia sintética justamente por suas vantagens envolvendo o caráter permanente de sua recombinação unidirecional, a utilização de sítios *att* menores e independência de fatores auxiliares mencionados. Sua aplicação enquanto ferramenta biotecnológica passa a ser ainda mais significativa quando manipulações nos desenhos dos pares de sítio *att* conferem novos usos para além da integração de moléculas de DNA. Como mencionado, a integração ocorre quando os sítios *attB* e *attP* se encontram em moléculas distintas, como é o caso dos genomas de bacteriófago e bactéria hospedeira. No entanto, desenhos racionais de construções permitem a utilização das integrases como ferramentas capazes de inverter, excisar ou até mesmo trocar sequências de DNA flanqueadas por sítios *att*. Conforme ilustrado na **Figura 2**, a orientação relativa entre os sítios *attB* e *attP* é a grande responsável por essas diferentes resultantes. Quando os sítios *attB* e *attP* de um par estão presentes em uma mesma molécula com a mesma orientação, a sua recombinação resultará na excisão do DNA por eles flanqueado, restando na molécula principal apenas o sítio *attL* recombinante. Já em uma construção onde os sítios sejam inseridos em orientação oposta em relação a seu par, a sua recombinação ocasionará a inversão 180° da orientação do DNA flanqueado por eles. Esta última aplicação é de grande importância por permitir a regulação da expressão de um gene se considerarmos que a inversão pode alterar a

orientação de uma sequência codificadora em relação a seu promotor, resultando em seu silenciamento ou ativação a depender do estado originalmente estabelecido na construção (Figura 2B). Uma outra importante propriedade das integrases, denominada ortogonalidade, é a especificidade de ação restrita ao seu respectivo par de sítios *attB/attP*, não sendo capaz de levar à recombinação de sítios *att* correspondentes à outra integrases. Diversos trabalhos já mostraram por meio de testes cruzados e elaboração de matrizes de ortogonalidade a especificidade destas recombinases²⁹⁻³¹, apesar de em alguns poucos casos ser possível observar certo grau de ativação cruzada entre integrases e sítios *att* de outra recombinase, como exemplo os pares Int10/Int7²⁹, HbiF/FimE²⁹ e Cre/Dre³¹.

A.



B.

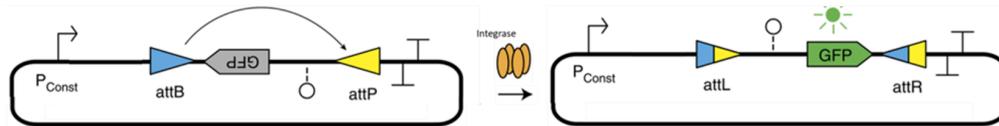


FIGURA 2. Esquematização de possíveis resultados de recombinação por serina-integrases. A) Integração, Excisão, Inversão e Troca de Cassete Mediada por Recombinase (RMCE - *Recombinase Mediated Cassette Exchange*). B) Inversão demonstrando em detalhes o funcionamento de um interruptor baseado em integrases. A inversão da sequência codificadora resulta em alinhamento de sua orientação com relação ao seu promotor Pconst, ativando a produção de GFP. B- adaptação de Yang et al.²⁹

Dadas as propriedades mencionadas, principalmente as várias possíveis formas de rearranjo de DNA, a irreversibilidade da recombinação e sua ortogonalidade (que assim permite o uso concomitante de diversas integrases em um mesmo sistema), tais recombinases ganharam grande destaque na criação de interruptores e circuitos gênicos. Empregando as serina-integrases Bxb1 e TP901-1 e um conjunto de módulos reguladores como promotores induzíveis e terminadores, Bonnet e colaboradores sistematizaram a criação e uso de portas lógicas na

construção de circuitos para regulação da taxa transcricional de genes³². Com as mesmas integrases reguladas por promotores induzíveis por glicose e nitrito Coubert e colaboradores construíram bactérias *Escherichia coli* capazes de identificar glicosúria em urina de pacientes diabéticos³³. Desde então as mais variadas aplicações de tais circuitos têm sido realizadas, seja na regulação de expressão de genes de interesse, construção de sensores e registro de eventos. Nesse contexto de desenvolvimento de circuitos biológicos cabe destaque a ferramentas de bioinformática para o auxílio e automação do desenho de circuitos complexos baseados em portas lógicas, como a plataforma Cello (www.cellocad.org) desenvolvida por Nielsen e colaboradores³⁴⁻³⁶. Outra aplicação promissora de integrases é a construção de sistemas de registro de memória e computação genética com base na sua capacidade de invertase unidirecional. Considerando os estados invertido e não-invertido de uma sequência de DNA como estados binários 1 e 0 resultantes da ativação ou não de 11 integrases, Yang e colaboradores foram capazes de construir um sistema de armazenamento permanente com capacidade proposta de 1.375 bytes (referentes às 2048 possíveis combinações de estados das 11 sequências reguladas)²⁹. Estudos anteriores já haviam demonstrado a aplicabilidade de integrases na computação de eventos com menor capacidade de armazenamento com a utilização de 2 integrases: Cre/Flpe²⁶ e Bxb1/TP901-1³².

Ainda que decorrido mais de 3 décadas desde os primeiros artigos com serina-integrases³⁷ e apesar de todo o potencial apresentado por elas, por muito tempo apenas um limitado grupo de integrases era aplicado nos estudos publicados, especialmente no caso de uso em organismos eucariontes, com largo predomínio do uso de phiC31 e Bxb1 e as integrases phiR4 e TP901 em menor frequência. Assim, o desenvolvimento de *pipelines* e plataformas de bioinformática para identificação de novas integrases e seus sítios, em conjunto com trabalhos de sistematização da caracterização funcional dessas moléculas em organismos diversos observados nos últimos anos é de extrema importância, especialmente considerando a demanda por um conjunto maior de integrases disponíveis e ortogonais entre si para a construção de cascatas de ativação e circuitos multi-componente cada vez mais intrincados. Dentre as ferramentas de identificação *in silico* de novas integrases se destacam os trabalhos realizados por Yang e colaboradores²⁹, conseguindo identificar 13 novas integrases funcionais e seus sítios e utilizando-as na construção de interruptores de memória permanente em *E. coli*, e o mais recente trabalho de Durrant e colaboradores³⁰, capaz de identificar mais de 60 novas integrases e caracterizá-las funcionalmente em células humanas. A identificação de novas integrases funcionais é ainda mais necessária quando consideramos sua aplicação em eucariotos, uma vez que a maior complexidade e compartimentalização das células criam barreiras funcionais que acabam por reduzir o leque de recombinases disponíveis. Xu e colaboradores³⁸, por exemplo, demonstraram que algumas integrases capazes de promover a recombinação em DNA plasmidial transformado em células humanas perderam a capacidade de fazê-lo quando o DNA alvo encontrava-se inserido no genoma, enquanto Andreas e colaboradores³⁹ conseguiram aumentar a eficiência de recombinação de phiC31 em células de camundongo por meio de alterações como a adição de sequências de localização nuclear (NLS) na porção C-terminal da proteína. Gomide e

colaboradores também demonstram esse efeito de transição procarioto-eucarioto ao aplicar em modelos variados de células animal e vegetal 6 integrases previamente caracterizadas em *E. coli*⁴⁰. Ainda que no estudo em bactéria todas as integrases tenham apresentado altos índices de eficiência na ativação de expressão do gene repórter, sem grandes variações entre si, no estudo com modelos eucariontes padrões diversos de eficiência foram observados, ainda que tenha sido possível a detecção molecular da recombinação mesmo para integrases que não levaram a um aumento de expressão do gene alvo.

Esta variação de eficiência entre diferentes modelos não ocorre apenas quando da transição de um modelo procarionte para um eucarionte, mas também entre organismos taxonomicamente mais próximos, como o caso de estudos com TP901, que em um mesmo trabalho se mostrou funcional em células humanas mas incapaz de executar a recombinação em células de camundongo³⁸, mesmo já tendo se mostrado funcional em *Saccharomyces cerevisiae*⁴¹. Mais recentemente Guizoui e colaboradores⁴² também não obtiveram sucesso com o uso desta integrase em *Arabidopsis*. Tais variações reforçam a importância de plataformas para sistematização da caracterização funcional de integrases em modelos diversos para melhor estabelecer o conjunto dessas ferramentas de edição gênica mais apropriado para cada organismo. Um levantamento dos estudos até então publicados com a aplicação de serina-integrases para edição de DNA em organismos eucariontes é apresentado na **Tabela 01**, incluindo não apenas as integrases e organismos estudados, mas também o mecanismo de rearranjo do DNA aplicado e as metodologias de entrega da integrase efetora. O artigo de descrição de nossa plataforma de protocolos para avaliação sistemática de serina-integrases em células animais, incluindo humanas, e vegetais encontra-se publicado no periódico PlosONE com título “*Protocol for the establishment of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells*” [doi: <https://doi.org/10.1371/journal.pone.0303999>], sendo apresentada uma cópia no **Anexo A**.

TABELA 01. Aplicações de serina-integrases na edição de DNA em sistemas eucariontes.

Int	Mecanismo	Organismo	Delivery	Ref
68 novas (Sh25 ; Si74 ; Bm99 ; Me99 ; Ma37 ; Nm60 ; Cc91 ; Vh19 ; Cs56 ; Bl24 ; No67 ; Fm04 ; Bu30 ; Ma05 ; Rh64 ; Cb16 ; uCb4 ; Ec03 ; Ec04 ; Ec05 ; Ec06 ; Ec07 ; Ef01 ; Ef02 ; Kp01 ; Kp03 ; Kp04 ; Kp05 ; Pa01 ; Pa03 ; Sa01 ; Sa02 ; Pf13 ; Td08 ; Se37 ; Ct03 ; Cd31 ; Ps40 ; Sa10 ; Td01 ; Enc3 ; Fp10 ; Ph43 ; Sm18 ; Cd16 ; Pf80 ; Bs46 ; Pf48 ; Rb27 ; Sa51 ; Bc30 ; Cd04 ; Cd15 ; Sa34 ; Pp20 ; RI09 ; Efs2 ; Pf15 ; Ps45 ; Sp56 ; Dn29 ; Vh73 ; Em12 ; Pc64 ; Vp82 ; Cp36 ; Pc01 ; Enc9)	integração	<i>Homo sapiens</i>	transfecção	30

Int	Mecanismo	Organismo	Delivery	Ref
A118	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens</i> ; <i>Mus musculus</i>	transfecção ; eletroporação ;	38
	integração ; excisão	<i>Homo sapiens</i>	transfecção	43
Bxb1	integração	<i>Homo sapiens</i>	transfecção ; eletroporação	44
	integração	<i>Plasmodium falciparum</i>	transfecção	45
	inversão ; excisão	<i>Mus musculus</i> ; <i>Drosophila melanogaster</i>		46
	excisão	<i>Homo sapiens</i>	transfecção	47
	inversão	<i>Homo sapiens</i> , <i>Bos indicus</i> , <i>Arabidopsis thaliana</i>	transfecção ; PEG transformação	40
	integração	<i>Homo sapiens</i>	transfecção	30
	integração	<i>Homo sapiens</i>	transfecção	48
	inversão	<i>Arabidopsis thaliana</i> ; <i>Nicotiana benthamiana</i>	agroinfiltração	42
	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens</i> ; <i>Mus musculus</i>	transfecção ; eletroporação	38
	integração ; excisão	<i>Homo sapiens</i>	transfecção	43
	inversão ; excisão ; integração ;	<i>Schizosaccharomyces pombe</i>	eletroporação	49
	integração	<i>Plasmodium falciparum</i>	transfecção	50
	integração	<i>Plasmodium falciparum</i>	transfecção	51
	biolistic	<i>Oryza sativa</i> ; <i>Nicotiana tabacum</i>	agroinfiltração	52
	RMCE	<i>Homo sapiens</i>	nucleofação	53
	RMCE	<i>Homo sapiens</i> ; <i>Mus musculus</i>		54
EFC-1	integração	<i>Homo sapiens</i>	transfecção	55

Int	Mecanismo	Organismo	Delivery	Ref
Int13	inversão	<i>Homo sapiens, Bos indicus, Arabidopsis thaliana</i>	transfecção ; PEG transformação	40
Int2	excisão	<i>Homo sapiens</i>	transfecção	47
	inversão	<i>Homo sapiens, Bos indicus, Arabidopsis thaliana</i>	transfecção ; PEG transformação	40
Int4	excisão	<i>Homo sapiens</i>	transfecção	47
	inversão	<i>Homo sapiens, Bos indicus, Arabidopsis thaliana</i>	transfecção ; PEG transformação	40
Int5	inversão	<i>Homo sapiens, Bos indicus, Arabidopsis thaliana</i>	transfecção ; PEG transformação	40
Int7	excisão	<i>Homo sapiens</i>	transfecção	47
	inversão	<i>Homo sapiens, Bos indicus, Arabidopsis thaliana</i>	transfecção ; PEG transformação	40
Int9	inversão	<i>Homo sapiens, Bos indicus, Arabidopsis thaliana</i>	transfecção ; PEG transformação	40
MR11	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens ; Mus musculus</i>	transfecção ; eletroporação ;	38
phi370.1	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens ; Mus musculus</i>	transfecção ; eletroporação	38
phiBT1	inversão ; integração	<i>Homo sapiens</i>	transfecção	56
	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens ; Mus musculus</i>	transfecção ; eletroporação	38
phiC1	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens ; Mus musculus</i>	transfecção ; eletroporação	38

Int	Mecanismo	Organismo	Delivery	Ref
phiC31	inversão ; excisão	<i>Homo sapiens ; Mus musculus</i>		57
	inversão ; integração	<i>Homo sapiens</i>	transfecção	56
	integração	<i>Cricetulus griseus</i>	transfecção	58
	integração	<i>Drosophila melanogaster</i>	injeção	59
	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	60
	RMCE	<i>Homo sapiens</i>	nucleofação	53
	integração	<i>Homo sapiens</i>	transfecção	30
	integração	<i>Homo sapiens</i>	transfecção	48
	inversão	<i>Arabidopsis thaliana ; Nicotiana benthamiana</i>	agroinfiltração	42
	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Danio rerio</i>	injeção	61
	RMCE	<i>Homo sapiens ; Mus musculus</i>	transfecção ; eletroporação ;	38
	RMCE	<i>Homo sapiens ; Mus musculus</i>		54
	RMCE	<i>Mus musculus</i>	eletroporação	62
	RMCE	<i>Drosophila melanogaster</i>	injeção	63
	inversão ; excisão	<i>Gallus gallus</i>	eletroporação	64
	integração ; excisão	<i>Anopheles gambiae ; Aedes aegypti ; Drosophila melanogaster ; Bombyx mori ; Spodoptera frugiperda</i>	transfecção	65
	integração	<i>Xenopus laevis</i>	injeção	66
	inversão	<i>Nicotiana benthamiana</i>	agroinfiltração	67
	integração	<i>Arabidopsis thaliana</i>	agrobacteria floral dip	68
excisão	<i>Homo sapiens</i>	transfecção	47	
inversão	<i>Homo sapiens, Bos indicus, Arabidopsis thaliana</i>	transfecção ; PEG transformação	40	

Int	Mecanismo	Organismo	Delivery	Ref
phiFC1	integração ; excisão	<i>Homo sapiens</i>	transfecção	43
phiK38	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens ; Mus musculus</i>	transfecção ; eletroporação ;	38
phiRV	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens ; Mus musculus</i>	transfecção ; eletroporação ;	38
	integração ; excisão	<i>Homo sapiens</i>	transfecção	43
R4	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens ; Mus musculus</i>	transfecção ; eletroporação ;	38
	integração ; excisão	<i>Anopheles gambiae ; Aedes aegypti ; Drosophila melanogaster ; Bombyx mori ; Spodoptera frugiperda</i>	transfecção	65
SPBC	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens ; Mus musculus</i>	transfecção ; eletroporação ;	38
TG1	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens ; Mus musculus</i>	transfecção ; eletroporação ;	38
	inversão ; integração	<i>Homo sapiens</i>	transfecção	56
TP901	excisão	<i>Homo sapiens</i>	transfecção	47
	inversão	<i>Arabidopsis thaliana ; Nicotiana benthamiana</i>	agroinfiltração	42
	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens ; Mus musculus</i>	transfecção ; eletroporação ;	38

Int	Mecanismo	Organismo	Delivery	Ref
	inversão ; excisão ; integração ;	<i>Schizosaccharomyces pombe</i>	eletroporação	49
U153	integração ; excisão	<i>Homo sapiens</i>	transfecção	43
	inversão ; excisão ; integração ;	<i>Schizosaccharomyces pombe</i>	eletroporação	49
Wβ	RMCE	<i>Saccharomyces cerevisiae</i>	transformação	41
	RMCE	<i>Homo sapiens ; Mus musculus</i>	transfecção ; eletroporação ;	38

A célula mínima Mycoplasma mycoides JCVI-Syn3

Para além da necessidade de desenvolvimento de ferramentas moleculares que permitam a manipulação genética de organismos, a busca por dominar os mecanismos moleculares da vida ao ponto de criá-la sinteticamente demandaria um conhecimento profundo, se não absoluto, do comportamento, interação e função de cada componente molecular em uma célula. Tal ideal, praticamente inconcebível em muitos aspectos dada a complexidade inerente aos sistemas biológicos, torna-se menos intangível à luz de uma visão embasada em um reducionismo científico, propondo a utilização de modelos mais simples possíveis para interpretação e explicação de fenômenos. Assim, a idealização de um modelo reducionista em termos moleculares culminaria na busca de uma célula contendo apenas um conjunto mínimo de genes teoricamente capazes de originar um organismo vivo e funcional, fornecendo um contexto celular simplificado a ser usado nestes estudos.

Tal debate se faz presente na literatura há muito tempo, com discussões teóricas e primeiras proposições já desde os anos 1950 com o interesse na identificação das menores unidades replicativas autônomas^{69,70}, com as discussões ganhando força e tomando novos formatos na busca direta de genes essenciais com a início da publicação de genomas completos em meados dos anos 1990⁷¹⁻⁷³. Tal conjunto mínimo autossuficiente de genes, denominados então como genes essenciais, definiriam as bases moleculares da própria existência e propagação do conceito de vida como conhecemos⁷². A existência de tal conjunto seria inclusive sugerida na própria natureza em eventos de perda de genes observados em relações de parasitismo, por exemplo, onde muitos dos componentes celulares e vias metabólicas deixam de ser necessários a partir do momento em que diversas necessidades metabólicas passam a ser atendidas pelo meio

externo. Um exemplo interessante deste tipo de minimização natural de genomas pode ser encontrado no caso de bactérias como as mycoplasmas. *M. genitalium*, uma bactéria patogênica parasita responsável por infecções dos tratos genital e respiratório humanos, é considerado o principal evento desse processo de redução natural de genomas, carregando um genoma de 580 kb consistindo em apenas 525 genes⁷² (a bactéria modelo *E. coli*, a fins de comparação, possui 4401 genes), já tendo sido teorizado e demonstrado por ensaios de perdas de função individual que tal número poderia chegar a ser reduzido para 382⁷⁴.

Esta busca, porém, não se limitou à simples observação e descrição de células presentes na natureza, dando lugar ao direcionamento de esforços voltados à criação de uma célula sintética capaz de carregar apenas genes essenciais. O primeiro organismo controlado por um genoma inteiramente sintetizado quimicamente foi publicado por Gibson e colaboradores em 2010, com a construção da bactéria *Mycoplasma mycoides* JCVI-syn1.0⁷⁵. A construção deste organismo consistiu na síntese química completa do genoma de *M. mycoides*, com adaptações para remoção de genes de virulência e inserção de sequências necessárias para montagem e identificação da nova molécula, e o posterior transplante desse genoma para células de *M. capricolum* cujo material genético havia sido previamente removido. Os avanços na capacidade de síntese e manipulação de um genoma completo tornaram então possível a edição de sequências extensas e reorganização dessas moléculas para criação de novos organismos, além de fornecer à comunidade científica uma nova plataforma para estudos de genômica funcional com a linhagem JCVI-syn1.0. Dominando assim um amplo arsenal de ferramentas relacionadas ao desenho, síntese, manipulação e transplante de genomas sintéticos desenvolvidas no âmbito da geração de Syn1.0, tais estudos se concentraram no uso desta linhagem para identificação dos chamados genes essenciais, uma vez que agora seria possível não apenas silenciar individualmente genes para avaliação de essencialidade em ensaios de perda de função, mas de fato construir um organismo contendo apenas genes essenciais em seu genoma.

Os estudos de essencialidade realizados envolveram o silenciamento de genes por meio de bombardeamento de transposon Tn5 a fim de provocar sua integração aleatória em regiões do genoma, resultando em quebra do gene afetado com consequente interrupção de expressão da proteína por ele codificada e perda de função exercida por esta^{73,76}. A **Figura 3** traz o diagrama com mapeamento dos sítios de inserção do transposon ao longo do genoma identificados. Após os bombardeamentos as células foram cultivadas e as sobreviventes tiveram seu genoma sequenciado para detecção do local de inserção do transposon. A presença do transposon em uma sequência identificada em uma célula viva indicaria, assim, a não essencialidade do gene silenciado, uma vez que sua perda de função não resultou em perda de viabilidade. Ausência de eventos de inserção em um determinado gene seria uma confirmação de sua essencialidade, pois indicaria que as células em que tal gene fosse silenciado não seriam capazes de sobreviver a ponto de aparecerem nos sequenciamentos. Uma terceira categoria de genes também foi classificada, denominados genes quasi-essenciais. Esses genes seriam identificados ao acompanhar culturas contendo *pools* de células carregando inserções em diferentes partes do

genoma individualmente. Ao longo das gerações a queda de representatividade (chegando à ausência) de células com um determinado gene silenciado inicialmente presente nos primeiros ciclos de divisão seria indicativo de uma perda de capacidade competitiva e prejuízo à capacidade de propagação. Ainda que não imediatamente deletéria, sua remoção traria danos significativos à célula, sendo então mantidos juntamente com os genes essenciais.

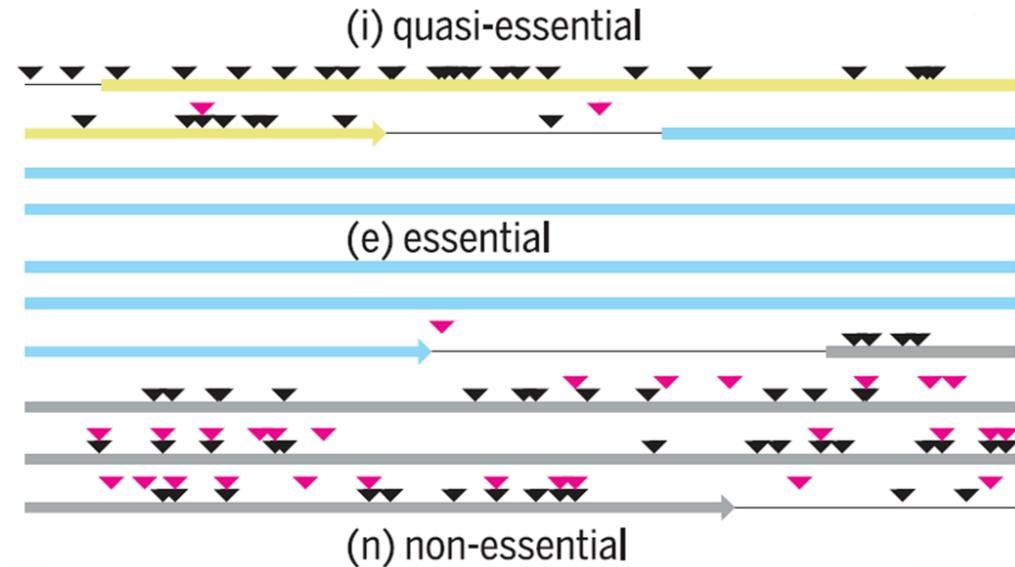


Figura 3. Mapeamento de ocorrências de inserção de Tn5 no genoma de Syn1.0 por meio de ensaios de bombardeamento de transposon para identificação de genes essenciais. Setas pretas representam inserções detectadas na primeira passagem. Setas rosa representam inserções detectadas após a quarta passagem. A ausência de inserções em um gene permite sua classificação como essencial, enquanto a detecção de inserções na primeira passagem mas sua ausência após a quarta passagem sinaliza perda de capacidade competitiva. Adaptado de Hutchinson et al. (2016)⁷⁶

Tendo em mãos o levantamento de genes afetados procedeu-se então a síntese química de um novo genoma de *M. mycoides*, agora sem os genes considerados não-essenciais. A síntese e implantação do novo genoma resulta na criação e publicação da nova linhagem *M. mycoides* JCVI-syn3.0, carregando um genoma de apenas 531 kbp, com manutenção de apenas 473 genes dos 910 originalmente presentes em Syn1⁷⁶. Apesar da construção bem sucedida e do importante avanço que representou a criação de uma célula sintética mínima viva e capaz de se replicar, Syn3.0 apresentou considerável redução nas taxas de replicação e certa instabilidade morfológica que levaram à necessidade de reinserção em seu genoma de 19 genes originalmente considerados não essenciais, dando origem à linhagem JCVI-syn3A^{77,78}. Uma nova linhagem foi posteriormente criada a partir de Syn3A visando facilitar sua utilização como plataforma experimental. A nova linhagem, denominada JCVI-syn3B, carrega os 492 genes de Syn3A e uma região de integração direcionada (*landing pad*) baseada em sítios loxP para inserção facilitada de

genes de interesse em seu genoma por meio de utilização do sistema Cre/Lox. Uma comparação entre os genomas de Syn1.0 e Syn3.0 é apresentada na **Figura 4**.

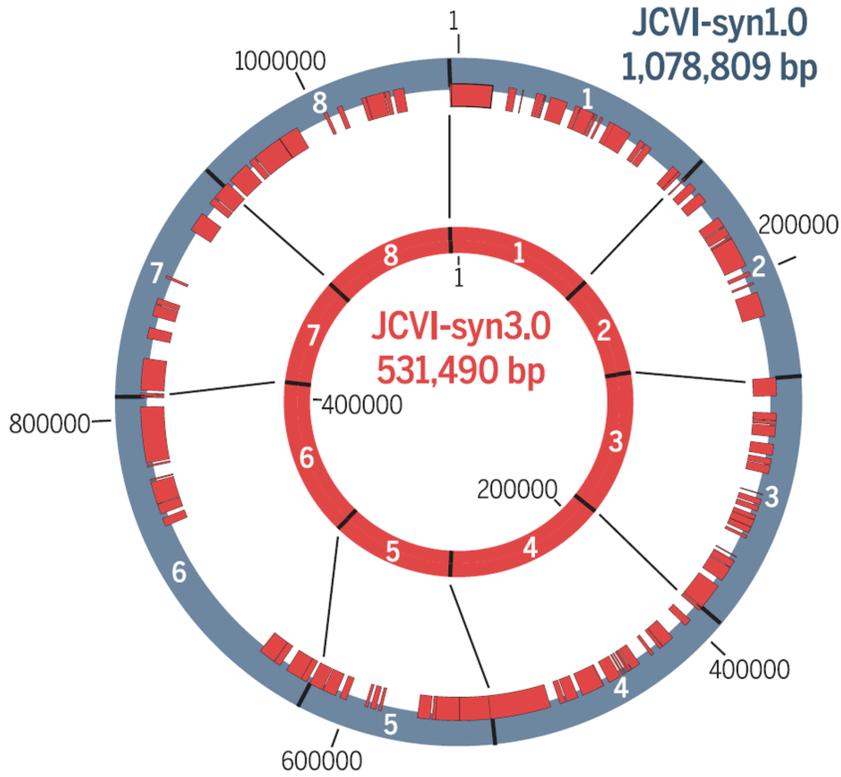


Figura 4. Representação dos genomas da célula sintética *M. mycoides* JCVI-syn1.0 e da célula minimizada *M. mycoides* JCVI-syn3.0. Regiões em vermelho representam genes mantidos durante o processo de minimização. Adaptado de Hutchinson et al. (2016)⁷⁶

Capítulo I

Construção de interruptores genéticos ativados por serina-integrases na célula sintética minimizada *Mycoplasma mycoides* JCVI-syn3B: regulação transcricional controlada por Int9 e sua aplicação no controle de expressão da endonuclease I-CeuI para produção de *SimCells*.

Introdução

Marco incontestável no campo da biologia sintética, a criação da célula controlada por um genoma minimizado *M. mycoides* JCVI-Syn3.0 e suas linhagens derivadas tem aberto caminho para significativos avanços em áreas diversas, da genômica funcional⁷⁸ à criação de modelos descritivos *in silico*⁷⁹⁻⁸¹, bem como sua aplicação enquanto chassi para expressão heteróloga de proteínas e montagem de mecanismos moleculares provenientes de outros organismos^{82,83}. A considerável redução de seu genoma facilita (ou ao menos torna menos complexo) o processamento computacional de informações globais do sistema, sendo uma das áreas de destaque na literatura envolvendo a célula mínima que tem se mostrado altamente prolífica em especial nos últimos anos. Dentre os vários avanços de bioinformática observados tem destaque a caracterização proteômica *in silico*⁸⁴, mapeamento de fluxos metabólicos⁷⁸ e descrição de arquitetura cromossomal e mecanismos de expressão gênica da célula, assim como a montagem do banco de dados Synwiki⁸⁵, voltado para o levantamento e disponibilização fácil de genes e proteínas de Syn3A. Outro campo de destaque é o de estudos evolutivos com a célula. O acompanhamento de culturas ao longo de centenas a milhares de gerações em estudos de evolução laboratorial adaptativa tem demonstrado, dentre outros, um aumento de fitness com redução do tempo de duplicação ao longo das passagens^{86,87}. Hossain e colaboradores também descrevem o desenvolvimento de resistência a antibióticos por meio de ensaios de evolução adaptativa com Syn3B⁸⁸.

A remoção de praticamente metade dos genes originalmente presentes fornece um ambiente citoplasmático com muito menos componentes moleculares capazes de interferir no funcionamento de vias metabólicas ou expressão heteróloga de proteínas, por exemplo. O

contexto molecular menos complexo também abre espaço para o estudo, desenvolvimento e aplicação de ferramentas moleculares e circuitos genéticos visando a construção *de novo* de componentes sintéticos que poderão vir a formar uma célula completamente sintética dentro da abordagem *Bottom-Up*. No entanto, este mesmo processo de minimização cria impedimentos ao uso de importantes ferramentas moleculares ao remover componentes celulares necessários.

A remoção de sistemas de reparo de DNA, por exemplo, interfere diretamente no uso do sistema CRISPR para edições de DNA em Syn3B e outras espécies de mycoplasma minimizadas naturalmente, uma vez que ainda que a enzima Cas seja capaz de ser direcionada a um sítio específico com uso de RNAs guia e catalisar a quebra das fitas de DNA, a ausência do um sistema de reparo e recombinação homóloga impede a inserção de mutações ou integração de sequências de DNA no genoma, interrompendo o processo inclusive por incapacidade de promover a religação nas regiões clivadas⁸⁹. Atualmente, para qualquer edição diretamente no genoma da célula mínima este precisa ser introduzido em células de levedura, onde serão editados por CRISPR ou recombinação homóloga, por exemplo, e posteriormente purificado e transplantado novamente para uma célula receptora, sendo um processo consideravelmente trabalhoso. Para contornar tais limitações, algumas metodologias vêm sendo desenvolvidas a fim de permitir edição ou ao menos regulação gênica nestes sistemas. Um dos exemplos é a metodologia de interferência por CRISPR, batizada de CRISPRi, utilizada para permitir o silenciamento de genes alvo nas células Syn1.0, Syn3.0 e derivadas⁹⁰. O método envolve a utilização de uma Cas9 defeitiva (dCAS9), capaz de se ligar à sequência alvo guiada pelo sistema clássico de gRNA, mas incapaz de causar quebra na dupla fita de DNA. Sua ligação à região alvo causa apenas um impedimento estérico à ligação de polimerases, bloqueando assim a transcrição do gene a ser silenciado. Já para edição direta de DNA em algumas espécies de mycoplasma metodologias como a fusão de dCas9 a editores de base⁹¹, reintrodução de vias de reparo que possibilitem o uso da tecnologia CRISPR e associação de recombinases já foram publicadas⁹².

Especialmente visando a criação de circuitos genéticos capazes de emular redes regulatórias sintéticas e o desenvolvimento de ferramentas de edição gênica funcionais para a célula mínima, serina-integrases surgem como importantes candidatas dada sua capacidade de realizar rearranjos de DNA a partir da recombinação direta de seus sítios de reconhecimento sem a necessidade de nenhum outro fator complementar do hospedeiro, o que se mostra uma considerável vantagem no contexto molecular reduzido da célula mínima.

Considerando o potencial representado pelos avanços alcançados na biologia sintética com a identificação de genes essenciais e a construção e disponibilização da célula mínima JCVI-syn3B, somado aos obstáculos impostos à aplicação de ferramentas moleculares amplamente utilizadas, como CRISPR e a escassez de promotores induzíveis funcionais no sistema, justamente em decorrência da minimização atingida, é proposto no presente trabalho a avaliação funcional e aplicação de serina-integrases na célula mínima *M. mycoides* JCVI-Syn3B visando ampliar a disponibilidade de ferramentas de edição gênica no organismo, bem como

permitir a construção de interruptores genéticos capazes de regular a expressão gênica em resposta a estímulos externos, o que abrirá caminho para a construção de circuitos genéticos complexos capazes de atuar como redes regulatórias sintéticas em possíveis construções *bottom-up* de uma célula verdadeiramente sintética em um futuro próximo.

Objetivo Geral

Construir interruptores genéticos controlados por serina-integrase capazes de modular a expressão de gene-alvo integrado ao genoma da célula sintética minimizada *M. mycoides* JCVI-syn3B por meio da inversão 180° de sua sequência codificadora. Além de funcionar como protótipo de redes regulatórias sintéticas, o interruptor deverá ser aplicado na geração de *SimCells* (ou *Simple Cells*, corpos celulares depletados de genoma utilizados como plataforma de expressão em biologia sintética) através do controle da transcrição da endonuclease I-CeuI.

Objetivos específicos

- Seleção de integrases e desenho de vetores integrativos contendo os módulos efetor (gene da integrase) e repórter (reverso-complemento do gene de proteína fluorescente *mcherry* flanqueado por sítios *att*);
- Construção dos vetores por meio de metodologia de montagem *in vivo* em *E. coli* de fragmentos com extremidades homólogas;
- Transformação integrativa de Syn3B para geração de linhagens contendo o interruptor em seu genoma
- Avaliação funcional do interruptor por indução com tetraciclina e quantificação de níveis de fluorescência
- Construção e avaliação de interruptor Int9-ICeuI para controle de expressão da endonuclease
- Geração de *SimCells*
- Testes de promotores induzíveis para construção de sistema *kill-switch* baseado em CRISPRi para seleção negativa de células escape
- Desenvolvimento de sistema de seleção negativa e produção de amostras de *SimCells* puras

Metodologia

Soluções e Meios de Cultura

Meio LB baixo sódio (para seleção com Zeocina)

<i>Tryptona</i>	10 g
<i>NaCl</i>	5 g
<i>Extrato de levedura</i>	5 g
<i>Bacto agar</i>	16 g
<i>Água destilada</i>	qsp 1 L

Ajuste para pH 7.5
Autoclave 15 min; 121°C

Meio SP4

1 litro

(Parte 1):

<i>Mycoplasma Broth Base (BBL)</i>	3.5 g
<i>Bacto Tryptone (Difco)</i>	10 g
<i>Bacto Peptone (Difco)</i>	5.3 g
<i>Água destilada</i>	qsp 600 mL
<i>(agar microbiológico , granulado)</i>	10 g apenas para meio sólido

Ajuste para pH 7.5 com 1-2 N NaOH
Autoclave 15 min.

(Parte 2):

<i>CMRL 1066 10x TC medium (Gibco 21540-026)</i>	50 mL
<i>Dextrose 20%, filtrada</i>	25 mL
<i>Yeast Extract Solution (Gibco 18180-059)</i>	35 mL
<i>Yeastolate, 4% autoclavado (Difco 5577-15)</i>	50 mL
<i>Água destilada autoclavada</i>	50 mL
<i>KnockOut* (Gibco 10828028)</i>	170 mL
<i>Penicillin G, premade solution, 400000 U/mL stock</i>	2.5 mL
<i>NaHCO₃, 7.5%, filtrado</i>	14.6 mL
<i>L-Glutamine, 200 mM, filtrada</i>	5 mL
<i>Phenol red solution, 0.5% (Sigma P0290)</i>	4 mL
<i>Total:</i>	403.7 mL

Recomenda-se filtrar com membrana de 0.22um a Parte 2 completa após preparo, mesmo com a utilização de componentes estéreis.

* *KnockOut* pode ser substituído por Soro Fetal Bovino (Inativado por calor a 56°C por 1h), mantendo-se o mesmo volume indicado.

Após preparo das duas frações (Parte 1 e Parte 2), e garantindo-se o resfriamento da Parte 1 após autoclavagem, misture-as para obtenção do meio SP4 completo. pH final do meio deve estar entre 7.6 e 7.8

Tampão S/T

0.5 M Sacarose

10 mM tris-HCl, pH 7.5

Solução Concentrada de Sacarose

0.5 M Sacarose

20 mM tris-HCl, pH 7.5

Solução 70% PEG6000

PEG6000 17.5 g

Tampão S/T 10.5 mL

Volume final: 25 mL

Incubar a 50°C com agitação para solubilização. Manter a 4°C após preparo

Obs.: antes de usar durante transformação, manter a 50-60°C para preparo de alíquotas e reequilibrar a 37°C por pelo menos 15 min antes de adicionar às células

Linhagens celulares, cultivo e manipulação

As linhagens de *M. mycoides* JCVI-syn1.0, JCVI-syn3.0 e JCVI-syn3B utilizadas neste trabalho foram gentilmente cedidas por Dr. John Glass através de Acordo de Transferência de Materiais entre EMBRAPA Recursos Genéticos e Biotecnologia e *J. Craig Venter Institute* (JCVI, La Jolla, CA, EUA). O material recebido foi utilizado na inoculação de culturas líquidas de meio SP4 puro mantidas em estufa de incubação a 37°C por pelo menos duas passagens para reativação e preparo de estoques a serem mantidos congelados em ultra-freezer a -80°C. Inóculos iniciais para experimentos eram realizados por meio de raspagem de culturas estoque congeladas e inoculação do material em 1 mL de meio SP4 - puro ou suplementado com os devidos antibióticos quando necessário - incubadas a 37°C *overnight*. Mudanças da coloração do meio SP4 de vermelho para alaranjado/amarelado em decorrência da acidificação do meio sinalizam a progressão dos estágios de desenvolvimento das culturas.

Os pares de oligonucleotídeos *SynID* apresentados na **Tabela 2** foram usados na confirmação das linhagens quando do preparo de estoques ou sempre que necessária a confirmação ao longo do trabalho. Toda manipulação de microrganismos foi realizada em capela de fluxo laminar presente em laboratório sob conformidade e registro de Nível de Biossegurança II.

TABELA 2. Oligonucleotídeos utilizados na genotipagem de linhagens *M. mycoides* JCVI-syn

Oligonucleotídeo	Sequência
SynID_All_Fwd*	cgtagatcaattggatagatcg
SynID_All_Rvs**	agagcttcagtctgctgctc
SynID_capri_Fwd	gcatcaaaaacaacacttgacg
SynID_capri_Rvs	gaagctgaagaaattgctgctg
SynID_S10_Fwd	tgaggattgcatgaaacacgac
SynID_S10_Rvs	ctcttgcttgacctaaactagc
SynID_S20_Fwd	ggctctggtagttatactggtg
SynID_S20_Rvs	ggtaaacttggttctgtagag
SynID_S3s_Fwd	gaaccaagtacaatgctaagtg
SynID_S3s_Rvs	gatctatacgctatgattgctg
SynID_S30_Fwd	taattgtaatggtgtagcaagc
SynID_S30_Rvs	cagtatgctactgctagatgctc
SynID_S3A30_Fwd	acgtggagggtaattctgctag
SynID_S3A30_Rvs	cttcattccgtaactcttctac
SynID_S3A3B_Fwd	tgataaccaccagttattcctg
SynID_S3A3B_Rvs	ccaattcaaggaactgcttcag
SynID_S3B_Fwd	ccttcgtttatcttgctgctc
SynID_S3B_Rvs	cagagtactgcaattgactg

*Fwd: *forward* (oligos com orientação senso)

**Rvs: *reverse* (oligos com orientação antisenso)

Design e construção de plasmídeos

Avaliação funcional de serina-integrases. Vetores utilizados nos ensaios de avaliação funcional de interruptores genéticos ativados por serina-integrases foram construídos utilizando-se como arcabouço o plasmídeo pSD080 cedido por Dr Yo Suzuki (JCVI, La Jolla, CA, EUA), derivado da família de plasmídeos desenvolvidos para aplicação do sistema CRISPRi em Syn3A⁹⁰. Tal plasmídeo carrega a maquinaria do sistema Cre/LoxP necessária para integração direcionada no *landing pad* presente no genoma da linhagem Syn3B, bem como genes de resistência a Zeocina e Puomicina para seleção de transformantes. O desenho dos plasmídeos de interesse inclui dois módulos presentes no mesmo vetor integrativo: um módulo efetor e um módulo repórter. O módulo efetor consiste no gene da serina-integrase selecionada sob controle transcricional do promotor induzível TetR-Pxyl/TetO2 ativado por tetraciclina. O módulo repórter consiste no gene *mcherry* silenciado, por estar inserido em orientação contrária à de seu promotor constitutivo de espiralina, flanqueado pelos sítios *attB* e *attP* da respectiva integrase avaliada. Os genes utilizados na construção foram desenhados e sintetizados na forma de *gBlocks* (IDT) com códon otimização baseada na tabela de frequência de códons de *Plasmodium falciparum* disponibilizada pela própria ferramenta. Foram selecionadas para avaliação da capacidade de inverter o gene repórter as integrases Int9 e Int13^{29,40}. Vetores intermediários diversos também foram construídos como controles para avaliar possíveis interferências como a presença de sítios pré- e pós- recombinação sobre os níveis de expressão do repórter. Diagramas das construções estão apresentados na **Figura 5**.

Os vetores utilizados receberam as seguintes denominações:

- pSD080_iX_pBP: gene repórter silenciado flanqueado pelos sítios *attB* e *attP*; ausência do gene da respectiva integrase (iX, onde ‘X’ representa o número referente à Int utilizada)
- pSD080_iX_pLR: gene repórter ativo flanqueado pelos sítios recombinados *attL* e *attR*; ausência do gene da respectiva integrase (iX, onde ‘X’ representa o número referente à Int utilizada)
- pSD080_iX_iBP: gene repórter silenciado flanqueado pelos sítios *attB* e *attP*; inclusão do gene da respectiva integrase (iX, onde ‘X’ representa o número referente à Int utilizada)
- pSD080_iX_iLR: gene repórter ativo flanqueado pelos sítios recombinados *attL* e *attR*; inclusão do gene da respectiva integrase (iX, onde ‘X’ representa o número referente à Int utilizada)
- pSD080_Tet*mcherry*: gene repórter ativo sob regulação por promotor induzível TetR-Pxyl/TetO2; ausência de quaisquer sítios *att* e integrase

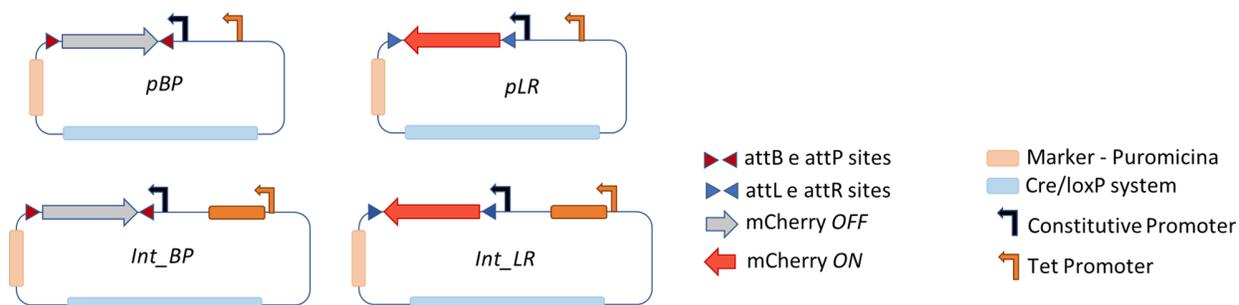


FIGURA 5. Mapas esquemáticos das construções realizadas

A construção dos vetores foi realizada através da metodologia de montagem *in vivo* de fragmentos com extremidades homólogas publicada por Kostylev e colaboradores⁹³ (Figura 6). Brevemente, oligonucleotídeos são utilizados para amplificação dos fragmentos de interesse e adição de regiões de homologia com 30-50 nt nas extremidades. Após amplificação as amostras são submetidas a tratamento com enzima de restrição DpnI (ThermoFisher) para eliminação do DNA molde e purificação por kit (Zymo) dos *amplicons* gerados. Os fragmentos obtidos são então misturados em proporção molar de 3:1 e adicionados diretamente em reações de transformação por choque térmico de células quimiocompetentes de *E. coli* DH5a (NEB).

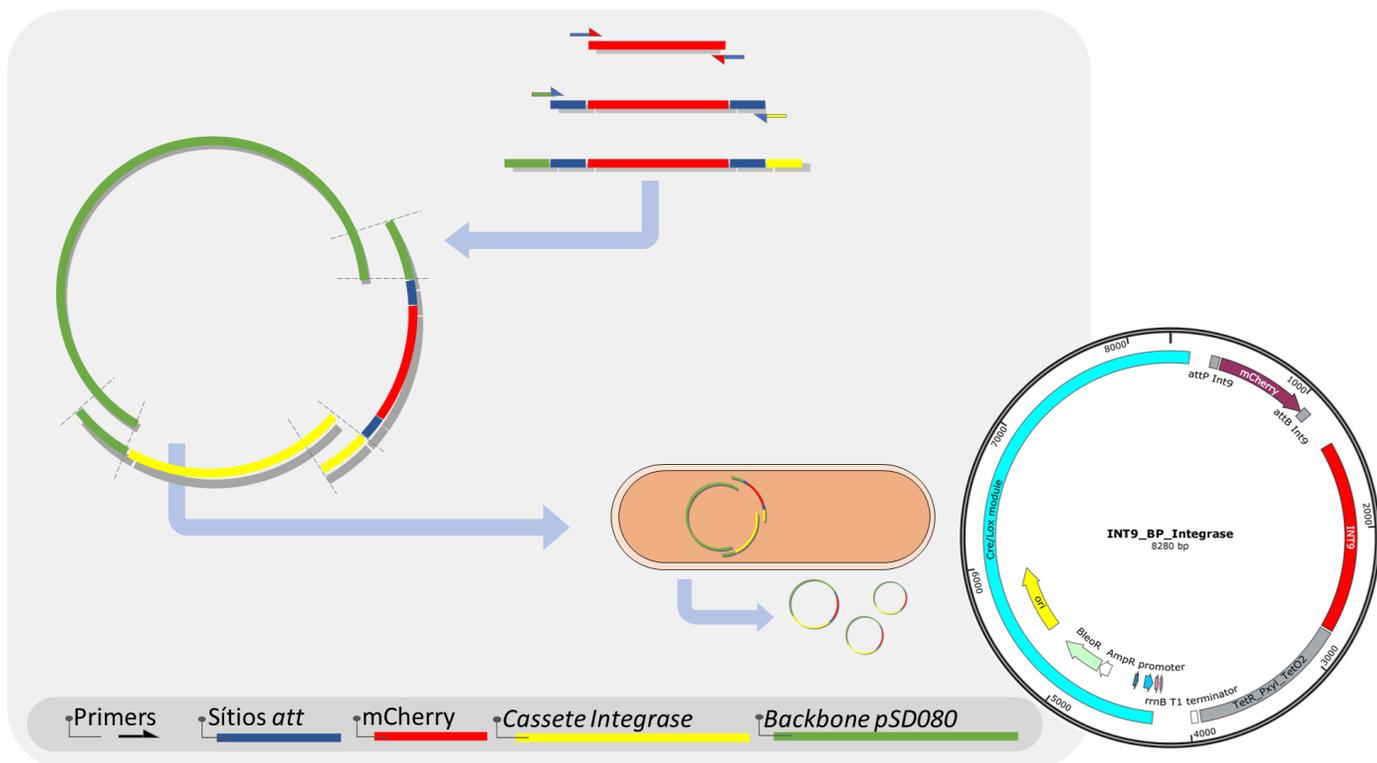


FIGURA 6. Modelo de construção de plasmídeos pelo método de montagem *in vivo* de fragmentos com extremidades homólogas

Avaliação de promotores induzíveis. Para os ensaios com promotores induzíveis em Syn3b foi utilizado como arcabouço o vetor Pmod2loxpurolox-cre-sp. Assim como pSD080, o vetor pMod também possui o sistema Cre/LoxP para integração no *landing pad* e genes de resistência a ampicilina e puromicina para seleção de transformantes em *E. coli* e mycoplasma, respectivamente. As sequências de promotores induzíveis pA13 (arabinose), pL52b62 e pLG64 (IPTG) e respectivos fatores de transcrição foram selecionadas após publicação de caracterização funcional em *Mycoplasma pneumoniae* por Broto e colaboradores⁹⁴ e utilizadas no desenho final das construções contendo os fatores de transcrição com expressão constitutiva controlada por promotor de espiralina e o gene repórter *mcherry* códon otimizado sob controle transcricional dos promotores induzíveis selecionados. Os mapas e instruções de construção dos vetores Pmod_pA13_mCh_TFAraR, Pmod_pA13_mCh_TFAraR_AraE, Pmod_pL52b62_mCh_TFL4 e Pmod_pLG64_mCh_TFL4 foram enviados para a empresa EPOCH Life Science Inc (Missouri City, TX, EUA) para síntese química.

Transformação de células Syn3b

Para obtenção de linhagens Syn3B transformantes de interesse, 1 mL de meio SP4 líquido puro era inoculado e incubado a 37°C por 24h. Após esse período, 1 µL, 2 µL, 5 µL, 10 µL e 20 µL da cultura inicial eram adicionados a novos tubos de centrifugação contendo mais 1.3 mL de meio SP4 puro e incubadas *overnight*, 37°C. Ao fim da incubação a cultura de células naturalmente competentes mais apropriada para início do processo de transformação é escolhida com base na cor da cultura, sendo o ideal a escolha da cultura apresentando um início de transição para uma coloração levemente alaranjada do meio. A cultura deve então ser centrifugada a 9000 RCF por 8 min a temperatura ambiente, lavada em 800 µL de Tampão S/T e após nova centrifugação ressuspender o precipitado em 70 µL de CaCl₂ e incubadas em gelo por 30 min. Após incubação, 17 µL de células são transferidas para um novo tubo contendo 3 µL da amostra de DNA a ser transformada e a mistura homogeneizada com uma pipeta e novamente incubada em gelo por 15 min. 133 µL de solução 70% de PEG6000 dissolvido em Tampão S/T são adicionados ao mix de células e DNA e incubados por 2 min a temperatura ambiente. Segue-se nova lavagem com 800 µL de Tampão S/T, centrifugação a 10000 RCF, 10°C por 15 min, ressuspensão do precipitado em 500 µL de meio SP4 e nova incubação por 4h a 37°C. Ao fim desse período as células devem ser espalhadas em superfície de meio SP4 sólido contendo 2 mg/mL de puomicina para seleção dos transformantes. As colônias estarão visíveis com auxílio de lupa estereoscópica após 2 a 4 dias de incubação a 37°C.

Sequenciamento Nanopore e análise de dados

Para a identificação da integração direcionada dos interruptores genéticos na região de *landing pad* presente no genoma de Syn3B, bem como a confirmação da edição gênica após ativação de Int9 ou Int13 - com consequente formação dos sítios *attL* e *attR* e a inversão e ativação do gene repórter - foi realizado sequenciamento de DNA genômico de culturas das linhagens Syn3B_Int9BP e Syn3B_Int13BP crescidas na ausência ou presença do indutor Tetraciclina por meio de tecnologia *nanopore* (Oxford Nanopore Technologies - ONT, UK). Para tal, 1 mL de culturas crescidas por 48h em meio SP4 na presença ou ausência do indutor foi utilizado para extração de gDNA por meio de uso do kit PureLink Genomic DNA Mini Kit (Invitrogen, USA) de acordo com recomendações do fabricante. O DNA obtido foi avaliado por eletroforese em gel de agarose e quantificação em Qubit com kit 1x dsDNA BR Assay (Invitrogen, USA). Para o preparo das bibliotecas a serem analisadas foi utilizado o protocolo de sequenciamento rápido para gDNA - Rapid Barcoding kit 96 V14 (SQK-RBK114.96 - ONT, UK) de acordo com orientações do fabricante. Este protocolo utiliza transposase para quebra e inserção das sequências de identificação Rapid Barcodes (RB01-96), sendo as etapas de lavagem realizadas por precipitação magnética com esferas AMPure XP beads (ONT, USA).

As bibliotecas foram sequenciadas em flow cell R10.4.1 FLO-PRO114M (Oxford Nanopore Technologies, UK) na unidade de sequenciamento PromethIon 2 (PRO-SEQ002) com uso do software MinKNOW [v23.11.5]. *Basecalling*, *demultiplexing* e remoção de sequências

adaptadoras foram realizados usando modelo de *High-accuracy basecalling* (HAC) 400bps na versão mais atualizada do *basecaller* Dorado [v7.2.13]. Para análise dos resultados, as sequências FastQ obtidas para cada *barcode* foram importadas para o software Geneious Prime [2024.0.2] e mapeadas contra as sequências referência apropriadas a cada caso utilizando a ferramenta de mapeamento Minimap2 [2.24].

Avaliação funcional de interruptores genéticos.

Para avaliação funcional das integrases foram realizados ensaios de indução e quantificação de fluorescência do gene repórter. Para tal, raspagens das culturas congeladas de transformantes foram inoculadas em meio SP4 + 2 mg/mL de Puromicina e incubadas *overnight*, 37°C. As culturas crescidas foram então utilizadas na inoculação de dois grupos cada: I- Grupo controle em SP4 + Puromicina e II - Grupo Induzido em meio SP4 + Puromicina + Tetraciclina. Em um primeiro momento todas as induções foram realizadas com 4 µg/mL de Tetraciclina, sendo testada a concentração de 0.4 µg/mL posteriormente apenas nos ensaios com Int9. Após 48h na presença do indutor, 800 µL de cada cultura dos grupos experimentais foram transferidos para um novo tubo de centrifugação contendo 400 µL de solução concentrada de sacarose (0.5 M sacarose, 20 mM Tris-HCl; pH 7.5) e centrifugadas a 16000 RCF, 10 min em temperatura ambiente. Com cuidado o meio na fase superior deve ser retirado, seguindo-se a remoção da solução de sacarose e a ressuspensão do precipitado em 100 µL de solução de sacarose nova.

Para aferição da intensidade de fluorescência de mCherry, os 100 µL contendo as células lavadas foram transferidos para placas de cultura de 96 poços pretas. As leituras foram realizadas em leitor SpectraMax M3 (Molecular Devices LLC, EUA) com medição em comprimentos de onda 575 nm para excitação e 620 nm para emissão. Após a leitura, 90 µL das culturas foram novamente transferidas para placas de cultura de 96 poços transparentes para leitura de absorbância em 600 nm utilizada na normalização dos dados de fluorescência obtidos.

Ensaio de Unidades de Mudança de Cor de Culturas

O ensaio de Unidades de Mudança de Cor⁹⁵ (*Colour Changing Units* ou CCU, em inglês) foi utilizado na avaliação funcional da endonuclease I-CeuI expressa após ativação de interruptor controlado por Int9 em Syn3B. O ensaio permite acompanhar alterações de viabilidade de culturas. Para tal, após 24h de indução a 37°C com 4 mg/mL de Anidrotetraciclina (aTc), diluições seriadas em fator 1:10 das culturas foram preparadas em poços de placa de cultura de 96 poços contendo meio SP4, volume final de 150 µL por poço, e incubadas por 48h a 37°C. Ao final da incubação a extensão de pontos de diluição apresentando alteração de cor do meio em decorrência do crescimento celular indica a intensidade do efeito citotóxico. 20 µL das culturas induzidas também foram plaqueadas em spots em placa de meio SP4 sólido + 4 µg/mL aTc para avaliação da formação de colônias.

Curva de Crescimento e Acúmulo de Conteúdo Genômico

Dada a atividade de endonuclease e consequente destruição do material genômico de I-CeuI nas células, outra forma aplicada para avaliar seu nível de atividade dependente de ativação pelo sistema Int9 foi a realização de curvas de crescimento em presença do marcador de DNA DAPI (4',6'-diamino-2-fenil-indol). Inicialmente, pré-inóculos com as células transformadas foram incubados por 24h a 37°C. Após o crescimento foi adicionado às culturas DAPI (*Ready made solution* 1 mg/mL, Sigma-Aldrich) na concentração final de 10 µg/mL por 30 minutos para garantir uma marcação inicial das células, seguindo-se lavagem com meio SP4 novo, centrifugação (4000 RCF, 4 min, temperatura ambiente) e ressuspensão das células em meio SP4 novo mantendo-se o volume inicial da cultura. Após a marcação inicial com DAPI as células foram inoculadas em placas pretas de 96 poços na presença ou ausência de 4 µg/mL de tetraciclina para ativação do interruptor genético e consequente indução da expressão de I-CeuI. Concentrações de 1:1 e 1:50 do pré-inóculo foram plaqueados em volume final de 150 µL de meio SP4 + 1 µg/mL DAPI em cada poço e a placa incubada em leitor SpectraMax M3 a 37°C por 48h em regime de cinética com leituras de fluorescência (Excitação: 360 nm, Emissão: 460 nm) a cada 30min precedidas de agitação de 10 segundos para obtenção das curvas.

Caracterização funcional de promotores induzíveis

A avaliação dos promotores induzíveis pA13, pL52b62 e pLG64 também foi realizada por meio de quantificação de intensidade de fluorescência em leitor SpectraMax M3, uma vez que nas construções sintetizadas tais promotores controlam a transcrição do gene repórter *mcherry*. As linhagens transformadas com cada construção foram incubadas por 24h na presença dos indutores apropriados. As culturas foram preparadas seguindo o mesmo protocolo acima descrito para avaliação funcional dos interruptores genéticos controlados por integrase, sendo a aferição de intensidade de fluorescência do repórter realizada também nas condições descritas.

As concentrações de indutor testadas para cada promotor foram: I- 0%, 0.2%, 0.5% e 1% de arabinose para indução das construções pA13_AraR e pA13_AraE_AraR ; II - 0 mM, 0.5 mM, 1 mM e 2 mM de IPTG para indução das construções pLG64_L4 e pL52b62_L4 ; III - 0 µg/mL e 0.4 µg/mL aTc para indução da construção *Tetmcherry*.

Resultados e Discussão

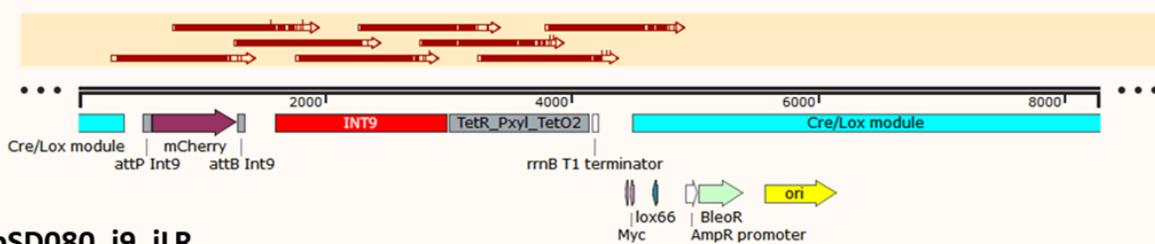
Os grandes avanços atingidos com a criação de uma célula altamente minimizada como Syn3 e os estudos sobre essencialidade gênica envolvidos, bem como as limitações e barreiras encontradas ao longo do projeto, ampliam consideravelmente as possibilidades de desenvolvimento de conhecimentos e ferramentas em diversas frentes da Biologia Sintética. Algumas dessas possibilidades incluem o uso da célula minimizada como chassi de biofábricas e meios de estudo e criação de vias e componentes moleculares em uma abordagem *Bottom-up* da

biologia sintética, uma vez que disponibiliza-se um ambiente biológico mais simplificado, com menos interferências possíveis nos sistemas em um contexto onde diversas vias metabólicas e regulatórias estão ausentes. Em contrapartida, limitações impostas como a possibilidade de silenciamento de apenas um gene por vez em cada célula utilizando-se do bombardeamento de transposons, a presença de genes identificados como essenciais ou quasi-essenciais ainda com funções desconhecidas e a possibilidade de ainda ampliar a minimização do genoma trazem grandes demandas para uma abordagem *Top-down* da biologia sintética. Dada a experiência de nosso grupo com aplicação de serina-integrases em diversos modelos eucarióticos e a plasticidade de tais ferramentas, seu estudo e aplicação na célula mínima gera bastante interesse nas duas abordagens.

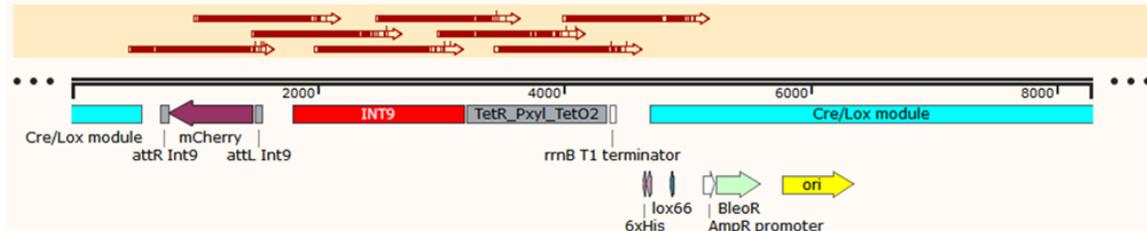
A falta de vias regulatórias abre espaço para a construção de vias sintéticas baseadas em cascatas de ativadores construindo interruptores e circuitos genéticos, bem como a falta de um sistema de reparo de DNA eficiente torna o organismo recalcitrante a muitas metodologias de edição gênica, especialmente através do sistema CRISPR tão amplamente utilizado atualmente. A independência de co-fatores endógenos e a direcionalidade de recombinação de sítios *att* específicos torna o uso de serina-integrases como ótimos candidatos a ferramentas de edição e controladores de vias regulatórias no sistema, contornando algumas das dificuldades impostas pela minimização. Ao mesmo tempo elas também podem enriquecer os esforços de identificação de genes essenciais, uma vez que o uso de combinações de diferentes integrases, alinhado à inserção de sítios *att* flanqueando grupos de diferentes genes relacionados em um mesmo cromossomo possibilitaria o estudo dos efeitos do silenciamento concomitante de mais de um gene ao mesmo tempo, permitindo um entendimento maior das relações inter-genes ofuscadas por redundâncias funcionais. Nesse contexto, nos propusemos à avaliação funcional e aplicação de algumas dessas serina-integrases na célula minimizada *M. mycoides* JCVI-syn3B.

A estratégia utilizada no desenho das construções se baseia na utilização de integrases para inverter a orientação de um gene de interesse como consequência da recombinação de seus sítios *attB* e *attP* inseridos antes e após a sequência codificadora do gene (**Figura 5**). Dado o desenho inicial, onde o gene de interesse se encontra posicionado em orientação oposta à de seu promotor, portanto silenciado, sua inversão pela integrase efetora deverá resultar na ativação da expressão. Com base em resultados previamente obtidos por nosso grupo⁴⁰ foram selecionadas as integrases Int9 e Int13. As construções realizadas por meio de montagem *in vivo* de fragmentos com extremidades homólogas ou Gibson Assembly foram verificadas por sequenciamento. Mutações presentes em fragmentos foram consideradas erros de sequenciamento quando houvesse existência de fragmentos corretos se sobrepondo (**Figura 7**). Após confirmação as construções foram utilizadas na transformação integrativa de Syn3B, originando as linhagens experimentais.

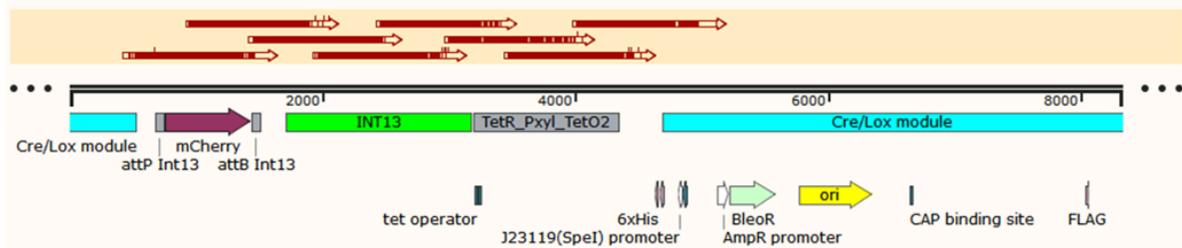
pSD080_i9_iBP



pSD080_i9_iLR



pSD080_i13_iBP



pSD080_i13_iLR

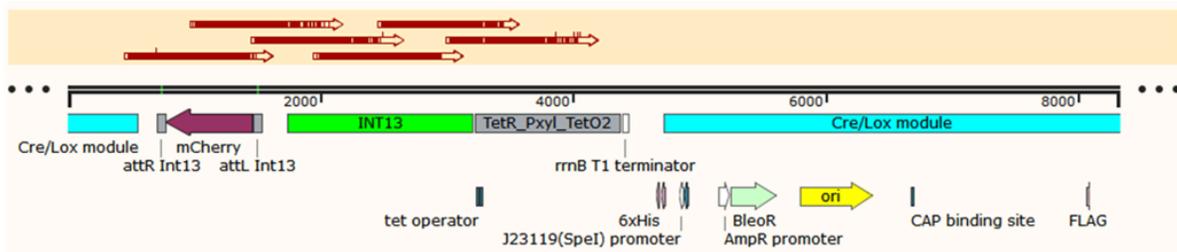
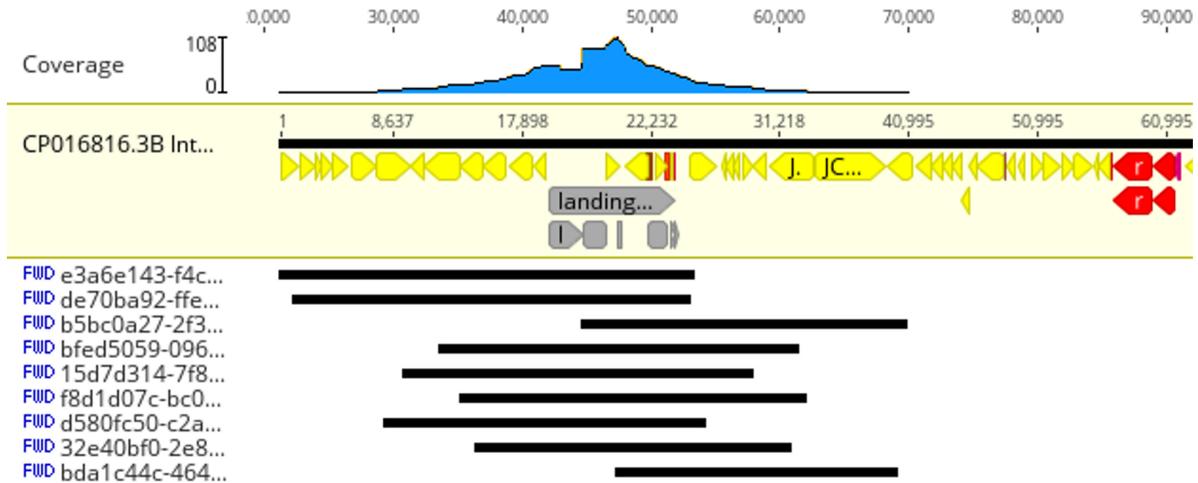


FIGURA 7. Mapa das construções utilizadas no estudo. As setas vermelhas representam alinhamento de seqüências obtidas por sequenciamento Sanger para confirmação das construções. Mutações corrigidas pela sobreposição de seqüências corretas foram consideradas erros de sequenciamento.

Os transformantes com resistência a puromicina também foram avaliados por sequenciamento de genoma utilizando tecnologia *nanopore* para confirmação da integração das construções na região de *landing pad* presente em Syn-3B. As seqüências obtidas foram filtradas por meio de alinhamento local utilizando a versão *plugin* do mapeador Minimap2 presente no software Geneious Prime, sendo utilizado como isca a seqüência do repórter *mcherry*. Os *reads* selecionados por conter seqüências do gene repórter foram então alinhados pelo mesmo método ao genoma de Syn3B com a inserção esperada. Conforme apresentado na **Figura 8**, é possível

observar a integração adequada da construção, com regiões flanqueadoras alinhadas à jusante e à montante do sítio de inserção.

A.



B.

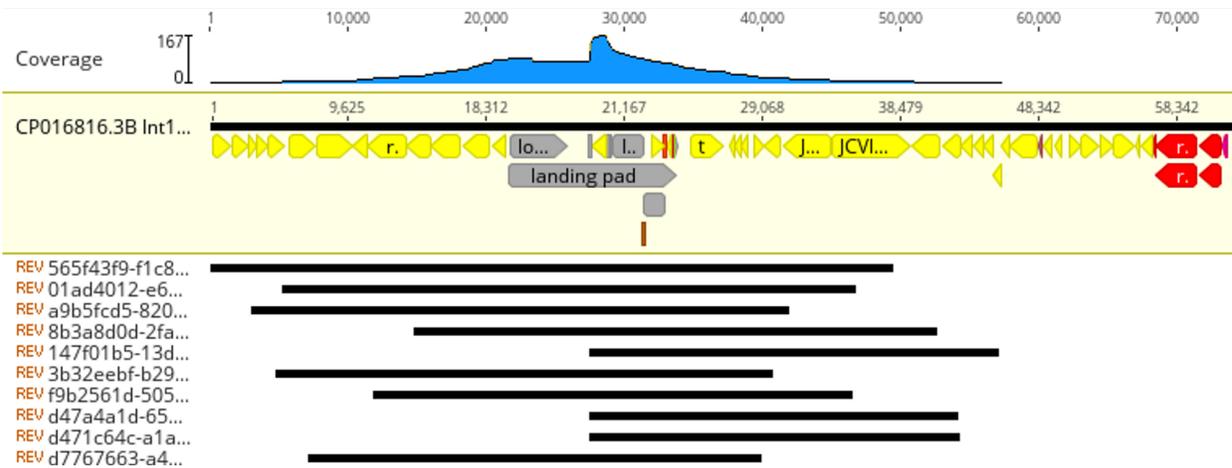


FIGURA 8. Mapeamento de integração dos interruptores genéticos controlados por Int9 ou Int13 ao genoma de Syn3B. A) Sequenciamento genômico da linhagem Syn3B_Int9iBP. B) Sequenciamento genômico da linhagem Syn3B_Int13iBP. Sequências de gDNA obtidas por tecnologia nanopore.

Inicialmente as induções foram realizadas crescendo as culturas na presença de tetraciclina na concentração de 4 µg/mL. Além da linhagem carregando o sistema de ativação completo, diversos outros controles foram avaliados. As linhagens pBP e pLR carregam o gene repórter flanqueado por sítios *att*, porém não apresentam o gene da respectiva integrase no

sistema, sendo esse um controle focado no comportamento do gene repórter inativado ou a resultante pós-recombinação no caso de pLR. Assim como o controle Int_LR já apresenta o gene repórter ativo na posição esperada pós-recombinação porém também conta com a presença da integrase no sistema, funcionando como controle positivo. Linhagens transformadas com o vetor pSD080_DelCas9 foram utilizadas como outro controle positivo, onde há expressão constitutiva de mCherry na ausência de sítios *att* o flanqueando. Já a construção gRNA853 utiliza o mesmo arcabouço, mas não possui o gene *mcherry*, sendo considerado como controle negativo.

Como indicado pelos resultados obtidos nas primeiras rodadas de indução apresentados nas **Figuras 9A e 9B**, a baixa intensidade de fluorescência medida com o grupo pBP de Int9 e Int13 em ambas condições, presença ou ausência de indutor, demonstram o silenciamento bem sucedido do gene repórter e ausência de possível atividade promotora dos sítios *att* como observado por Yang e colaboradores para o sítio *attP* de Int13 em *E. coli*²⁹. A ausência de fluorescência em Int_BP para ambas integrases em condições não induzidas também aponta uma boa robustez do sistema, sem ocorrência de vazamentos (*leaking*) e auto-indução significativos. Esse comportamento ganha destaque quando considerados os resultados obtidos por Mariscal e colaboradores⁹⁰ com o modelo de silenciamento por dCas9 em seu sistema de CRISPRi nas células sintéticas *M. mycoides* JCVI-syn1.0 e Syn3.0, onde um grau de vazamento do sistema é observado.

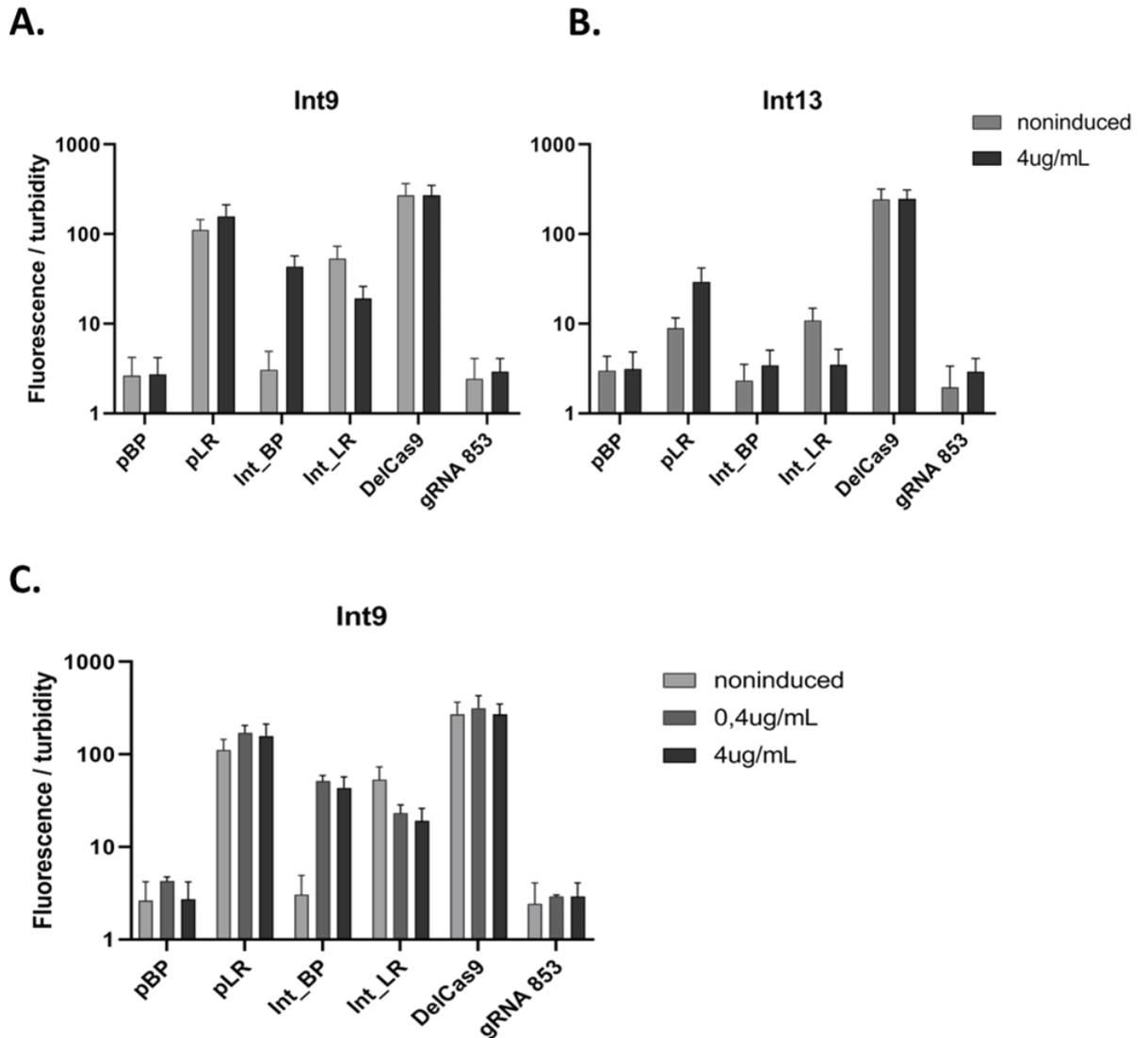


FIGURA 9. Ensaios de ativação do interruptor controlado por integrase por meio da indução do efector por tetraciclina. A) e B) Intensidade de fluorescência do repórter mCherry após 48h de incubação em presença ou ausência de indutor para transformantes contendo interruptores ativados por Int9 ou Int13, respectivamente. C) Avaliação de diferentes concentrações de indutor para ativação de interruptor controlado por Int9 após 48h de incubação. DelCas9: controle positivo com expressão constituída de *mcherry*; gRNA853: controle negativo sem presença do gene *mcherry*.

O segundo grupo controle com resultados importantes são as linhagens pLR. Nessas construções o comportamento difere entre Int9 e Int13. No primeiro caso os níveis de fluorescência observados se aproximam do controle positivo DelCas9, indicando pouca interferência causada pela presença dos sítios recombinado *attL* e *attR* sobre a expressão do gene repórter. Já com a presença dos sítios *attL* e *attR* de Int13 flanqueando o gene ativado nos grupos pLR e Int_LR há um considerável efeito negativo nas taxas de expressão de mCherry. Tal efeito

pode ser uma das razões por trás do aparente insucesso da ativação do gene alvo após indução de Int13 no grupo Int_BP.

O significativo aumento nos níveis de fluorescência registrados com a linhagem Int_BP de Int9 após indução da expressão de integrase na presença de tetraciclina demonstra a capacidade da integrase em recombinar seus sítios *attB* e *attP* integrados ao genoma de Syn3B, levando à inversão da sequência codificadora do gene *mcherry* e sua consequente ativação e expressão.

Visando avaliar a sensibilidade do sistema à tetraciclina, foram realizados novos ciclos de induções com a concentração de 4 µg/mL de tetraciclina inicialmente utilizada e uma concentração 10 vezes menor (0.4 µg/mL) do indutor. Os níveis de expressão de mCherry se mostraram semelhantes nas duas condições (**Figura 9C**). A possibilidade de uso de menores concentrações do indutor pode ser interessante, por exemplo, quando da aplicação do sistema em condições menos favoráveis ao crescimento celular, permitindo um menor desgaste das células, ou contextos que limitem a disponibilidade da tetraciclina até às células ou demandem uma economia de recursos, como a utilização em culturas de maior volume e até produções em larga escala.

A detecção molecular da recombinação promovida também foi realizada por meio de sequenciamento *nanopore*, tendo sido possível identificar a formação adequada esperada dos sítios *attL* e *attR* após indução da integrase, tanto para Int9 quanto para Int13 (**Figura 10**). O contraste entre a ausência de aumento dos níveis de fluorescência mesmo com a ocorrência da inversão e formação de sítios com a indução de Int13 se assemelha a resultados obtidos por Gomide e colaboradores⁴⁰ com outras integrases em nosso estudo com células eucariotes.

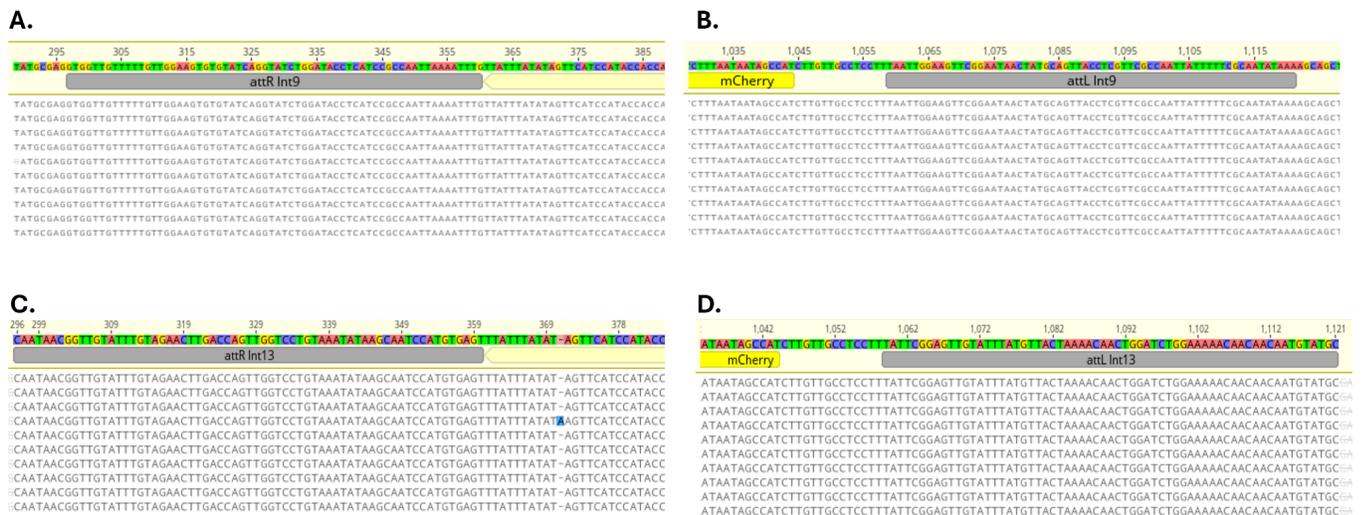


FIGURA 10. Sequenciamento e detecção da formação de sítios *attL* e *attR* após 48h de indução da expressão da respectiva integrase efetora. A) *attR* Int9; B) *attL* Int9; C) *attR* Int13 e D) *attL* Int13. Sequenciamento de gDNA extraído de células induzidas realizado com uso de tecnologia *nanopore*.

O sequenciamento também possibilita a quantificação do número de sequências contendo cada um dos sítios pré- e pós- recombinação. Ainda que seja necessário aplicar metodologias de análise dos dados de sequenciamento mais apropriadas para obtenção de dados quantitativos, a observação do número bruto de *reads* contendo cada sítio reforça algumas características identificadas nos ensaios de quantificação de fluorescência realizados. Conforme pode-se observar na **Tabela 3**, não houve detecção de sítios *attL* ou *attR* de Int9 na ausência de tetraciclina, reforçando a ausência de vazamentos e escape no controle da expressão da integrase e ativação indevida do repórter. Já considerando o grupo induzido, a ausência de *reads* contendo sítios *attB* ou *attP* é indicativo da alta eficiência de conversão do sistema nas culturas induzidas. Como forma de normalização, também foram feitos alinhamentos com três sequências de aproximadamente 60 nucleotídeos (semelhante aos sítios *att*) retiradas de regiões diferentes do gene *mcherry* presente na construção.

TABELA 3. Quantificação de *reads* contendo sequências dos sítios *att* de Int9 obtidas por meio de sequenciamento *nanopore*. As sequências mCh pertencem ao gene repórter *mcherry*.

	<i>attB</i>	<i>attP</i>	<i>attL</i>	<i>attR</i>	mCh1	mCh2	mCh3
i9iBP_noTet	75	76	0	0	80	74	86
i9iBP_Tet	0	0	132	148	148	147	149

Considerado em conjunto com o comportamento observado nos demais controles, esses dados comprovam a capacidade do sistema interruptor com Int9 de causar alterações diretamente no genoma hospedeiro funcionando como uma eficiente via regulatória de expressão de genes de interesse e edição do genoma.

Outro ponto importante a ser avaliado foi a comprovação da estabilidade das modificações de orientação e ativação do repórter. Para isso, culturas induzidas e não induzidas após 48h de incubação foram lavadas e submetidas a um novo ciclo de indução por 48h (**Figura 11**). Os termos “_0” e “_4” no final dos nomes das amostras indicam as condições “não induzido” ou “induzido com 4 µg/mL de tetraciclina” no primeiro ciclo. Todos os grupos foram então submetidos a nova condição de indução ou não-indução em uma segunda etapa após lavagens do meio. A manutenção da capacidade de ativação do sistema em um segundo ciclo de indução observada em Int_BP_0 demonstra a estabilidade do sistema após diversos ciclos de multiplicação celular, mantendo-se o sistema responsivo ao indutor. Já os altos níveis de

fluorescência observados em Int_BP_4 mesmo em condições de ausência do indutor demonstram o caráter permanente da edição e consequente ativação do gene alvo ocorrida ainda no primeiro ciclo de indução. O comportamento desse grupo na presença do indutor se assemelha ao observado para Int_LR, Int_LR_0 e Int_LR_4, onde na presença do indutor houve uma leve diminuição da intensidade de fluorescência, possivelmente em decorrência de interferência estérica causada por interações da integrase com os sítios recombinados *attL* e/ou *attR* e potencialmente atrapalhando a atividade transcricional da polimerase. Comportamento semelhante quanto ao efeito da presença de sítios *attL* e *attR* e sua interação com integrases já foi observado em outros trabalhos^{47,96}, sendo inclusive um fator levado em consideração nos modelos matemáticos de descrição de cinética elaborados por Chao e colaboradores. Outro possível fenômeno seria uma interferência no recrutamento da maquinaria de transcrição em decorrência do aumento de atividade transcricional na unidade transcricional responsável pela expressão da integrase na presença do indutor, por esta se encontrar imediatamente à montante da unidade de transcrição do gene repórter (a ser investigado).

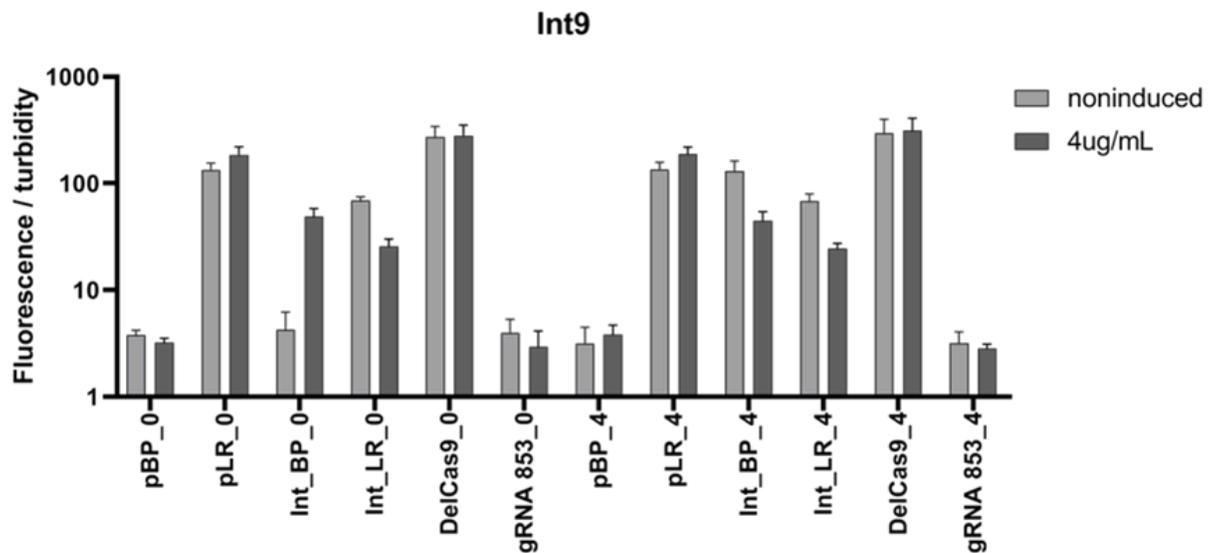


FIGURA 11. Ensaio de indução sequencial de culturas previamente expostas à tetraciclina. Uma segunda rodada de indução foi realizada após lavagens das células utilizadas no primeiro ciclo de 48h de indução. Grupos nomeados com o indicador “0” se referem às amostras controle não induzidas no primeiro ciclo, enquanto o marcador “_4” indica indução com 4 µg/mL de tetraciclina durante a primeira etapa.

Após a identificação de um sistema regulatório sintético baseado em integrases funcional no sistema minimizado, seguiu-se sua aplicação na construção de um interruptor genético com Int9 responsável pela ativação de um gene alvo com aplicações práticas, tendo sido selecionado o gene exógeno da endonuclease I-CeuI. Dada a presença de sítios de restrição de I-CeuI em

cópias do gene *rrl* do RNA ribossomal 23S, a ativação desta nuclease gera danos irreversíveis ao genoma da célula, iniciando cascatas de degradação do material genético. Os corpúsculos resultantes, apresentando membrana celular, citoplasma e toda maquinaria molecular intactos, porém agora sem um genoma, é conhecido como *SimCell*⁹⁷, do inglês *Simple Cells*. As *SimCells* apresentam grande valor dentro da biologia sintética justamente por ainda possuírem toda a maquinaria necessária para expressão de proteínas de interesse, montagem de vias metabólicas ou cascatas de ativação e testes de ferramentas moleculares, por exemplo, enquanto a ausência de um genoma contorna questões relacionadas à biocontenção de organismos geneticamente modificados e biossegurança. Nesse contexto consideramos o grande impacto da produção de *SimCells* a partir da célula mínima por esta apresentar um ambiente celular mais simplificado, uma vez que a falta dos diversos componentes moleculares não essenciais reduz o potencial de interferências indesejadas sobre os sistemas construídos e produtos esperados.

A construção do sistema de regulação de I-CeuI por Int9 apresenta a mesma estrutura geral do interruptor Int9 testado com o gene repórter *mcherry*. O cassete integrado ao genoma de Syn3B apresenta dois módulos, um contendo o gene de Int9 sob controle transcricional do promotor induzível por tetraciclina e outro contendo o gene da nuclease silenciado em orientação oposta a seu promotor de espiralina com expressão constitutiva, flanqueado pelos sítios *attB* e *attP* de Int9.

Uma vez que a ação de I-CeuI resulta na perda de material genético e viabilidade celular, as avaliações realizadas focaram no acompanhamento de curvas de crescimento em presença do marcador fluorescente de DNA DAPI e no impacto da ativação da nuclease sobre o crescimento de culturas. A **Figura 12** traz as curvas registradas para dois clones carregando o sistema Int9/I-CeuI e dois controles, sendo um a linhagem contendo o sistema repórter mCherry previamente avaliado e a linhagem selvagem Syn3B. As induções foram realizadas com o análogo de tetraciclina aTc (anidrotetraciclina) por sua maior estabilidade em cultura.

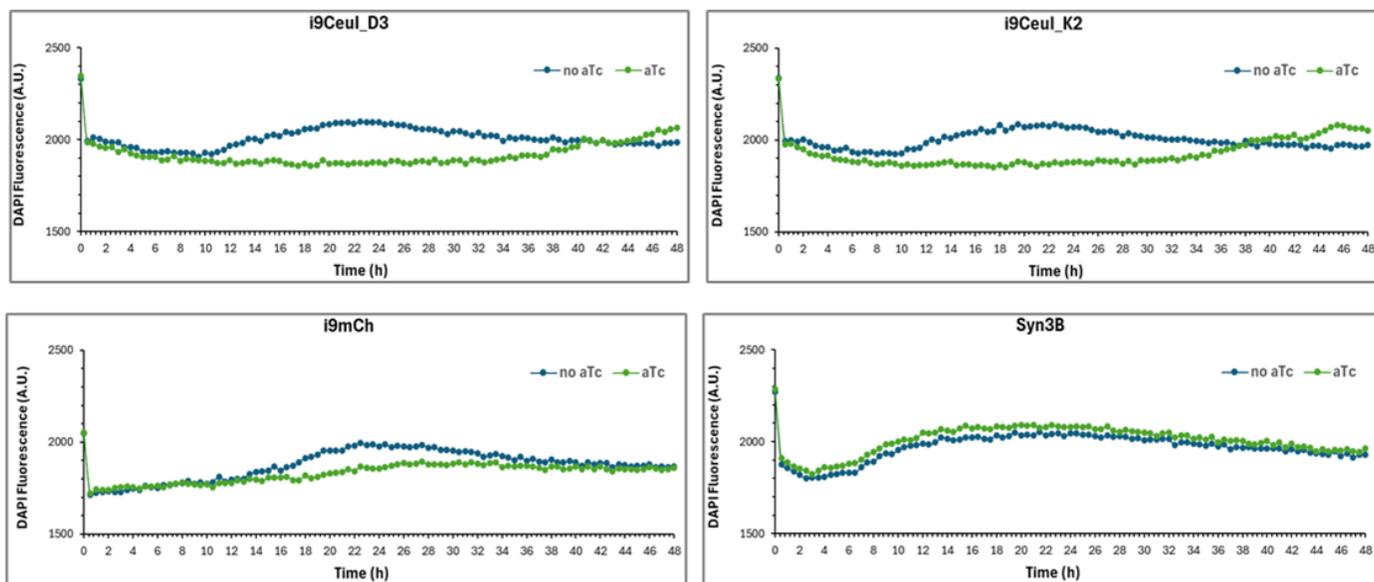


FIGURA 12. Curvas de crescimento com marcação de DNA por DAPI de linhagens contendo a endonuclease I-CeuI com expressão controlada por Int9. O acúmulo de DNA celular foi acompanhado por 48 horas em leitor de placa (Excitação: 360 nm, Emissão: 460 nm).

Como podemos observar, a presença de aTc não interfere no crescimento de Syn3B. Já as curvas obtidas com os clones Int9/I-CeuI apresentam uma clara queda, sem a retomada exponencial observada nos grupos não induzidos, possivelmente em decorrência da degradação do genoma e consequente morte celular das bactérias. A recuperação após aproximadamente 30 h de incubação está provavelmente relacionada à já reportada taxa de escape, com retomada de crescimento por parte de células sobreviventes.

Ainda que haja claramente uma diferença entre as curvas observadas entre as amostras induzidas e não induzidas de Int9/mCherry, a cinética difere do efeito apresentado pelos clones Int9/I-CeuI, se assemelhando mais a um atraso do crescimento do que um efeito deletério. Para melhor avaliar esse comportamento e melhor corroborar a hipótese de que as quedas no acúmulo de DNA obtidas com a indução de I-CeuI se deve ao apropriado funcionamento do sistema e ativação do efeito de endonuclease e não a algum efeito deletério da expressão de Int9, foram realizados o plaqueamento em meio sólido para observação de surgimento de colônias e avaliação de Unidades de Mudança de Cor (CCU) das culturas.

O crescimento em placas de CCU de culturas pré-induzidas e controles mostra que já nas primeiras 24h de incubação os clones da linhagem Int9/I-CeuI não apresentam crescimento em nenhum dos pontos de diluição quando tratados com aTc, enquanto as culturas das linhagens

selvagem e Int9/mCherry não apresentam diferenças na indicação de crescimento entre os grupos controle e tratado . Este comportamento se torna ainda mais evidente após 48h de incubação das placas, onde as linhagens controle não demonstram qualquer diferença no número de pontos de diluição seriada com alteração de cor indicativa de crescimento (7 pontos de diluição cada) entre grupos induzidos e não induzidos, enquanto clones Int9/I-CeuI induzidos apresentam crescimento em até apenas 4 pontos de diluição (**Figura 13A**). O entendimento da alteração nas curvas de crescimento do controle Int9/mCherry observada como sendo efeito apenas de um atraso na taxa de crescimento também fica indicado na (**Figura 13B**) com a formação abundante de colônias em meio sólido após plaqueamento de culturas induzidas, diferentemente do observado após indução e plaqueamento dos clones Int9/I-CeuI, onde não foi possível detectar formação de colônias.

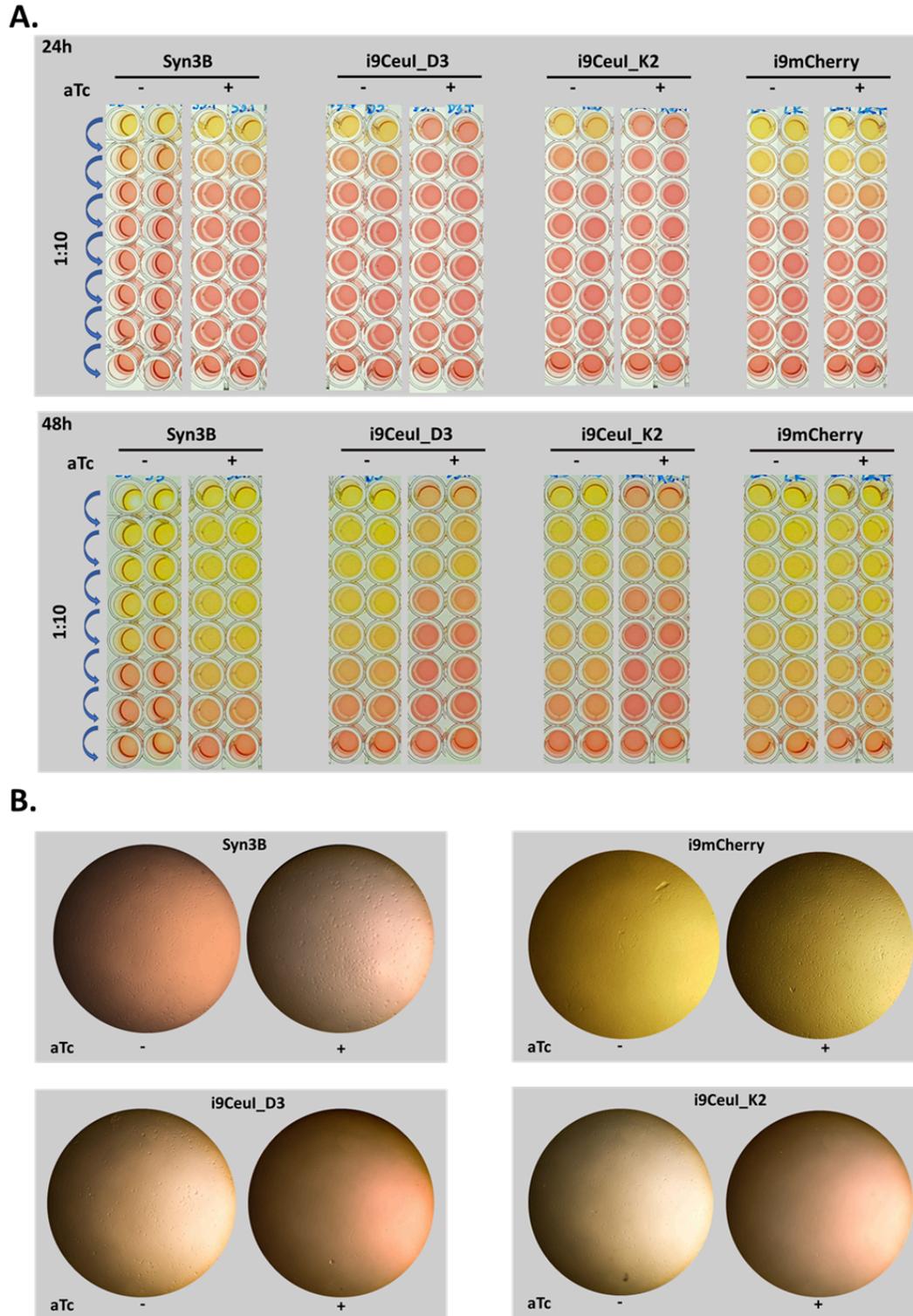


FIGURA 13. A) Ensaios *CCU* de linhagens carregando a endonuclease I-CeuI com expressão controlada por Int9 após indução por tetraciclina. B) Plaqueamento dos mesmos tratamentos para acompanhamento de formação de colônias em meio SP4 sólido.

Apesar de os resultados iniciais obtidos apontarem o funcionamento desejado do sistema, replicatas adicionais ainda são necessárias para a apropriada caracterização do interruptor Int9/I-CeuI e produção de *SimCells* derivadas de Syn3B. Ainda assim, eles apontam na direção de um sistema robusto de regulação transcricional da nuclease. Ensaio anteriormente realizados por membros do grupo com sistemas utilizando apenas o promotor induzível TetR-Pxyl/TetO2 como única camada regulatória de I-CeuI indicavam problemas possivelmente provenientes dos níveis de vazamento mesmo sem a presença do indutor, condição já descrita durante o estabelecimento do sistema de silenciamento por CRISPRi nas células JCVI-syn⁹⁰. Ao que tudo indica, a expressão de I-CeuI em condições não-indutoras leva a uma seleção negativa precoce do sistema, favorecendo o predomínio de escapes nas culturas.

A utilização do sistema Int9/I-CeuI parece permitir contornar essa questão. No entanto, a seleção das *SimCells* formadas ou seleção negativa dos escapes ainda se coloca como um desafio. Uma possível alternativa a ser testada é o acoplar ao sistema Int9/I-CeuI o sistema CRISPRi de forma que, após ação da nuclease e degradação do genoma, rodadas de ativação do sistema de silenciamento tendo como alvo genes essenciais levaria à morte das células que eventualmente ainda apresentassem o genoma intacto, funcionando assim o sistema CRISPRi como um *killswitch* responsável pela seleção negativa dos eventos de escape.

O principal gargalo para o acoplamento dos dois sistemas é a ausência de outros promotores induzíveis funcionais caracterizados nas diversas linhagens de *M. mycoides* JCVI-Syn além do promotor de tetraciclina atualmente usado. Nesse sentido nos propusemos a identificar novos promotores induzíveis funcionais em Syn3B, tendo iniciado as buscas selecionando alguns promotores recentemente caracterizados em *M. pneumoniae*. O promotor pA13 é induzido por L-Arabinose, sendo testado na presença do regulador AraR com ou sem adição da permease codificada por AraE, enquanto os promotores pLG64 e pL52b62, induzíveis por IPTG, foram testados juntamente com o fator de transcrição LacI4. As construções desenhadas para os testes carregam os genes de cada fator de transcrição sob controle de um promotor constitutivo, enquanto o promotor induzível a ser testado regula a expressão do gene repórter *mcherry*. A avaliação foi realizada utilizando-se três transformantes independentes para cada combinação de promotores e efetores, e três concentrações do respectivo indutor, bem como um grupo controle não induzido. Como podemos observar na **Figura 14A**, os resultados obtidos indicam comportamentos opostos dos dois promotores induzidos por IPTG, ainda que ambos indesejáveis: enquanto a utilização do promotor pLG64 não promoveu a expressão de mCherry independentemente da concentração de indutor utilizada, o uso do promotor pL52b62 resultou em expressão semelhante do repórter em todos os grupos, mesmo na ausência de IPTG. Já o promotor Lac pA13, induzido por L-Arabinose, apresentou melhor comportamento, especialmente na presença de AraE (**Figura 14B**). No entanto, ainda que o comportamento com padrão dose-resposta obtido com o aumento das concentrações do indutor seja um bom indicativo de funcionalidade, os níveis de fluorescência registrados ainda foram

consideravelmente mais baixos do que os resultados obtidos com o já previamente utilizado promotor TetR-PxyI/TetO2, induzido por tetraciclina (**Figura 14C**).

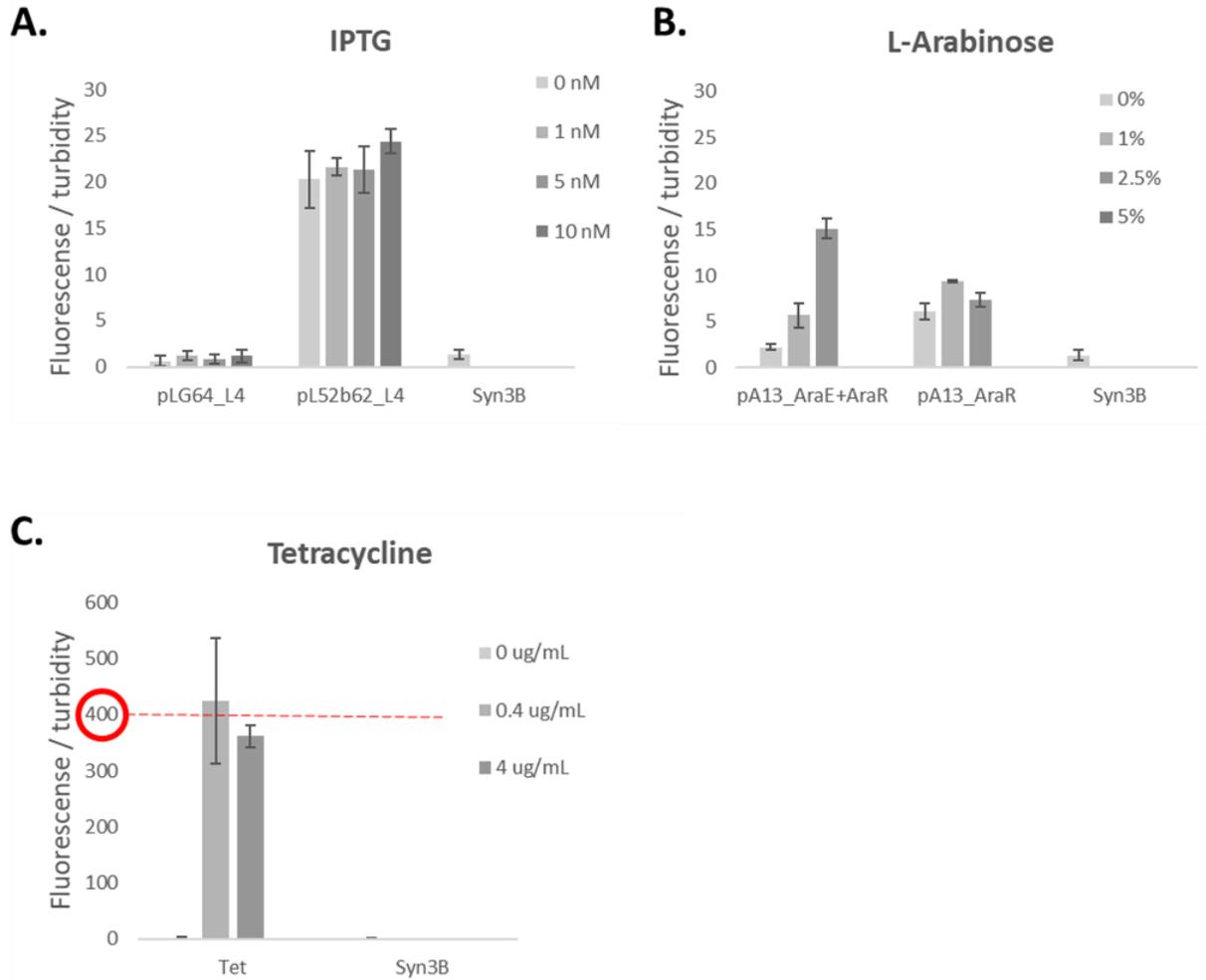


FIGURA 14. Quantificação de expressão do repórter *mcherry* sobre regulação de diferentes promotores induzíveis na presença de variadas concentrações dos respectivos indutores. A) Níveis de fluorescência de mCherry sob controle transcricional de promotores induzíveis por IPTG. B) Níveis de fluorescência de mCherry sob controle transcricional de promotores induzíveis por L-Arabinose. C) Níveis de fluorescência de mCherry sob controle transcricional de promotor induzível por Tetraciclina.

É importante considerar que a transição entre espécies para uso de promotores induzíveis normalmente demanda ajustes nas concentrações de indutores usadas. A ativação do promotor induzido por tetraciclina utilizado, por exemplo, demanda concentração de 400 ng/mL para indução na célula mínima, ainda que, em *M. pneumoniae*, concentrações de até 10 ng/mL já sejam capazes de ativar a expressão com esse promotor⁹⁰. No entanto, especialmente no caso da construção mais promissora pA13_AraR+AraE, o aumento da concentração de indutor representa

um desafio talvez inatingível, uma vez que, ao menos nas condições testadas, 5% de indutor no meio SP4 já foi o suficiente para inibição de crescimento das células.

Nesse contexto, a utilização de promotores induzíveis por fatores outros que não químicos, como temperatura ou comprimentos de onda de luz específicos, pode ser uma boa alternativa para tentar contornar a inibição de crescimento causada por altas concentrações de indutores químicos.

É importante considerar também que, mesmo levando a níveis de expressão mais baixos no sistema, promotores como pA13 poderiam ser utilizados na regulação da expressão do gene de integrase valendo-se do princípio de cascata de amplificação de sinal, uma vez que a inversão de um gene pela integrase só necessitaria ocorrer uma única vez no genoma de uma célula para ativá-lo e levar à produção da proteína de interesse. Após a inversão os níveis de proteína produzidos dependeriam apenas das taxas de expressão permitidas por seu próprio promotor, tornando possível o uso de promotores mais fracos na construção de um sistema de seleção negativa dos escapes observados durante a geração de *SimCells*.

Conclusões e Perspectivas

A proposta deste trabalho de utilização de integrases no sistema sintético representado pela célula mínima *M. mycoides* JCVI-syn3B visa principalmente ampliar suas possibilidades de estudo e aplicação no campo da biologia sintética por meio do desenvolvimento de novas ferramentas moleculares com plasticidade funcional para tanto contornar dificuldades práticas enfrentadas no uso de metodologias importantes na edição genética, quanto permitir a criação de vias regulatórias sintéticas baseadas em circuitos genéticos em um sistema em que grande parte dos componentes moleculares originalmente presentes foram eliminados, o que potencializa a aplicação de tais circuitos como parte constituinte de uma célula verdadeiramente sintética no futuro. Os esforços aqui apresentados também são de grande relevância por constituírem os primeiros trabalhos envolvendo a célula mínima JCVI-Syn3B realizados no país.

Tendo testado inicialmente em Syn3B duas das integrases de melhor funcionamento utilizadas em nosso grupo, Int9 se mostrou capaz de controlar a expressão do gene repórter integrado ao genoma mínimo. No interruptor genético construído, a adição de Tetraciclina como indutor externo ativa a expressão da integrase efetora, que por sua vez age invertendo a sequência silenciada do gene alvo através da recombinação de seus sítios de reconhecimento inseridos nas extremidades do gene. Essa inversão do DNA alinha a orientação da sequência codificadora com a de seu promotor e permite a sua transcrição. Com a avaliação de diversos dos grupos controles utilizados, uma das importantes conclusões observadas é a robustez do sistema, capaz de operar com concentrações reduzidas do indutor, mas ainda sim apresentando um escape mínimo em condições não induzidas. Esse resultado favorece o uso da ferramenta no sistema, uma vez que dentre as poucas ferramentas já disponíveis para uso na célula mínima, algumas importantes como CRISPRi e o próprio promotor induzível por Tetraciclina apresentam taxas de escape que podem prejudicar sua utilização. O sucesso na inversão da sequência de DNA alvo

sem a necessidade de outros componentes celulares auxiliares também reforça a importância do uso de Int9 no sistema, uma vez que o uso de CRISPR-Cas9 na célula, a principal ferramenta atualmente utilizada na edição de genomas, é impedido pela dependência de sistemas de reparo e recombinação homóloga ausentes em Syn3B.

Essas características observadas no interruptor controlado por Int9 também o tornaram um interessante candidato para otimização do controle de expressão da nuclease I-CeuI para a produção de *SimCells*. Ensaios anteriores realizados pelo grupo com a nuclease vinham enfrentando dificuldades pois, por se tratar de um sistema sensível, o vazamento transcricional observado quando utilizado somente o promotor TetR-Pxyl/TetO2 interfere fortemente no controle da ação da nuclease, favorecendo a seleção de células imunes à degradação e perda do gene na cultura. Os resultados iniciais obtidos até o momento com a utilização do interruptor Int9 para controle da expressão de I-CeuI indicam o sucesso do sistema.

Uma segunda etapa no processo de produção de *SimCells* originadas de Syn3B com o uso do interruptor ativado por Int9 envolve a seleção das células que tiveram seu genoma degradado. Uma forma proposta para alcançar tal resultado é a utilização de mecanismos de contenção biológica baseados em CRISPRi para realizar a seleção negativa em decorrência do silenciamento de genes essenciais em células que sobreviverem à ativação da nuclease. Uma segunda alternativa de seleção negativa seria a indução de expressão de um repórter fluorescente nas células escape e posterior exclusão destas com uso de citometria de fluxo acoplado a sistema de seleção direcionada. Em ambos cenários, uma segunda classe de promotor induzível será necessária, uma vez que a expressão de dCas9 para ativação do sistema CRISPRi ou do gene repórter deve ocorrer após a ativação de I-CeuI. Os resultados iniciais obtidos em Syn3B com promotores induzíveis previamente caracterizados em *M. pneumoniae* demonstram o potencial de utilização pelo menos do promotor pA13 induzido por L-Arabinose, porém investigações incluindo otimização das condições de indução ainda precisam ser realizadas a fim de se construir o sistema de seleção negativa para produção de *SimCells*.

Uma vez que a construção de circuitos complexos é uma possibilidade com a concatenação de múltiplas integrases e seus respectivos sítios *att* e tais circuitos seriam de grande interesse em aplicações dentro da biologia sintética, esforços em busca da caracterização e identificação de integrases adicionais funcionais em Syn3B deverão ser priorizados nas etapas futuras. Como candidatos se destacam as demais serina-integrases já testadas pelo grupo em modelos animais e vegetais. Seu teste em Syn1.0 também deverá ser realizado visando expandir a aplicação dessas recombinases em ensaios de genômica funcional, uma vez que a inserção de sítios *att* de uma mesma integrase flanqueando genes diversos permitirá, por exemplo, o silenciamento simultâneo de grupos de genes em uma mesma célula. Tais ensaios de perda de função em grupo deverão enriquecer os estudos de essencialidade já realizados e poderão contribuir para a otimização do processo de minimização do genoma de Syn1.0, uma vez que as metodologias de silenciamento por bombardeamento de transposon utilizadas permitem apenas a avaliação de perda de função de um gene por vez em uma célula.

Capítulo II

Desenvolvimento da plataforma Int-Plex@: interruptor de memória genética bimodular controlado pelas serina-integrases Int9, Int13, phiC31 e Bxb1 capaz de regulação transcricional por meio de inversão ou excisão de sequências integradas ao genoma de *Nicotiana benthamiana*.

Introdução

Ao longo das últimas décadas, grande parte dos avanços de maior impacto em Biologia Sintética de Plantas⁹⁸ surgiram do aprofundamento no entendimento das questões envolvendo os processos de coevolução entre vírus e seus hospedeiros e o consequente desenvolvimento de ferramentas biotecnológicas a partir dos aparatos moleculares selecionados por milhões de anos dessa corrida armamentista. Alguns exemplos dessas ferramentas biotecnológicas baseadas em uma abordagem *bottom-up* da biologia sintética incluem (I) endonucleases de restrição bacterianas e DNA ligases oriundas de fagos, que abriram os caminhos para a revolução das tecnologias do DNA recombinante utilizadas atualmente em metodologias de montagem de *BioBricks* e construções complexas de DNA em uma única reação (do inglês *one-step assembly*); (II) Serina-Integrases⁹⁹ (Ints) derivadas de fagos filamentosos, recombinases multifuncionais com potencial de reorganizar sequências de DNA a partir da recombinação de sítios pequenos pré-definidos, servindo também como ferramentas de montagem *in vitro* de moléculas de DNA como plasmídeos; (III) a tecnologia do RNA interferente (RNAi)^{100,101} amplamente utilizada em estudos de perda de função; (IV) Sistema CRISPR-Cas¹⁰², a tecnologia mais utilizada atualmente na edição gênica de plantas. Uma grande desvantagem na utilização das duas últimas metodologias mencionadas é o caráter permanente das alterações por elas causadas, uma vez silenciado ou editado um gene não é possível retornar ao seu estado original. Nesse contexto ganha destaque a aplicação de serina-integrases na superação desse gargalo, uma vez que seu uso permite a construção de aparatos genéticos modulares baseados em sistemas de portas lógicas e lógica booleana a fim de concatenar diferentes integrases efetoras ortogonais cuja ação em cascata possa culminar em edições e posterior regresso ao estado original de um gene de interesse^{29,32}. Tal circuito genético permitiria a modulação dinâmica e reversível de seções do genoma de organismos escolhidos. Provenientes de bacteriófagos, Ints são recombinases capazes de catalizar de maneira direta e unidirecional eventos sítio-específicos de inserção, deleção ou inversão 180° de sequências de DNA flanqueadas por sítios de reconhecimento denominados sítios *attB* e *attP*, sendo o tipo de reorganização resultante determinado pela orientação

direcional de um sítio em relação a seu par¹¹. A recombinação dos sítios gera dois novos sítios recombinantes denominados sítios *attL* e *attR*.

Visando aumentar o número de Ints disponíveis como ferramentas moleculares, Yang e colaboradores²⁹ desenvolveram abordagens de bioinformática que permitiram a identificação de novas 34 Ints e respectivos sítios *att* em bancos de dados, sendo em seguida 11 deles funcionalmente caracterizadas em *E. coli*. Uma vez que a limitação de Ints funcionais caracterizadas é ainda mais proeminente em sistemas eucarióticos, Gomide e colaboradores⁴⁰ avaliaram a atividade das novas integrases em células animais e vegetais, identificando Int9 e Int13 dentre as de maior eficiência na inversão da sequência codificadora de eGFP em células de planta em um contexto de transformação transiente de protoplastos. Neste trabalho é proposta e avaliada a construção do Interruptor Genético de Memória Int-Plex@ (INTEGRase PLant EXpression), um sistema formado por dois módulos funcionais e quatro efetores constituído pela sequência reverso complementar do gene *mgf* flanqueado por um arranjo dos quatro sítios *attB* e outro dos quatro sítios *attP* das integrases phiC31, Bxb1, Int13 e Int9 imediatamente à montante e à jusante do gene, respectivamente. A construção, integrada ao genoma de *Nicotiana benthamiana*, responde à ação das respectivas integrases como inputs de sinal responsáveis pelo output (edição do genoma) definido pelo usuário: ativação ou inativação por inversão do gene de interesse ou excisão do gene, podendo ser entendido como um circuito de portas lógicas. O módulo de excisão, ativado pelos inputs phiC31 ou Bxb1, é desenhado como uma porta NOR, onde a presença de qualquer um dos inputs impede a expressão do gene de interesse, uma vez que este será excisado do genoma. Já o módulo de inversão pode ser descrito como uma porta XOR onde apenas um dos inputs pode estar presente para que haja ativação do sinal. Caso Int9 e Int13 participem ativamente no sistema, independentemente da ordem ou simultaneamente, o sistema regressará ao seu estado original inativado. A combinação de estados dos dois módulos dita o comportamento final do interruptor com um todo, forçando-os a se comportar como dois inputs de uma porta final AND a fim de que seja obtida a expressão do gene alvo (**Figuras 15A e 15B**).

Objetivo Geral

Desenvolvimento da plataforma Int-Plex@ (*Integrase - Plant Expression*), um interruptor de memória genética bimodular controlado pelas serina-integrases Int9, Int13, phiC31 e Bxb1 capaz de modular expressão de genes alvo através de inversão 180° reversível de sua sequência codificadora flanqueada por sítios *att* ou excisão do cassete integrado ao genoma de *N. benthamiana*.

Objetivos específicos

- Construção dos interruptores genéticos constituintes da plataforma Int-Plex@;
- Teste inicial dos circuitos genéticos em sistema de expressão *in vitro cell-free* TxTI;

- Geração da linhagem transgênica *N. benthamiana* - IntPlex@;
- Cultivo e seleção de eventos transformantes positivos de *N. benthamiana* - IntPlex@, com caracterização das inserções obtidas;
- Agroinfiltração de eventos positivos para entrega dos genes efetores ao sistema.
- Extração de DNA genômico de folhas infiltradas para detecção molecular de edições decorrentes de ativação do sistema.
- Avaliação funcional do módulo de inversão por meio de PCR, sequenciamento de DNA e microscopia confocal de folhas agroinfiltradas.
- Avaliação funcional do módulo de excisão por meio de PCR e sequenciamento de DNA de folhas agroinfiltradas.
- Avaliação da capacidade de disseminação sistêmica com uso de vetor viral PVX por meio de PCR e sequenciamento de DNA de folhas apicais não infiltradas 21 dias após inoculação original.

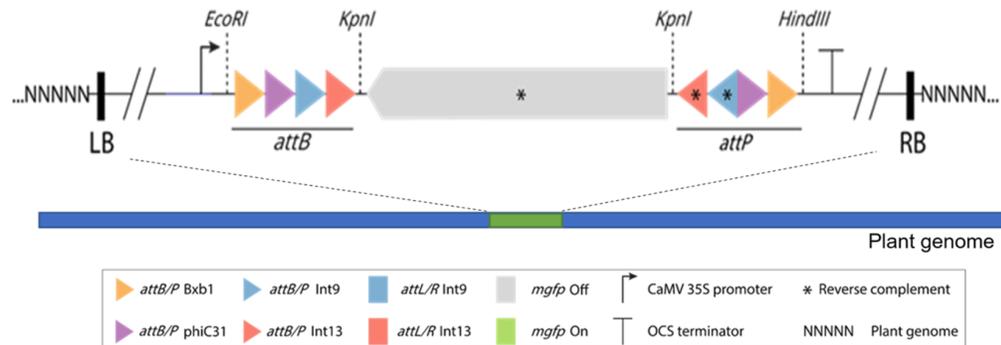
Metodologia

Design e construção de plasmídeos

Os vetores utilizados no trabalho podem ser divididos em dois grupos, uma construção repórter a ser integrada ao genoma de *N. benthamiana* e as construções efetoras carregando os genes de integrase.

Plasmídeo repórter. O desenho da unidade repórter do interruptor genético construído consiste na sequência reverso-complementar do gene fluorescente *mgfp*, dessa forma silenciado por se encontrar em orientação oposta a de seu promotor, flanqueado pelos sítios *attB* das integrases Bxb1, phiC31, Int9 e Int13 nessa ordem à montante do gene e os sítios *attP* de Bxb1, phiC31, Int9 e Int13 nessa ordem à jusante (**Figura 15A**).

A.



B.

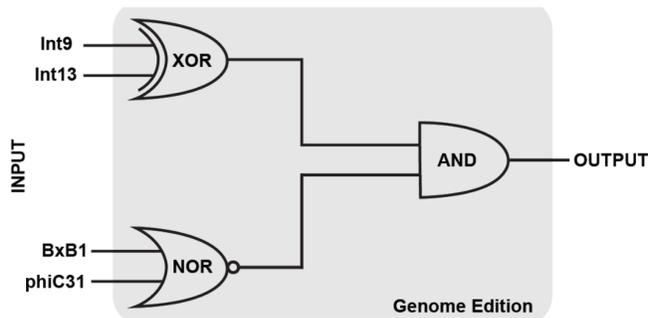


FIGURA 15. A) Mapa da construção repórter pART27_mgfp(rc) inserida no genoma de *N. benthamiana*. LB (left border - borda esquerda) e RB (right border - borda direita) indicam as extremidades do inserto no genoma da planta. “NNNNNNN” ilustra seqüências de DNA genômico no sítio de integração. B) Portas lógicas constituídas pelos módulos de inversão e excisão do sistema Int-Plex@.

Enquanto os sítios *attB* e *attP* de Int9 e Int13 se encontram em orientação oposta entre si, resultando em inversão da seqüência por eles flanqueada quando há recombinação dos sítios pelas respectivas integrase (**Figura 16A**), os sítios *attB* e *attP* mais externos, de Bxb1 e phiC31, encontram-se na mesma orientação. Tal configuração resultará na deleção de todo o cassete por eles flanqueado quando houver a recombinação dos sítios por alguma dessas integrase, removendo a construção do genoma (**Figura 16B**). Outra possibilidade de ação do interruptor é a ativação combinada de mais de uma integrase. A introdução de Int9 no sistema, por exemplo, ativará a expressão de mGFP ao causar inversão de sua seqüência codificadora. Uma posterior ação de Int13 sobre esse cassete ativado ocasionará no retorno ao estado original silenciado da seqüência codificadora, permitindo assim uma regulação temporal da expressão do gene alvo. (**Figura 16C**)

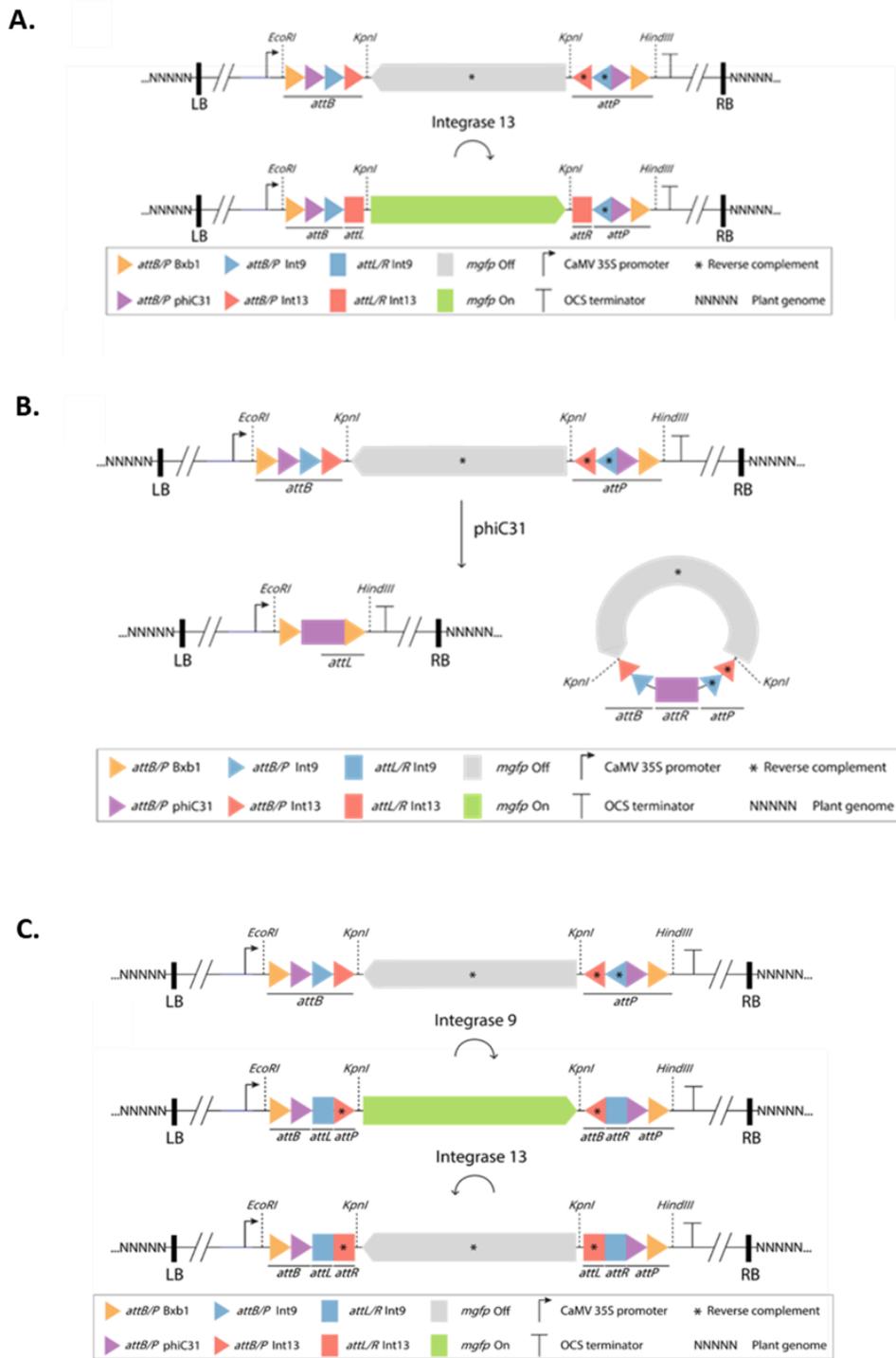


FIGURA 16. Esquema de edições esperadas do cassete integrado pART27_*mgfp*(rc) após ativação de integrases. A) Módulo de inversão editado por Int13. B) Módulo de excisão editado por phiC31. C) Módulo de inversão após edição sequencial por Int9 e Int13. A dupla inversão retorna o gene *mgfp* a seu estado inicial silenciado

Foi realizada síntese química da construção e o fragmento gerado foi clonado no vetor primário pART7 entre os sítios de restrição de XhoI a fim de completar a unidade transcricional com o promotor CaMV 35S e região terminadora do gene Octopina Sintetase (OCS) presentes no plasmídeo. Uma nova rodada de digestão e subclonagem foi realizada para inserção da unidade transcricional no vetor binário pART27. Este vetor possui o replicon mínimo RK@ para replicação em *E. coli* e *Agrobacterium tumefaciens*, gene Tn7 de resistência a espectinomicina/estreptomicina para seleção de transformantes e, fazendo parte do T-DNA a ser integrado no genoma vegetal, o gene NPTII de resistência a kanamicina para seleção das plantas transgênicas geradas. A construção final foi nomeada pLSB_pART27_mGFP(rc), tendo sido todas as etapas de síntese e montagens realizadas pela empresa EPOCH Life Science Inc (Missouri City, TX, EUA).

Plasmídeos efetores. Os genes das serina-integrases Int9, Int13, PhiC31 e Bxb1 utilizados foram selecionados a partir do trabalho previamente publicado⁴⁰. As unidades transcricionais compostas pelo promotor constitutivo actin2, o gene da integrase e terminador NOS foram recuperados do vetor pUC57 utilizado e subclonados no vetor binário pCAMBIA2300¹⁰³, gerando o conjunto de plasmídeos efetores nomeados pLSB_pCAMBIA2300_X (onde X indica a integrase presente). Os genes das integrases também foram inseridos em vetor binário PVX-GW¹⁰⁴ para criação do conjunto de plasmídeos efetores virais nomeados pLSB_PVX_GW_X (novamente, X = Int9, Int13, PhiC31 ou Bxb1). Versões dos vetores com os mesmos genes de Int fusionados a uma Sequência de Localização Nuclear (NLS) foram construídas a fim de avaliar seu possível impacto na atividade das recombinases, dado que sua atividade deve ser exercida diretamente no material genômico localizado no núcleo das células.

Geração de Plantas Transgênicas

A transformação das plantas para geração de *N. benthamiana* transgênica foi realizada por Martha L. Orozxo-Cardenas, MSc., PhD no Centro de Pesquisa em Transformação Vegetal da Universidade da Califórnia Riverside. Células competentes de *A. tumefaciens* GV3101 foram transformadas por eletroporação com o plasmídeo PLSB_pART27_mgfp(rc) e os transformantes positivos selecionado com espectinomicina e estreptomicina. Os clones resistentes aos antibióticos ainda foram confirmados por PCR de colônia com oligonucleotídeos para detecção de genes de resistência e fragmentos internos do cassete. A linhagem de agrobactéria transformada foi então utilizada na agroinfiltração de folhas de *N. benthamiana* com seis semanas de cultivo. Após infiltração calos e brotos foram induzidas com uso de meio de indução contendo sais MS, Vitaminas MS, 100 mg/L de Myo, 2 mg/L BAP, 0.1 mg/L NAA, 3% sacarose, 100 mg/L kanamicina, 250 mg/L Cefotaxima e 500 mg/L Carbenicilina. Trinta brotos transformados com pLSB_pART27_mgfp(rc) e dez contendo o vetor vazio pART27 foram

transferidos para meio de enraizamento contendo ½ concentração de Sais MS, vitaminas MS, 100 mg/L Myo-inositol, 0.2 mg/L IBA, 1.5% sacarose, 500 mg/L Vancomicina, 500 mg/L Cefotaxima e 500 mg/L Carbenicilina.

As plantas resistentes foram confirmadas por PCR utilizando oligonucleotídeos para detecção dos genes *mgfp* e *nptii*, este último em especial no caso das plantas portadoras do vetor pART27 sem o cassete *att_mgfp(rc)*.

Condições de cultivo e manuseio em casa de vegetação

As sementes de *N. benthamiana* transgênicas recebidas foram germinadas em solo comercial sob condições ambientais de luminosidade, temperatura e umidade controladas em sala de cultivo. Plântulas regeneradas transferidas de meio de cultura *in vitro* para solo foram mantidas sob as mesmas condições controladas. Para os ensaios de agroinfiltração foram utilizadas sempre plantas de geração F2 com oito semanas de crescimento.

Sequenciamento Nanopore e análise de dados

Para a identificação dos pontos de integração do interruptor genético no genoma de *N. benthamiana* foi realizado sequenciamento de genoma completo dos quatro eventos selecionados por meio de tecnologia nanopore (Oxford Nanopore Technologies, UK). Para tal, DNA genômico total das plantas foi obtido por meio de extração pelo método CTAB como descrito acima. O DNA obtido foi avaliado por eletroforese em gel de agarose e quantificação em Quibit com kit 1x dsDNA HS Assay (Invitrogen, USA). Para o preparo das bibliotecas a serem analisadas foi utilizado o protocolo de sequenciamento por ligação para gDNA - Native Barcoding kit 96 V14 (SQK-NBD114-96 - Oxford Nanopore Technologies, UK) de acordo com orientações do fabricante. Brevemente, as amostras foram submetidas a tratamento com NEBNext FFPE Repair Mix (NEB, M6630) e NEBNext Ultra II End repair/dA-tailing Module (NEB, E7546) para reparo de quebras (*nicks*) e preparo de extremidades para posterior ligação de oligos contendo os sistemas de identificação Native Barcodes (NB01-96) por meio dos kits NEB Blunt/TA Ligase Master Mix (NEB, M0367) e NEBNext Quick Ligation Module (NEB, E6056), sendo as etapas de lavagem realizadas por precipitação magnética com esferas AMPure XP beads (ONT, USA).

Para sequenciamento de amplicons a fim de confirmar a edição do genoma após ação das integrases, reações de PCR realizadas conforme descrito anteriormente foram purificadas por meio de uso do kit ReliaPrep DNA Clean-up and Concentration System (Promega, USA). As bibliotecas foram preparadas utilizando o protocolo de sequenciamento por ligação para amplicons - Native Barcoding kit 96 V14 (SQK-NBD114-96) seguindo orientações do fabricante. Neste protocolo também há todas as etapas de preparo de extremidades, ligação de adaptadores e purificação por precipitação magnética conforme descrito no parágrafo acima, sendo uma das principais diferenças a ausência de etapa de reparo de quebras.

Independentemente do método de preparo utilizado, as bibliotecas foram sequenciadas em *flow cell* R10.4.1 FLO-PRO114M (Oxford Nanopore Technologies, UK) na unidade de

sequenciamento PromethION 2 (PRO-SEQ002) com uso do software MinKNOW [v23.11.5]. *Basecalling*, *demultiplexing* e remoção de sequências adaptadoras foram realizados usando modelo de High-accuracy basecalling (HAC) 400bps na versão mais atualizada do *basecaller* Dorado [v7.2.13]. Para análise dos resultados, as sequências FastQ obtidas para cada barcode foram importadas para o software Geneious Prime [2024.0.2] e mapeadas contra as sequências referência apropriadas a cada caso utilizando a ferramenta de mapeamento Minimap2 [2.24].

Agroinfiltração

Células de *A. tumefaciens* GV3101 foram transformadas com os vetores binários pLSB_pCAMBIA2300_X ou pLSB_PVX_GW_X por eletroporação e selecionadas em meio LB sólido contendo 50 µg/mL de kanamicina e 15 µg/mL gentamicina. Após dois dias de incubação, conjuntos de colônias eram recuperados das placas e inoculados em 2 mL de meio LB líquido e crescidas *overnight* em agitador orbital a 28°C, 200 rpm. As células foram recuperadas por centrifugação a 4000 RCF por 15 min a temperatura ambiente e ressuspensas em Tampão de Infiltração 1x MMA (10 mM MES - pH 5.6, 10 mM MgCl₂, 180 µM acetoseringona) em densidade OD₆₀₀= 0.3. Células contendo os plasmídeos com cada integrase e o gene 35S:p19 foram misturadas em volumes iguais para formulação da suspensão de infiltração. Folhas de plantas com 8 semanas de cultivo foram então inoculadas com a suspensão de agrobactéria por meio de seringa sem agulha pressionada diretamente em sua superfície. Testes de co-infiltração contendo ou não transformantes com o gene 35S:p19 foram realizados sem observação de alterações no resultado. Infiltrações com cada uma das integrases foram realizadas em diferentes folhas de uma mesma planta, sendo feitas duplicatas da infiltração em cada planta, com pelo menos 3 plantas por ensaio independente.

Extração de DNA e PCR

A extração total de DNA genômico foi realizada de acordo com o protocolo modificado do método CTAB estabelecido por Lacorte e colaboradores () com adaptações. Secções com aproximadamente 20 mg foram obtidas das folhas tratadas ou grupos controles 5 dias após infiltração, armazenadas em tubos de centrifugação de 1.5 mL e imediatamente congeladas em nitrogênio líquido. As amostras congeladas foram então maceradas com auxílio de pistilos de vidro dentro dos tubos até atingirem um estado de pó fino, ao qual foi adicionado 200 µL de CTAB dando sequência ao protocolo.

As amostras de DNA obtidas foram então submetidas a reações de amplificação por PCR. Para detecção molecular da flipagem do gene repórter *mgfp* foram utilizados os pares de oligonucleotídeos 35S_282F + *attR*_Int9_R e *attL*_Int9_F + IDP_OCSt_349_Rv para identificação dos sítios recombinados *attL* e *attR* de Int9 e os pares de oligonucleotídeos 35S_282F + *attR*_Int13_R e *attL*_Int13_F + IDP_OCSt_349_Rv para identificação dos sítios recombinados *attL* e *attR* de Int13. Para detecção dos fragmentos OnOff resultantes de dupla

edição Int9/Int3 foram utilizados os pares de oligonucleotídeos 35S_282short_Fw + OnOff_attR9_core_Rv, OnOff_attL9_core_Fw + IDP_OCSt_349_Rv, 35S_282short_Fw + OnOff_attL13_core_Rv, OnOff_attR13_core_Fw + IDP_OCSt_349_Rv. Já a detecção molecular da deleção do cassete realizada por phiC31 ou Bxb1 foi utilizado o par 35S_282F + IDP_OCSt_349_Rv para amplificação da sequência mantida no genoma, e o par NbDel_mGFP_493_Fw + NbDel_mGFP_300_Rv para detecção da molécula circular excisada contendo o gene *mgfp* e os sítios *att* removidos.

Para avaliação das recombinações em sistema TxTl foram utilizados os oligonucleotídeos 2041GbA_ck_5p_FW + deGFP_Hd_Ws_Rv para detecção de inversões, 2041GbA_ck_5p_FW + 2041_ck_3p_RV1 para as deleções e o par deGFP_circ_Cw2 + deGFP_circ_Cw2 para amplificação das moléculas circular excisada. As sequências dos oligonucleotídeos utilizados podem ser encontradas na **Tabela 4**.

TABELA 4. Oligonucleotídeos utilizados na detecção das edições realizadas no cassete repórter pART27_*mgfp*(rc) ou pART27_*degfp*(rc)

Oligonucleotídeo	Sequência
35S_282F	attgatgtgatatctccactgacgtaagggatgacgcac
35S_282short_Fw	tgacgtaagggatgacgcac
IDP_OCSt_349_Rv	taggtttgaccggttctgcc
<i>attL</i> _Int9_F	ataattggcgaacgaggtatctgcatagttattccgaac
<i>attR</i> _Int9_R	tggaaagtgtgtatcaggttaactggataacctcatc
<i>attL</i> _Int13_F	tccagatccagttgttttagtaacataaataca
<i>attR</i> _Int13_R	gtagaacttgaccagttggctctgtaaatataagcaatcc
OnOff_attL9_core_Fw	ttggcgaacgaggtatctgc
OnOff_attR9_core_Rv	gaagtgtgtatcaggttaactgg
OnOff_attL13_core_Rv	gttttccagatccagttgtt
OnOff_attR13_core_Fw	tgtagaacttgaccagttggt
NbDel_mGFP_493_Fw	tcaagaccgccacaacatcg
NbDel_mGFP_300_Rv	gaagaagatggctcctctctgc
2041GbA_ck_5p_FW	atgacccaaatcccttaacgtgag
deGFP_Hd_Ws_Rv	cccgggaagcttgatcccggcgcgcg

2041_ck_3p_RV1	tggettaactatgcggcatcag
deGFP_circ_Cw2	ccttgatgccgttcttctgcttg
deGFP_circ_Ccw2	cgacaaccactacctgagcac

As Reações de amplificação foram realizadas em Termociclador Veriti 96-well Fast (Applied Biosystems, CA, EUA) com uso de 1U da enzima Platinum Taq DNA Polymerase (Invitrogen) e mix de reação composto por 2.5 µL de Tampão de PCR 10X (- MgCl₂) e 1.5 mM MgCl₂ fornecidos juntamente com a polimerase, 0.4 mM mix de dNTPs, 0.2 uM de cada oligo e 1 µL de DNA para um volume final de 25 µL. Condições de ciclagem consistiram em uma incubação inicial a 94°C por 2 min seguida de 40 ciclos compostos por desnaturação a 94°C, 30s; Anelamento em temperatura apropriada a cada par de oligos, 30s e Extensão a 72°C na taxa de 1 min/kb do *amplicon* esperado e uma incubação final a 4°C. Uma vez que o par de oligos utilizado na detecção da deleção por phiC31 ou Bxb1 se liga e amplifica tanto o cassete completo quanto o deletado, e levando em consideração a coexistência de ambas as formas na amostra de DNA avaliada, o tempo da etapa de extensão nas reações de PCR foi reduzido para 30s com intuito de priorizar a amplificação do fragmento menor.

Os fragmentos obtidos foram separados em gel de agarose 1% corado com SYBR Safe DNA Gel Stain (Invitrogen) e as bandas excisadas com auxílio de transluminador de luz LED azul a fim de evitar danos causado às moléculas de DNA quando da utilização de luz UV. O DNA recuperado em gel foi purificado com o kit comercial ReliaPrep™ DNA Clean-Up and Concentration System (Promega) e ligado ao vetor de clonagem pGEM-t-Easy de acordo com as instruções do fabricante. Após transformação das ligações em células quimiocompetentes *E. coli* DH10b e seleção em meio sólido das colônias positivas, os plasmídeos foram recuperados por miniprep com kit comercial e enviados para sequenciamento (Macrogen, Seul, KR).

Ensaio de expressão in vitro

Para as avaliações funcionais iniciais foram realizados ensaios de transcrição e tradução *in vitro* com extrato celular livre de células (*cell-free TxTl*) utilizando o kit myTxTl T7 expression (Arbor Biosciences, USA) de acordo com recomendações do fabricante. Reações de 24 µL foram preparadas em tubo de centrifugação de 1,5 mL contendo 10 nM de cada plasmídeo (repórter e efetor) e incubadas a 29°C por 16h em termobloco. Os resultados foram avaliados primeiramente em transluminador de luz azul para detecção de fluorescência do repórter deGFP. Posteriormente 1 µL de cada reação foi recuperado para análise molecular por PCR e sequenciamento como previamente descrito. Como controles foram realizadas reações contendo

apenas o plasmídeo repórter (negativo), apenas o plasmídeo de expressão constitutiva pTXTL-T7p14-deGFP (positivo) e na ausência de plasmídeos (branco).

Microscopia Confocal

As imagens de microscopia confocal foram obtidas com uso do Microscópio Confocal de Varredura a Laser Leica TCS SP8. A detecção de mGFP foi realizada utilizando-se os comprimentos de onda de 488 nm e 500-550 nm para excitação com laser de argônio e filtro *bandpass* para emissão, respectivamente. Sinais oriundos de autofluorescência por cloroplastos foram detectados, em paralelo à aquisição de fluorescência de mGFP, usando filtro *bandpass* de comprimentos de onda 650-710 nm. Os parâmetros de aquisição de imagem (como potência de laser, diâmetro de abertura de *pinhole*, ganho, entre outros), preparo de amostras e período pós-infiltração foram mantidos constantes entre os diversos grupos ao longo de cada ensaio. Os dados brutos foram processados com o software LAS X (Leica Application Suite, www.leica-microsystems.com)

Resultados e Discussão

Os esforços iniciais de aplicação em sistemas eucariontes das novas serina-integrases descritas por Yang e colaboradores²⁹ buscaram ampliar o leque de ferramentas moleculares disponíveis nesses sistemas para a criação de circuitos genéticos capazes de modular a expressão gênica em organismos complexos tanto de origem animal quanto vegetal⁴⁰. No entanto, as avaliações em sistema vegetal se restringiram a transformações transientes de protoplastos obtidos de *Arabidopsis thaliana*, o que, apesar de ainda permitir avaliar o comportamento das integrases em um ambiente molecular eucariótico, remove os obstáculos principais representados pelo núcleo e o acesso à cromatina quando o gene alvo flanqueado pelos sítios de reconhecimento se encontra integrado ao genoma do organismo.

Visando dar continuidade ao trabalho inicial e abordar as questões sobre os obstáculos destacados, um sistema composto por dois módulos funcionais operados pelas integrases de maior eficiência em célula vegetal identificadas no primeiro trabalho (Int9 e Int13) e duas integrases de uso já obstatante descrito plantas (*phiC31* e *Bxb1*) foi construído e integrado ao genoma de *N. benthamiana*. A estrutura geral do cassete repórter consiste na sequência do gene *mgfp* em orientação reverso-complementar por quatro sítios *attB* e quatro sítios *attP* em cada uma das extremidades. A orientação relativa entre cada par de sítios *att* é o que define o efeito resultante da ação de cada uma das integrases e confere o caráter bi-modular do sistema. Enquanto os sítios *attB* e *attP* das integrases Int9 e Int13 se encontram em orientação oposta em relação a seu par, resultando na inversão do gene por eles flanqueado, os sítios de *phiC31* e *Bxb1* se apresentam na mesma orientação, o que gera como resultante da atividade de recombinação

dos sítios a deleção da sequência interna. Outra característica importante é a de que, devido ao caráter ortogonal das integrases, isto é, sua especificidade de ação apenas sobre seus sítios cognatos, recombinações sequenciais com Int9 e Int13 resultarão na inversão e ativação de mGFP seguida pelo regresso ao estado original do sistema.

A fim de avaliar a possibilidade de interferência na atividade isolada de cada integrase decorrente da presença dos demais sítios *att*, bem como a capacidade de uma segunda integrase inverter o cassete já previamente editado por outra integrase no caso do módulo On/Off, foram inicialmente realizadas reações em sistema de expressão *in vitro* (TxTI) com o kit myTxTI (Arbor Bioscience, MI, EUA) utilizando os vetores p2041_pART27_*degfp*(rc) e p2041_Int. Após 24h de incubação das reações a 29°C os tubos foram visualizados em câmara de luz azul para detecção de fluorescência de deGFP. Como pode-se observar na **Figura 17A**, foi possível identificar a expressão de GFP na presença tanto de Int9 quanto de Int13. Nas reações com co-incubação de p2041_pART27_*degfp*(rc) com os dois vetores p2041_Int9 e p2041_Int13 também foi detectada a expressão de GFP, ainda que o esperado seja o regresso da construção ao estado silenciado. Neste caso a manutenção da fluorescência se deve à grande estabilidade da molécula nas reações TxTI, podendo ser detectada na reação por vários dias. No entanto, a ação combinada das integrases sobre o vetor repórter pode ser comprovada por meio de amplificação e sequenciamento do DNA presente nas reações após incubação (**Figura 17B**). A orientação do gene *degfp* oposta à do promotor e a formação dos dois pares de sítios *attL* e *attR* de Int9 e Int13, com a presença do sítio *attR Int13* à esquerda e não à direita do gene, confirmam a dupla edição On/Off. Todos os rearranjos esperados puderam ser detectados por PCR, como indicado na **Figura 17C**, incluindo a detecção da molécula circular gerada após os eventos de deleção do alvo.

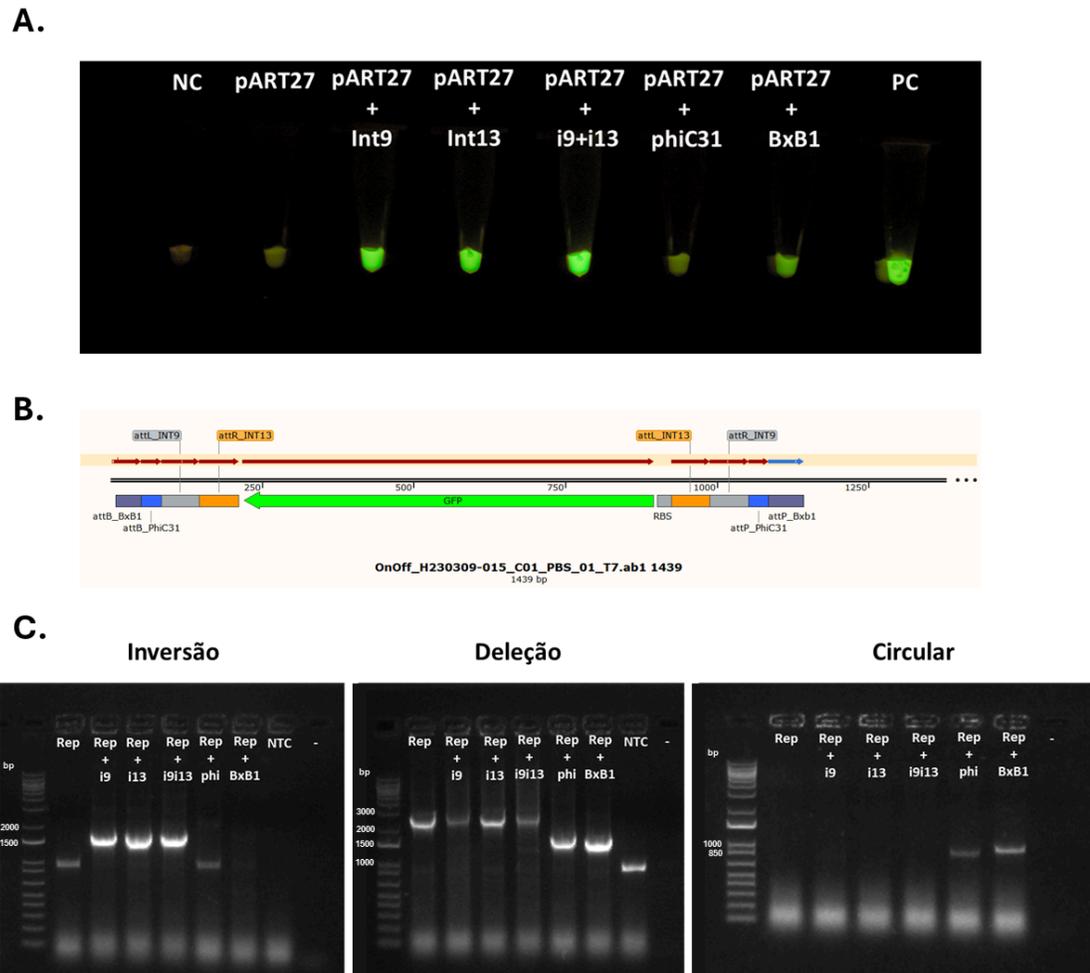


FIGURA 17. Teste em sistema de expressão *in vitro* TxTI do módulo de inversão de Int-Plex@. A) Detecção da expressão de deGFP em transluminador de luz Azul. As legendas indicam os plasmídeos adicionados em cada reação. pART27: vetor repórter não ativado; NC: controle negativo sem DNA adicionado; PC: controle positivo com adição de vetor apresentando expressão constitutiva de *degfp*. B) Confirmação por sequenciamento Sanger de ocorrência de dupla edição do material submetido a sistema TxTI na presença de vetores efetores de Int9 e Int13. C) Amplificações para detecção de todos os eventos de rearranjo observados no sistema TxTI. Rep: repórter (vetor pART27_deGFP); NTC: *no-template control* (Controle negativo sem adição de DNA à per). Marcador de peso molecular: 1kb Plus (invitrogen).

Tendo validado o interruptor em sistema cell-free deu-se início às avaliações *in vivo* com a geração de plantas *N. benthamiana* transgênicas carregando o sistema repórter bi-modular em seu genoma. As plantas recebidas da *facility* de transformação vegetal da Universidade da Califórnia Riverside foram submetidas a processamento em quarentena vegetal de acordo com protocolos de biossegurança estabelecidos pela Unidade de Quarentena da Embrapa Recursos genéticos, sendo posteriormente submetidas a seleção por kanamicina e detecção molecular do cassette *attB_mgfp(rc)_attP* por PCR. Quatro plantas derivadas de eventos de transformação distintos foram confirmadas e selecionadas ao final do processo.

Para a caracterização das inserções do cassete no genoma de *N. benthamiana* foi utilizada a técnica de sequenciamento NGS com tecnologia *nanopore* por meio do sequenciador PromethION. A capacidade de gerar sequências longas em larga escala, permitindo o sequenciamento de genomas completos em um curto espaço de tempo, abre espaço para avaliações diretas de eventos de integração de transgenes de uma maneira mais precisa e extremamente menos trabalhosa que a utilização de metodologias clássicas, como a realização de *southern blot* para confirmação e quantificação de eventos de integração somados a técnicas complexas de PCR dependentes de primers degenerados, como *tail-PCR*¹¹⁵, para detecção da região em que tais interações se deram. As claras vantagens da aplicação de sequenciamento *nanopore* para tal fim têm contribuído para que cada vez mais trabalhos façam uso desta técnica, já tendo sido aplicada na avaliação de transgenes inseridos em genomas de diversas espécies vegetais para caracterização e rastreamento de plantas geneticamente modificadas^{116, 117, 118}.

No presente trabalho, análises iniciais de identificação dos locais de integração do cassete repórter do sistema Int-Plex@ por meio de sequenciamento *nanopore* permitiram o mapeamento de pelo menos um ponto de integração em cada um dos eventos selecionados, com eventos E1 e E15 apresentando integração de duas cópias cada, em regiões distintas do genoma. A **tabela 5** traz informações gerais, qualitativas e quantitativas, a respeito das corridas de sequenciamento realizadas.

TABELA 5. Dados descritivos do sequenciamento *nanopore* realizado para mapeamento de sítios de integração do sistema Int-Plex@

Evento	Reads (M)	Bases (Gb)	N50 (kb)	Qscore	Hits*
E1	1.01	6.12	13,8	15	6
E3	0.87	6.22			5
E10	0.54	3.68			2
E15	1.62	15			18

*para classificação de hits foi estabelecido o critério de *reads* contendo ao menos uma parte do cassete repórter complementada de sequência não pertencente à construção integrada. *Reads* contendo apenas sequências da construção foram desconsiderados.

É possível observar uma maior cobertura obtida no sequenciamento dos eventos E1 e E15, bem como a obtenção de um maior número de sequências contendo uma parte do cassete inserido e pelo menos uma das regiões genômicas flanqueadoras (coluna *Hits*). Este padrão possivelmente apresenta correlação com a identificação de mais de um evento de integração apenas nos indivíduos E1 e E15, sendo necessário complementar as análises para obtenção de maior cobertura. Giraldo e colaboradores¹¹⁷ e Malmberg¹¹⁸ e colaboradores, por exemplo, estimam a necessidade de uma cobertura entre 5 a 10 vezes para uma resolução satisfatória na identificação de transgenes.

Para identificação dos sítios de inserção, inicialmente foram realizados alinhamentos dos dados brutos para mapeamento de *reads* contendo ao menos uma parte do cassete repórter e flancos não alinhados indicativos de sequência genômica. Com essas regiões candidatas foram então realizados alinhamentos contra o genoma de *N. benthamiana* utilizando a ferramenta Blast do próprio bando de dados. Como referência foi selecionada a versão v2.6.1 do genoma de *N. benthamiana* depositado na plataforma SolGenomics Network (<https://solgenomics.net/>), um banco de dados e plataforma dedicada exclusivamente à dados genômicos de solanáceas¹¹⁹. Um último alinhamento local utilizando o mapeador Minimap2 no software Geneious Prime foi realizado diretamente contra os cromossomos identificados durante os Blast para confirmação de colocação de flancos direito e esquerdo de cada sítio de inserção. O mapeamento das inserções até agora identificadas nos eventos de *N. benthamiana* avaliados neste trabalho estão indicados na **Figura 18**.

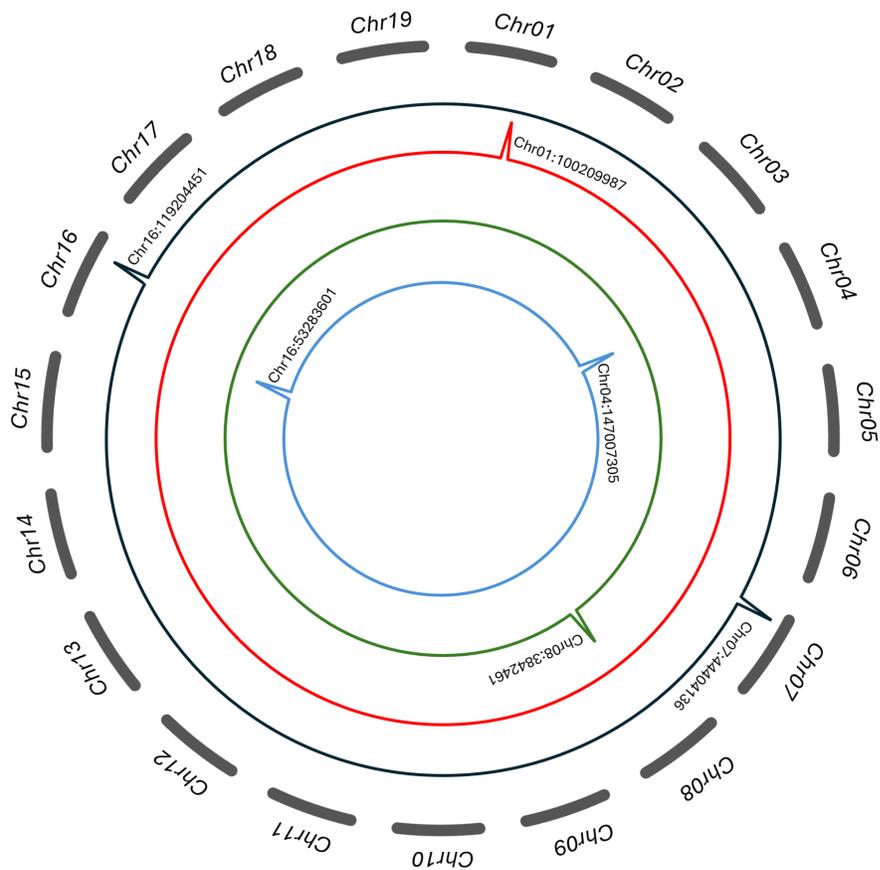


FIGURA 18. Localização dos sítios de integração do sistema Int-Plex@ identificados no genoma de *N. benthamiana* para os quatro eventos independentes avaliados. Os círculos concêntricos coloridos representam o genoma de cada um dos eventos alinhado ao genoma selvagem Niben2.6.1 utilizado como referência. Picos representam a integração. Do mais externo para o mais interno: eventos E1(azul escuro) > E3(vermelho) > E10(verde) > E15(azul claro).

Os ensaios de avaliação funcional do sistema Int-Plex@ foram realizados por meio de agroinfiltração das plantas com células de *A. tumefaciens* transformadas com os vetores efetores pLSB_pCAMBIA2300_X ou pLSB_PVX_GW_X carregando as integrases. Dessa forma a transferência do T-DNA contendo o gene de uma das quatro integrases garantirá sua expressão na célula vegetal e, se bem-sucedida, a edição do sistema repórter presente no genoma, resultando em expressão de mGF ou excisão do cassete a depender do efetor utilizado. No quinto dia pós-infiltração as folhas infiltradas eram recolhidas para remoção de seções destinadas à extração de DNA genômico para detecção por PCR seguida de sequenciamento dos fragmentos do cassete editado ou microscopia confocal de fluorescência. As sequências obtidas após transformação com Int9 ou Int13 confirmam a inversão do gene e ativação da expressão de mGFP, bem como não apontam existência de dano ao DNA ou inserção de mutações no sítio de clivagem e recombinação ou nos demais sítios *att* adjacentes (**Figura 19A** para Int9; **Figura 20A** para Int13).

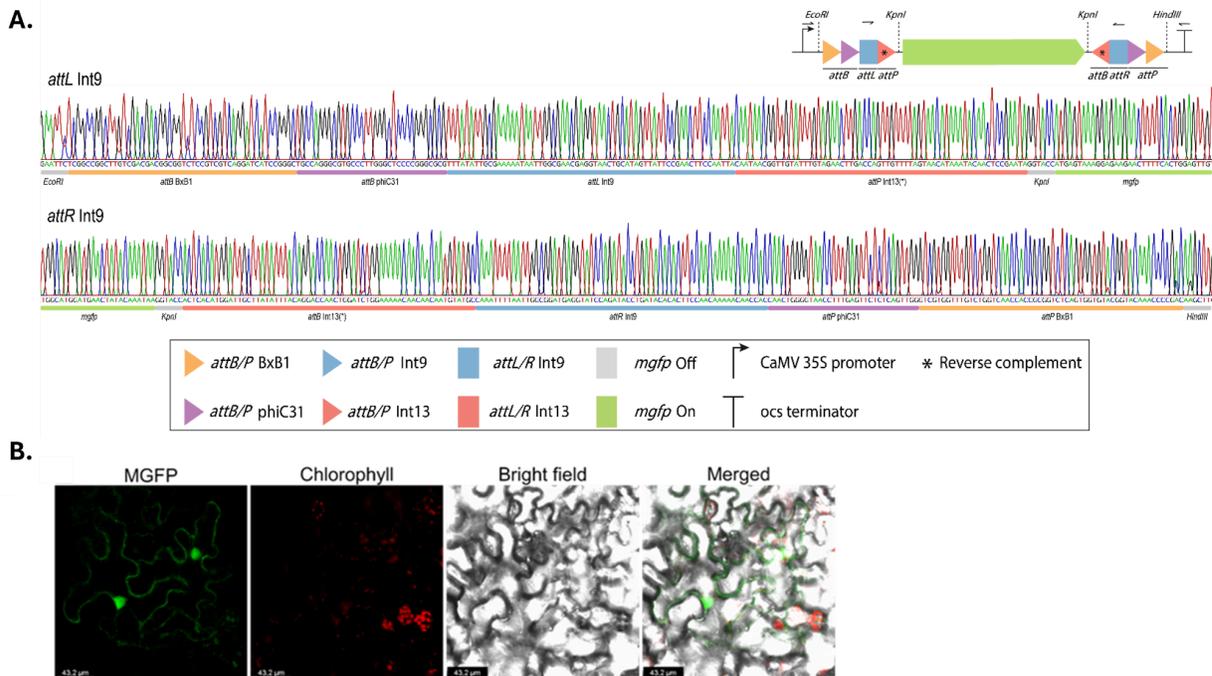


FIGURA 19. Resultados representativos obtidos para um dos eventos avaliados após infiltração com Int9. A) Sequenciamento de fragmentos contendo sítio *attL* Int9 ou sítio *attR* Int9 amplificados após flipagem de *mgfp*. Resultados obtidos por sequenciamento Sanger de *amplicons* purificados. B) Microscopia de fluorescência a laser confocal realizada em região infiltrada mostrando ativação da expressão de mGFP.

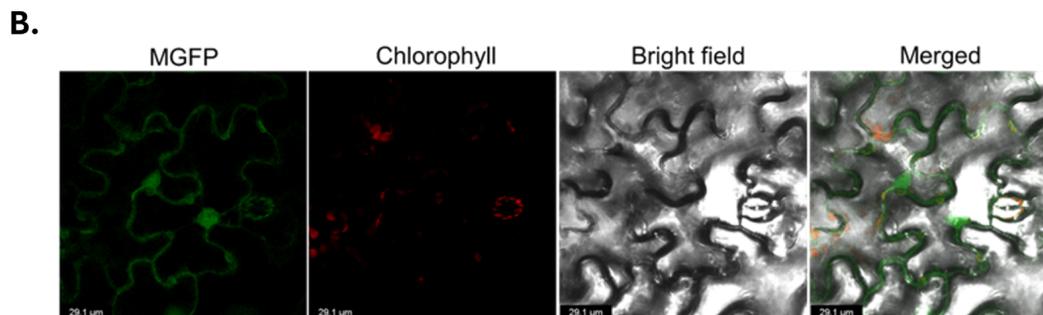
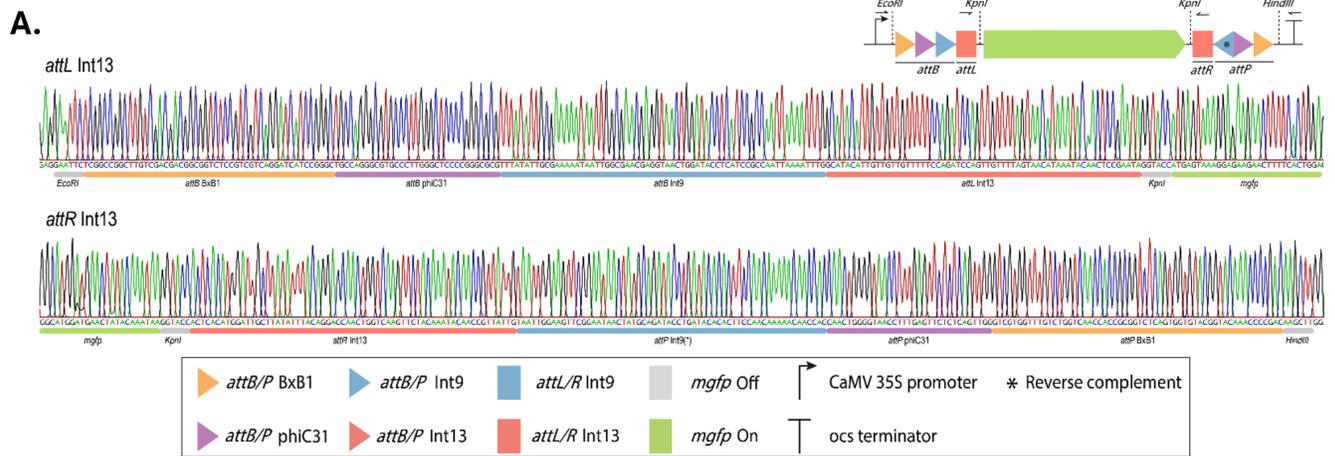


FIGURA 20. Resultados representativos obtidos para um dos eventos avaliados após infiltração com Int13. A) Sequenciamento de fragmentos contendo sítio *attL* Int13 ou sítio *attR* Int13 amplificados após flipagem de *mgfp*. Resultados obtidos por sequenciamento Sanger de *amplicons* purificados. B) Microscopia de fluorescência a laser confocal realizada em região infiltrada mostrando ativação da expressão de mGFP.

As imagens de microscopia confocal permitem observar o surgimento de emissão de fluorescência em decorrência da ativação da expressão de mGFP, porém de maneira heterogênea ao longo da superfície das folhas infiltradas, compatível com taxas de eficiência relatadas previamente por nós e outros trabalhos presentes na literatura (**Figuras 19B e 20B** para Int9 e Int13, respectivamente).

Outro dado significativo obtido é a detecção de inversão e ativação do sistema em folhas distantes das inoculadas passados 21 dias da infiltração com os vetores pLSB_PVX_GW_X. Sendo derivado de vetor viral PVX_GW¹⁰⁴ proveniente do vírus da batata *Potato virus X*, uma das habilidades esperadas com esse método de entrega do DNA efetor é justamente sua

capacidade de movimentação entre as células vegetais, disseminando o gene da integrase para o restante da planta. Após o período de disseminação sistêmica é possível observar o desenvolvimento de sintomas indicativos de infecção viral nas folhas apicais, com o surgimento de mosaico foliar com branqueamento de nervuras, encarquilhamento foliar e má formação de extremidades e folhas jovens (**Figura 21**). A detecção de inversão do gene repórter em plantas apicais não infiltradas reforça a capacidade de disseminação sistêmica e estabilidade da construção efetora quando utilizado um vetor viral.

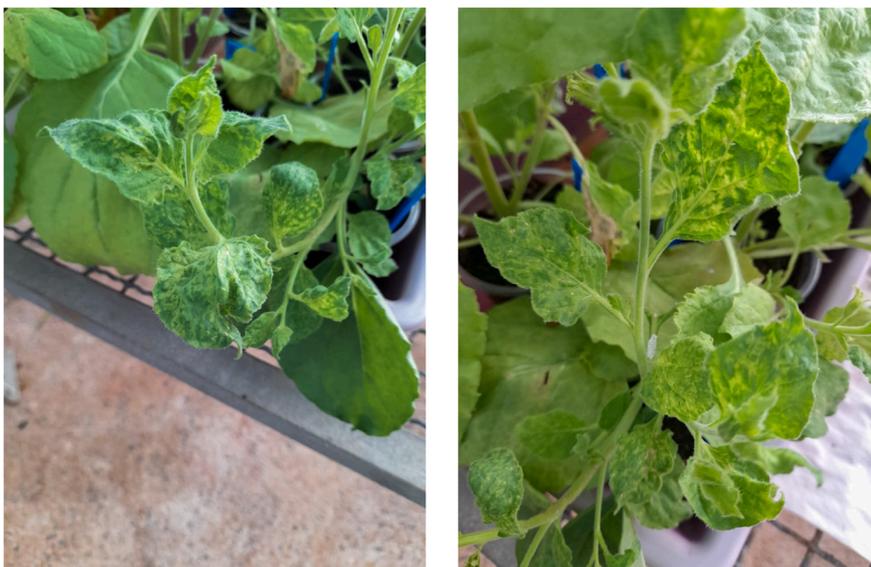


FIGURA 21. Identificação de sintomas foliares indicativos de disseminação sistêmica do vetor viral PVX_GW contendo integrase efetora.

Já com relação à introdução sequencial de Int9 e Int13 visando ativação e posterior retorno ao estado inicial silenciado do gene repórter, foram avaliados dois intervalos entre o primeiro e o segundo evento de infiltração, com introdução da segunda integrase três ou cinco dias após agroinfiltração com a primeira. No entanto, diferentemente do obtido nos ensaios iniciais em sistema TxT1 de expressão *in vitro*, ainda não foi possível detectar o fragmento com dupla edição em ampliações do DNA genômico das plantas transgênicas. Uma possível razão para a não detecção é a baixa representatividade da molécula On/Off em relação a todas as cópias não editadas ou editadas por apenas uma das integrases no pool de DNA genômico analisado. A dependência de um evento prévio de baixa eficiência (primeira edição) para a ocorrência do segundo evento também pouco eficiente torna necessária uma otimização da técnica de detecção aplicada.

As infiltrações para entrega das integrases phiC31 e Bxb1 foram realizadas seguindo o mesmo protocolo já discutido. Uma vez expressas, estas integrases ativam o módulo de excisão do sistema repórter. A recombinação dos sítios *attB* e *attP* deverá resultar na liberação de uma molécula circular contendo o gene *mgfp* e os sítios *att* das demais integrases, com o surgimento do sítio recombinante *attR* no ponto de circularização, permanecendo no genoma as regiões de promotor 35S CaMV e terminador OCS com o sítio *attL* formado na região de ligação do DNA, único vestígio remanescente no genoma do cassete removido. Como pode-se observar no diagrama apresentado na **Figura 22A**, a utilização do par de oligos 35S_282F e IDP_OCSt_349_Rv resulta na amplificação de um fragmento de aproximadamente 1.6 kb da construção repórter original ou um fragmento de aproximadamente 0.6 kb em caso de deleção por ação de integrase. A obtenção desse padrão de mudança de tamanho de banda em separação por eletroforese apresentada na **Figura 22B** em gel de agarose indica o sucesso da deleção em todos os eventos avaliados, porém a presença de bandas indicativas tanto do cassete editado quanto da construção original em uma mesma amostra reforça a natureza heterogênea da amostra. A fim de permitir uma melhor resolução dos resultados de sequenciamento molecular dos fragmentos, reduções no tempo de extensão da polimerase nas reações de PCR foram realizadas para favorecer o enriquecimento dos fragmentos de 0.6 kb (**Figura 22B**).

A.

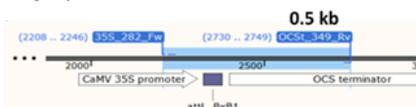
Estrutura inicial da construção



Deleção por phiC31



Deleção por BxB1



B.

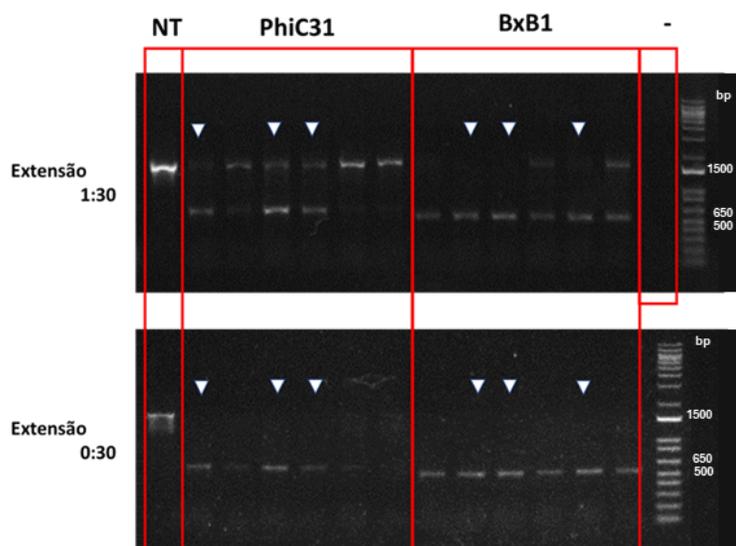


FIGURA 22. Ativação do módulo de excisão do sistema Int-Plex@. A) Fragmentos esperados em amplificação após excisão do gene *mgfp* em decorrência de infiltração com integrases phiC31 ou Bxb1. B) Amplificações obtidas após tratamentos. Obtenção de fragmentos dos dois tamanhos indicam heterogeneidade da amostra de DNA, com redução do tempo e extensão da polimerase selecionando preferencialmente a detecção do fragmento editado. Amostras indicadas por triângulos foram tratadas com a versão da integrase contendo a sequência NLS. NT, evento positivo não transfectado com integrase.

É interessante notar que, pelo menos qualitativamente considerando o padrão de intensidade das bandas, a utilização de phiC31 otimizada com uma sequência NLS parece ser mais eficiente na excisão do cassete, o que deverá ser futuramente avaliado por métodos quantitativos apropriados. Resultados semelhantes com relação ao aumento de eficiência de phiC31 em sistema eucariótico com a adição de sequências de localização nuclear já foram demonstrados³⁹. Os fragmentos amplificados foram submetidos a sequenciamento para

confirmação da deleção da região flanqueada pelos sítios e permanência de apenas o sítio *attL* formado pela ação da integrase utilizada. Também foram realizados amplificação e sequenciamento com o objetivo de detectar a molécula circular resultante da excisão do gene *mgfp* e os sítios *att* internos da construção utilizando o par de oligonucleotídeos NbDel_mGFP_493_Fw e NbDel_mGFP_300_Rv, que ao promoverem a extensão em direção às extremidades do gene *mgfp* resultarão em um *amplicon* apenas se houver a circularização da molécula excisada. Como indicado na **Figura 23**, os sequenciamentos de amostras obtidas após ativação do sistema com Bxb1 conseguiu identificar ambas moléculas esperadas, confirmando a geração e persistência nas células, pelo menos após 5 dias, da molécula circular contendo o transgene deletado.

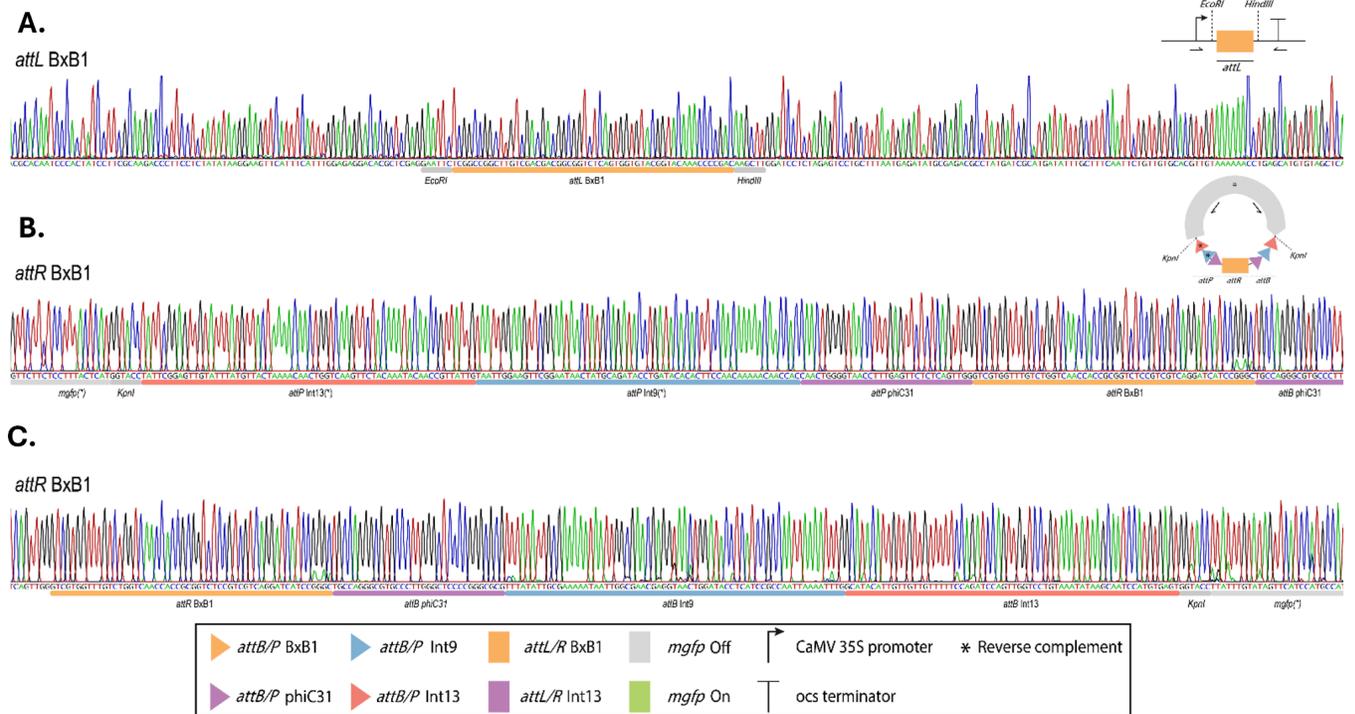


FIGURA 23. Alinhamento de seqüências obtidas a partir das amplificações do DNA de amostras tratadas com integrases Bxb1. (A) Detecção do sítio *attL* mantido no genoma vegetal resultante da deleção do cassette por ação da integrase. (B) e (C) Detecção do sítio *attR* de Bxb1 presente na molécula circular gerada pela excissão do cassette do genoma, juntamente com os sítios *attP* das demais Int à montante (B) e sítios *attB* à jusante (C). Resultados obtidos por seqüenciamento Sanger de *amplicons* purificados.

Quanto à edição por phiC31, foi possível identificar a marca de deleção no genoma (**Figura 24**), porém a molécula circular deletada ainda não foi identificada (apesar de obtenção, ainda que fraca, de *amplicon* de tamanho esperado em gel de agarose). Avaliações adicionais serão realizadas para melhor esclarecer tal ausência, se por questões de baixa eficiência da

integrase no modelo somada à necessidade de métodos de detecção de maior sensibilidade ou se há possibilidade de uma menor estabilidade da molécula circular, fazendo com que 5 dias após deleção já não seja possível sua detecção

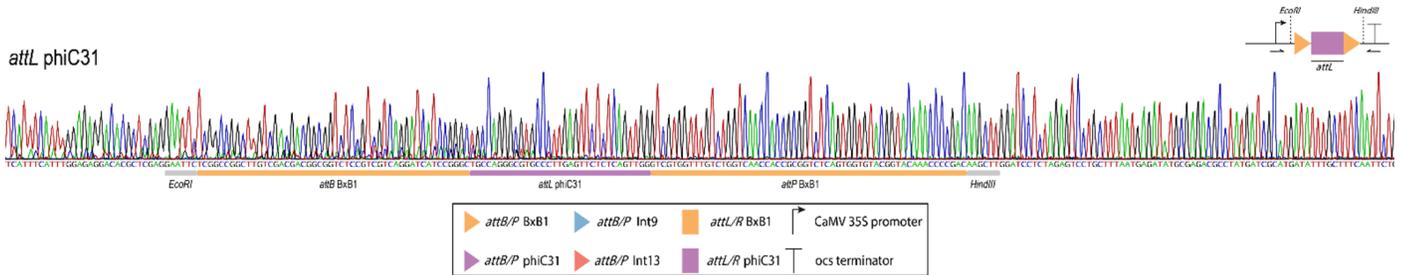


FIGURA 24. Alinhamento de sequências obtidas a partir das amplificações do DNA de amostras tratadas com integrases phiC31. (A) Detecção do sitio *attL* mantido no genoma vegetal resultante da deleção do cassette por ação da integrase. Resultados obtidos por sequenciamento Sanger de *amplicons* purificados.

Conclusões e Perspectivas

Sendo um avanço de nosso trabalho envolvendo o uso de serina-integrases em sistemas eucarióticos previamente publicado⁴⁰, o sistema Int-Plex@ foi construído buscando-se avaliar principalmente se as integrases seriam capazes de atuar em alvos integrados ao genoma do organismo, em contraste com o sistema transiente inicialmente analisado. A integração ao genoma eucariótico impõe desafios principalmente de acesso ao alvo, tanto pela barreira representada pela membrana nuclear quanto pelas questões de acesso à cromatina e imposições topográficas impostas. A composição de dois módulos funcionais, um de inversão e outro de excisão do DNA, também representa a evolução de complexidade em relação aos interruptores simples testados anteriormente. Além disso, a possibilidade de controle temporal da expressão de algum traço, principalmente em plantas, é bastante valiosa em um cenário em que o acúmulo de um produto apenas em um momento ou contexto específico pode representar um ganho de eficiência energética importante nas etapas iniciais de desenvolvimento ou de maior susceptibilidade a algum possível efeito citotóxico do produto em questão, com significativo impacto em termos de produtividade

Os resultados obtidos demonstram o sucesso da plataforma tanto na modulação da expressão do gene de interesse quanto na possibilidade de biocontenção de transgenes com o módulo de excisão da construção integrada ao genoma de *N. benthamiana*. As avaliações moleculares por amplificação e sequenciamento após introdução das integrases no sistemas indicam a capacidade de tanto Int9 quanto Int13 promoverem a inversão do gene alvo por meio da recombinação de seus sítios *att* sem que haja dano ao DNA ou interferência em decorrência

da presença dos sítios *att* das demais integrases. A ocorrência do segundo evento de inversão, que deverá retornar o sistema a seu estado original após inserção sequencial das duas integrases com um espaçamento temporal, ainda necessita ser melhor avaliada, com otimização de sensibilidade dos métodos escolhidos para a avaliação, tendo sido até o momento identificada apenas nos testes com a plataforma em sistema de transcrição e tradução *cell-free in vitro*.

O módulo de excisão controlado pelas integrases phiC31 e Bxb1 também se mostrou bem sucedido após integração da construção ao genoma vegetal. A detecção da molécula circular resultante da deleção reforça o sucesso, porém será importante realizar avaliação temporal de sua detecção a fim de melhor acompanhar o processo de degradação e eliminação do transgene pelo organismo.

Com relação à capacidade das integrases de acessarem o núcleo, todas as quatro integrases avaliadas obtiveram sucesso em sua edição independente da presença ou não da Sequência de Localização Nuclear inserida em seu gene, indicando a possibilidade da existência de NLS intrínsecas nas sequências originais, ainda que de um ponto de vista evolutivo tais sequências não se teriam feito necessárias considerado o contexto natural de interação entre bacteriófagos portadores das integrases e um hospedeiro procarioto, conseqüentemente sem barreiras nucleares a serem ultrapassadas. Vale ressaltar o aparente aumento de eficiência observado em phiC31 após inserção da sequência NLS, ainda que tal observação tenha sido feita até o momento de maneira indireta não sistematizada, sendo necessário ainda aprofundar a avaliação por meio do emprego de métodos quantitativos apropriados.

Por fim, tanto o método de entrega das integrases com o vetor binário pCAMBIA quanto a utilização do vetor viral PVX também se mostraram bem sucedidos. Em termos de aplicação em campo, a utilização de um vetor viral merece destaque dado a sua capacidade de disseminação sistêmica a partir do ponto de infiltração, atingindo todos os tecidos vegetais. Tal disseminação pode ser observada quando da aplicação do vetor PVX, com alcance e edição de folhas apicais distantes do ponto de aplicação, inclusive folhas formadas semanas após a inoculação dos efetores. No entanto, o uso de um vetor viral impõe dificuldades principalmente com relação à contenção de transmissão entre indivíduos e ao desenvolvimento de sintomas que podem comprometer a saúde e produtividade do cultivo. O acompanhamento das plantas inoculadas com o vetor viral PVX, por exemplo, demonstrou o surgimento de sintomas nas folhas a partir da terceira semana. No entanto, ainda que seja importante para otimizar a entrega do gene efetor de maneira eficiente, a otimização e engenharia de vetores virais não-infectivos demanda intensos esforços em virologia vegetal e fitopatologia não condizentes com o escopo do projeto desenvolvido no presente trabalho, sendo uma possibilidade de estudo no futuro, mas sem integrar as proposições de perspectivas imediatas do trabalho.

A primeira versão do manuscrito para publicação do sistema Int-Plex@ já encontra-se depositada no repositório BioRxiv [doi: <https://doi.org/10.1101/2024.01.11.575089>]. Uma cópia é apresentada ao final do presente trabalho (**Anexo B**).

Capítulo III

Implementação de metodologias chave da Biologia Sintética: Sistema de expressão *cell-free* TxTI (Transcription-Translation), Produção de Lipossomos e Encapsulamento como ponto de partida na construção de Células Sintéticas.

Introdução

A abordagem *Bottom-Up* da biologia sintética tem como objetivo final ideal a criação de células completamente constituídas por partes geradas sinteticamente por cientistas capazes de, uma vez devidamente agrupadas, originar um organismo que apresente características que emulem aquelas apresentadas por seres naturais como auto-suficiência energética, replicação e disseminação de informação genética, por exemplo. No entanto, grandes avanços se fazem ainda necessários para se alcançar tal meta, demandando a concentração de esforços em áreas abrangentes como biologia de membranas lipídicas, constituintes citoplasmáticos e componentes moleculares envolvidos em vias metabólicas e de replicação do DNA. Um campo intermediário neste percurso é a montagem de sistemas novos baseados em componentes celulares orgânicos pré-existentes, como a utilização de extratos celulares para criação de sistemas de expressão *in vitro* e a utilização de lipídeos na construção de estruturas de membrana na forma de lipossomos. A encapsulação nessas vesículas dos extratos celulares capazes de expressar proteínas a partir da introdução apenas de plasmídeos de DNA seria uma forma de mimetizar funções celulares em organismos não vivos. Outro exemplo dessas unidades sintéticas intermediárias, capazes de executar funções moleculares e carregar informações genéticas porém não vivas, seriam as *SimCells* discutidas no Capítulo I do presente trabalho.

Neste contexto, parte dos esforços realizados durante a execução dos projetos em biologia sintética realizados no âmbito do desenvolvimento da presente tese incluíram a implementação de metodologias de preparo de meios de expressão *in vitro* baseados na utilização de extratos celulares e a geração de lipossomos para encapsulação de extratos e produtos protéicos de interesse.

Objetivo

Implementar metodologias base em biologia sintética para construção de organelas e células sintéticas, envolvendo protocolos de produção e aplicação de sistema de expressão de proteínas *in vitro* baseado em extrato celular de *E. coli* (TxTI), produção de lipossomos unilamelares e encapsulamento do sistema de expressão *cell-free* em vesículas de bicamada lipídica.

Objetivos Específicos

- Desenho e síntese de plasmídeos para expressão no sistema TxTI
- Produção de extrato de lisado celular de *E. coli*
- Expressão e purificação da RNA polimerase T7 para uso em sistema TxTI
- Montagem do sistema TxTI de expressão *in vitro* com todos seus componentes e produção de proteínas de interesse
- Preparo de lipossomos constituídos por fosfolipídios e colesterol pelo método de emulsão inversa
- Encapsulamento em lipossomos de proteínas produzidas no sistema TxTI

Metodologia

Os protocolos utilizados foram adaptados para uso a partir de protocolos e treinamentos gentilmente cedidos através de colaboração com o Laboratório de Biologia Sintética da Universidade de Minnesota, sob orientação da Dra. Kate Adamala. Os protocolos referentes a preparo de extrato celular e montagem de reações de expressão *in vitro* são baseados em adaptações dos trabalhos desenvolvidos por Noireaux¹⁰⁵ e Jewett¹⁰⁶

Protocolo de produção de extrato celular para sistema TxTI

Meio 2x YTPG

Para 1 L:

- 31 g de 2xYT-Broth
- 3 g Fosfato de Potássio monobásico K₂HPO₄ (22 mM) – 22 mL de estoque 1 M
- 7 g Fosfato de Potássio dibásico K₂HPO₄ (40 mM), 40 mL de estoque 1 M

- em 800 mL de H₂O
- Autoclavar 15 min ; 121C
- Solução de glicose: 18 g de Glicose em 200 mL H₂O
- Autoclavar

- Acrescentar a solução de glicose estéril ao meio autoclavado. Armazenar na geladeira

Tampão de Lavagem A (Wash Buffer A)

Para 1 L:

10 mL 1 M Tris Acetato, pH8.2

14 mL 1 M Acetato de Magnésio (estoque a 1 M ou 1.99 g de MgAce)

5.89 g de Acetato de Potássio (98.14 g/mol)

Adicionar aproximadamente 974 mL de água destilada estéril

Autoclavar

Adicionar 2 mL de estoque 1 M DTT apenas antes de usar (concentração final de 2 mM DTT)

Dia 0: Crescer pré-inóculo em 50 mL de 2xYTPG, raspar células da placa riscada ou estoque de glicerol overnight a 37 C. Células Rosetta devem ser inoculadas em meio com chloramphenicol para manutenção do plasmídeo pLys (códon raro).

Dia 1:

- inocular 750 mL de 2xYTPG (sem antibiótico) com 5 mL do pré-inóculo.

- Incubar a 30C ; 180 rpm

Certifique-se de pre-resfriar a centrifuga, tubos falcon 50 mL e garrafas de centrifugação 500 mL.

Pesar tubos falcon vazios e anotar o resultado

- Coletar as células quando a cultura atingir OD600 de 0.5. Centrifugar a 3.000 rpm ; 30 min ; 4C

- Lavar precipitado com 200 mL de Tampão de Lavagem A duas vezes. Baixar células centrifugando a 3.000 rpm ; 30 min ; 4C.

- Realizar última lavagem com 40 mL de Tampão e juntar os precipitados. Remover tampão após centrifugação

- Pesar o tubo falcon e calcular a massa do precipitado obtido.

O precipitado pode ser congelado em nitrogênio líquido nesta etapa e mantido a -80C para retomada do protocolo no dia seguinte.

- Ressuspender precipitado em Tampão de Lavagem A (1.1 mL Tampão para cada 1 g de precipitado)

- Para sonicator: preparar banho de gelo em Erlenmeyer com agitador magnético.

- Sonicar 4.5 mL de células em tubo falcon de 15 mL. Ciclos: 10s ON, 15s OFF, 50% amplitude até atingir 2.7kJ (aprox. 14 minutos)
- Após sonicar, transferir lisado para tubos eppendorf e centrifugar por 30 min ; 15.000 g ; 4C
- Coletar sobrenadante e distribuir 500 µL em tubos falcon 15 mL.
- Incubar tubos por 1h a 37° C sob agitação para runoff
- Após a incubação, transferir extrato para tubos eppendorf e centrifugar por 30 min ; 15.000 g ; 4C
- recuperar sobrenadante e distribuir em alíquotas de 50 ul
- Congelar imediatamente em nitrogênio líquido ou banho de álcool e armazenar a -80C.

Protocolo de produção e purificação de RNA polimerase T7 para uso em sistema TxTI

Materiais a serem preparados de antemão:

- 2 L meio LB, autoclavados
- 4 Erlenmeyer de 2 L, autoclavados
- 2x Garrafas de Centrifuga e Tampas, autoclavados
- 2 mL de Estoque 1 M IPTG (mantido a -20°C)
- 0.25 mL PMSF 200 mM (alíquotas mantidas e -20°C. !! Reagente tóxico, manusear com cuidado !!

Células transformadas para expressão de T7 RNAPol 6HisTag: Rosetta2-1396p, transformada com o plasmídeo UMN 1396p (Pt7-911Q Seelig Lab) e mantidas em estoque glicerol a -80°C

Soluções

Os tampões devem ser feitos e mantidos sem os Inibidores de Protease ou BME

Tampão de Lise: 50 mM HEPES KOH pH 7.6, 1 M NH₄Cl, 10 mM MgCl₂

Para 50 mL de tampão:

2.5 mL 1 M HEPES KOH pH 7.6

2.68 g Ammonium Chloride

0.5 mL 1 M MgCl₂

24.8 µL 14.3 M BME

Tampão de Lavagem: 50 mM HEPES pH 7.6, 1 M NH₄Cl, 10 mM MgCl₂, 15 mM Imz

Para 20 mL de tampão:

1 mL of 1 M HEPES KOH pH 7.6

1.07 g of Ammonium Chloride

0.2 mL of 1 M MgCl₂
300 µL of 1 M Imidazole

Tampão de Eluição: 50 mM HEPES KOH pH 7.6, 100 mM KCl, 10 mM MgCl₂, 300 mM Imz

Para 20 mL de tampão:

1 mL of 1 M HEPES KOH pH 7.6
1 mL of 2 M KCl
0.2 mL of 1 M MgCl₂
6 mL of 1 M Imidazole

DIA 1

1. Prepare 2 pré-inóculos de 10 mL LB (+10 µL Ampicilina), raspando células do estoque glicérol
2. Incube overnight a 37°C sob agitação

DIA 2

1. Inicie culturas grandes: 20 mL de pré-inóculo para 2 L de LB+Amp
2. Distribua 500 mL da cultura em cada um dos 4 Erlenmeyer de 2 L e incube a 37°C, 250 rpm
3. Realize medições periódicas de Absorbância até a cultura atingir OD~0.5 (entre 6h e 7h para Rosetta ; aprox. 3h para BL21)
4. Ao atingir a OD inicie a indução por IPTG, concentração final 1 mM (Adicione 1 mL do estoque 1 M IPTG para cada litro de cultura)
5. Mantenha as garrafas de centrifugação vazias em câmara fria ou geladeira para resfriarem, bem como resfrie a centrífuga com rotor basculante para as garrafas
6. Após 3h de incubação pós-indução, resfrie as culturas em gelo por 20 min
7. Centrifugue as células em duas garrafas de centrífuga a 3700 rpm, 10 min, 4°C (centrifugue 500 mL ou menos por vez em cada garrafa, descartando o sobrenadante e repetindo a centrifugação de 10 min para o restante do volume)
8. O precipitado presente nas duas garrafas pode ser processado e purificado no mesmo dia (próxima seção) ou congelados em nitrogênio líquido e mantidos a -80°C.

Purificação (para precipitado de 1 L de cultura)

1. Descongele as células na câmara fria (aprox.. 30 min)
2. Finalize o Tampão de Lise: acrescente 0.25 mL PMSF 200 mM e 24.5 µL BME para 50 mL de tampão (Aqui recomendo o preparo de apenas 25 mL de tampão para cada precipitado de 1 L de cultura)

3. Após descongelado, ressuspender o precipitado em 20 mL de Tampão de Lise, agitando em movimentos circulares vigorosos ou com auxílio de uma p5000
4. Transfira o volume para um tubo Falcon de 50 mL e incube em gelo por 30 min
5. Durante a incubação refrigere a centrífuga e monte o setup do sonicador na câmara fria (fotos)
6. Após os 30 min de incubação prossiga com a sonicação: 15% Amplitude (~7W); ciclos de 15s ON/15s OFF até atingir 2kJ (aprox. 8 min). Deixe a amostra resfriar por 5 min. Repita o processo 4 vezes.
7. Centrifugue o lisado a 15000 g, 45min, 4°C
8. Durante a centrifugação prepare a resina: lave 1.2 mL de Ni-NTA agarose slurry (deve conter ~0.6 mL de beads de agarose) 2 vezes com água e 2 vezes com tampão de lise usando um eppendorf e centrífuga
9. Após a centrifugação, recupere o sobrenadante e mantenha no gelo (100 µL podem ser separados para posterior análise em SDS-PAGE)
10. Adicione a resina lavada ao falcon contendo o sobrenadante recuperado
11. Incube a mistura 'sobrenadante + resina' por 1h na mesa agitadora (rocker) na câmara fria
12. Durante a incubação prepare e finalize as soluções:

Tampão de Armazenamento 2X (Storage buffer) (1 L): 100 mM TrisHCl pH 7.6, 200 mM KCl, 20 mM MgCl₂, 14 mM BME:

100 mL of 1 M TrisHCl 7.6

100 mL of 2 M KCl

20 mL of 1 M MgCl₂

980 µL of 14.3 M BME

Tampão de Lavagem: Acrescente 4.9 µL de BME em 10 mL do Tampão de Lavagem estoque

Tampão de Eluição: Acrescente 4.9 µL de BME em 10 mL do Tampão de Eluição estoque

13. Finalizada a incubação transfira, ainda na câmara fria, a resina para uma coluna cromatográfica plástica de 10 mL, drenando o sobrenadante com auxílio de uma seringa acoplada à coluna (Fotos). Lave com 10 mL de Tampão de Lavagem

14. Elua a proteína com 10 mL de Tampão de Eluição, coletando as 3 primeiras frações de 1 mL por gravidade. A maior concentração da proteína estará nas duas primeiras frações.

15. Realize a quantificação por leitura de A280 para cada fração, coletando as com maior absorbância.

16. Dialise as frações contendo a enzima (frações com quantificações semelhantes podem ser misturadas) contra o Tampão de Armazenamento 2X, 1 L no total (Dialise em 500 mL da solução por 3h ; Troque o tampão por mais 500 mL novos e deixe overnight. Sempre na câmara fria)

17. Para a diálise utilize cartuchos Slyde-A-lyzer MWCO 30 kDa de volume apropriado para a quantidade de amostra recuperada

18. Colete a enzima dos cartuchos, acrescente 1 volume de glicerol estéril e mantenha em tubos de 1.5 mL a -20°C

19. Realize a quantificação final. A concentração será: [Concentração](em M)= A280/ (Coeficiente de Extinção da T7 x Peso Molecular T7), sendo Coeficiente de Extinção T7 = 1.42 e Peso Molecular T7 = 99000 Da. Concentrações finais para purificações a partir de precipitado de 1 L de cultura deverão ficar entre 20-50 uM

20. Em gel SDS-PAGE a T7 RNAPol proveniente do vetor pT7-911Q terá ~106 kDa, podendo ser marcada por anticorpo Anti-His em *Western Blot*.

Protocolo de preparo de mix de aminoácidos para reações TxTI

O estoque é preparado na concentração 10X, com 20 mM de cada aminoácido.

- Pese e adicione cada aminoácido de acordo com a **Tabela 6** abaixo em um Falcon 50 mL (para 40 mL de solução).

- Solubilize em 40 mL ou 100 mL de solução 400 mM KOH

- Ajuste para aproximadamente pH 6.6 com Ácido Acético (solução inicial deve ter pH ~10)

- **Para 40mL de solução → aproximadamente 750 µL de Ácido Acético glacial**

- Aferições de pH devem ser realizadas por meio de fita reagente

- Filtrar solução e preparar alíquotas de 500 µL em tubos 1.5 mL rotulados (AA mix)

- Congelar imediatamente ao final do processo e manter a -80°C

TABELA 6. Quantidades de cada aminoácido para preparo de solução estoque 20 mM do Mix de aminoácidos para reações TxTI

Aminoácido		Peso molecular	Para 40 mL de estoque 20 mM	para 100 mL de estoque 20 mM
Alanina	A	89.09	71.3 mg	178.18
Arginina	R	174.29	139.4	348.58
Asparagina	N	132.12	105.7	264.24

Aspartato	D	133.1	106.5	266.2
Cisteína	C	121.16	96.9	242.32
Glutamato	E	147.13	117.7	294.26
Glutamina	Q	146.15	116.9	292.3
Glicina	G	75.07	60.1	150.14
Histidina	H	155.15	124.1	310.3
Isoleucina	I	131.18	104.9	262.36
Leucina	L	131.18	104.9	262.36
Lisina	K	164.21	131.4	328.42
Metionina	M	149.21	119.4	298.42
Fenilalanina	F	165.19	132.2	330.38
Prolina	P	115.13	92.1	230.26
Serina	S	105.9	84.7	211.8
Treonina	T	119.12	95.3	238.24
Triptofano	W	204.23	163.4	408.46
Tirosina	Y	181.19	145.0	362.38
Valina	V	117.15	93.7	234.3

Protocolo de preparo de Mix Energético

Composição Mix Energético 10x

15 mM ATP e GTP	15 mM espermidina
9 mM CTP e UTP	40 mM oxalato de sódio
0.68 mM ácido folínico	7.5 mM cAMP
2 mg/mL de mix de tRNAs de <i>E. coli</i>	300 mM 3-PGA
3.3 mM nicotinamida adenina dinucleotídeo (NAD)	500 mM HEPES pH 8
2.6 mM coenzima-A (CoA)	

- Descongele as soluções estoque em gelo e prepare o mix na câmara fria

- Os reagentes devem ser misturados seguindo a ordem apresentada na **Tabela 7** ,
misturados com vortex após a adição de cada componente

-Após o preparo, separar em alíquotas de 0.5 mL e congelar imediatamente

TABELA 7. Reagentes e concentrações para preparo de estoque 10X do Mix Energético para reações TxTI

Reagentes e concentração final em Mix 10X	Estoque*	Volume do estoque para 10 mL de Mix **
500 mM HEPES pH 8	2 M	2.5 mL
Água milliQ	-	0.94 mL
15 mM ATP e GTP, 9 mM CTP e UTP ***		961uL
2 mg/mL mistura de tRNAs de E. coli	50 mg/mL	400uL
0.68 mM ácido folínico	33.9 mM	200uL
3.3 mM nicotinamida adenina dinucleotídeo (NAD)	175 mM	188uL
2.6 mM coenzima-A (CoA)	65 mM	400uL
15 mM espermidina	1 M	150uL
40 mM oxalate de sódio	200 mM	2 mL
7.5 mM cAMP	650 mM	115.4uL
300 mM 3-PGA (energy)	1.4 M	2.14 mL

Observações:

* soluções estoque devem ser mantidas a -80°C

** Volumes menores podem ser preparados

Preparo de reações de expressão in vitro com sistema TxTI

Para o preparo de reações de 50 µL seguir as concentrações e volumes indicados na **Tabela 8.**

TABELA 8. Composição e ordem de preparo de reação TxTI completa

Reagente	Estoque	Concentração final	Volume (ul)
H2O qsp			15.65

Glutamato de Magnésio (mM)	1000 mM	12 mM	0.6
Glutamato de Potássio (mM)	3000 mM	140 mM	2.3
DTT (mM)	100 mM	1 mM	0.5
Mix Energético	10x	1x	5
Mix de Aminoácidos (10x)	20 mM	2 mM	5
DNA	200 nM	10 nM	2.5
Inibidor de RNase murino 40U/ul	50x	1x	1
T7 polimerase	100uM	1.5uM	0.75
Extrato celular	3x	1x	16.7

- É possível substituir a RNA polimerase T7 purificada pelo uso direto do plasmídeo p70a_T7rnap (Arbor Bioscience). Seu uso facilita a purificação das proteínas produzidas com cauda de histidina pela ausência de polimerase contendo His-tag. Porém, é esperado uma redução na eficiência da reação.
- Descongelar e manter todos os componentes em gelo durante o preparo das reações. Retornar aos freezers o mais rápido possível.
- Alíquotas do Extrato Cell Free são de ~55 µL, o suficiente para TRÊS reações de 50 µL, já considerados desvios da pipetagem e preparo das alíquotas. !! EVITE DESCONGELAR ALIQUOTAS EXCEDENTES !!
- As reações devem ser incubadas a 30°C de 4h a 16h
- Reações podem ser incubadas em termociclador, com incubação overnight a 4°C ao final se necessário

Produção de Lipossomos por Emulsão Inversa

A produção de lipossomos foi realizada por método de emulsão inversa baseada em método publicado por Fujii e colaboradores¹⁰⁷. Esse método resulta em estruturas de 1 a 10 µm de diâmetro. Como base lipídica foram utilizados a fosfatidilcolina POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine), colesterol e 18:1 Liss Rhod PE, um fosfolipídio conjugado ao marcador fluorescente Rodamina.

O protocolo envolve, em linhas gerais, a suspensão dos lipídios em clorofórmio, deposição da mistura em frascos de vidro para formação de um filme lipídico em decorrência da evaporação do solvente, ressuspensão do filme lipídico em óleo mineral e emulsão da reações de TxTl a serem encapsuladas. A emulsão é então depositada sobre tampão aquoso e centrifugada para formação dos lipossomos, seguindo uma etapa de lavagem e concentração dos lipossomos.

Deposição e preparo de filme lipídico. Utilize seringas de vidro para manipulação das soluções lipídicas. Preparar a mistura de lipídios nas seguintes proporções em frasco de vidro:

- POPC/Colesterol/Lyss-Rhod-PE
 - 2.67 mL (66.7 mg) de solução 25 mg/mL de POPC em clorofórmio
 - 1.33 mL (33.3 mg) de solução 25 mg/mL de Colesterol em clorofórmio
 - 200 μ L (0.2 mg) de solução 1 mg/mL de Liss-Rhod-PE em clorofórmio
- Agite o frasco em movimentos circulares para homogeneização e aliquote 700 μ L em frascos de vidro de 2 mL (volume suficiente para preparo de 6 alíquotas)
- Mantenha os frascos abertos em capela de exaustão *overnight* para evaporação do solvente e formação dos filmes lipídicos
- Após preparo, manter os frascos selados com parafilme em freezer -80C até o uso, caso não faça uso imediato das preparações

Ressuspensão do filme lipídico. Para reconstituir o filme em solução, utilize óleo mineral pesado puro.

- Adicione 0.5 mL de óleo mineral em cada frasco contendo os filmes lipídicos
- Incube a 60°C por 10 min
- Agite em vortex por 10 min
- Incube novamente a 60°C por 3h
- Envolve os vidros em papel laminado e mantenha-os em banho sonicator a 60°C por 30 min
 - Esta etapa pode ser substituída por repetições dos ciclos de vortex e incubação a 60°C por 10 min até que o filme seja completamente dissolvido

Preparo por auto-montagem dos lipossomos.

- Adicione 225 μ L de Tampão de Centrifugação (HEPES 100 mM, glicose 200 mM, pH 8) em tubos de centrifugação de 1.5 mL
- Nos vidros contendo a solução de lipídios em óleo mineral, adicionar 30 μ L de Tampão interno (HEPES 100 mM, sacarose 200 mM, pH 8 - é possível acrescentar marcadores fluorescentes hidrossolúveis nesta etapa) ou 30 μ L de reação TxTl a ser encapsulada.
 - Agite em vortex por 30s para preparo da emulsão, equilibrando a 4°C por 10min em seguida
- Adicione a emulsão equilibrada cuidadosamente sobre os 225 μ L de Tampão de Centrifugação pipetando-a na parede interna do tubo. Espere pelo menos 1 min para estabilização da interface criada entre a emulsão em óleo e a fase aquosa, garantindo uma separação bem demarcada

- Centrifugue a 18000 g a 4°C por 15 min
- Retire com cuidado e descarte com auxílio de uma pipeta o máximo possível da fase oleosa
 - Os lipídios formados deverão estar depositados no fundo do tubo, na fase aquosa
- Adicione 225 µL de Tampão de Lavagem (HEPES 100 mM, glicose 250 mM, pH 8) em um novo tubo de centrifugação
- Cuidadosamente transfira o Tampão de Centrifugação contendo os lipossomos para o tubo preparado com o Tampão de Lavagem, evitando contaminação com a fase oleosa remanescente no tubo
- Centrifugue a 12000 g a 4°C, 5min
- Transfira 225 µL da fase inferior contendo os lipossomos para um tubo novo

Os lipossomos formados podem ser observados por microscopia de luz ou fluorescência quando utilizado lipídio conjugado a Rodamina, assim como quando encapsulado algum outro marcador. A solução contendo os lipossomos também pode ser submetida a processo de extrusão para garantir uma menor polidispersividade referente à variação de diâmetro da população de lipossomos formados.

Vetores de expressão em sistema TxTI

As construções realizadas para expressão *in vitro* em sistema TxTI utilizaram como arcabouço, majoritariamente, o plasmídeo 2041p, gentilmente cedido por Dra Adamala (Universidade de Minnesota, Mn, EUA). De modo geral, diversos dos vetores comerciais disponíveis para expressão em *E. coli* podem ser utilizados no sistema, desde que utilizem promotores de fator sigma como p70a ou σ^{28} . Dada a complementação do sistema com adição da RNA polimerase T7, plasmídeos contendo o promotor T7 também podem ser utilizados, como a família de vetores de expressão pET. No entanto, vetores como 2041p apresentam otimizações para melhor desempenho no sistema, como exemplo a redução em seu comprimento e utilização de sequências de ligação do ribossomo (RBS) e regiões UTR selecionadas. A unidade transcricional de 2041p também conta com terminador T500 e uma isoforma otimizada do promotor T7 recentemente caracterizada denominada promotor T7Max¹⁰⁸ **Figura 25**. Para detecção e purificação das proteínas de interesse, também foi adicionado às construções inseridas no plasmídeo 2041p cauda de histidina 6xHisTag na extremidade C-terminal. Outro vetor bastante utilizado no sistema é o vetor 2008p, apresentando basicamente o mesmo arcabouço e componentes da unidade transcricional principal (pT7Max, RBS e T500), no entanto carregando o gene repórter *degfp* (uma isoforma de GFP otimizada para expressão em sistema TxTI) com cauda 8xHisTag C-terminal.



FIGURA 25. Mapa da unidade transcripcional principal do plasmídeo 2041p para uso em sistema TxTI, contendo promotor T7Max, sequência RBS e terminador T500.

Western Blotting

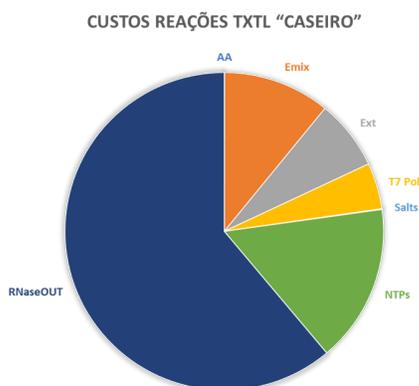
Para avaliação das proteínas produzidas por reações TxTI, bem como confirmação de encapsulamento, foram realizados ensaios de *Western Blot* utilizando anticorpos anti-His conjugados a fosfatase alcalina.

Microscopia Óptica

Para observação inicial dos lipossomos produzidos, alíquotas de 5 μ L foram depositadas diretamente sobre lâminas de vidro e retângulos paralelos de fita adesiva foram utilizados como espaçadores para evitar colapso das estruturas em decorrência do peso das lamínulas de vidro utilizadas. Não foi realizado tratamento para fixação ou contraste das amostras. As lâminas preparadas foram observadas em microscópio AxioPhot (Zeiss) sob luz branca com filtro de contraste de fase Ph1. As imagens adquiridas foram analisadas com auxílio do software Zen v3.3 *Blue Edition* (Zeiss) para medição.

Resultados e Discussão

Atualmente é possível encontrar sistemas de transcrição e tradução *in vitro* desenvolvidos e comercializados por diversas empresas, com composições e aplicações bastante variadas. Porém alto custo, longos períodos de envio e taxas de importação podem ser alguns dos fatores limitantes no uso destas metodologias. Desta forma, a possibilidade de produção *in loco* de extrato livre de células apresenta-se como um meio de contornar várias destas questões e facilitar seu acesso. Como pode-se observar no painel apresentado na **Figura 26**, há considerável redução de custo por reação quando utilizado o sistema TxTI produzido com os protocolos aqui apresentados em comparação a kits comerciais. Algumas adaptações ainda são possíveis a fim de uma redução maior de custos, como o uso de inibidores de RNase alternativos, a inclusão de NTPs diretamente ao mix energético e uso de plasmídeo para expressão de T7 RNA polimerase diretamente nas reações, em substituição à expressão e purificação da proteína.



CUSTO POR PROTOCOLO

Componente	Volume Final	Custo Produção
AA mix 20mM 10X	40 mL	R\$ 22.36
Energy Mix 3-PGA 10X	10 mL	R\$ 2578.75
Extrato Celular (1.5L Cultura)	2.7 mL	R\$ 139.49
T7 RNAPol (1L cultura Rosetta-p911Q)	2 mL (20uM)	R\$ 297.20

COMPARATIVO DE CUSTOS

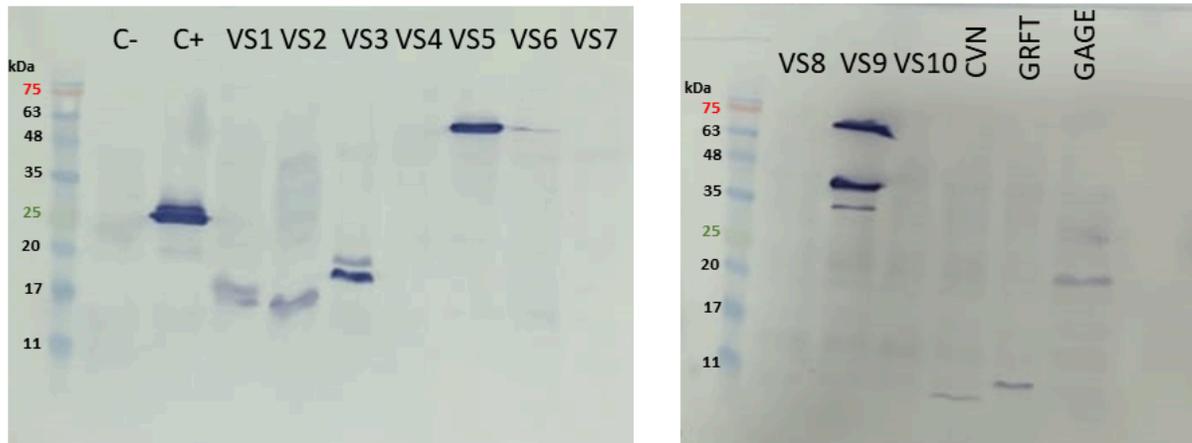
Produto	Custo Rxn 12ul	Custo ul de Rxn
TxTI_Caseiro	R\$ 2.88	R\$ 0.24
Promega (E coli T7 S30)	R\$ 25.64	R\$ 2.14
Arbor (myTxTI T7 Expression)	US\$ 11	US\$ 0.92
Murray/Noireaux (2012)*	US\$ 0.31	US\$ 0.03

* Protocolo caseiro. Sem gastos de T7pol e RNase Inhibitor. Considera mão de obra no cálculo (60% do total).

FIGURA 26. Levantamento de custos de produção de sistema TxTI e comparação com sistemas comerciais. Preços levantados em novembro/2021.

Após os esforços iniciais para implementação e adequação dos protocolos de produção do sistema TxTI, diversas proteínas já encontram-se em processo de expressão e teste na plataforma em nossos laboratórios. Dentre as primeiras proteínas expressas estão alguns objetos de estudo no âmbito de variados projetos atualmente desenvolvidos no laboratório de Biologia Sintética da EMBRAPA Cenargen, como as lectinas de atividade antiviral Cianovirina^{109,110} (CVN) e Griftisina¹¹¹ (GRFT), o peptídeo com atividade antitumoral GAGE¹¹², espidroínas Masp2¹¹³ de teia de aranha e mais recentemente diversas proteínas derivadas de soja retiradas do proteoma da cultivar BRS 537¹¹⁴ (**Tabela 9**). *Western blots* representativos da expressão dessas proteínas se encontram na **Figura 27**.

A.



B.

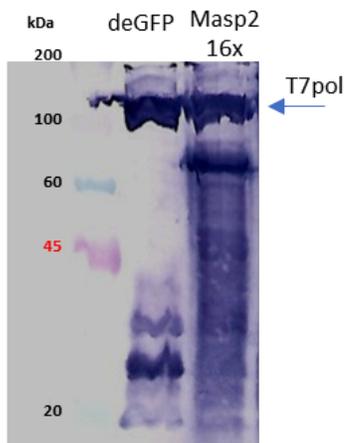


FIGURA 27. *Western Blotting* de expressão de proteínas diversas em sistema TxTI. A) Proteínas de soja (VS1-VS10), CVN, GRFT e GAGE, todas apresentando cauda de 6x-histidina para detecção com anticorpo anti-his. C-: controle negativo sem inclusão de proteína no poço; C+: controle positivo GFP-his. B) Proteína de teia de aranha Masp2 16x, produzida em TxTI com uso de T7pol-his purificada adicionada à reação (seta azul). A revelação foi realizada com uso de anticorpo anti-His conjugado a fosfatase alcalina e substrato cromogênico NBT/BCIP

Com relação à produção de lipossomos por emulsão inversa, é possível observar a formação das estruturas esféricas esperadas, porém com um alto grau de polidispersidade (**Figura 28B**). Uma maior homogeneidade pode ser alcançada após passagem dos lipossomos em sistema de extrusão usando membranas com poros de 0,1 μm , apesar de a redução de dimensões interferir na capacidade de visualização das estruturas (**Figuras 28A e 28C**).

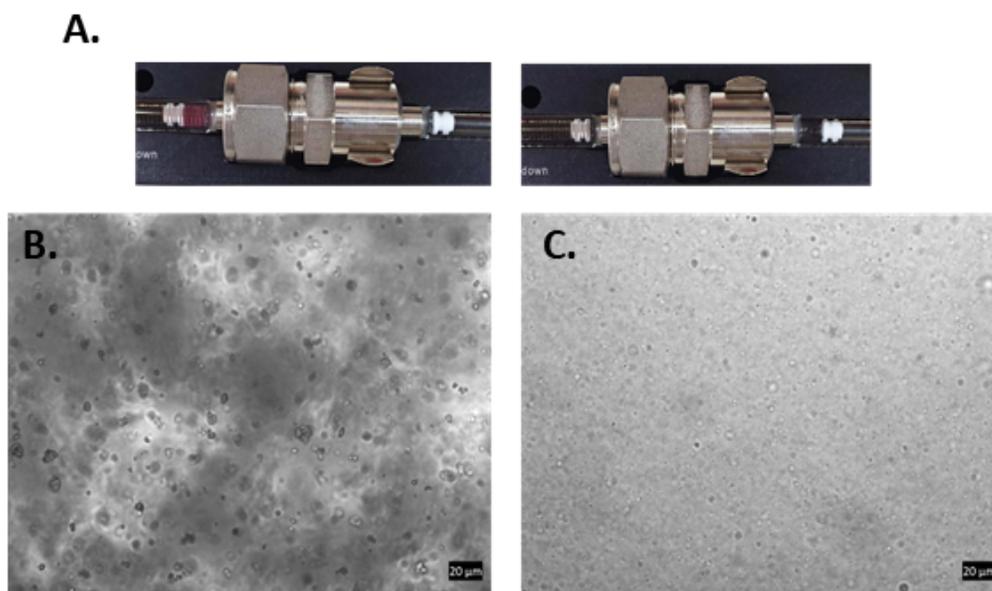


FIGURA 28. Produção de lipossomos por emulsão inversa. A) Sistema de extrusão manual com membrana de 0.1 µm. B) Microscopia de campo claro dos lipossomos gerados. C) Microscopia de campo claro após extrusão dos lipossomos demonstrando redução de diâmetro e menor índice de polidispersividade aparente.

Para os ensaios iniciais de encapsulação de reações TxTI e proteínas produzidas no sistema foram utilizados os vetores 2041p com genes de proteínas selecionadas a partir do genoma de soja para teste com um espectro variado de tamanho das proteínas (**Tabela 9**).

TABELA 9. Proteínas de soja selecionadas para processo de expressão em sistema TxTI e encapsulamento em lipossomos.

Código	Proteína	Massa (kDa)
VS1	Bowman Birk type proteinase inhibitor C II	13.13
VS2	Bowman Birk type proteinase inhibitor D II	13.31
VS3	S2 albumin	19.53
VS4	Lectin 1	31.99
VS5	Tubulin β -chain Fragment	51.23
VS6	β -amylase	57.14
VS7	Glycinin G4 (GY4)	64.86
VS8	α -subunit of β - Conglycinin Fragment	71.37
VS9	Heat shock 70 kDa protein	71.93

VS10	Lipoxygenase	95.48
------	--------------	-------

As reações de expressão *in vitro* foram realizadas previamente a 29°C *overnight*, sendo então adicionadas à etapa de emulsão para encapsulamento reações já contendo as proteínas de interesse expressas divididas em dois *pools* denominados *Odd* (VS1, VS3, VS5, VS7 e VS9) e *Even* (VS2, VS4, VS6, VS8 e VS10). Também não foi realizada extrusão dos lipossomos formados a fim de facilitar a visualização e diminuir a lise dos lipossomos com consequente perda de material encapsulado (**Figura 29**), tendo sido identificados lipossomos de diâmetros entre 1 µm e 6 µm no geral, como esperado de acordo com o protocolo utilizado.

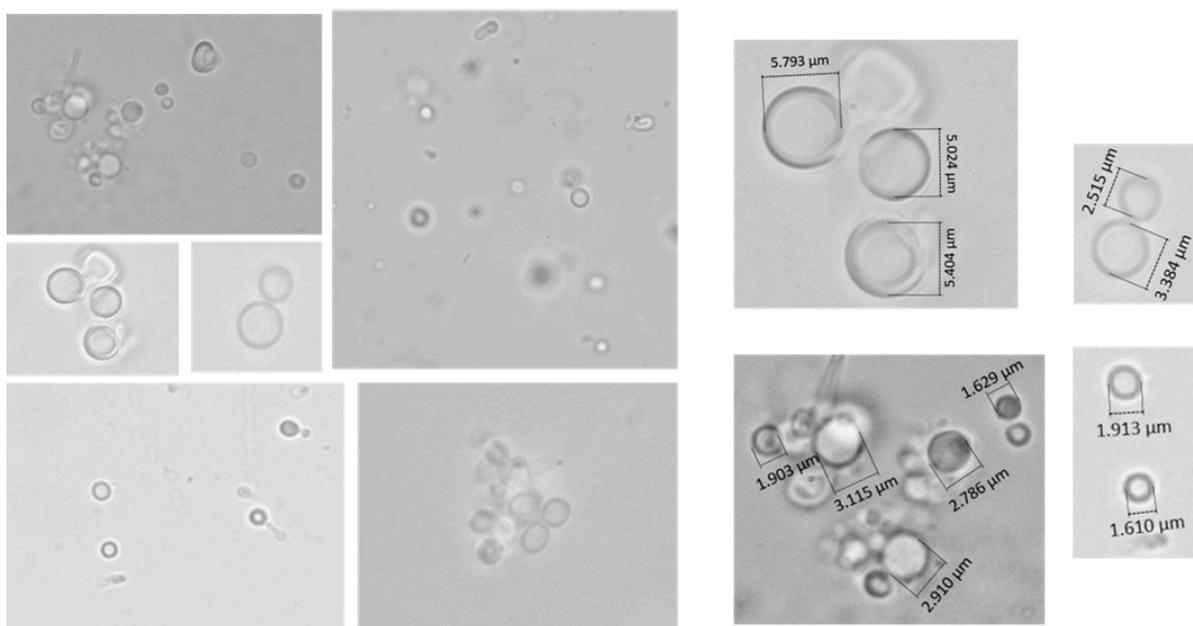


FIGURA 29. Microscopia de campo claro para medição de lipossomos produzidos por emulsão inversa utilizados no encapsulamento de proteínas.

Para detecção indireta da encapsulação, a solução de lipossomos produzidos foi centrifugada e parte do sobrenadante recuperada. Em seguida, os lipossomos foram ressuspensos e uma segunda alíquota da solução foi removida. O método assegura que no caso de detecção de proteínas expressas em *western blot* apenas após lise dos lipossomos em tampão de amostra contendo SDS fervido a 95°C, mas não no sobrenadante centrifugado há indícios positivos do encapsulamento das proteínas, não apenas detecção de material contaminante externo aos lipossomos. Como podemos observar na **Figura 30**, a ausência de proteínas no sobrenadante indica o sucesso das etapas de lavagem e ausência de proteínas livres no sistema. No entanto, de todas as proteínas adicionadas com as reações a serem encapsuladas, apenas a

proteína SV6 foi detectada após lise dos lipossomos (**Figura 30B, amostra I**). Também há indícios de encapsulamento de SV3 (**Figura 30A, seta vermelha**), porém o sinal muito fraco demanda melhor caracterização para comprovação.

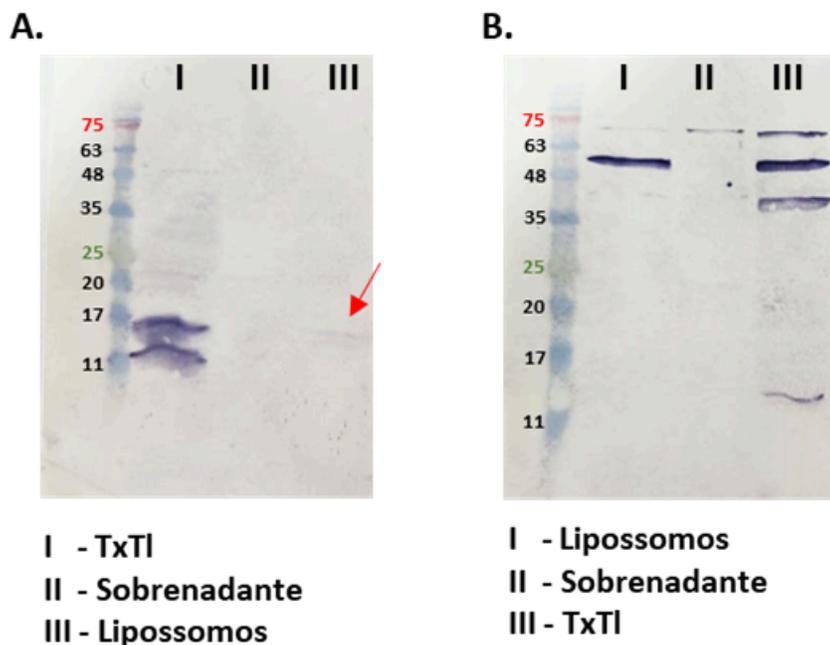


FIGURA 30. *Western blotting* para detecção de expressão de proteínas de soja em sistema TxTI e encapsulamento. A) Encapsulamento de pool de proteínas “ímpar” (SV1/3/5/7/9). B) Encapsulamento de pool de proteínas “par” (SV2/4/6/8/10), conforme apresentado na tabela 09. TxTI: reação *cell-free* pura após produção das proteínas; Sobrenadante: tampão recuperado após centrifugação para retirada de lipossomos; Lipossomos: lisado preparado a partir de precipitado de lipossomos contendo proteínas encapsuladas. A revelação foi realizada com uso de anticorpo anti-His conjugado a fosfatase alcalina e substrato cromogênico NBT/BCIP

Ainda que o resultado indique o sucesso de encapsulamento de SV6, com sugestão positiva também para SV3, contida nas reações de TxTI, estudos ainda serão necessários para identificação das características determinantes do sucesso ou falha envolvidas no processo de encapsulamento, além da otimização dos protocolos para garantir o encapsulamento de um maior número de proteínas nos lipossomos.

Conclusões e Perspectivas

Ainda que existam separadamente e suas aplicações individuais sejam extremamente amplas, a junção de um sistema de expressão *in vitro* baseado em extrato celular e lipossomos apontam uma direção bastante clara de biomimética para o desenvolvimento de novas plataformas sintéticas de protótipos celulares dentro da abordagem *bottom-up* de construção de um organismo sintético. O objetivo inicial de implementação de ambos processos no portfólio

metodológico do grupo se mostra bastante bem sucedido, não só pela capacidade de expressão de diversas proteínas de interesse e produção de lipossomos demonstradas, mas principalmente pelos diversos projetos derivados da aplicação conjunta ou individual das técnicas, incluindo a utilização de sistema TxTl na testagem inicial do Int-Plex®, interruptor genético desenvolvido em *N. benthamiana* apresentado no Capítulo II do presente trabalho.

Com relação às capacidades de encapsulamento em estruturas lipídicas, nosso processo de formação de lipossomos está atualmente sendo aprimorado por meio da inclusão de mecanismos de produção automatizada através de sistema microfluídico. Seu uso permite a geração de lipossomos com diâmetros variados, ainda que mantendo um índice de polidispersidade consideravelmente baixo, isto é, alta homogeneidade na distribuição de diâmetros em uma população. Além de que com a automatização passa ser possível a produção de lipossomos em escala consideravelmente aumentada, favorecendo não apenas a utilização destas plataformas no estudo de condições favoráveis ao desenvolvimento de organismos sintéticos, como a aplicação prática de lipossomos carreadores de cargas de interesse em áreas diversas.

Referências Bibliográficas

1. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12**, 381–390 (2014).
2. Meng, F. & Ellis, T. The second decade of synthetic biology: 2010-2020. *Nat. Commun.* **11**, 5174 (2020).
3. Gardner, T. S. Synthetic biology: from hype to impact. *Trends Biotechnol.* **31**, 123–125 (2013).
4. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nat. Rev. Genet.* **11**, 367–379 (2010).
5. Fritz, B. R., Timmerman, L. E., Daringer, N. M., Leonard, J. N. & Jewett, M. C. Biology by design: from top to bottom and back. *J. Biomed. Biotechnol.* **2010**, 232016 (2010).
6. Arber, W. Host-controlled modification of bacteriophage. *Annu. Rev. Microbiol.* **19**, 365–378 (1965).
7. Fogg, P. C. M., Colloms, S., Rosser, S., Stark, M. & Smith, M. C. M. New applications for phage integrases. *J. Mol. Biol.* **426**, 2703–2716 (2014).
8. Sheth, R. U. & Wang, H. H. DNA-based memory devices for recording cellular events. *Nat. Rev. Genet.* **19**, 718–732 (2018).
9. Tian, X. & Zhou, B. Strategies for site-specific recombination with high efficiency and precise spatiotemporal resolution. *J. Biol. Chem.* **296**, 100509 (2021).
10. Yuan, P., Gupta, K. & Van Duyne, G. D. Tetrameric structure of a serine integrase catalytic

- domain. *Struct. Lond. Engl.* **1993** **16**, 1275–1286 (2008).
11. Merrick, C. A., Zhao, J. & Rosser, S. J. Serine Integrases: Advancing Synthetic Biology. *ACS Synth. Biol.* **7**, 299–310 (2018).
 12. 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 Res.* **46**, 1308–1320 (2018).
 13. Ghosh, P., Wasil, L. R. & Hatfull, G. F. Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biol.* **4**, e186 (2006).
 14. Breüner, A., Brøndsted, L. & Hammer, K. Novel organization of genes involved in prophage excision identified in the temperate lactococcal bacteriophage TP901-1. *J. Bacteriol.* **181**, 7291–7297 (1999).
 15. Bibb, L. A. & Hatfull, G. F. Integration and excision of the Mycobacterium tuberculosis prophage-like element, phiRv1. *Mol. Microbiol.* **45**, 1515–1526 (2002).
 16. Abe, K. *et al.* Developmentally-regulated excision of the SP β prophage reconstitutes a gene required for spore envelope maturation in Bacillus subtilis. *PLoS Genet.* **10**, e1004636 (2014).
 17. 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**, e02767-16 (2017).
 18. Mandali, S., Gupta, K., Dawson, A. R., Van Duyne, G. D. & Johnson, R. C. Control of Recombination Directionality by the Listeria Phage A118 Protein Gp44 and the Coiled-Coil Motif of Its Serine Integrase. *J. Bacteriol.* **199**, e00019-17 (2017).
 19. Zhang, L., Zhu, B., Dai, R., Zhao, G. & Ding, X. Control of directionality in Streptomyces

- phage ϕ BT1 integrase-mediated site-specific recombination. *PloS One* **8**, e80434 (2013).
20. Fan, H.-F., Su, B.-Y., Ma, C.-H., Rowley, P. A. & Jayaram, M. A bipartite thermodynamic-kinetic contribution by an activating mutation to RDF-independent excision by a phage serine integrase. *Nucleic Acids Res.* **48**, 6413–6430 (2020).
 21. Rowley, P. A., Smith, M. C. A., Younger, E. & Smith, M. C. M. A motif in the C-terminal domain of phiC31 integrase controls the directionality of recombination. *Nucleic Acids Res.* **36**, 3879–3891 (2008).
 22. Meinke, G., Bohm, A., Hauber, J., Pisabarro, M. T. & Buchholz, F. Cre Recombinase and Other Tyrosine Recombinases. *Chem. Rev.* **116**, 12785–12820 (2016).
 23. Nakayama, M. VCre/VloxP and SCre/SloxP as Reliable Site-Specific Recombination Systems for Genome Engineering. *Methods Mol. Biol. Clifton NJ* **2637**, 161–180 (2023).
 24. Kim, H., Kim, M., Im, S.-K. & Fang, S. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. *Lab. Anim. Res.* **34**, 147–159 (2018).
 25. Golic, K. G. & Lindquist, S. The FLP recombinase of yeast catalyzes site-specific recombination in the *Drosophila* genome. *Cell* **59**, 499–509 (1989).
 26. Friedland, A. E. *et al.* Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009).
 27. Hartley, J. L., Temple, G. F. & Brasch, M. A. DNA cloning using in vitro site-specific recombination. *Genome Res.* **10**, 1788–1795 (2000).
 28. Walhout, A. J. *et al.* GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* **328**, 575–592 (2000).
 29. Yang, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nat. Methods* **11**, 1261–1266 (2014).
 30. Durrant, M. G. *et al.* Systematic discovery of recombinases for efficient integration of large

- DNA sequences into the human genome. *Nat. Biotechnol.* **41**, 488–499 (2023).
31. 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).
 32. Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P. & Endy, D. Amplifying genetic logic gates. *Science* **340**, 599–603 (2013).
 33. Courbet, A., Endy, D., Renard, E., Molina, F. & Bonnet, J. Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates. *Sci. Transl. Med.* **7**, 289ra83 (2015).
 34. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352**, aac7341 (2016).
 35. Jones, T. S., Oliveira, S. M. D., Myers, C. J., Voigt, C. A. & Densmore, D. Genetic circuit design automation with Cello 2.0. *Nat. Protoc.* **17**, 1097–1113 (2022).
 36. Park, Y., Espah Borujeni, A., Gorochofski, T. E., Shin, J. & Voigt, C. A. Precision design of stable genetic circuits carried in highly-insulated *E. coli* genomic landing pads. *Mol. Syst. Biol.* **16**, e9584 (2020).
 37. Kuhstoss, S., Richardson, M. A. & Rao, R. N. Plasmid cloning vectors that integrate site-specifically in *Streptomyces* spp. *Gene* **97**, 143–146 (1991).
 38. 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).
 39. Andreas, S., Schwenk, F., Küter-Luks, B., Faust, N. & Kühn, R. Enhanced efficiency through nuclear localization signal fusion on phage PhiC31-integrase: activity comparison with Cre and FLPe recombinase in mammalian cells. *Nucleic Acids Res.* **30**, 2299–2306 (2002).

40. Gomide, M. S. *et al.* Genetic switches designed for eukaryotic cells and controlled by serine integrases. *Commun. Biol.* **3**, 255 (2020).
41. Z, X. & Wr, B. Comparison and optimization of ten phage encoded serine integrases for genome engineering in *Saccharomyces cerevisiae*. *BMC Biotechnol.* **16**, (2016).
42. Guiziou, S., Maranas, C. J., Chu, J. C. & Nemhauser, J. L. An integrase toolbox to record gene-expression during plant development. *Nat. Commun.* **14**, 1844 (2023).
43. Keravala, A. *et al.* A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. *Mol. Genet. Genomics MGG* **276**, 135–146 (2006).
44. Yarnall, M. T. N. *et al.* Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nat. Biotechnol.* **41**, 500–512 (2023).
45. Adjalley, S. H., Lee, M. C. S. & Fidock, D. A. A method for rapid genetic integration into *Plasmodium falciparum* utilizing mycobacteriophage Bxb1 integrase. *Methods Mol. Biol. Clifton NJ* **634**, 87–100 (2010).
46. Chow, K.-H. K. *et al.* Imaging cell lineage with a synthetic digital recording system. *Science* **372**, eabb3099 (2021).
47. Chao, G., Travis, C. & Church, G. Measurement of large serine integrase enzymatic characteristics in HEK293 cells reveals variability and influence on downstream reporter expression. *FEBS J.* **288**, 6410–6427 (2021).
48. Blanch-Asensio, A. *et al.* STRAIGHT-IN enables high-throughput targeting of large DNA payloads in human pluripotent stem cells. *Cell Rep. Methods* **2**, 100300 (2022).
49. Thomson, J. G. & Ow, D. W. Site-specific recombination systems for the genetic manipulation of eukaryotic genomes. *Genes. N. Y. N 2000* **44**, 465–476 (2006).

50. Nkrumah, L. J. *et al.* Efficient site-specific integration in *Plasmodium falciparum* chromosomes mediated by mycobacteriophage Bxb1 integrase. *Nat. Methods* **3**, 615–621 (2006).
51. Spalding, M. D., Allary, M., Gallagher, J. R. & Prigge, S. T. Validation of a modified method for Bxb1 mycobacteriophage integrase-mediated recombination in *Plasmodium falciparum* by localization of the H-protein of the glycine cleavage complex to the mitochondrion. *Mol. Biochem. Parasitol.* **172**, 156–160 (2010).
52. Li, R. *et al.* Method for Biolistic Site-Specific Integration in Plants Catalyzed by Bxb1 Integrase. *Methods Mol. Biol. Clifton NJ* **1469**, 15–30 (2016).
53. Farruggio, A. P., Bhakta, M. S. & Calos, M. P. Use of the DICE (Dual Integrase Cassette Exchange) System. *Methods Mol. Biol. Clifton NJ* **1642**, 69–85 (2017).
54. Geisinger, J. M. & Calos, M. P. Using phage integrases in a site-specific dual integrase cassette exchange strategy. *Methods Mol. Biol. Clifton NJ* **1239**, 29–38 (2015).
55. Yoon, B., Kim, I., Nam, J.-A., Chang, H.-I. & Ha, C. H. In vivo and in vitro characterization of site-specific recombination of a novel serine integrase from the temperate phage EFC-1. *Biochem. Biophys. Res. Commun.* **473**, 336–341 (2016).
56. Farruggio, A. P. & Calos, M. P. Serine integrase chimeras with activity in *E. coli* and HeLa cells. *Biol. Open* **3**, 895–903 (2014).
57. Farruggio, A. P., Chavez, C. L., Mikell, C. L. & Calos, M. P. Efficient reversal of phiC31 integrase recombination in mammalian cells. *Biotechnol. J.* **7**, 1332–1336 (2012).
58. Ahmadi, M. *et al.* Evaluating the efficiency of phiC31 integrase-mediated monoclonal antibody expression in CHO cells. *Biotechnol. Prog.* **32**, 1570–1576 (2016).
59. Blanco-Redondo, B. & Langenhan, T. Parallel Genomic Engineering of Two *Drosophila*

- Genes Using Orthogonal attB/attP Sites. *G3 Bethesda Md* **8**, 3109–3118 (2018).
60. Snoeck, N. *et al.* Serine integrase recombinational engineering (SIRE): A versatile toolbox for genome editing. *Biotechnol. Bioeng.* **116**, 364–374 (2019).
 61. Hu, G., Goll, M. G. & Fisher, S. Φ C31 integrase mediates efficient cassette exchange in the zebrafish germline. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **240**, 2101–2107 (2011).
 62. Wei, Q.-X., Odell, A. F., van der Hoeven, F. & Hollstein, M. Rapid derivation of genetically related mutants from embryonic cells harboring a recombinase-specific Trp53 platform. *Cell Cycle Georget. Tex* **10**, 1261–1270 (2011).
 63. Gohl, D. M. *et al.* A versatile in vivo system for directed dissection of gene expression patterns. *Nat. Methods* **8**, 231–237 (2011).
 64. Malla, S., Dafhnis-Calas, F., Brookfield, J. F. Y., Smith, M. C. M. & Brown, W. R. A. Rearranging the centromere of the human Y chromosome with phiC31 integrase. *Nucleic Acids Res.* **33**, 6101–6113 (2005).
 65. Chomposri, J. *et al.* Intramolecular integration assay validates integrase phi C31 and R4 potential in a variety of insect cells. *Southeast Asian J. Trop. Med. Public Health* **40**, 1235–1253 (2009).
 66. Allen, B. G. & Weeks, D. L. Transgenic *Xenopus laevis* embryos can be generated using phiC31 integrase. *Nat. Methods* **2**, 975–979 (2005).
 67. Bernabé-Orts, J. M. *et al.* A memory switch for plant synthetic biology based on the phage ϕ C31 integration system. *Nucleic Acids Res.* **48**, 3379–3394 (2020).
 68. De Paepe, A., De Buck, S., Nolf, J., Van Lerberge, E. & Depicker, A. Site-specific T-DNA integration in *Arabidopsis thaliana* mediated by the combined action of CRE recombinase and ϕ C31 integrase. *Plant J. Cell Mol. Biol.* **75**, 172–184 (2013).

69. Maniloff, J. The minimal cell genome: 'on being the right size'. *Proc. Natl. Acad. Sci.* **93**, 10004–10006 (1996).
70. Razin, S. Peculiar properties of mycoplasmas: The smallest self-replicating prokaryotes. *FEMS Microbiol. Lett.* **100**, 423–431 (1992).
71. Fleischmann, R. D. *et al.* Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
72. Fraser, C. M. *et al.* The Minimal Gene Complement of *Mycoplasma genitalium*. *Science* **270**, 397–404 (1995).
73. Hutchison, C. A. *et al.* Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**, 2165–2169 (1999).
74. Glass, J. I. *et al.* Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 425–430 (2006).
75. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
76. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).
77. Pelletier, J. F. *et al.* Genetic requirements for cell division in a genomically minimal cell. *Cell* **184**, 2430-2440.e16 (2021).
78. Breuer, M. *et al.* Essential metabolism for a minimal cell. *eLife* **8**, e36842 (2019).
79. Goodsell, D. S. Integrative illustration of a JCVI-syn3A minimal cell. *J. Integr. Bioinforma.* **19**, 20220013.
80. Thornburg, Z. R. *et al.* Kinetic Modeling of the Genetic Information Processes in a Minimal Cell. *Front. Mol. Biosci.* **6**, 130 (2019).

81. Goodsell, D. S. & Autin, L. Integrative modeling of JCVI-Syn3A nucleoids with a modular approach. *Curr. Res. Struct. Biol.* **7**, 100121 (2023).
82. Kiyama, H., Kakizawa, S., Sasajima, Y., Tahara, Y. O. & Miyata, M. Reconstitution of a minimal motility system based on Spiroplasma swimming by two bacterial actins in a synthetic minimal bacterium. *Sci. Adv.* **8**, eabo7490 (2022).
83. Nishiumi, F. *et al.* Blockade of endoplasmic reticulum stress-induced cell death by Ureaplasma parvum vacuolating factor. *Cell. Microbiol.* **23**, e13392 (2021).
84. Bianchi, D. M., Pelletier, J. F., Hutchison, C. A., Glass, J. I. & Luthey-Schulten, Z. Toward the Complete Functional Characterization of a Minimal Bacterial Proteome. *J. Phys. Chem. B* **126**, 6820–6834 (2022).
85. Pedreira, T., Elfmann, C., Singh, N. & Stülke, J. SynWiki: Functional annotation of the first artificial organism Mycoplasma mycoides JCVI-syn3A. *Protein Sci. Publ. Protein Soc.* **31**, 54–62 (2022).
86. Moger-Reischer, R. Z. *et al.* Evolution of a minimal cell. *Nature* **620**, 122–127 (2023).
87. Sandberg, T. E. *et al.* Adaptive evolution of a minimal organism with a synthetic genome. *iScience* **26**, 107500 (2023).
88. Hossain, T., Deter, H. S., Peters, E. J. & Butzin, N. C. Antibiotic tolerance, persistence, and resistance of the evolved minimal cell, Mycoplasma mycoides JCVI-Syn3B. *iScience* **24**, 102391 (2021).
89. Carvalho, F. M. *et al.* DNA repair in reduced genome: the Mycoplasma model. *Gene* **360**, 111–119 (2005).
90. Mariscal, A. M. *et al.* Tuning Gene Activity by Inducible and Targeted Regulation of Gene Expression in Minimal Bacterial Cells. *ACS Synth. Biol.* **7**, 1538–1552 (2018).

91. Ipoutcha, T. *et al.* Genome Editing of Veterinary Relevant Mycoplasmas Using a CRISPR-Cas Base Editor System. *Appl. Environ. Microbiol.* **88**, e0099622 (2022).
92. Ipoutcha, T., Gourgues, G., Lartigue, C., Blanchard, A. & Sirand-Pugnet, P. Genome Engineering in *Mycoplasma gallisepticum* Using Exogenous Recombination Systems. *ACS Synth. Biol.* **11**, 1060–1067 (2022).
93. Kostylev, M., Otwell, A. E., Richardson, R. E. & Suzuki, Y. Cloning Should Be Simple: *Escherichia coli* DH5 α -Mediated Assembly of Multiple DNA Fragments with Short End Homologies. *PloS One* **10**, e0137466 (2015).
94. Broto, A., Gaspari, E., Miravet-Verde, S., dos Santos, V. A. P. M. & Isalan, M. A genetic toolkit and gene switches to limit *Mycoplasma* growth for biosafety applications. *Nat. Commun.* **13**, 1910 (2022).
95. Stemke, G. W. & Robertson, J. A. Comparison of two methods for enumeration of mycoplasmas. *J. Clin. Microbiol.* **16**, 959–961 (1982).
96. Li, H., Sharp, R., Rutherford, K., Gupta, K. & Van Duyne, G. D. Serine Integrase attP Binding and Specificity. *J. Mol. Biol.* **430**, 4401–4418 (2018).
97. Fan, C. *et al.* Chromosome-free bacterial cells are safe and programmable platforms for synthetic biology. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 6752–6761 (2020).
98. Wang, Y. & Demirer, G. S. Synthetic biology for plant genetic engineering and molecular farming. *Trends Biotechnol.* S0167-7799(23)00088–4 (2023)
doi:10.1016/j.tibtech.2023.03.007.
99. Stark, W. M. Making serine integrases work for us. *Curr. Opin. Microbiol.* **38**, 130–136 (2017).
100. Schwab, R., Ossowski, S., Riester, M., Warthmann, N. & Weigel, D. Highly Specific

- Gene Silencing by Artificial MicroRNAs in Arabidopsis. *Plant Cell* **18**, 1121–1133 (2006).
101. Pandey, P., Mysore, K. S. & Senthil-Kumar, M. Recent Advances in Plant Gene Silencing Methods. *Methods Mol. Biol. Clifton NJ* **2408**, 1–22 (2022).
 102. Zhu, H., Li, C. & Gao, C. Applications of CRISPR–Cas in agriculture and plant biotechnology. *Nat. Rev. Mol. Cell Biol.* **21**, 661–677 (2020).
 103. Hajdukiewicz, P., Svab, Z. & Maliga, P. The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. *Plant Mol. Biol.* **25**, 989–994 (1994).
 104. Lacorte, C., Ribeiro, S. G., Lohuis, D., Goldbach, R. & Prins, M. Potatovirus X and Tobacco mosaic virus-based vectors compatible with the Gateway cloning system. *J. Virol. Methods* **164**, 7–13 (2010).
 105. Sun, Z. Z. *et al.* Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology. *JoVE J. Vis. Exp.* e50762 (2013)
doi:10.3791/50762.
 106. Kwon, Y.-C. & Jewett, M. C. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci. Rep.* **5**, 8663 (2015).
 107. Fujii, S. *et al.* Liposome display for in vitro selection and evolution of membrane proteins. *Nat. Protoc.* **9**, 1578–1591 (2014).
 108. Deich, C. *et al.* T7Max transcription system. *J. Biol. Eng.* **17**, 4 (2023).
 109. Boyd, M. R. *et al.* Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob. Agents Chemother.* **41**, 1521–1530 (1997).
 110. Muñoz-Basagoiti, J. *et al.* Cyanovirin-N binds to select SARS-CoV-2 spike

- oligosaccharides outside of the receptor binding domain and blocks infection by SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2214561120 (2023).
111. Isolation and characterization of griffithsin, a novel HIV-inactivating protein, from the red alga *Griffithsia* sp - PubMed. <https://pubmed.ncbi.nlm.nih.gov/15613479/>.
112. Nagel, H., Laskawi, R., Eiffert, H. & Schlott, T. Analysis of the tumour suppressor genes, FHIT and WT-1, and the tumour rejection genes, BAGE, GAGE-1/2, HAGE, MAGE-1, and MAGE-3, in benign and malignant neoplasms of the salivary glands. *Mol. Pathol. MP* **56**, 226–231 (2003).
113. Silva, L. P. & Rech, E. L. Unravelling the biodiversity of nanoscale signatures of spider silk fibres. *Nat. Commun.* **4**, 3014 (2013).
114. Molinari, M. D. C. et al. Exploring the Proteomic Profile of Soybean Bran: Unlocking the Potential for Improving Protein Quality and Quantity. *Plants* **12**, 2704 (2023).
115. Liu, Y.-G., Mitsukawa, N., Oosumi, T. & Whittier, R. F. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *The Plant Journal* **8**, 457–463 (1995).
116. Pucker, B., Kleinbölting, N. & Weisshaar, B. Large scale genomic rearrangements in selected *Arabidopsis thaliana* T-DNA lines are caused by T-DNA insertion mutagenesis. *BMC Genomics* **22**, 599 (2021).
117. Giraldo, P. A., Shinozuka, H., Spangenberg, G. C., Smith, K. F. & Cogan, N. O. I. Rapid and Detailed Characterization of Transgene Insertion Sites in Genetically Modified Plants via Nanopore Sequencing. *Front. Plant Sci.* **11**, (2021).

118. Malmberg, M. M., Spangenberg, G. C., Daetwyler, H. D. & Cogan, N. O. I. Assessment of low-coverage nanopore long read sequencing for SNP genotyping in doubled haploid canola (*Brassica napus* L.). *Sci Rep* **9**, 8688 (2019).
119. Fernandez-Pozo, N. et al. The Sol Genomics Network (SGN)--from genotype to phenotype to breeding. *Nucleic Acids Res* **43**, D1036-1041 (2015).

Anexo A

Artigo Científico

Protocol for the establishment of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells.

LAB PROTOCOL

Protocol for the establishment of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells

Marco A. de Oliveira^{1,2}, Lilian H. Florentino^{1,2,3}, Thais T. Sales^{1,2,3}, Rayane N. Lima^{2,3}, Luciana R. C. Barros⁴, Cintia G. Limia⁵, Mariana S. M. Almeida^{2,3}, Maria L. Robledo⁵, Leila M. G. Barros^{2,3}, Eduardo O. Melo^{2,3}, Daniela M. Bittencourt^{2,3}, Stevens K. Rehen^{6,7}, Martín H. Bonamino^{8,9}, Elibio Rech^{2,3*}

1 Department of Cell Biology, Institute of Biological Science, University of Brasília, Brasília, Distrito Federal, Brazil, **2** National Institute of Science and Technology in Synthetic Biology (INCT BioSyn), Brasília, Distrito Federal, Brazil, **3** Embrapa Genetic Resources and Biotechnology, Brasília, Distrito Federal, Brazil, **4** Center for Translational Research in Oncology, Instituto do Câncer do Estado de São Paulo, Hospital das Clínicas da Faculdade de Medicina de Universidade de São Paulo, São Paulo, Brazil, **5** Molecular Carcinogenesis Program, Research Coordination, National Cancer Institute (INCA), Rio de Janeiro, Brazil, **6** D'Or Institute for Research and Education (IDOR), Rio de Janeiro, Brazil, **7** Institute of Biomedical Sciences, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, **8** Cell and Gene Therapy Program, Research Coordination, National Cancer Institute (INCA), Rio de Janeiro, Brazil, **9** Vice-Presidency of Research and Biological Collections (VPPCB), FIOCRUZ – Oswaldo Cruz Foundation Institute, Rio de Janeiro, Brazil

* elibio.rech@embrapa.br



OPEN ACCESS

Citation: de Oliveira MA, Florentino LH, Sales TT, Lima RN, Barros LRC, Limia CG, et al. (2024) Protocol for the establishment of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells. PLoS ONE 19(5): e0303999. <https://doi.org/10.1371/journal.pone.0303999>

Editor: Chen Ling, Fudan University, CHINA

Received: September 6, 2023

Accepted: May 4, 2024

Published: May 23, 2024

Copyright: © 2024 de Oliveira et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All results and dataset generated in the development and use of the present protocol are available in previous publication and supplementary information (DOI <https://doi.org/10.1038/s42003-020-0971-8>) Plasmid maps and sequences are available from Addgene database (<https://www.addgene.org/browse/article/28203482/>).

Funding: This research was funded by Embrapa Genetic Resources and Biotechnology/National Institute of Science and Technology in Synthetic

Abstract

Serine integrases (Ints) are a family of site-specific recombinases (SSRs) encoded by some bacteriophages to integrate their genetic material into the genome of a host. Their ability to rearrange DNA sequences in different ways including inversion, excision, or insertion with no help from endogenous molecular machinery, confers important biotechnological value as genetic editing tools with high host plasticity. Despite advances in their use in prokaryotic cells, only a few Ints are currently used as gene editors in eukaryotes, partly due to the functional loss and cytotoxicity presented by some candidates in more complex organisms. To help expand the number of Ints available for the assembly of more complex multifunctional circuits in eukaryotic cells, this protocol describes a platform for the assembly and functional screening of serine-integrase-based genetic switches designed to control gene expression by directional inversions of DNA sequence orientation. The system consists of two sets of plasmids, an effector module and a reporter module, both sets assembled with regulatory components (as promoter and terminator regions) appropriate for expression in mammals, including humans, and plants. The complete method involves plasmid design, DNA delivery, testing and both molecular and phenotypical assessment of results. This platform presents a suitable workflow for the identification and functional validation of new tools for the genetic regulation and reprogramming of organisms with importance in different fields, from medical applications to crop enhancement, as shown by the initial results obtained. This protocol can be

Biology, National Council for Scientific and Technological Development (465603/2014-9) and Research Support Foundation of the Federal District (0193.001.262/2017).

Competing interests: The authors have declared that no competing interests exist.

completed in 4 weeks for mammalian cells or up to 8 weeks for plant cells, considering cell culture or plant growth time.

Introduction

The ability to regulate gene expression in response to external cues is one of the central mechanisms of the differentiation and maintenance of life in nature, as well as one of the main goals of scientists in efforts to control and reprogram organisms. Therefore, the availability of molecular tools that allow genetic manipulation is crucial for advances in synthetic biology, especially in creating intricate genetic circuits and activation cascades to work as synthetic regulatory networks. A prominent group of effectors used to that end are integrases, a superfamily of site-specific recombinases (SSRs) capable of directed and controlled rearrangement of DNA sequences [1–4]. Although tyrosine integrases such as Cre/LoxP [5–8] and λ Int [5] systems have historically been predominantly used, recently, another family of SSRs known as serine integrases has received attention, in great part because of the advantages they present over their counterparts. The main advantages for application in synthetic biology include the shorter length of their attachment sites, unidirectional recombination and nondependence on auxiliary effectors to work [9–14]. Despite the advantages, only a very limited number of functional serine integrases have been available for use in eukaryotic organisms. We have recently used 6 out of the 13 newly described functional serine integrases previously identified and characterized in *Escherichia coli* by Yang and collaborators [15] to assemble and test genetic switches capable of modulating gene expression upon Int activation in eukaryotic systems, including plant-, bovine- and human-derived cells [16]. While in the original work in *E. coli* all integrases showed similar efficiency, their efficiency in plant and mammal cells varied considerably, yet successful recombination was detected for all integrases used.

Originally present in nature as a mechanism used by some bacteriophages to integrate their DNA into the genome of a prokaryotic host, the serine-integrase mechanism of action involves recognition and physical interaction with a specific pair of attachment sites present in both phage and bacterial host DNA known as *attP* and *attB* sites, respectively [17, 18]. Upon binding to DNA, conformational alterations of the complex help to expose the central part of the *att* sites, known as the core sequence, to the catalytic site responsible for the double break in the DNA strands. Subsequent conformational changes lead to the recombination of cut ends from complementary *att* sites and final ligation to form the newly recombined *attL* and *attR* sites in a permanent, unidirectional way [19–21]. Since this whole process occurs without the need for any additional endogenous effector, serine integrases exhibit valuable host plasticity when considered as a gene editing tool. However, it is the rational manipulation of *att* site combination and orientation that truly broadens the use possibilities of these recombinases. When cognate *att* sites are present in different DNA molecules, the natural integration process occurs. Notwithstanding, if *att* sites are synthetically designed to be present at the same molecule, the recombination of these sites can have varying outcomes according to their orientation, namely, excision of a portion of the DNA molecule, a 180° inversion of a DNA sequence flanked by the *att* sites [15] and recombinase-mediated cassette exchange (RMCE) [22, 23]. Diagrams to better illustrate these possible outcomes are presented in Fig 1.

Genetic switches designed to invert a given sequence are especially interesting for expression control. In this type of system, a gene of interest or the promoter regulating its expression is assembled in an opposing orientation relative to the other and flanked by a pair of *att* sites. Upon integrase activation, *att* sites are recombined with consequent inversion of the DNA

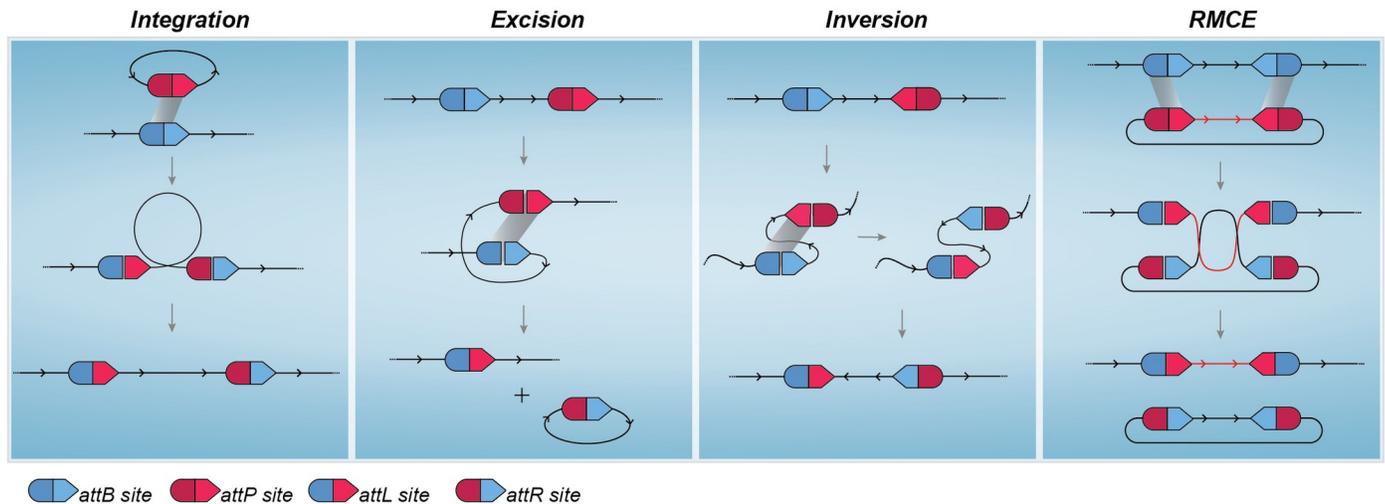


Fig 1. Serine integrase mediated DNA rearrangement. (A) Although the original function of this class of recombinases in nature is the integration of bacteriophage DNA into a host genome upon recombination of *attB* and *attP* sites, synthetic manipulation of *att* sites can lead to different outcomes. Recombination of *att* sites located in the same molecule will result in (B) Excision of flanked DNA region if both sites have the same orientation or (C) Inversion when sites are present in opposite orientations. (D) Recombinase Mediated Cassette Exchange (RMCE) allows the exchange of a large sequence of DNA when a molecule harbors two copies of an *attP* site and a second molecule harbors two copies of the cognate *attB* site.

<https://doi.org/10.1371/journal.pone.0303999.g001>

sequence in between, therefore allowing proper gene transcription [15, 16]. Recombination by serine-integrase is unidirectional and permanent unless a cognate recombination directional-ity factor (RDF; an accessory protein encoded in the same phage genome that contains the one integrase with which it can interact) [24] or bidirectional activating mutations [25] are present.

Although the most prominent applications of recombinases to date involve the use of tyro-sine-integrases, serine-integrases were first identified and used to deliver DNA in *Streptomyces spp.* over 30 years ago [26]. Since then, many works have taken advantage of their recombina-tion capabilities in prokaryotic models [15], unicellular eukaryotic organisms [23], and in a few cases, complex multicellular organisms, including mouse and human cells [16, 27], and they have even been used as molecular tools for in vitro DNA plasmid assembly [14, 28]. More recently, the serine-integrase ability to excise or flip terminators and genes has been more widely applied in the assembly of genetic circuits designed to either modulate gene expression or compute events by permanently rearranging DNA parts based on binary logic [15, 29–32]. The higher intricacy of activation cascades and genetic circuits with multiple integrases as effectors intensifies the demand for new functional integrases and corresponding *att* sites, encouraging efforts to describe new members and to develop systematic models for the identi-fication and testing of candidates, such as bioinformatic tools to identify potential serine-inte-grases and *att* sites in genome databases [15, 27] and *in silico* platforms to help with the design and assembly of complex genetic circuits [33–35]. That need is even more accentuated when considering the use of genetic circuits based on integrases in eukaryotic organisms, since the higher complexity and compartmentalization create a functional bottleneck in which many integrases that function in bacteria do not work or present cytotoxic effects in eukaryotic sys-tems [16, 23]. Xu et al. [36] showed that some integrases able to rearrange sequences on plas-mid DNA in human cells lost their ability to do so when the target DNA was integrated into the genome, while Andreas et al. [37] showed that alterations such as adding a nuclear localiza-tion signal (NLS) at the Φ C31 C-terminus region enhanced its editing capabilities in mouse

cells. As mentioned, we also showed that integrases with similar efficiency levels in bacteria showed varying degrees of efficiency when tested in plant and animal cells. Moreover, despite the increasing levels of interest and publications using serine-integrases in eukaryotic organisms, very few new members have been validated in these systems, with many groups always working with the same limited pool of candidates, mainly ϕ C31, BxB1, and less frequently, ϕ R4 and TP901. Considering this scenario, we propose in this protocol a platform to systematically test and identify functional serine integrases for the assembly of genetic circuits in different plant and mammalian organisms.

The platform can be seen as a workflow comprising six main stages: I) *in silico* design and synthesis of plasmid constituents of each effector and reporter module; II) cell acquisition and culture maintenance, which can be subdivided into *in vitro* culture of established mammalian cell lines, patient cell isolation and plant growth for protoplast isolation according to the model selected; III) DNA delivery; IV) cytotoxicity evaluation; V) molecular analysis, including DNA preparation, primer design, PCR setup and sequencing; and VI) phenotypic analyses by fluorescence microscopy and flow cytometry. Fig 2 shows a schematic overview of the workflow proposed.

Materials and methods

The protocol described in this peer-reviewed article is published on [protocols.io](https://doi.org/10.17504/protocols.io.rm7vzx945gx1/v1) [dx.doi.org/10.17504/protocols.io.rm7vzx945gx1/v1] and is included for printing purposes as S1 File. Sub-protocols specific for each organism group used in the present work are included for printing purposes as S2–S5 Files.

Ethical statement

All work involving cells derived from mammalian organisms must comply with and be performed under strict ethical guidelines. Experiments with human materials must conform to all

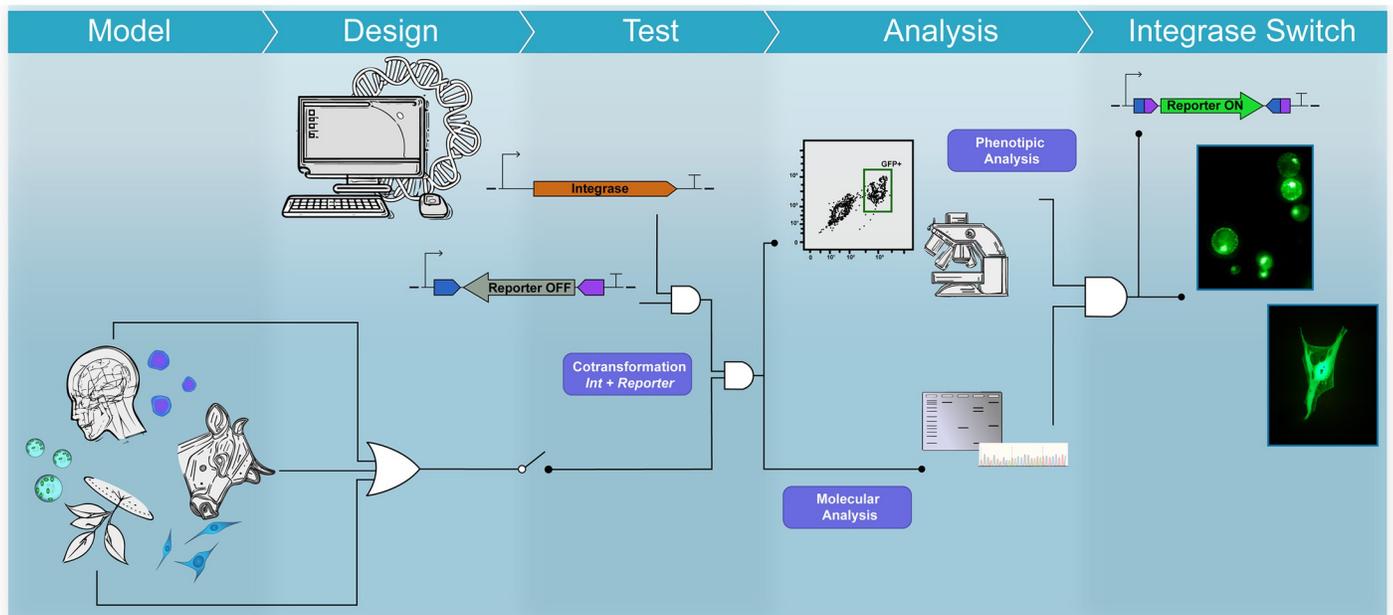


Fig 2. Strategic overview of the serine integrase-based platform for the functional characterization of genetic switch controllers in eukaryotic cells.

<https://doi.org/10.1371/journal.pone.0303999.g002>

relevant institutional and governmental ethics regulations, and appropriate informed consent must be obtained for the use of human blood or patient-derived materials. For the development of the present protocol, Peripheral Blood Mononuclear Cells from healthy donors were collected and used upon approval by the Brazilian National Cancer Institute (INCA) Ethics Committee and signing of Review board-approved informed consent forms by the donors. Experiments involving the use of human-derived stem cells were approved by the ethics committee of Copa D'Or Hospital (CAAE number 60944916.5.0000.5249, approval number 1.791.182). Regarding the acquisition and use of bovine primary fibroblasts, all experimentation performed was approved by the Ethics Committee on the Use of Animals (CEUA) of Embrapa Genetic Resources and Biotechnology (Brasilia, Brazil) in March 2013 (approval reference no. 001/2013).

Experimental design

***In silico* design of plasmids.** These initial steps are crucial for a successful experiment since they include the definition of all the regulatory parts to be used, as well as the rational design of the switch to be tested. The modular aspect of the construction allows the easy and fast assembly of numerous plasmid sets carrying different integrases and their respective reporter cassettes. As indicated in the schematics presented in Fig 3, the basic expression unit for the effector plasmids consists of the promoter sequence most appropriate to the organism in which the test will be performed, followed by the integrase gene and a terminator sequence downstream. For better expression rates, the integrase gene sequence may be codon optimized before synthesis to match the codon usage patterns of the organism or group of interest. This

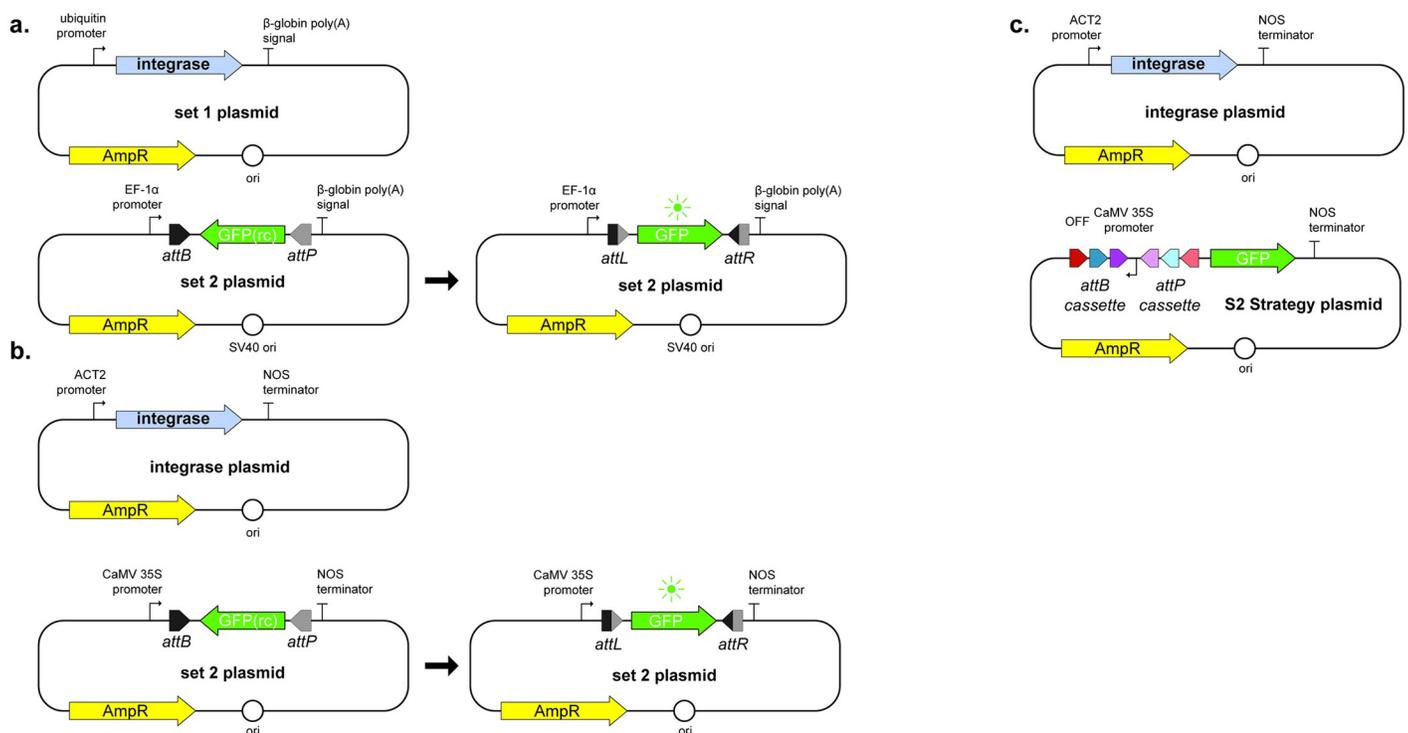


Fig 3. Designed unidirectional genetic switches composed of two sets of synthesized plasmids. (a) and (b) Plasmid sets for S1 strategy in mammalian and plant models, respectively. (c) Plasmid sets for S2 strategy applied for plant models. Here, the promoter sequence, not the *gfp* gene, is under the control of integrase-mediated inversion and activation. Figure adaptation from Gomide et. al., 2020 [16].

<https://doi.org/10.1371/journal.pone.0303999.g003>

expression unit must be inserted into an appropriate backbone that allows cloning and transformant selection for the preparation of DNA in bacteria before assays and proper behavior once delivered into the eukaryotic model. The promoter sequence must render constitutive expression, and weak promoters, i.e. sequences known to promote low transcription rates—therefore affecting mRNA accumulation—, are avoided to reduce the chance of falsely identifying an integrase as nonfunctional in the system when in reality it is due to low expression levels. Additional elements, like RBS and enhancer sequences can be screened and used to fine-tune transcription and increase integrase expression. In this protocol, we selected the promoters pUbc [38] and pAct [39] for use with mammalian and plant cells, respectively.

The reporter unit has a more complex design but still follows the modular aspect of the effector plasmids with the classic downstream sequence of promoter, gene of interest and terminator sequence. The promoter and terminator sequences selected for this protocol in the reporter module were pEF1a [38] and B-globin poly(A) signal for use with mammalian cells and pCaMV35S and NOS terminator for use with plant cells. *gfp* was used as the reporter gene here, although the choice of reporter must take into consideration background fluorescence wavelength and possible emissions from cell constituents such as chloroplasts in the plant cells to avoid emission overlap. Two reporter strategies were designed and named S1 and S2 Strategy. S1 strategy (Fig 3A and 3B) is designed to assess the integrase's capacity for rearranging a coding DNA sequence orientation, for which the reporter gene sequence must be inserted into the construct in an anti-sense position in relation to the promoter, thus in a silenced state—here referred to as the OFF STATE—and flanked by the integrase recombination sites. It is crucial that recombination sites are inserted in opposite directions to each other to ensure inversion of the DNA sequence between them given the recombination dynamics discussed above. The selection of integrases and their cognate *att* site pair should involve considering the eventual presence of starting codons in the *att* sites or recombined *attL* and *attR* site sequences that can interfere with gene expression and reading frames. In the occurrence of such codons, swapping *attB* and *attP* positions can be considered, as long as the final design contains the most downstream site in its reverse complement form.

The genetic switch can be assembled to direct the reorientation of genetic parts other than a gene of interest. In the S2 strategy designed for this platform, the promoter is initially inactive due to its opposite orientation in relation to the reporter gene (Fig 3C). Another improvement in this strategy is the use of a cassette of *att* sites in tandem, which allows the use of the same reporter plasmid to test the activity of different integrases or sequential induction events with different integrases to invert the flanked sequence back to its initial position. Despite broadening the applications of the genetic switch, this strategy demands prior knowledge of the integrases selected or a more thorough evaluation and proper use of controls to ensure their orthogonality.

All plasmids used in the development of this protocol are listed in Table 1 below, with Addgene database accession number for more information including annotated sequence maps. The genetic parts in the constructions can be changed to better suit different organisms.

Cell acquisition and culture maintenance. Another important aspect of the presented protocol is the plasticity in regard to eukaryotic cell models with which it can be performed. Primary bovine fibroblasts were obtained from 14-month-old *Bos indicus* bull oxtail by biopsies following the protocol defined by Freshney [40] with modifications. Briefly, recovered pieces of tissue are washed with 0.05% trypsin to detach cells, which are then cultured in Dulbecco's Modified Eagle Medium supplemented with fetal bovine serum 10% at 37°C and 5% CO₂ atmosphere in culture flasks for at least three passages or until a homogeneous culture of attached fibroblasts can be observed. HEK293T cells can be purchased from cell line culture collections and stored, although some culture passages are recommended prior to cell

Table 1. Plasmids used in the assembly of S1 and S2 strategy sets.

Plasmid Set	Name	Addgene ID
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

<https://doi.org/10.1371/journal.pone.0303999.t001>

transfection to ensure that metabolically fit cultures are being used. For human donor primary or derived cells used to pursue potential medical applications of the genetic switches, PBMCs were isolated from donor blood via Ficoll density gradient centrifugation, while neural stem cells and human embryonic stem cells from the BR-1 cell line were cultured in neural induction medium (NEM, Advanced DMEM/F12 and neurobasal medium (1:1) with neural induction supplement). All animal and human donor-derived cells and experimentation performed with them must undergo appropriate ethics committee evaluation and approval.

Plant cells derived from *A. thaliana*. Briefly, protoplasts were isolated from young leaves of 5-week-old plants cultivated at 22°C under a 12 h light/12 h dark regimen according to the protocol described by Yoo et al. [41] with modifications. The collected leaves were prepared by making shallow cuts on the adaxial face with a scalpel to enhance internal exposure to the enzymatic solution. In this work, *A. thaliana* leaf cells were disaggregated using an enzymatic

cocktail of 0.2% pectolyase, 0.5% driselase and 1.5% cellulase, but other combinations and adaptations are possible to better digest the cellular walls of cells from other species. The age and overall morphology and health of leaves can have a great impact on protoplast isolation; therefore, the use of young specimens of up to 60 days is recommended.

DNA delivery and controls. Given the particularities of each model, different plasmid delivery strategies are applied depending on the cell line. In general, nonintegrative transfection resulting in transient expression of the effector integrase and reporter gene is sufficient for the functional screening proposed in this protocol. All strategies applied here have in common the concurrent transfection of effector and corresponding reporter plasmids in equimolar ratio and a period of at least 24 h post-transfection before assessing the outcome responses. Different delivery strategies were used depending on the model. Transfection of HEK293T cells and primary bovine fibroblasts was performed by lipid-mediated delivery using Lipofectamine LTX and Plus reagent (Invitrogen) according to the manufacturer's recommendations. Despite being widely used and considered a gold standard method for DNA delivery [42], lipofectamine transformation can result in low transformation efficiency and high cytotoxicity for some cell lines, including human primary cells and non-adherent cultures. Hence, PBMC, NCS and hES cell transformation was carried out by electroporation using an optimized adaptation of the nucleofection method established by Chicaybam et al. [43]. This method yields higher transformation efficiency for some human cells, including primary and stem cells, with low cytotoxic effects at a lower price than the classical nucleoporation method. For plant protoplasts, CaCl₂-PEG chemical transfection [44, 45] was used to carefully deliver the plasmid sets. Although other methods exist for DNA delivery in protoplasts, including agrobacteria-mediated transformation [46] and electroporation [47], PEG-mediated transfection is the most used delivery system due to its easy manipulation, low cost and no need for specific equipment. Another important advantage of this method is that it allows for more gentle manipulation of cells, an important aspect when handling fragile protoplasts following cell wall digestion.

Although the cotransformation of an effector and corresponding reporter plasmid pair is the central aspect of the functional evaluation of an integrase, many control groups containing different plasmid combinations are necessary to validate the obtained results. The first set of controls will be the samples to which only one member from the plasmid pair is delivered to eliminate any background signal or endogenous machinery activity interference that could lead to false-positive results. With these samples, we will have the opportunity to assess fluorescence levels coming from a non-activated reporter in the absence of the effector integrase, for instance, which could be indicative of a leaky system or unwanted promoter activity from the downstream *att* site present. One important example of such occurrence is the observed promoter activity of Int13 *attP* site described by Yang et al. in *E. coli* [15]. An unlikely, although not impossible, fluorescence emission by the integrase itself could also be ruled out in this context. These control groups are also important to identify potential cytotoxic effects from any of the components of the system, especially from the integrase being tested, which may cause DNA damage or integration at endogenous pseudosites [36, 48, 49], disrupting the proper expression of important unrelated genes. It is important to note that for these controls, the plasmid absent from the original pair must be replaced by the same amount of a mock plasmid, normally the empty backbone used in the constructions, to maintain the DNA load concentration and molar ratio. Another set of controls is especially necessary in the case of our S2 strategy or when future applications of the integrases being tested involve the simultaneous presence of more than one in a genetic switch or cascade: in this case, combinations of an effector plasmid with one integrase should be cotransformed with the reporter plasmid

carrying recombination sites of a different integrase to assess orthogonality, i.e., one integrase is not capable of interacting with or rearranging *att* sites other than their cognate site pair.

Positive controls included plasmids carrying unaltered reporter gene sequences under constitutive expression and the design and synthesis of the expected rearranged cassette carrying the reporter gene sequence in its forward orientation flanked by the expected recombined *attL* and *attR* sites. The first group is a classic positive control to ensure cell behaviour and allow equipment calibration. As for the constructions with *attL* and *attR* sites, it is an important group to help evaluate if an observed lowered level or lack of signal from the reporter following integrase activation indicates a negative result or inhibitory transcriptional interference due to the presence of the post-recombination sites. Although serine-integrases are not able to recombine *attL/R* sites, Chao et al. demonstrated that integrase interaction with these sites can alter target expression [50].

As mentioned, the use of the first set of control groups is suitable for evaluating cytotoxicity. Given that one of the steps in the recombination of *att* sites by an integrase involves double-strand breaks in the DNA, the occurrence of off-target activity at pseud-sites causing unintended DNA damage can be a concerning source of toxicity and limit the use of such integrase in the assembly of a genetic switch. Cell integrity markers and colorimetric assays designed to measure cell viability are an easy way to check for deleterious effects. PBMCs and stem cells used were stained with 7-AAD, a live cell impermeable fluorophore that allows the quantification of stained dead cells in the same flow cytometry run used to assess reporter expression. A similar strategy was applied to protoplasts by staining cells with fluorescein diacetate (FDA), although in this case, live cells were stained. For HEK293T cells and primary bovine fibroblasts, an enzymatic MTT assay was performed.

Molecular analyses

To confirm integrase activity, both molecular and phenotypical analysis methodologies should be applied. As we observed previously [16], when screening for integrases capable of rearranging DNA in eukaryotic cells, sometimes the confirmation of sequence inversion by amplification and sequencing was not accompanied by a positive reporter signal. DNA extraction and amplification should follow standard practices established for Sanger sequencing, but primer design must be carefully performed. In addition to confirming DNA inversion, amplification and sequencing are means to confirm proper recombination of *att* sites by directed editing of the integrase not causing point mutations or DNA damage at the core sequencing, gene of interest or surrounding regions. A good strategy is to amplify the 5' end of the gene and upstream formed *attL* site as one fragment and the 3' end of the gene and the downstream *attR* formed site as another fragment, keeping an overlap region between them to ensure good coverage of the whole cassette, as shown by the diagram in Fig 4. It is important to note that when opting to use primers annealing to post recombination *attL* and *attR* sites to increase specificity, the primers must have a high T_m point and PCR setup performed with higher annealing temperatures since part of the primer will inevitably anneal to respective *attB* and *attP* sites to a certain extent.

Obtained amplicons must be sequenced to ensure amplification specificity and correct formation of *attL* and *attR* sites with target coding sequence inversion. For most standard model organisms, especially cell cultures with high transformation efficiency, amplification should be no problem and Sanger sequencing will provide good results whilst still easily accessible in facilities or sequencing companies at low costs. A recommendation is to have fragments sequenced with forward and reverse primers to allow better coverage of both ends of the region of interest. Overlap of these sequences can also help identify mutations from

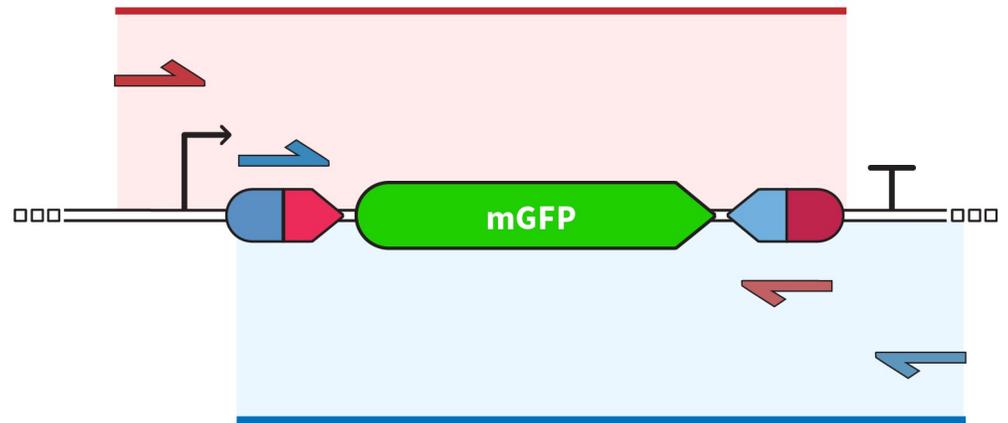


Fig 4. Oligonucleotide pairs and expected amplification regions in the two-amplicon strategy. Figure adaptation from Gomide et. al., 2020 [16].

<https://doi.org/10.1371/journal.pone.0303999.g004>

sequencing error instead of integrase-cause DNA damage. However, sequencing method selection can be affected by other factors. When working with cells of low transformation efficiency that could impair DNA delivery, tissues instead of cell culture that can result in mixed population with more non-transformed than transformed cells, or screening integrases with very low recombination efficiency, a more robust and sensitive sequencing method may be used. Nanopore DNA sequencing is a good alternative in these scenarios [51], more easily detecting low copy numbers of edited sequences in a mixed population. This methodology present other advantages when compared to Sanger sequencing, including the possibility to better evaluate the occurrence and rate of integrase-related mutations and DNA damage due to its high accuracy and possibility of reads quantification for integrated stoichiometry analysis.

Phenotypical analyses

Reporter activation can be assessed by fluorescence visualization or quantification. Fluorescence microscopy is a fast way to observe reporter expression activation but will result only in qualitative data. For accurate quantification to allow efficiency ranking between all integrases tested flow cytometry must be performed. Multiple time points should be evaluated to identify the best incubation period for each model. For the models used in our work, 48h was defined as standard incubation time, with eGFP signal decrease after 72h post-incubation in mammalian cells. Due to the short survival in culture conditions, flow cytometry of treated protoplasts was performed after 24 hours following transformation. High event count must be used, and all control groups must be considered to ensure a proper definition of gates to differentiate between eGFP positive and negative cell populations.

Assays must be done with at least three independent biological replicates; technical replicates should be included to account for measurement and equipment variations. Statistical method choice must be done considering the specificities of each new model and experimental design selected. For our platform, quantitative data were analyzed using the nonparametric Kruskal—Wallis test, with Dunn's test as post hoc for pairwise comparisons. Although One-way ANOVA is typically used for multiple comparisons, nonparametric tests do not rely on assumptions regarding normal distribution and variance homogeneity, which can be hard to confirm when working with small sample sets.

Expected results

We first developed our platform to assess the functionality of 6 integrases previously characterized only in *E. coli*, named Int2, Int4, Int5, Int7, Int9 and Int13 [15]. Original bacterial host and sequence of *attB/attP* sites for each Int are presented at Table 2. However, the plasticity of our workflow makes it suitable for the initial screening in eukaryotic cells of any other new serine integrases identified in silico from genome databases or with ones previously tested only in prokaryotic cells with very little adaptation needed.

In our work [16], we were able to confirm the successful recombination of recognition sites and formation of correct *attL* and *attR* sites with consequent inversion of the *egfp* gene used as a reporter for all the integrases studied. Despite the positive results at the molecular level, *mgfp* expression levels after switch activation varied depending on the integrase and organism used when cells were analyzed by flow cytometry or fluorescence microscopy. Overall, for the S1 experimental design Int13 switches resulted in higher levels of fluorescence emission for all the organisms used, followed by Int9 and Int4 in bovine fibroblasts and protoplasts. Int5 switch yielded the lowest levels of signal despite molecular confirmation of DNA inversion, possibly an indication of low frequencies of recombination in the systems (Fig 5). Integrases phiC31 and BxB1, used for comparison given their well established applications in mammal and plant systems, also resulted in varying signal levels depending on the cell used, although overall results were similar to the ones obtained when using Int13. Interestingly, the S2 promoter switch construction led to a result compatible with the S1 GFP switch in plant protoplasts. However, Int2 exhibited a much higher number of EGFP-positive cells when the promoter was flipped than when the targeted gene was [16].

These analyses have proven the efficiency of our Int-based platform for the functional characterization of these enzymes, as well as its robustness for the further investigation of Ints as genetic switch controllers in eukaryotic cells. Considering that all six integrases showed similar efficiency compared to each other when first used in *E. coli*, the broader range of efficiency obtained in eukaryotic cells can be advantageous when considering the assembly of multicomponent circuits, where variations on potency can help regulate flow rates at defined points in a cascade, allowing a fine-tuning of the system.

Although not further investigated in the present work, the variations in *gfp* expression levels observed both amongst integrases or between mammalian and plant models were partially unexpected. Some works studying integrase applications on eukaryotic models concluded that such variations could occur not only between distant groups like mammalian cells and plants or yeast but also when comparing more closely related organisms, as reported by Xu et al. (2013) [36] when comparing the use of more than ten Ints in human and murine cells. One example can be easily observed with Int TP901: In the mentioned work, this Int was found functional in human cells but incapable of rearranging DNA in mouse stem cells. The same group reported its functionality in *Saccharomyces cerevisiae* with considerably higher efficiency [23], while Guiziou et al. (2023) [52] found no TP901 activity in *Arabidopsis*. Despite reporting and discussing these variations, none of the works presented a deeper investigation into their biological causes. Possible clues can be found in the work published by Chao and collaborators (2021) [50]. They measured recombination kinetic parameters for six Ints, including TP901, to propose a prediction model for systematically reporting Int recombination activity. One important finding in their work is that different Ints can present varying expression levels under the same promoter, and Int concentration can also impact the recombination rate. For instance, phiC31 showed higher recombination rates when under transcriptional control of a weaker Ub promoter than when controlled by the strong CMV promoter.

Table 2. Bacterial host and sequence of attachment sites for the six integrases used by Gomide and collaborators (2020) [16].

Integrase	Host	attB (5' - 3')	attP (5' - 3')
Int2	<i>Streptomyces scabiei</i> 87.22	ggacggcgcagaaaggaggtagctcttcgcccgaccgctc gacatactgctcagctcgtc	gctcaatgtatgtgtctctacgcgcgagattctt cgcccgagaacttctgcaaggcactgctctcttggct
Int4	<i>Streptococcus equi subsp. equi</i> 4047	ttccaaagagcgcaccaacgcgacccctgaaatttgaat aagactgctgcttgtgtaaaaggcagatgatt	caaaaattacaagttttcaacccttgatttga attagcgggtcaaaataaatttgtaattcgtttt
Int5	<i>Streptomyces phage PhiK38-1</i>	gagcggcggatcaggagtggaacggcctgggagc gctaacgctgtgctgcggtcgtgc	ccctaatacgcgaagtgcataaactctcc tgggagcgttgacaacttgcgcaacctgatctg
Int7	<i>Geobacillus sp</i> Y412MC61	agacgagaaaacgttccgtccgtctgggtcagttgggcaa agttgatgaccgggtcgtccgttccctt	ggtgtataaaacctgtgtgagagttaaagtttac atgcctaaccttaactttacgcaggttcagctt
Int9	<i>Staphylococcus aureus</i> str. Newman	tttatattgcgaaaaataatggcgaacagggtaact ggatacctcatcccaataaaattt	gtggttgtttttgttggaaagtgtgtatcaggtatc tgcatagttattccgaaacttccaatta
Int13	<i>Bacillus cytotoxicus</i> NVH 391-98	cgcatacattgttgttttttccagatccagtttggtcc tgtaaatataagcaatcccatgtgagt	caataacggttgtatttttagaacttgac cagttgttttagtaacataaaatacaactccgaata

Adaptation from Yang et al (2014) [15].

<https://doi.org/10.1371/journal.pone.0303999.1002>

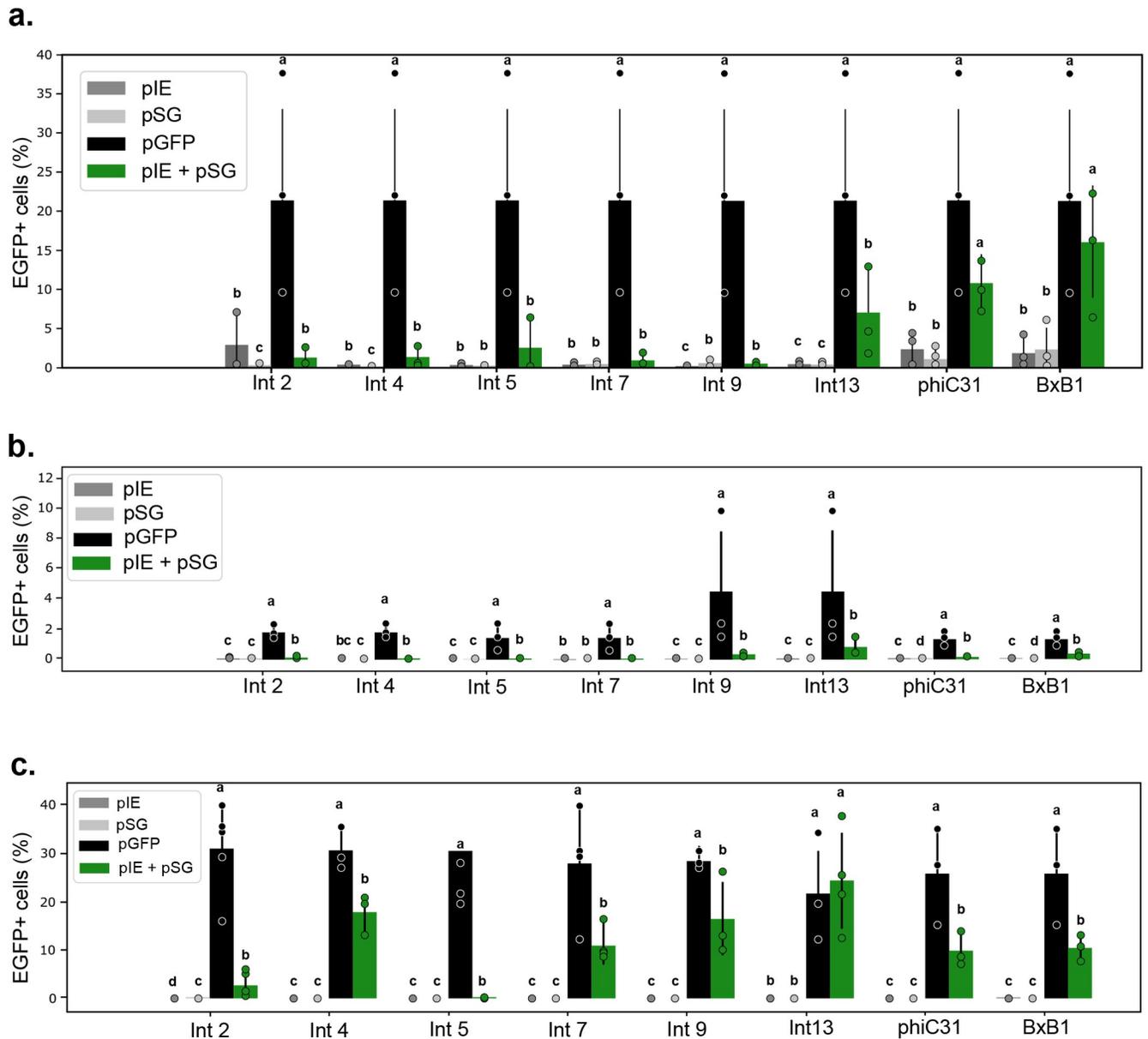


Fig 5. EGFP fluorescent cell ratio between Int S1 treatments in the different eukaryotic systems analyzed (a, human cell; b, bovine fibroblasts cells; and c, *A. thaliana* protoplasts). pIE: effector Integrase Expression plasmid; pSG: reporter Switch GFP plasmid; pGFP: positive control with constitutive expression of *gfp*. Figure adaptation from Gomide et. al., 2020 [16].

<https://doi.org/10.1371/journal.pone.0303999.g005>

These results highlight the complexity of Int efficiency control and how organism-dependent molecular context can affect Int behavior.

Limitations

Although the Int-based platform proved efficient and robust for the functional validation of integrases in eukaryotic cells, some limitations exist. Studies have revealed that Ints are typically very specific to their cognate *att* sites. However, Int shows low levels of off-target activity at pseudo-site sequences similar to their natural *att* sites [17, 53–55]. Additionally, Ints can

cause residual DNA damage or mutations when acting in cells that are not their natural hosts. Various types of damage have been found, including mutations and deletions of the *att* sites, which can make them refractory to later reactions [56]. Cell cytotoxicity is another limiting factor observed when using Ints [23, 36].

As for technical limitations in our platform, the main one we should consider is the discovery of new Ints to be tested since a complete system involves not only the integrase itself but also the identification of the cognate *attB/attP* site pair, which can be more complex to identify than the recombinase gene [15]. The use of RDFs to allow controlled bidirectional recombination is another factor that can broaden Int applications and should be considered. However, limitations including the need of one specific RDF protein for each new serine-integrase, the availability of only a few Int-RDF pairs identified at the moment and laborious validation and optimization required to establish new tools present significant challenges for their use. Regarding analytics, implementing quantitative methods for molecular detection of DNA rearrangement like qPCR assays can bring another valuable parameter for Int efficiency classification. The use of a more robust method for sequencing, like nanopore NGS, can be an alternative in some situations, especially when dealing with low recombination efficiency or a mixed population of edited and not edited targets if an inadequate representation of edited molecules could be preventing identification of DNA rearrangement.

Supporting information

S1 File. Step-by-step protocol collection.

(PDF)

S2 File. Step-by-step protocol for assembly of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells-human.

(PDF)

S3 File. Step-by-step protocol for assembly of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells-animal.

(PDF)

S4 File. Step-by-step protocol for assembly of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells-plant.

(PDF)

S5 File. Step-by-step protocol molecular analyses.

(PDF)

Author Contributions

Conceptualization: Martín H. Bonamino, Elibio Rech.

Funding acquisition: Elibio Rech.

Investigation: Marco A. de Oliveira, Lilian H. Florentino, Thais T. Sales, Rayane N. Lima, Luciana R. C. Barros, Cintia G. Limia, Mariana S. M. Almeida, Maria L. Robledo, Leila M. G. Barros, Eduardo O. Melo, Stevens K. Rehen.

Resources: Eduardo O. Melo, Stevens K. Rehen, Martín H. Bonamino, Elibio Rech.

Writing – original draft: Marco A. de Oliveira, Lilian H. Florentino, Thais T. Sales, Rayane N. Lima, Luciana R. C. Barros, Cintia G. Limia, Eduardo O. Melo, Daniela M. Bittencourt, Stevens K. Rehen, Martín H. Bonamino, Elibio Rech.

Writing – review & editing: Marco A. de Oliveira, Rayane N. Lima, Luciana R. C. Barros, Daniela M. Bittencourt, Martín H. Bonamino, Elibio Rech.

References

1. Arber W. Host-controlled modification of bacteriophage. *Annu Rev Microbiol.* 1965; 19: 365–378. <https://doi.org/10.1146/annurev.mi.19.100165.002053> PMID: 5318444
2. Fogg PCM, Colloms S, Rosser S, Stark M, Smith MCM. New applications for phage integrases. *J Mol Biol.* 2014; 426: 2703–2716. <https://doi.org/10.1016/j.jmb.2014.05.014> PMID: 24857859
3. Sheth RU, Wang HH. DNA-based memory devices for recording cellular events. *Nat Rev Genet.* 2018; 19: 718–732. <https://doi.org/10.1038/s41576-018-0052-8> PMID: 30237447
4. Tian X, Zhou B. Strategies for site-specific recombination with high efficiency and precise spatiotemporal resolution. *J Biol Chem.* 2021; 296: 100509. <https://doi.org/10.1016/j.jbc.2021.100509> PMID: 33676891
5. Meinke G, Bohm A, Hauber J, Pisabarro MT, Buchholz F. Cre recombinase and other tyrosine recombinases. *Chem Rev.* 2016; 116: 12785–12820. <https://doi.org/10.1021/acs.chemrev.6b00077> PMID: 27163859
6. Kim H, Kim M, Im S-K, Fang S. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. *Lab Anim Res.* 2018; 34: 147–159. <https://doi.org/10.5625/lar.2018.34.4.147> PMID: 30671100
7. Abremski K, Hoess R, Sternberg N. Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell.* 1983; 32: 1301–1311. [https://doi.org/10.1016/0092-8674\(83\)90311-2](https://doi.org/10.1016/0092-8674(83)90311-2) PMID: 6220808
8. Nakayama M. VCre/VloxP and SCre/SloxP as reliable site-specific recombination systems for genome engineering. *Methods Mol Biol Clifton NJ.* 2023; 2637: 161–180. https://doi.org/10.1007/978-1-0716-3016-7_13 PMID: 36773146
9. Merrick CA, Zhao J, Rosser SJ. Serine integrases: advancing synthetic biology. *ACS Synth Biol.* 2018; 7: 299–310. <https://doi.org/10.1021/acssynbio.7b00308> PMID: 29316791
10. Stark WM. Making serine integrases work for us. *Curr Opin Microbiol.* 2017; 38: 130–136. <https://doi.org/10.1016/j.mib.2017.04.006> PMID: 28599144
11. Brown WRA, Lee NCO, Xu Z, Smith MCM. Serine recombinases as tools for genome engineering. *Methods.* 2011; 53: 372–379. <https://doi.org/10.1016/j.ymeth.2010.12.031> PMID: 21195181
12. Snoeck N, De Mol ML, Van Herpe D, Goormans A, Maryns I, Coussement P, et al. Serine integrase recombinational engineering (SIRE): A versatile toolbox for genome editing. *Biotechnol Bioeng.* 2019; 116: 364–374. <https://doi.org/10.1002/bit.26854> PMID: 30345503
13. Zhao J, Pokhilkov A, Ebenhöf O, Rosser SJ, Colloms SD. A single-input binary counting module based on serine integrase site-specific recombination. *Nucleic Acids Res.* 2019; 47: 4896–4909. <https://doi.org/10.1093/nar/gkz245> PMID: 30957849
14. Abioye J, Lawson-Williams M, Lecanda A, Calhoun B, McQue AL, Colloms SD, et al. High fidelity one-pot DNA assembly using orthogonal serine integrases. *Biotechnol J.* 2023; 18: e2200411. <https://doi.org/10.1002/biot.202200411> PMID: 36504358
15. Yang L, Nielsen AAK, Fernandez-Rodriguez J, McClune CJ, Laub MT, Lu TK, et al. Permanent genetic memory with >1-byte capacity. *Nat Methods.* 2014; 11: 1261–1266. <https://doi.org/10.1038/nmeth.3147> PMID: 25344638
16. Gomide MS, Sales TT, Barros LRC, Limia CG, de Oliveira MA, Florentino LH, et al. Genetic switches designed for eukaryotic cells and controlled by serine integrases. *Commun Biol.* 2020; 3: 255. <https://doi.org/10.1038/s42003-020-0971-8> PMID: 32444777
17. Gupta K, Sharp R, Yuan JB, Li H, Van Duyne GD. Coiled-coil interactions mediate serine integrase directionality. *Nucleic Acids Res.* 2017; 45: 7339–7353. <https://doi.org/10.1093/nar/gkx474> PMID: 28549184
18. Li H, Sharp R, Rutherford K, Gupta K, Van Duyne GD. Serine integrase attP binding and specificity. *J Mol Biol.* 2018; 430: 4401–4418. <https://doi.org/10.1016/j.jmb.2018.09.007> PMID: 30227134
19. Mandali S, Johnson RC. Control of the serine integrase reaction: roles of the coiled-coil and helix e regions in DNA site synapsis and recombination. *J Bacteriol.* 2021; 203: e00703–20. <https://doi.org/10.1128/JB.00703-20> PMID: 34060907
20. Yuan P, Gupta K, Van Duyne GD. Tetrameric structure of a serine integrase catalytic domain. *Struct Lond Engl* 1993. 2008; 16: 1275–1286. <https://doi.org/10.1016/j.str.2008.04.018> PMID: 18682229

21. Fan H-F, Hsieh T-S, Ma C-H, Jayaram M. Single-molecule analysis of ϕ C31 integrase-mediated site-specific recombination by tethered particle motion. *Nucleic Acids Res.* 2016; 44: 10804–10823. <https://doi.org/10.1093/nar/gkw861> PMID: 27986956
22. Turan S, Zehe C, Kuehle J, Qiao J, Bode J. Recombinase-mediated cassette exchange (RMCE)—a rapidly-expanding toolbox for targeted genomic modifications. *Gene.* 2013; 515: 1–27. <https://doi.org/10.1016/j.gene.2012.11.016> PMID: 23201421
23. Z X, Wr B. Comparison and optimization of ten phage encoded serine integrases for genome engineering in *Saccharomyces cerevisiae*. *BMC Biotechnol.* 2016; 16. <https://doi.org/10.1186/s12896-016-0241-5> PMID: 26860416
24. Fogg PCM, Younger E, Fernando BD, Khaleel T, Stark WM, Smith MCM. Recombination directionality factor gp3 binds ϕ C31 integrase via the zinc domain, potentially affecting the trajectory of the coiled-coil motif. *Nucleic Acids Res.* 2018; 46: 1308–1320. <https://doi.org/10.1093/nar/gkx1233> PMID: 29228292
25. Fan H-F, Su B-Y, Ma C-H, Rowley PA, Jayaram M. A bipartite thermodynamic-kinetic contribution by an activating mutation to RDF-independent excision by a phage serine integrase. *Nucleic Acids Res.* 2020; 48: 6413–6430. <https://doi.org/10.1093/nar/gkaa401> PMID: 32479633
26. Kuhstoss S, Richardson MA, Rao RN. Plasmid cloning vectors that integrate site-specifically in *Streptomyces* spp. *Gene.* 1991; 97: 143–146. [https://doi.org/10.1016/0378-1119\(91\)90022-4](https://doi.org/10.1016/0378-1119(91)90022-4) PMID: 1995427
27. Durrant MG, Fanton A, Tycko J, Hinks M, Chandrasekaran SS, Perry NT, et al. Systematic discovery of recombinases for efficient integration of large DNA sequences into the human genome. *Nat Biotechnol.* 2022. <https://doi.org/10.1038/s41587-022-01494-w> PMID: 36217031
28. Ba F, Liu Y, Liu W-Q, Tian X, Li J. SYMBIOSIS: synthetic manipulable biobricks via orthogonal serine integrase systems. *Nucleic Acids Res.* 2022; 50: 2973–2985. <https://doi.org/10.1093/nar/gkac124> PMID: 35191490
29. Brophy JAN, Voigt CA. Principles of genetic circuit design. *Nat Methods.* 2014; 11: 508–520. <https://doi.org/10.1038/nmeth.2926> PMID: 24781324
30. Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Endy D. Amplifying genetic logic gates. *Science.* 2013; 340: 599–603. <https://doi.org/10.1126/science.1232758> PMID: 23539178
31. Roquet N, Soleimany AP, Ferris AC, Aaronson S, Lu TK. Synthetic recombinase-based state machines in living cells. *Science.* 2016; 353: aad8559. <https://doi.org/10.1126/science.aad8559> PMID: 27463678
32. Courbet A, Endy D, Renard E, Molina F, Bonnet J. Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates. *Sci Transl Med.* 2015; 7: 289ra83. <https://doi.org/10.1126/scitranslmed.aaa3601> PMID: 26019219
33. Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, et al. Genetic circuit design automation. *Science.* 2016; 352: aac7341. <https://doi.org/10.1126/science.aac7341> PMID: 27034378
34. Jones TS, Oliveira SMD, Myers CJ, Voigt CA, Densmore D. Genetic circuit design automation with Cello 2.0. *Nat Protoc.* 2022; 17: 1097–1113. <https://doi.org/10.1038/s41596-021-00675-2> PMID: 35197606
35. Park Y, Espah Borujeni A, Gorochofski TE, Shin J, Voigt CA. Precision design of stable genetic circuits carried in highly-insulated *E. coli* genomic landing pads. *Mol Syst Biol.* 2020; 16: e9584. <https://doi.org/10.15252/msb.20209584> PMID: 32812710
36. Xu Z, Thomas L, Davies B, Chalmers R, Smith M, Brown W. 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.* 2013; 13: 87. <https://doi.org/10.1186/1472-6750-13-87> PMID: 24139482
37. Andreas S, Schwenk F, Küter-Luks B, Faust N, Kühn R. Enhanced efficiency through nuclear localization signal fusion on phage PhiC31-integrase: activity comparison with Cre and FLP ϵ recombinase in mammalian cells. *Nucleic Acids Res.* 2002; 30: 2299–2306. <https://doi.org/10.1093/nar/30.11.2299> PMID: 12034816
38. Matsuda T, Cepko CL. Electroporation and RNA interference in the rodent retina in vivo and in vitro. *Proc Natl Acad Sci U S A.* 2004; 101: 16–22. <https://doi.org/10.1073/pnas.2235688100> PMID: 14603031
39. Rech EL, Vianna GR, Aragão FJL. High-efficiency transformation by biolistics of soybean, common bean and cotton transgenic plants. *Nat Protoc.* 2008; 3: 410–418. <https://doi.org/10.1038/nprot.2008.9> PMID: 18323812
40. Wickramasinghe SN. Culture of animal cells. A manual of basic technique, 3rd edn. R. Ian Freshney, Wiley-Liss, Inc: New York. xxiv + 486 pages (1994). *Cell Biochem Funct.* 1996; 14: 75–76.
41. Yoo S-D, Cho Y-H, Sheen J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc.* 2007; 2: 1565–1572. <https://doi.org/10.1038/nprot.2007.199> PMID: 17585298

42. Cardarelli F, Digiacomo L, Marchini C, Amici A, Salomone F, Fiume G, et al. The intracellular trafficking mechanism of Lipofectamine-based transfection reagents and its implication for gene delivery. *Sci Rep*. 2016; 6: 25879. <https://doi.org/10.1038/srep25879> PMID: 27165510
43. Chicaybam L, Barcelos C, Peixoto B, Carneiro M, Limia CG, Redondo P, et al. An efficient electroporation protocol for the genetic modification of mammalian cells. *Front Bioeng Biotechnol*. 2016; 4: 99. <https://doi.org/10.3389/fbioe.2016.00099> PMID: 28168187
44. Krens FA, Molendijk L, Wullems GJ, Schilperoort RA. In vitro transformation of plant protoplasts with Ti-plasmid DNA. *Nature*. 1982; 296: 72–74. <https://doi.org/10.1038/296072a0>
45. Paszkowski J, Shillito RD, Saul M, Mandák V, Hohn T, Hohn B, et al. Direct gene transfer to plants. *EMBO J*. 1984; 3: 2717–2722. <https://doi.org/10.1002/j.1460-2075.1984.tb02201.x> PMID: 16453573
46. De Buck S, Jacobs A, Van Montagu M, Depicker A. *Agrobacterium tumefaciens* transformation and cotransformation frequencies of *Arabidopsis thaliana* root explants and tobacco protoplasts. *Mol Plant-Microbe Interact MPMI*. 1998; 11: 449–457. <https://doi.org/10.1094/MPMI.1998.11.6.449> PMID: 9612943
47. Hauptmann RM, Ozias-Akins P, Vasil V, Tabaeizadeh Z, Rogers SG, Horsch RB, et al. Transient expression of electroporated DNA in monocotyledonous and dicotyledonous species. *Plant Cell Rep*. 1987; 6: 265–270. <https://doi.org/10.1007/BF00271995> PMID: 24248756
48. Malla S, Dafhnis-Calas F, Brookfield JFY, Smith MCM, Brown WRA. Rearranging the centromere of the human Y chromosome with phiC31 integrase. *Nucleic Acids Res*. 2005; 33: 6101–6113. <https://doi.org/10.1093/nar/gki922> PMID: 16246911
49. Sangiorgi E, Shuhua Z, Capecchi MR. In vivo evaluation of PhiC31 recombinase activity using a self-excision cassette. *Nucleic Acids Res*. 2008; 36: e134. <https://doi.org/10.1093/nar/gkn627> PMID: 18829714
50. Chao G, Travis C, Church G. Measurement of large serine integrase enzymatic characteristics in HEK293 cells reveals variability and influence on downstream reporter expression. *FEBS J*. 2021; 288: 6410–6427. <https://doi.org/10.1111/febs.16037> PMID: 34043859
51. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol*. 2021; 39: 1348–1365. <https://doi.org/10.1038/s41587-021-01108-x> PMID: 34750572
52. Guiziou S, Maranas CJ, Chu JC, Nemhauser JL. An integrase toolbox to record gene-expression during plant development. *Nat Commun*. 2023; 14: 1844. <https://doi.org/10.1038/s41467-023-37607-5> PMID: 37012288
53. Hu Z, Chen L, Jia C, Zhu H, Wang W, Zhong J. Screening of potential pseudo att sites of *Streptomyces* phage Φ C31 integrase in the human genome. *Acta Pharmacol Sin*. 2013; 34: 561–569. <https://doi.org/10.1038/aps.2012.173> PMID: 23416928
54. Herisse M, Porter JL, Guerillot R, Tomita T, Goncalves Da Silva A, Seemann T, et al. The Φ BT1 large serine recombinase catalyzes DNA integration at pseudo-attB sites in the genus *Nocardia*. *PeerJ*. 2018; 6: e4784. <https://doi.org/10.7717/peerj.4784> PMID: 29740520
55. Bessen JL, Afeyan LK, Dančik V, Koblan LW, Thompson DB, Leichner C, et al. High-resolution specificity profiling and off-target prediction for site-specific DNA recombinases. *Nat Commun*. 2019; 10: 1937. <https://doi.org/10.1038/s41467-019-09987-0> PMID: 31028261
56. Calos MP. Phage integrases for genome editing. In: Cathomen T, Hirsch M, Porteus M, editors. *Genome editing: the next step in gene therapy*. New York, NY: Springer; 2016. pp. 81–91.

Anexo B

Artigo Científico

Development of Int-Plex@ binary memory switch system: plant genome modulation driven by large serine-integrases.

1 **Development of Int-Plex@ binary memory switch system: plant genome**
2 **modulation driven by large serine-integrases.**

3

4 Marco Antônio de Oliveira^{1*}, Rayane Nunes Lima^{1*}, Lilian Hasegawa Florentino¹,
5 Mariana Martins Severo de Almeida¹, Fernando Lucas Melo², Raquel Vasquez Bonnet
6 ¹, Cristiano Lacorte¹, Elibio Rech^{1†}.

7

8 ¹Embrapa Genetic Resources and Biotechnology/National Institute of Science and
9 Technology in Synthetic Biology, Brasilia 70770-917, Brazil.

10

11 ² Onsite Genomics, Sig Q.3 Bl.C SN Lt. 42 Lj. 37, Brasilia, 70610-433, Brazil.

12

13 *Contributed equally as first author.

14

14 †Corresponding author:

15

15 Elibio Rech, Phone: +55 61 9 8404 2559, e-mail: elibio.rech@embrapa.br

16

17 **ABSTRACT**

18 The comprehension of virus-host interactions has allowed numerous advances in
19 developing biotechnological methodologies for plant genome editions, constituting a
20 promising path for plant genetic engineering. Among these advancements, phage-
21 encoded large serine-integrases have emerged as noteworthy tools to modulate plant
22 metabolic pathways by inserting, excising, or inverting DNA stretches in a reversible
23 and specific way. The present work shows the foundation of the Int-Plex@ (INTEGRase
24 PLant EXpression) binary memory switch system, which consists of the application of
25 four distinct orthogonal prophage large serine-integrases (Int) (BxB1, phiC31, Int13,
26 and Int9) as an input trigger mechanism for the inversion or excision of genomic DNA.
27 The memory genetic switch is divided into the excision module and the inversion
28 module. The excision module is activated by BxB1 or phiC31 enzymes (input). In this
29 case, the DNA sequence flanked by its attachment sites is excised from the genome
30 (output). The inversion module is activated by Int9 or Int13 (input). The inverted
31 *mgf* gene sequence is flipped to its functional coding sequence, and the switch output is
32 mGFP. Moreover, prokaryotic-based cell-free *in vitro* transcription-translation reactions
33 (TxTI) were used as a fast platform for testing Ints *attB/P in tandem* site activity.
34 Furthermore different plasmid delivery strategies for plant cell Int heterologous
35 expression were tested: leaf tissue agroinfiltration of *Agrobacterium*
36 *tumefaciens* transformed with binary plasmids and a biolistic system. After each
37 treatment, the edited genomic DNA sequences were amplified and verified by Sanger
38 and Nanopore sequencing. Despite the challenges of using Ints, the potential benefits
39 are significant and deserve deeper exploration and development. The Int-Plex@ binary
40 genome memory switch system can be applied to produce genetic circuits combined
41 with omics tools and sgRNAs to engineer and modulate plant metabolic pathways
42 temporally and reversibly.

43

44 **Keywords:** orthogonal phage serine integrase, genome editing, nanopore sequencing,
45 plant biotechnology, synthetic biology.

46 1. INTRODUCTION

47

48 Over the last several decades, major advances in plant synthetic biology have emerged
49 from significant discoveries based on the understanding of the co-evolutionary arms
50 race between viruses and their hosts. These ongoing battles have led to the evolution of
51 intricate and complex molecular interactions that were exploited to create novel
52 biotechnological tools based on the bottom-up approach for plant genome modulation.
53 Examples include (I) bacteria restriction endonucleases and phage DNA ligases, which
54 allowed the revolution of recombinant DNA technology and are used in
55 BioBrick™ assembly and modern one-step DNA assembly techniques¹⁻⁷; (II) viral
56 vectors for gene delivery^{8,9}; (III) virus-induced gene silencing (VIGS)^{10,11}; (IV) the
57 RNA interference (RNAi) technology, which is widely used in gene down-regulation
58 studies¹²⁻¹⁴; and (V) CRISPR-Cas systems, which has recently been widely applied in
59 crop genome editing technology^{12,15-18}. However, these last two technologies are
60 irreversible and permanent; once the plant gene is silenced or edited, it cannot return to
61 its original state. Therefore, to overcome these limitations, a promising approach,
62 originating from bacteriophage-prokaryotic interactions, is the application of
63 filamentous phage-encoded large serine integrases (Ints), which are multifunctional
64 recombinases with the potential to edit DNA sequences and also serve as efficient *in*
65 *vitro* DNA assemblers¹⁹⁻²². Therefore, considering the development of a dynamic and
66 reversible genome modulation system in plant cells, it is possible to include the
67 orthogonal Ints in modular synthetic genetic devices based on Boolean logic gates²³⁻²⁹.
68 Ints catalyze directly and unidirectionally a site-specific insertion, deletion or 180°
69 inversion of DNA sequence flanked by the two attachment sites (*att*), *attB* (bacteria)
70 and *attP* (phage) depending on the design of their orientation relative to each
71 other, which results in the formation of two new post-recombination attachment
72 sites, *attL* (left) and *attR* (right)^{19,20,25-28,30}. BxB1 and phiC31 Ints have been
73 successfully used in plant engineering and gene stacking^{29,31-34}.
74 Furthermore, using a bioinformatics approach, Yang *et al.* prospected 11 new Ints with
75 their respective functional *att* sites²⁰. Subsequently, Gomide *et al.* showed that those
76 Ints function in eukaryotic cells and ranked Int9 and Int13 among them as the most
77 effective Ints for plasmid flipping in the plant cell context²⁵. Thus, the IntPlex@ binary
78 memory switch takes BxB1, phiC31, Int9, and Int13 as input signals to control an
79 output (genome edition) in a user-defined manner (site-specific excision or inversion).
80 Lastly, the possibility of using distinct plasmid delivery as a general method for
81 integrase expression was examined.

82

83 2. RESULTS

84

85 2.1 Int-Plex@ binary memory switch system design

86 We have assembled a multifunctional genetic switch system based on *att* site pairs from
87 4 Ints, namely BxB1, phiC31, Int9, and Int13. The cassette constructs consist of the
88 reverse complement sequence of green fluorescent protein reporter gene *gfp* (*mgfp* or
89 *degfp*) flanked by in tandem *attB* or *attP* sites from the Ints selected (Figure 1A).
90 Promoter and terminator sequences to complete the transcription unit varied according
91 to the *in vitro* or *in vivo* experimental model. Effector Ints were delivered individually
92 in a separate plasmid under transcriptional control of a constitutive promoter^{25,35}. This
93 configuration allows the formation of a binary system with an Excision Module
94 controlled by BxB1 or phiC31 recombinases designed as a NOR logic gate, in which the
95 presence of either Int (input) prevents reporter expression due to its deletion (output)

96 and an Inversion Module controlled by Int9 or Int13 with the structure of an XOR gate
97 (Figure 1B). To achieve varying outcomes in DNA rearrangement depending on the Int
98 acting on the system, we inserted the *att* sites in the construction either in the same or
99 opposing orientation relative to their respective site counterpart. While for BxB1 and
100 phiC31, *attB* and *attP* sites are in sense orientation, with their recombination resulting in
101 excision of the DNA sequence flanked by them from the construction (Figure 1, C and
102 D), Int9 and Int13 *att* sites are in reverse complement orientation, which means that
103 upon recombination, the target DNA sequence will be 180° flipped (output), modulating
104 the *gfp* expression (Figure 1, E and F). Although both Int9 or Int13 are capable of
105 unidirectionally flipping the reporter gene coding sequence and activating its
106 transcription, an essential feature of this type of logic gate is the gene set/reset, that
107 upon concurrent or sequential introduction of both enzymes, a second inversion event
108 will result in the return of the target DNA orientation to the original OFF state, silencing
109 reporter expression (Figure 1, G and H). In this double-editing event, one Int will
110 recognize its *attB/P* sites and catalyze the inversion of the reporter from its OFF state to
111 ON state, forming the intermediate plasmid with *attL/R* sites which results in the GFP
112 production. Instantaneously, this intermediate plasmid is the substrate for the action of
113 the second Int, which, upon recombination of its *attB/P* sites, will result in the formation
114 of the final plasmid with both *attL/R* sites for Int9 and Int13 with consequent recovery
115 from the ON state to initial OFF state, silencing *gfp* expression (Figure 1, G and H).

116

117 **2.2 *In vitro* function of Int-Plex@ binary memory switch system**

118 Our first goal was to understand the behavior of this new binary memory system and
119 ensure that the presence of *att* sites *in tandem* would not interfere with expected Int
120 activity. For the functional evaluation, we chose cell-free *in vitro* transcription-
121 translation reactions (TxTI), given its valuable application as a fast platform for testing
122 Int activity, DNA editing tools and genetic circuits^{35,36}. The switch was assembled using
123 TxTI optimized parts, including the *degfp* gene as a reporter, Ribosome Binding Site
124 (RBS) and the T7Max transcription system³⁵ (Figure 2A). Reactions containing the
125 reporter construction alone or with plasmids for Int expression were incubated overnight
126 and analyzed for DNA rearrangement and deGFP fluorescence emission. This enabled
127 us to test the efficacy of each reaction independently and transiently *in vitro* for the first
128 time before integrating the memory switch into the plant genome. Besides deGFP
129 expression, DNA excisions by BxB1 or phiC31 were detected, as predicted, by the PCR
130 amplifications (Figure 2, B and C). The excised circular DNA molecule was confirmed
131 by Sanger sequencing, where post-recombination *attR* site for BxB1 is present (Figure
132 3, D and E). Sanger sequencing of phiC31 treatments is in progress. All sequences are
133 available for consultation at the link provided in the Data availability section. In
134 addition, the introduction of Int9 or Int13 could invert the DNA and allowed deGFP
135 production *in vitro* (Figure 3, A and B). Although the presence of both Int9 and Int13 in
136 the same reaction must result in the re-establishment of the initial OFF state with *degfp*
137 silencing, the closed nature of the reaction environment and relatively short incubation
138 time preserved the protein produced while the single-edited intermediate was present,
139 maintaining a high level of fluorescence. The double-editing event was confirmed by
140 Sanger sequencing of the reporter construct, where both post-
141 recombination *attL/attR* site pairs for Int9 and Int13 are present in the same molecule
142 (Figure 3, C-E).

143

144 **2.3 Int-Plex@ binary memory switch system design for plant genome modulation**

145 To demonstrate the applicability of this tool in DNA modulation, we also verified the
146 functionality of the switch stably integrated into the plant genome. Here, we assembled
147 the reverse complement sequence of *mgfp* flanked by four in tandem BxB1, phiC31,
148 Int9, and Int13 *attB* and *attP* sites in a synthetic genetic switch, denoted IntPlex@
149 binary memory switch and inserted into the *Nicotiana benthamiana* genome. Each
150 serine integrase was plant codon optimized and inserted in a different set of plasmids
151 for two systems of gene delivery: agroinfiltration by *Agrobacterium tumefaciens* and
152 biolistic. The IntPlex@ binary memory switch takes these four serine integrase input
153 signals to control an output (genome edition) in a user-defined manner (excision or
154 inversion).

155

156 **2.3.1 Plant genome excision triggered by phiC31 or BxB1**

157 BxB1 or phiC31 catalyzed the excision of *mgfp* from the genome. Upon delivery of the
158 effector plasmids containing the integrases by agroinfiltration, molecular analyses of
159 plant DNA within five d.p.i (days post infiltration) showed successful deletion, with
160 genome sequencing confirmation of proper formation of respective *attL* sites for each
161 Int (Figure 4A and Figure 5) and loss of DNA sequence flanked by its attachment sites
162 (output) (Figure 4, B and C). Sanger sequencing of phiC31 treatments is in progress.
163 Moreover, Nanopore sequencing was utilized to identify the excised circular DNA
164 molecule containing *mgfp*, BxB1 *attR* site and remaining *attB/P* sites from the other
165 integrases (Figure 6). The sequencing run lasted four hours, yielding roughly 800,000
166 reads. As depicted in Figure 7, the basecalled read lengths exhibited a bimodal
167 distribution, with the median read length positioned within the distribution's upper
168 mode. Quality scores ranged along the x-axis, predominantly showing mid to high
169 values indicative of reliable base calling accuracy. This pattern reflects a diverse set of
170 read lengths, with shorter reads tending to have higher quality scores. For subsequent
171 analysis, only reads with quality scores of 10 or above were maintained (383,245 reads).
172 However, the detection of phiC31 output and stability and processing time for the
173 circular DNA excised degradation remains to be evaluated. Regarding the delivery of
174 the Int effector by biolistic, neither genome editing, nor the output molecule were
175 detected (data not shown). Nanopore sequencing will be applied to detect the output
176 molecules and analyze the low detection rate. All sequences are available for
177 consultation at the link provided in the Data availability section.

178

179 **2.3.2 Plant genome inversion triggered by Int 9 or Int 13**

180 This switch operates as an ON switch, where MGFP is expressed in the ON state and
181 not in the OFF state. Upon delivery of the effector plasmids containing the integrases by
182 agroinfiltration, molecular analyses of plant DNA within five d.p.i showed successful
183 DNA inversion. Int9 recognized its *attB/P* sites and catalyzed the inversion
184 of *mgfp* from its OFF state to ON state, resulting in the formation of the intermediate
185 genome with *attL/Int9* and *attR/Int9* sites and the output MGFP after int delivery via
186 agroinfiltration (Figure 8A). Likewise, Int13 recognized its *attB/P* sites and catalyzed
187 the inversion of *mgfp* from its OFF state to ON state, resulting in the formation of the
188 intermediate genome with *attL/Int13* and *attR/Int13* sites and the output MGFP (Figure
189 8B). As for effector Int delivery by biolistic, phenotypical analysis showed no increase
190 in mGFP fluorescence (data not shown). However, fluorescence could be detected in the
191 positive control of plants bombarded with microparticles coated with a plasmid
192 constitutively expressing eGFP³⁷. Despite the lack of signal, DNA inversion could be
193 detected for both Ints by PCR amplification, with confirmation of inversion and
194 recombined *attL/R* sites by DNA sequencing.

195 **2.3.3 Plant genome modulation: Turning DNA ON and OFF**

196 While a one-way memory switch is functional, reversibly turning a gene on and off
197 would allow interesting bidirectional regulation. Thus, Int9 and Int13 were
198 agroinfiltrated into the plant five days apart to turn *mgfp* on and off in the same cell.
199 Initially, Int 9 recognized its *attB/P* sites and catalyzed the inversion of *mgfp* from its
200 OFF state to ON state, resulting in the formation of the intermediate genome with
201 *attL/Int9* and *attR/Int9* sites and the output MGFP. Five days after, this intermediate
202 genome was the substrate for the action of Int13 that recognized its *attB/P* sites and
203 catalyzed the inversion of *mgfp* from its ON state to OFF state, resulting in the
204 formation of the final genome with both *attL/Int9* and *attR/Int9* and *attL/Int13* and
205 *attR/Int13* sites. Thus, the system presents the MGFP until the ON state mRNA
206 molecule can no longer be translated according to the MGFP half-life time. The Sanger
207 sequencing of Int9 and Int13 treatments are in progress. All sequences are available for
208 consultation at the link provided in the Data availability section.
209

210

211 **3. DISCUSSION**

212

213 In this study, our primary goal was to broaden the plant genome modulation toolbox by
214 creating a binary memory switch controlled by serine integrases. Currently, the
215 available array of tools for constructing synthetic memory systems in plants is quite
216 limited, mainly relying on phiC31 and BxB1^{29,31,32,38,39}. Efforts for the systematic
217 identification of new Ints and their *att* sites, including metagenomic studies,
218 development of new bioinformatic approaches, and functional characterization,
219 contribute to expanding the availability of such tools and enhance the possibilities for
220 assembly of more complex multicomponent genetic circuits⁴⁰. In that context, we
221 applied phiC31, BxB1, and two more recent Ints (Int9 and Int13) previously
222 characterized in a plant model to create a bimodular switch for DNA excision and
223 bidirectional inversion²⁵. All four Int successfully rearranged the target DNA inserted
224 into the genome of *N. benthamiana* as programmed, confirming the functionality of
225 both modules in our system.

226 More than assembling a new switch system for gene expression regulation, this proof-
227 of-concept demonstration paves the way for applying new Int candidates other than
228 phiC31 and BxB1 in plants. Beyond orthogonality, which allows for various effectors in
229 the same circuit, different integrases can show a range of efficiency degrees beneficial
230 for implementing flow control nodes in a synthetic regulatory pathway. As
231 demonstrated by Gomide et al. (2020), for instance, Int13 and Int5 both could invert a
232 target DNA sequence in *Arabidopsis thaliana* protoplasts, although resulting in different
233 levels of reporter expression²⁵.

234 Another essential advantage point of IntPlex@ is the use of independent Ints to control
235 inversion directionality. After the first inversion event by Int9 or Int13, restoration to
236 the initial OFF state depends exclusively on introducing the second enzyme, with no
237 chance of interference. In contrast, the RESET system published by Bernabé-Orts et al.
238 (2020) applies phiC31 with its correspondent RDF protein to revert unidirectionality²⁹.
239 Although aiming at somewhat different outcomes, it is noteworthy that with RDF, they
240 observed an increase in system instability. Recombination of *attL/R* sites requires direct
241 interaction of Int-RDF, but the presence of free phiC31 led to new recombination of the
242 *attB/P* sites just restored, resulting in mixed states of the switch at the same time²⁹.

243 The Sanger Sequencing method is widely used for identifying plant genome editions.
244 Nevertheless, it comes with some limitations, including the prerequisite for pre-existing

245 knowledge of the DNA sequence and a constrained ability to simultaneously detect
246 various DNA fragments. On the other hand, the utilization of Nanopore DNA
247 sequencing has enabled determination of total DNA products of the plant cells genome
248 edition reactions in a short time and with high-throughput data analysis in real time.
249 Nanopore sequencing holds great promise for detecting edited DNA and when
250 combined with stoichiometry analysis, this technology provides a powerful toolset to
251 access genomic diversity for studying edited DNA at a quantitative level. The
252 integration of nanopore sequencing and stoichiometry analysis not only enhances the
253 ability to detect edits accurately but also facilitates a deeper understanding of the
254 dynamics and efficiency of serine integrase editing reactions.

255 It is worth mentioning that when a system is imported from the prokaryotic to the
256 eukaryotic cell, as is the case for bacteriophage-derived recombinases, it is necessary to
257 consider the presence of the nucleus as a physical barrier to the performance of enzymes
258 that act on DNA. Nevertheless, all four integrases tested in this work were able to
259 overcome the nuclear barrier and interact with chromatin without the need to add NLS
260 or other protein engineering. It is possible that the passage through the nuclear pore
261 occurred due to the presence of intrinsic NLS in its native structure after exposure of
262 positively charged amino acids during protein folding²⁵. Notably, the system functioned
263 consistently over two plant generations. Employing large serine-integrases in modular
264 switch devices exemplifies a sophisticated and controlled approach, affording precision
265 in genetic alterations with the potential for intricate modulation of plant physiological
266 processes. Consequently, the integration of large serine-integrases into the
267 biotechnological toolkit heralds a promising trajectory for refining and expanding
268 strategies in plant genetic engineering. However, further research is needed to fully
269 understand the potential of this system in plant biotechnology, the potential for
270 unintended off-target effects, and to optimize its use.

271

272 **4. CONCLUSIONS**

273

274 PhiC31, BxB1, Int9, and Int13 promoted efficient directional DNA site-specific
275 recombination of complex chromatin of the plant cell, and thus have major potential
276 applications in genome engineering and metabolic pathway engineering. This memory
277 genetic switch system is dynamic and may be combined with other technologies like
278 inducible promoters and advanced sequencing tools. For example, implementing the
279 sequence barcoding with Nanopore sequencing amplifies the capabilities of genomic
280 analysis and allows for comprehensive profiling of the entire edited genome, capturing
281 structural variations or ensuring the fidelity of the editing process. Together, this
282 combined approach not only validates the accuracy of serine integrase-mediated edits at
283 a detailed level but also offers a holistic view of the entire genome modulation,
284 fostering a deeper understanding of the genomic landscape post-editing. Lastly, it is
285 more than evident that we are living in an era marked by the development of tools that
286 allow crop genomes to be edited to provide resistance to environmental challenges,
287 improve nutritional content and optimize defenses against pathogens. However, the lack
288 of efficacy of plasmid delivery systems has hampered advancements in the precision of
289 gene editing. This is a crucial moment that requires the convergence of efforts to
290 develop studies that aim to refine plasmid or gene delivery systems, ensuring that
291 genetic modifications reach the maximum number of cells efficiently.

292

293 **5. MATERIAL AND METHODS**

294

295 **5.1 mGFP genetic switch plasmid design**

296 In pLSB_pCAMBIA2300_IntX or pUC_IntX plasmids, both *actin2* gene promoter and
297 *NOS* terminator drive (a) plant codon-optimized serine integrases transcription for
298 heterologous expression or (b) serine integrases fused with NLS signal. In
299 pLSB_PVX_GW_IntX plasmids, both subgenomic CP promoter and *NOS* terminator
300 drive serine integrase transcription for heterologous expression with or without the NLS
301 signal.

302

303 **5.2 Serine-Integrase Plasmid Design**

304 The *Int9*, *Int13*, *phiC31*, and *BxB1* genes were retrieved from pUC57²⁵ and cloned into
305 PVX-GW binary plasmid³⁷ by Epoch Life Science Inc. These plasmids resulted in a set
306 of serine integrase expression plasmids, individually called pLSB_PVX_GW_X (X =
307 *Int9*, *Int13*, *phiC31*, or *BxB1*). The *Int9*, *Int13*, *phiC31*, and *BxB1* genes controlled by
308 the *actin2* gene promoter and *NOS* terminator were retrieved from pUC57 and cloned in
309 pCAMBIA2300 plasmid by Epoch Life Science Inc²⁵. These plasmids resulted in a set
310 of serine integrase expression plasmids, individually called pLSB_pCAMBIA2300_X
311 (X = *Int9*, *Int13*, *phiC31*, or *BxB1*).

312

313 **5.3 Plant transformation**

314 Transgenic plants of *Nicotiana benthamiana* were produced at University of California
315 Riverside's Department of Botany and Plant Science performed the plant
316 transformation. pART27 plasmid containing the genes BXB1_PHIC31_9_13_mGFP(rc)
317 and nptII were delivered into GV3101 *Agrobacterium tumefaciens* via electroporation.
318 Antibiotic-resistant colonies were selected and confirmed by colony PCR for glycerol
319 stocks. The transformed *A. tumefaciens* strain was used to inoculate the leaf tissue of
320 six-week-old plants. Callus induction and shoot induction media containing MS salts,
321 MS vitamins, 0.56 0.56mM Myo, 8.84 8.84 μ M BAP, 0.54 0.54 μ M NAA, 3% sucrose,
322 0.21 0.21 mM kanamycin, 0.55 0.55 mM Cefatoxime and 1.321.32 mM Carbenecillin
323 were used to induce calli and shoots. Thirty and ten shoots transformed with
324 BXB1_PHIC31_9_13_mGFP(rc) and pART27-nptII (empty plasmid) respectively were
325 transferred to rooting media containing ½ MS salts, MS vitamins, 0.56 0.56 mM Myo-
326 inositol, 0.98 0.98 μ M IBA, 1.5% sucrose, 0.34 0.34 mM vancomycin, 0.55 0.55 mM
327 cefotaxime, and 1.32 mM Carbenicillin. A total of 40 *N. benthamiana* plants were
328 analyzed by end-PCR using specific primers for each construct. Twenty-one plants
329 amplified an expected band of 717 bp with the specific primers FwmGFP/RvmGFP, and
330 nine plants transformed with pART27 (empty plasmid) amplified a band of 699 bp for
331 the nptII gene. *Nicotiana benthamiana* transgenic seeds germinated in vermiculite, and
332 plantlets were grown in a greenhouse. Eight-week-old plants were used for the
333 *Agrobacterium* infiltration or biolistic procedure.

334

335 **5.4 In vitro Transcription/Translation Reactions**

336 *In vitro*, Transcription/Translation Reactions, or TxTl reactions, were performed with
337 myTxTl T7 Expression Kit (Daicel Arbor Biosciences) following manufacturer
338 instructions. Briefly, 24 μ L reactions were assembled in 1.5 mL centrifugation tubes
339 with 18 μ L Sigma 70 Master Mix, 0.1 nM P70a-T7rnep HP, and 10 nM of each plasmid
340 (reporter + effector Int). Reactions were incubated at 29°C for 16 h. After incubation,
341 GFP production was visually assessed with an LED blue light Transilluminator
342 (KASVI), and 2 μ L were used as a template for PCR amplification.

343

344 **5.5 Delivery Systems**

345 **5.5.1 Agroinfiltration**

346 The binary plasmids were electroporated into *Agrobacterium tumefaciens* strain
347 GV3101 and selected on LB agar supplemented supplemented with 0.1 mM kanamycin
348 and 0.1 mM gentamicin.. After two days of growth, several colonies were carefully
349 harvested from the plate and suspended in LB medium (2 mL). The bacteria were grown
350 overnight on an orbital shaker at 28 °C at 200 rpm. The cells were collected by
351 centrifugation for 15 min at 4000 g, and the precipitate was resuspended in the 1x
352 MMA infiltration buffer (10 mM MES, pH = 5.6, 10 mM MgCl₂, 180 μM
353 acetosyringone) to OD₆₀₀ = 0.3. Agrobacteria harboring plasmids with each integrase
354 and 35S: p19 gene were mixed in equal volumes. Leaves of 8-week-old *N. benthamiana*
355 plants were injected with agrobacteria suspension using a syringe without a needle.
356 Agroinfiltration assays were assessed with and without co-infiltration with 35S: p19
357 with no changes in results. The infiltrations were performed on the same plant, and
358 duplicate infiltrations were performed on at least three plants within a single
359 experimental assay.

360

361 **5.5.2 Biolistic**

362 Genetic modified *N. benthamiana* plant leaves carrying inverted *mgfp* gene flanked by
363 integrases attachment sites were directly shot with gold (1.5-3.0 μm) and tungsten
364 (approx. 1μm) microparticles as previously described⁴¹. The microparticles were coated
365 with pUC27 plasmids carrying *Int9*, *Int13*, *phiC31*, and *BxB1* integrase genes
366 individually²⁵. As a positive control of gene delivery, a plasmid carrying a constitutive
367 promoter controlling GFP gene was shot under the same conditions at wild-type *N.*
368 *benthamiana* leaves. Microparticles mixed with 8 μL DNA (1μg.μL⁻¹) were accelerated
369 at 650 psi into plant leaves (2,5-3,5 cm length) at a 2 cm distance. Eight shots of each
370 DNA construction were bombarded in 8-week-old plants. 72h after foreign DNA was
371 delivered, shoot leaves were analyzed on Confocal laser scanning microscope (Leica
372 TCS SP8, GER) to visualize *mgfp* expression. PCR screening has been performed to
373 detect *mgfp* and the *attL* and *attR* post-recombination sites.

374

375 **5.6 Plant total DNA extraction and PCR**

376 Total plant DNA isolation was based on the modified CTAB Method Protocol described
377 by Lacorte et al. (2010) with adaptations³⁹. Firstly, plants were screened by PCR using
378 primer pairs DEP_mGFP_FW and DEP_mGFP_RV - designed to detect *mgfp* - and
379 *nptII_Fw* and *npt_II_Rv* - for the amplification of *nptII*, a kanamycin resistance marker
380 - to identify positive transformants. Following the integrase delivery method and
381 incubation period, we collected slices of approximately 20mg from leaves of treated and
382 control plants and proceeded with DNA extraction, as mentioned above. DNA samples
383 were then used as templates in PCR reactions for detection of either *attL* and *attR* sites
384 resulting from *Int9* or *Int13* recombination or the deletion of *mgfp* gene flanked by
385 *phiC31* and *BxB1 att* sites upon introduction of these integrases in the system. For *attL*
386 and *attR* sites amplification, we have used the primer combinations 35S_282F +
387 *attR_IntX_R* and *attL_IntX_F* + IDP_OCSt_349_Rv, respectively, with “IntX” on the
388 primer name denoting the use of specific primers for each Int. As for deletion
389 confirmation, primer pair 35S_282F + IDP_OCSt_349_Rv was used to amplify the
390 whole cassette between the 35S promoter and OCS terminator, with a band shift of
391 about 1kb indicating *mgfp* excision. The circular molecule containing *the mgfp* gene
392 excised from the system by *phiC31* or *BxB1* was amplified using the primer pair
393 NbDel_mGFP_493_Fw + NbDel_mGFP_300_Rv. Two PCR reactions were necessary
394 for the biolistic method to amplify *attL* and *attR* sites. The primer pair 35S_282F +

395 IDP_OCST_349_RV performed the first round, which amplified the cassette. The first
 396 reaction was then used as a template for the specific reaction with primers attached on
 397 the *attL* and *attR* sites, paired with terminator and promoter primers, respectively. That
 398 way, the second round of amplification for detection of the Int9 *attL* site used primers
 399 35S_282F + OnOff_attR9_core_Rv, for Int9 *attR* site primers OnOff_attL9_core_Fw +
 400 IDP_OCST_349_RV, for Int13 *attL* site primers 35S_282F + OnOff_attR13_core_Rv
 401 and to amplify Int13 *attR* site primers OnOff_attL13_core_Fw + IDP_OCST_349_RV.
 402 As for the primers used for molecular detection in TxTI reactions, inversion of the
 403 reporter was amplified with primer pair 2041GbA_ck_5p_FW + deGFP_Hd_Fw while
 404 excision was detected with primers 2041GbA_ck_5p_FW and 2041_ck_3p_RV1.
 405 Primers used in this study are listed in Table 1.

406
 407

Table 1.

PRIMER NAME	PRIMER SEQUENCE
DEP_MGFP_FW	atgagtaaaggagaagaact
DEP_MGFP_RV	ttattgtatagttcatccatg
NPT_II_FW	atggcaattaccttatccgcaacttc
NPT_II_RV	cagaagaactcgtcaagaaggcg
35S_282F	attgatgtgatatctccactgacgtaagggatgacgcac
ATTR_INT9_R	tggaagtgtgatcaggttaactggatacctcatc
ATTR_INT13_R	gtagaacttgaccagttggtcctgtaaataaagcaatcc
ATTL_INT9_F	ataattggcgaacgaggtatctgcatagttattccgaac
ATTL_INT13_F	tccagatccagttgttttagtaacataaataca
IDP_OCST_349_RV	taggtttgaccggttctgccc
ONOFF_ATT9_CORE_FW	ttggcgaacgaggtatctgc
ONOFF_ATTR9_CORE_RV	gaagtgtgatcaggttaactgg
ONOFF_ATT13_CORE_FW	tgtagaacttgaccagttggt
ONOFF_ATTR13_CORE_RV	gttttccagatccagttgt
NBDEL_MGFP_493_FW	tcaagaccgccacaacatcg
NBDEL_MGFP_300_RV	gaagaagatggtcctctcctgc
2041GBA_CK_5P_FW	atgacaaaatcccttaacgtgag
DEGFP HD FW	cccgggaagcttatggagcttttactggcggtg
2041_CK_3P_RV1	tggcttaactatcgccatcag

408

409 Amplification reactions were performed in a Veriti™ 96-Well Fast Thermal
 410 Cycler (Applied Biosystems, USA). Reaction set-up included 2.5ul of 10X PCR Buffer
 411 (Thermo Scientific, USA), 1.5 mM MgCl₂, 0.4 mM dNTP mix, 1 U Platinum™ Taq
 412 DNA Polymerase (5 U/ μL) (Thermo Scientific, USA), 0.2 μM each primer and 1 μL of
 413 purified DNA in a 25 μL final volume. PCR cycling consisted of one initial incubation
 414 at 94°C for 2 minutes, followed by 40 cycles of denaturation at 94°C, 30s; Annealing
 415 TEMP for 30s; Extension at 72°C for 1 min/kb, with final incubation at 4°C. Given that
 416 primer pair 35S_282F + IDP_OCSt_349_Rv will amplify both intact and disrupted
 417 cassette resulting from *mgfp* gene deletion by phiC31 or BxB1, an extension time of just
 418 30 seconds were used in this case to favor amplification of the 1kb smaller amplicon.
 419 All PCR products were detected in 1% agarose gel electrophoresis stained with
 420 SYBR™ Safe DNA Gel Stain (Invitrogen, USA). Gel slices containing the specific
 421 bands were excised under LED Blue Light transilluminator (KASVI, Taiwan) to prevent
 422 nicking of the DNA and its purification was carried out using the ReliaPrep™ DNA
 423 Clean-Up and Concentration System (Promega, USA) according to manufacturer

424 specifications. Isolated amplicons were cloned in pGEM-T-Easy (Promega, USA) and
425 sequenced (Macrogen, USA) after DNA propagation in *Escherichia coli* DH10 β strain.

426

427 **5.7 Confocal laser scanning Microscopy**

428 Leaf samples were analyzed using the Confocal laser scanning microscope (Leica TCS
429 SP8, GER) after 5 d.p.i. Argon laser line excitation wavelength and emission bandpass
430 filter wave lengths for MGFP were 484 nm and 500-550 nm. Chlorophyll
431 autofluorescence was detected, in parallel with MGFP acquisition, using a 650–710 nm
432 bandpass filter. Image acquisition parameters (e.g. laser power, pinhole, detector gain,
433 etc.) and sampling time post-infiltration were held constant within an experiment (i.e.
434 within each figure). Raw data were processed in the LAS X (Leica Application Suite,
435 GER) software (www.leica-microsystems.com, GER).

436

437 **5.8 Nanopore sequencing and Data analysis**

438 Library preparation was carried out using the ligation sequencing kit (Oxford Nanopore
439 Technologies, UK) SQK-NBD112-24, according to the manufacturer's instructions.
440 Briefly, all amplicons were purified, and approximately 200 fmol per sample were
441 submitted to end-prep and native barcode ligation, using NEBNext Ultra II End
442 repair/dA-tailing Module (New England Biolabs, USA) and NEB Blunt/TA Ligase
443 Master Mix (New England Biolabs, USA) respectively. Next, native adapter ligation
444 and cleaning steps were performed using NEBNext Quick Ligation Module (New
445 England Biolabs, USA) kit and Agencourt AMPure XP beads (Beckman Coulter, USA),
446 respectively. The library was sequenced on a R10.4 flow cell (FLO-MIN112) (Oxford
447 Nanopore Technologies, UK) on a MinION Mk1b sequencer with MinKNOW
448 (23.04.6). Basecalling, demultiplexing and adapter/barcode trimming were performed
449 using Guppy [v6.5.7]⁴², with a Super accurate (SUP) basecalling model. The reads with
450 barcodes on both ends and quality scores ≥ 10 were mapped against all the in silico
451 predicted genome-edited sequences, including the predicted circular molecules
452 containing the excised *mgfp* gene, using minimap2⁴³.

453

454 **Competing interests**

455 The authors declare no conflict of interest.

456

457 **Acknowledgements**

458 We acknowledge funding support from Embrapa Genetic Resources and
459 Biotechnology/National Institute of Science and Technology in Synthetic Biology,
460 National Council for Scientific and Technological Development/ Ministry of
461 Agriculture Livestock and Supply (465603/2014-9; 400145/2023-5), Research Support
462 Foundation of the Federal District (0193.001.262/2017), and Coordination for the
463 Improvement of Higher Education Personnel. We thank Plant Transformation Research
464 Center, University of California-Riverside, USA, for producing the transgenic plants of
465 *N. benthamiana*; Plant Germplasm Quarantine Station, Embrapa Genetic Resources and
466 Biotechnology, Brazil, for regulating the importation of *N. benthamiana*
467 (21016.000015/2020-21; 21016.004393/2020-83).

468

469 **Author contributions**

470 M.A.O.: conceptualization, investigation, data curation, and writing - original draft.

471 R.N.L.: conceptualization, investigation, data curation, and writing - original draft.

472 L.H.F.: investigation, data curation, and analysis. M.M.S.A.: investigation, data

473 curation, and analysis. F.L.M.: investigation, data curation, and analysis. R.V.B.:

474 investigation, data curation, and analysis. C.L.: investigation, data curation, and
475 analysis. E.L.R.: funding acquisition, supervision, project administration, and writing -
476 review & editing. All authors revised the manuscript and approved the final version.

477

478 **Data availability**

479 The sequencing data obtained in this study are available in: [https://github.com/Rech-](https://github.com/Rech-PBSyn/Serine_Integrases)
480 [PBSyn/Serine_Integrases](https://github.com/Rech-PBSyn/Serine_Integrases)

481

482 **Additional information**

483 Additional Supporting Information may be found in the online version of this article at
484 the publisher's website or under request. Correspondence and requests for materials
485 should be addressed to Elibio Rech (elibio.rech@embrapa.br).

486

487 **REFERENCES**

488

489

490

- 491 1 Ortega, C., Abreu, C., Oppezzo, P. & Correa, A. Overview of High-Throughput Cloning
492 Methods for the Post-genomic Era. *Methods in molecular biology (Clifton, N.J.)* **2025**,
493 3-32, doi:10.1007/978-1-4939-9624-7_1 (2019).
- 494 2 Casini, A., Storch, M., Baldwin, G. S. & Ellis, T. Bricks and blueprints: methods and
495 standards for DNA assembly. *Nature reviews. Molecular cell biology* **16**, 568-576,
496 doi:10.1038/nrm4014 (2015).
- 497 3 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred
498 kilobases. *Nature methods* **6**, 343-345, doi:10.1038/nmeth.1318 (2009).
- 499 4 Gibson, D. G. Enzymatic assembly of overlapping DNA fragments. *Methods in*
500 *enzymology* **498**, 349-361, doi:10.1016/b978-0-12-385120-8.00015-2 (2011).
- 501 5 Sleight, S. C., Bartley, B. A., Lieviant, J. A. & Sauro, H. M. In-Fusion BioBrick assembly
502 and re-engineering. *Nucleic acids research* **38**, 2624-2636, doi:10.1093/nar/gkq179
503 (2010).
- 504 6 Vazquez-Vilar, M. *et al.* GB3.0: a platform for plant bio-design that connects functional
505 DNA elements with associated biological data. *Nucleic acids research* **45**, 2196-2209,
506 doi:10.1093/nar/gkw1326 (2017).
- 507 7 Arber, W. Polylysogeny for bacteriophage lambda. *Virology* **11**, 250-272,
508 doi:[https://doi.org/10.1016/0042-6822\(60\)90065-9](https://doi.org/10.1016/0042-6822(60)90065-9) (1960).
- 509 8 Mahmood, M. A., Naqvi, R. Z., Rahman, S. U., Amin, I. & Mansoor, S. Plant Virus-
510 Derived Vectors for Plant Genome Engineering. *Viruses* **15**, doi:10.3390/v15020531
511 (2023).
- 512 9 Khakhar, A. & Voytas, D. F. RNA Viral Vectors for Accelerating Plant Synthetic Biology.
513 *Frontiers in plant science* **12**, 668580, doi:10.3389/fpls.2021.668580 (2021).
- 514 10 Zulfiqar, S. *et al.* Virus-Induced Gene Silencing (VIGS): A Powerful Tool for Crop
515 Improvement and Its Advancement towards Epigenetics. *International journal of*
516 *molecular sciences* **24**, doi:10.3390/ijms24065608 (2023).
- 517 11 Jagram, N. & Dasgupta, I. Principles and practice of virus induced gene silencing for
518 functional genomics in plants. *Virus genes* **59**, 173-187, doi:10.1007/s11262-022-
519 01941-5 (2023).
- 520 12 Rajput, M. *et al.* RNA Interference and CRISPR/Cas Gene Editing for Crop
521 Improvement: Paradigm Shift towards Sustainable Agriculture. *Plants (Basel,*
522 *Switzerland)* **10**, doi:10.3390/plants10091914 (2021).

- 523 13 Zhao, J. H. & Guo, H. S. RNA silencing: From discovery and elucidation to application
524 and perspectives. *Journal of integrative plant biology* **64**, 476-498,
525 doi:10.1111/jipb.13213 (2022).
- 526 14 Jin, L., Chen, M., Xiang, M. & Guo, Z. RNAi-Based Antiviral Innate Immunity in Plants.
527 *Viruses* **14**, doi:10.3390/v14020432 (2022).
- 528 15 Metje-Sprink, J., Menz, J., Modrzejewski, D. & Sprink, T. DNA-Free Genome Editing:
529 Past, Present and Future. *Frontiers in plant science* **9**, 1957,
530 doi:10.3389/fpls.2018.01957 (2018).
- 531 16 Pandey, P. K. *et al.* Versatile and multifaceted CRISPR/Cas gene editing tool for plant
532 research. *Seminars in cell & developmental biology* **96**, 107-114,
533 doi:10.1016/j.semcd.2019.04.012 (2019).
- 534 17 Adli, M. The CRISPR tool kit for genome editing and beyond. *Nature communications* **9**,
535 1911, doi:10.1038/s41467-018-04252-2 (2018).
- 536 18 Lin, Q. *et al.* Genome editing in plants with MAD7 nuclease. *Journal of genetics and*
537 *genomics = Yi chuan xue bao* **48**, 444-451, doi:10.1016/j.jgg.2021.04.003 (2021).
- 538 19 Smith, M. C. M. Phage-encoded Serine Integrases and Other Large Serine
539 Recombinases. *Microbiology spectrum* **3**, doi:10.1128/microbiolspec.MDNA3-0059-
540 2014 (2015).
- 541 20 Yang, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nature methods* **11**,
542 1261-1266, doi:10.1038/nmeth.3147 (2014).
- 543 21 Olorunniji, F. J. *et al.* Multipart DNA Assembly Using Site-Specific Recombinases from
544 the Large Serine Integrase Family. *Methods in molecular biology (Clifton, N.J.)* **1642**,
545 303-323, doi:10.1007/978-1-4939-7169-5_19 (2017).
- 546 22 Wang, X. *et al.* Bxb1 integrase serves as a highly efficient DNA recombinase in rapid
547 metabolite pathway assembly. *Acta biochimica et biophysica Sinica* **49**, 44-50,
548 doi:10.1093/abbs/gmw115 (2017).
- 549 23 Andres, J., Blomeier, T. & Zurbriggen, M. D. Synthetic Switches and Regulatory Circuits
550 in Plants. *Plant physiology* **179**, 862-884, doi:10.1104/pp.18.01362 (2019).
- 551 24 Kassaw, T. K., Donayre-Torres, A. J., Antunes, M. S., Morey, K. J. & Medford, J. I.
552 Engineering synthetic regulatory circuits in plants. *Plant science : an international*
553 *journal of experimental plant biology* **273**, 13-22, doi:10.1016/j.plantsci.2018.04.005
554 (2018).
- 555 25 Gomide, M. S. *et al.* Genetic switches designed for eukaryotic cells and controlled by
556 serine integrases. *Communications Biology* **3**, 255, doi:10.1038/s42003-020-0971-8
557 (2020).
- 558 26 Stark, W. M. Making serine integrases work for us. *Current opinion in microbiology* **38**,
559 130-136, doi:10.1016/j.mib.2017.04.006 (2017).
- 560 27 Lloyd, J. P. B. *et al.* Synthetic memory circuits for stable cell reprogramming in plants.
561 *Nature biotechnology* **40**, 1862-1872, doi:10.1038/s41587-022-01383-2 (2022).
- 562 28 Bonnet, J., Subsoontorn, P. & Endy, D. Rewritable digital data storage in live cells via
563 engineered control of recombination directionality. *Proceedings of the National*
564 *Academy of Sciences of the United States of America* **109**, 8884-8889,
565 doi:10.1073/pnas.1202344109 (2012).
- 566 29 Bernabé-Orts, J. M. *et al.* A memory switch for plant synthetic biology based on the
567 phage ϕ C31 integration system. *Nucleic acids research* **48**, 3379-3394,
568 doi:10.1093/nar/gkaa104 (2020).
- 569 30 Merrick, C. A., Zhao, J. & Rosser, S. J. Serine Integrases: Advancing Synthetic Biology.
570 *ACS synthetic biology* **7**, 299-310, doi:10.1021/acssynbio.7b00308 (2018).
- 571 31 Thomson, J. G. *et al.* The Bxb1 recombination system demonstrates heritable
572 transmission of site-specific excision in Arabidopsis. *BMC biotechnology* **12**, 9,
573 doi:10.1186/1472-6750-12-9 (2012).

- 574 32 Guiziou, S., Maranas, C. J., Chu, J. C. & Nemhauser, J. L. An integrase toolbox to record
575 gene-expression during plant development. *Nature communications* **14**, 1844,
576 doi:10.1038/s41467-023-37607-5 (2023).
- 577 33 Chen, W., Kaur, G., Hou, L., Li, R. & Ow, D. W. Replacement of stacked transgenes in
578 planta. *Plant biotechnology journal* **17**, 2029-2031, doi:10.1111/pbi.13172 (2019).
- 579 34 Li, Y. *et al.* Recombinase-mediated gene stacking in cotton. *Plant physiology* **188**, 1852-
580 1865, doi:10.1093/plphys/kiac005 (2022).
- 581 35 Deich, C. *et al.* T7Max transcription system. *Journal of Biological Engineering* **17**, 4,
582 doi:10.1186/s13036-023-00323-1 (2023).
- 583 36 Franco, R. A. L. *et al.* Signal Amplification for Cell-Free Biosensors, an Analog-to-Digital
584 Converter. *ACS synthetic biology* **12**, 2819-2826, doi:10.1021/acssynbio.3c00227
585 (2023).
- 586 37 Chiu, W. *et al.* Engineered GFP as a vital reporter in plants. *Current biology : CB* **6**, 325-
587 330, doi:10.1016/s0960-9822(02)00483-9 (1996).
- 588 38 Müller, K. *et al.* A red light-controlled synthetic gene expression switch for plant
589 systems. *Molecular bioSystems* **10**, 1679-1688, doi:10.1039/c3mb70579j (2014).
- 590 39 Thomson, J. G., Chan, R., Thilmony, R., Yau, Y. Y. & Ow, D. W. PhiC31 recombination
591 system demonstrates heritable germinal transmission of site-specific excision from the
592 Arabidopsis genome. *BMC biotechnology* **10**, 17, doi:10.1186/1472-6750-10-17 (2010).
- 593 40 Durrant, M. G. *et al.* Systematic discovery of recombinases for efficient integration of
594 large DNA sequences into the human genome. *Nature biotechnology* **41**, 488-499,
595 doi:10.1038/s41587-022-01494-w (2023).
- 596 41 Rech, E. L., Vianna, G. R. & Aragão, F. J. High-efficiency transformation by biolistics of
597 soybean, common bean and cotton transgenic plants. *Nature protocols* **3**, 410-418,
598 doi:10.1038/nprot.2008.9 (2008).
- 599 42 Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools
600 for Oxford Nanopore sequencing. *Genome biology* **20**, 129, doi:10.1186/s13059-019-
601 1727-y (2019).
- 602 43 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford,*
603 *England)* **34**, 3094-3100, doi:10.1093/bioinformatics/bty191 (2018).

604

605 Legend of Figures

606

607 **Figure 1.** Schematic overview of the IntPlex@ memory genetic switch architecture and
608 Boolean-based logic gates. (A) The diagram describes the architecture of the bimodular
609 genetic switch, which consists of a *gfp* sequence in reverse orientation (gray)
610 surrounded by four *attB* and *attP* sites (triangles) of four serine integrases: BxB1
611 (yellow), phiC31 (lilac), Int9 (blue) and Int13 (red). For in vitro assays, the switch
612 contains the *degfp* gene and the switch integrated into the plant genome contains the
613 *mgfp* gene. (B) Int9/Int13-based XOR logic gate for *gfp* orientation flipping (turn *gfp*
614 on). Int9/Int 13 based OR logic gate for turning *gfp* on and off. BxB1/phiC31-based
615 NOR logic gate for *gfp* excision. In the presence of Int9 or Int13, the *gfp* is inverted to
616 its coding sequence (ON state). Moreover, in the presence of Int9 and Int 13, the *gfp* is
617 modulated and oscillates between the on and off states. In the presence of BxB1 or
618 phiC31 input, the *gfp* is excised. The system output is edited DNA: plasmid or plant
619 genome. (C) The diagram shows IntPlex@ memory genetic switch operation driven by
620 BxB1. The resulting edited DNA with the BxB1 *attL* scar (left) and the excised circular
621 DNA molecule with BxB1 *attR* (right). (D) The diagram shows IntPlex@ memory
622 genetic switch operation driven by phiC31. The resulting edited DNA with the phiC31
623 *attL* scar (left) and the excised circular DNA molecule with phiC31 *attR* (right). (E) The

624 diagram shows IntPlex@ memory genetic switch operation driven by Int9. The
625 sequence flanked by the Int9 *attB* and *attP* sites and, consequently, the inversion of the
626 *gfp* to its coding sequence (GFP output). The Int13 *att* sites are also inverted and
627 oriented towards their reverse complement. (F) The diagram shows IntPlex@ memory
628 genetic switch operation driven by Int13. The *gfp* flanked by the Int13 *attB* and *attP*
629 sites is flipped to its coding sequence (GFP output). (G and H) The diagram shows
630 IntPlex@ memory genetic switch operation driven by Int9 and Int13. The *gfp* is flipped
631 in two steps according to which integrase is added first. (G) In the first step, Int9
632 recognizes its *attB/P* sites and catalyzes the inversion of *gfp* from its OFF state to ON
633 state, resulting in the formation of the intermediate sequence with *attL/Int9* and
634 *attR/Int9* sites and the output GFP. Instantaneously, this intermediate sequence is the
635 substrate for the action of Int13 that recognizes its *attB/P* sites and catalyzes the
636 inversion of *gfp* from its ON state to OFF state, resulting in the formation of the final
637 plasmid with both *attL/R Int9* and *attL/R Int13* sites. (H) In the first step, Int13
638 recognizes its *attB/P* sites and catalyzes the inversion of *gfp* from its OFF state to ON
639 state, resulting in the formation of the intermediate sequence with *attL/Int13* and
640 *attR/Int13* sites and the output GFP. Instantaneously, this intermediate sequence is the
641 substrate for the action of Int9 that recognizes its *attB/P* sites and catalyzes the
642 inversion of *gfp* from its ON state to OFF state, resulting in the formation of the final
643 plasmid with both *attL/R Int9* and *attL/R Int13* sites. The corresponding colors and
644 shapes indicate the genetic components of the switch.

645

646 **Figure 2.** In vitro function of Int-Plex@ memory switch driven by serine integrases in
647 cell-free TX-TL system. (A) Schematic overview of the IntPlex@ memory bimodular
648 genetic switch architecture, which consists of a RBS sequence in reverse orientation
649 (black) and a *degfp* sequence in reverse orientation (gray) surrounded by four *attB* and
650 *attP* sites (triangles) of four serine integrases: BxB1 (yellow), phiC31 (lilac), Int9 (blue)
651 and Int13 (red). (B) Agarose 1% gel of PCR reaction for each primer combination used
652 to confirm the presence of edited DNA. (C) Agarose 1% gel of PCR reaction for each
653 primer combination used to confirm the presence of the excised DNA with *attR* scar. (D
654 and E) Sanger sequencing chromatograms of excised circular DNA with *attR* scar. The
655 DNA fragments were obtained through PCR using pairs of primers indicated in Table 1.
656 The position of the primers is indicated by black arrows. All sequences corresponded to
657 the in silico predicted genome-edited sequences. The corresponding colors and shapes
658 indicate the genetic components of the switch. First lane is 1 kb plus ladder (Life
659 Technologies, USA). Reverse complementary sequences are indicated by a black
660 asterisk. No off-target effects were detected in the analyzed sequences.

661

662 **Figure 3.** In vitro function of Int-Plex@ memory switch driven by serine integrases in
663 cell-free TX-TL system. (A) The diagram shows the flipping of the *degfp* in two steps.
664 In the first step, Int 9 recognizes its *attB/P* sites and catalyzes the inversion of *degfp*
665 from its OFF state to ON state, resulting in the formation of the intermediate plasmid
666 with *attL/Int9* and *attR/Int9* sites and the output deGFP. Instantaneously, this
667 intermediate plasmid is the substrate for the action of Int13 that recognizes its *attB/P*
668 sites and catalyzes the inversion of *degfp* from its ON state to OFF state, resulting in the
669 formation of the final plasmid with both *attL/R Int9* and sites *attL/R Int13*. (B) *In vitro*
670 reactions after 24h post incubation of Int-Plex@ cassette plasmid with Int 9 and Int 13
671 plasmids in Arbor Biosciences™ myTXTL system. Furthermore, it is noteworthy that
672 since both enzymes were added simultaneously (pART27+Int9+Int13), the ON state
673 varies according to the first integrase to catalyze the plasmid inversion. Thus, the

674 system presents the deGFP for several days according to the ON state mRNA molecules
675 and protein half-lives. (C) Agarose 1% gel of PCR reaction for each primer combination
676 used to confirm the presence of the inverted DNA with *attL/R* sites. (D and E) Sanger
677 sequencing chromatograms of regions flanking the *degfp* obtained after simultaneous
678 treatment with Integrases 9 and 13. The DNA fragments were obtained through PCR
679 using pairs of primers indicated in Table 1. The position of the primers is indicated by
680 black arrows. All sequences corresponded to the *in silico* predicted genome-edited
681 sequences. The corresponding colors and shapes indicate the genetic components of the
682 switch. First lane is 1 kb plus ladder (Life Technologies, USA). Reverse complementary
683 sequences are indicated by a black asterisk. No off-target effects were detected in the
684 analyzed sequences.

685
686 **Figure 4.** IntPlex@ memory genetic switch for DNA excision driven by BxB1. (A)
687 Sanger sequencing chromatogram of edited plant genome with *attL/BxB1* scar. (B and
688 C) Sanger sequencing chromatograms of excised circular DNA with *attR* scar. The
689 DNA fragments were obtained through PCR using pairs of primers indicated in Table 1.
690 The position of the primers is indicated by black arrows. All sequences corresponded to
691 the *in silico* predicted genome-edited sequences. The corresponding colors and shapes
692 indicate the genetic components of the switch.

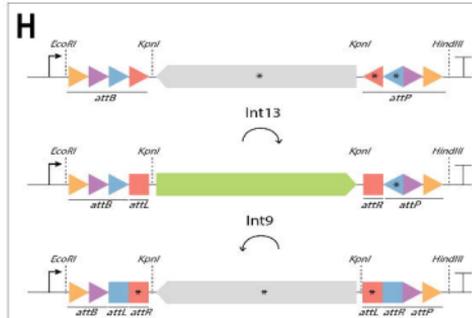
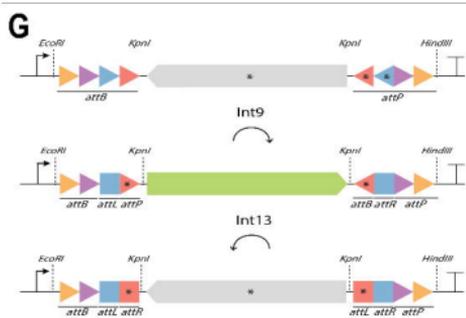
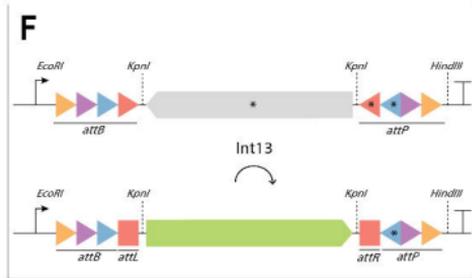
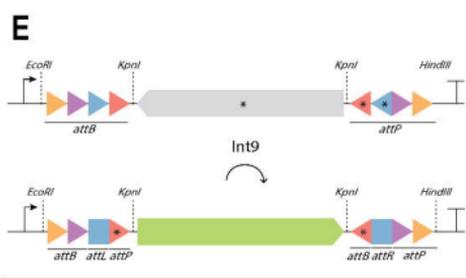
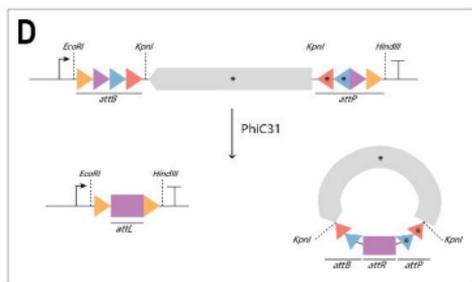
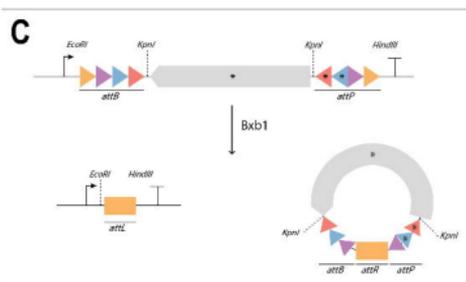
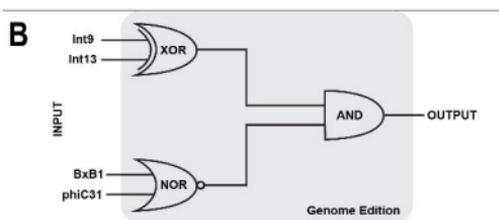
693
694 **Figure 5.** IntPlex@ memory genetic switch for DNA excision driven by phiC31. Sanger
695 sequencing chromatogram of edited plant genome with phiC31 *attL*. The DNA
696 fragments were obtained through PCR using pairs of primers indicated in Table 1. The
697 position of the primers is indicated by black arrows. The sequence corresponded to the
698 *in silico* predicted genome-edited sequence. The corresponding colors and shapes
699 indicate the genetic components of the switch.

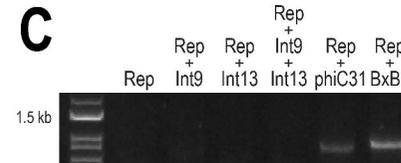
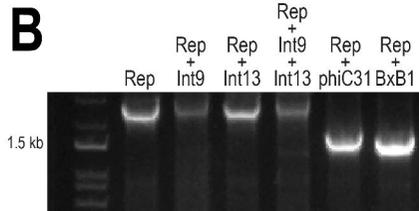
700
701 **Figure 6.** Nanopore sequencing evidence of genetic switch operation driven by BxB1.
702 (A) The diagram illustrates the nanopore raw reads mapped to the predicted excised
703 circular DNA molecule with BxB1 *attR*. The corresponding colors and shapes indicate
704 the genetic components of the switch. (B) A detailed view of the excision region,
705 showing the resulting BxB1 *attR* and the flanking phiC31 *attP/attB* sites.

706
707 **Figure 7.** Basecalled reads length vs reads quality score. The color gradient indicates
708 the density of reads, with warmer colors representing higher densities.

709
710 **Figure 8.** IntPlex@ memory genetic switch for gene modulation (turn gene on) driven
711 by Int9 or Int13. (A) Sanger sequencing chromatogram of edited plant genome with
712 Int9 *attL* and *attR* post recombination sites. The sequence shows the flipping of the
713 DNA flanked by the Int9 *attB* and *attP* sites and, consequently, the inversion of the *mgfp*
714 to its coding sequence (MGFP output). The Int13 *att* sites are also inverted and oriented
715 towards their reverse complement. (B) Sanger sequencing chromatogram of edited plant
716 genome with Int13 *attL* and *attR* post recombination sites and the inversion of the *mgfp*
717 to its coding sequence (MGFP output). The fragments were obtained through PCR using
718 pairs of primers indicated in table 1. The position of the primers is indicated by black
719 arrows. A pair of primers was used to amplify the complete *attL* site (black upper
720 arrow) and a second pair of primers to amplify the *attR* site (black lower arrow). The
721 sequences corresponded to the *in silico* predicted genome-edited sequences. The
722 corresponding colors and shapes indicate the genetic components of the switch.

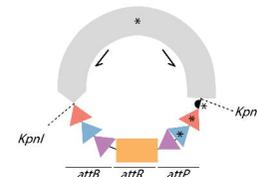
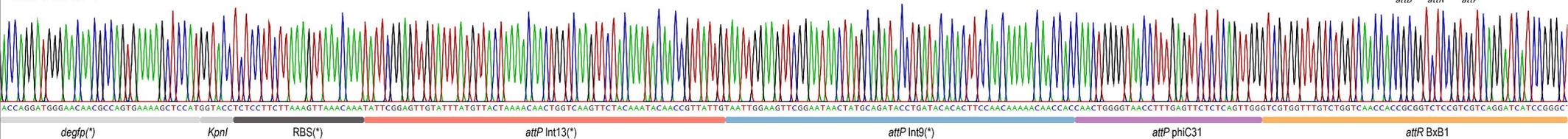
723





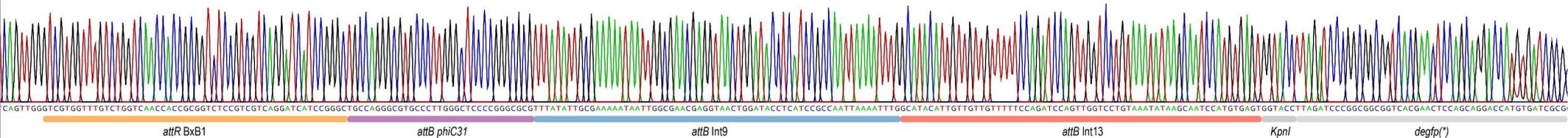
D

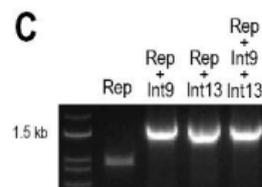
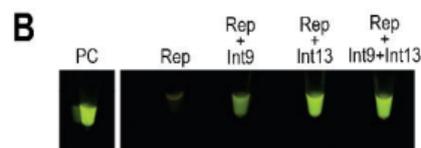
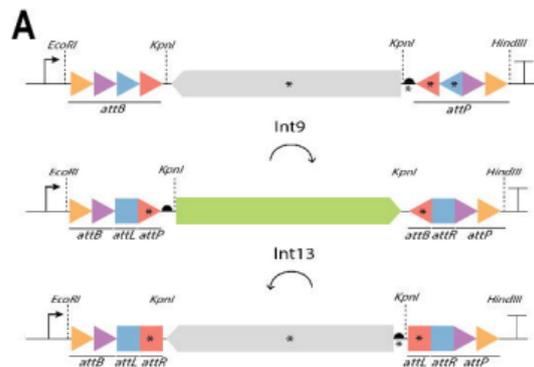
attR BxB1



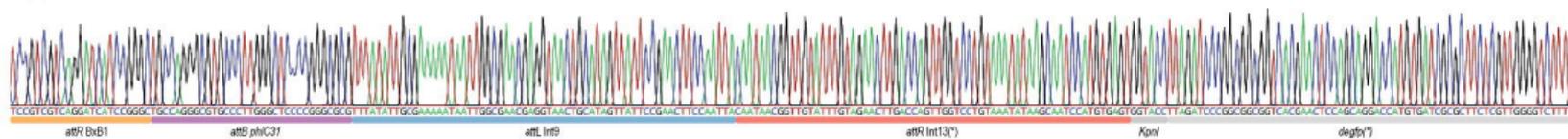
E

attR BxB1

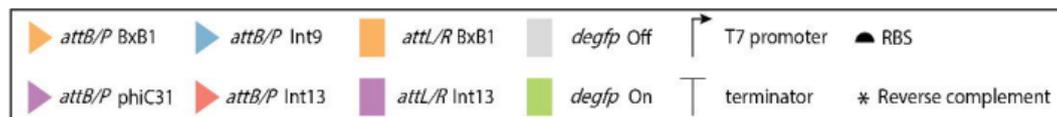
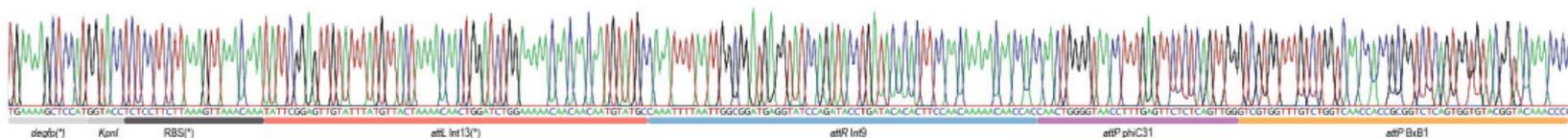


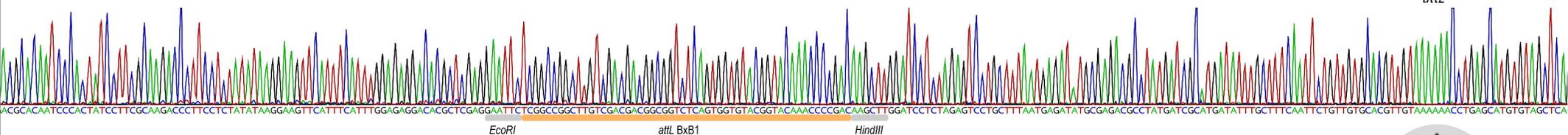
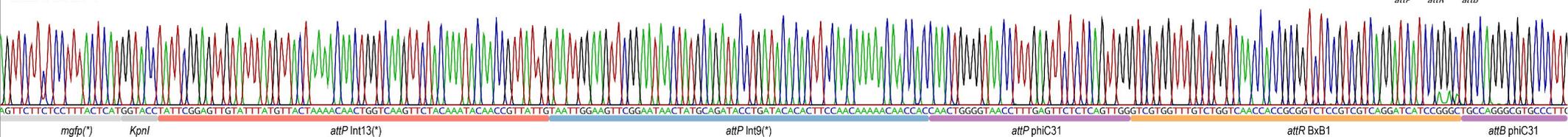
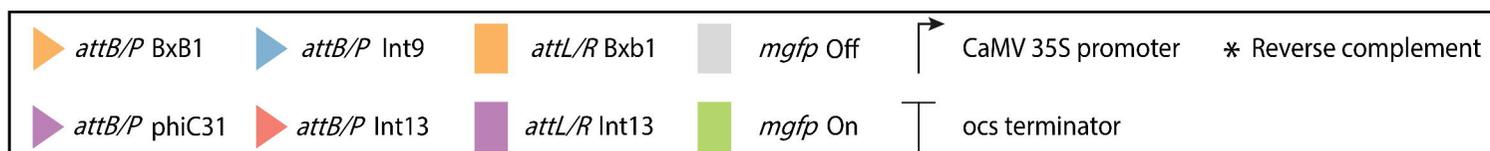
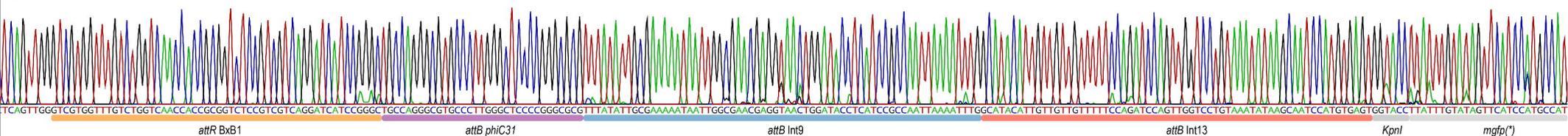


D

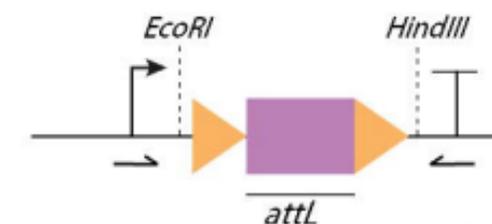
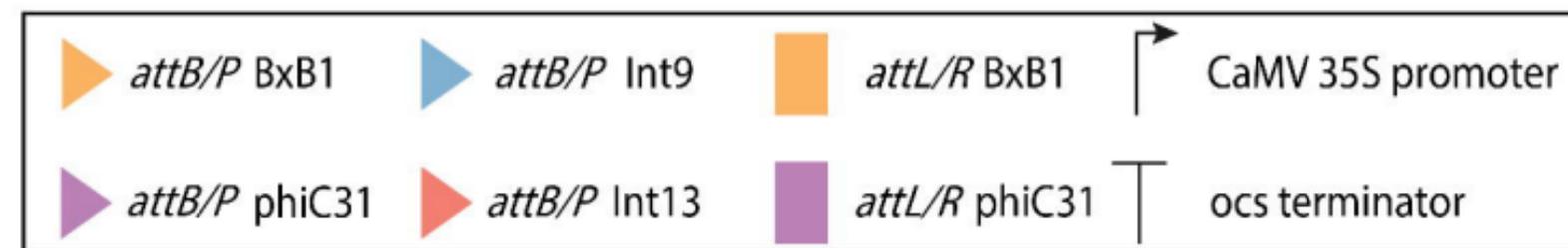
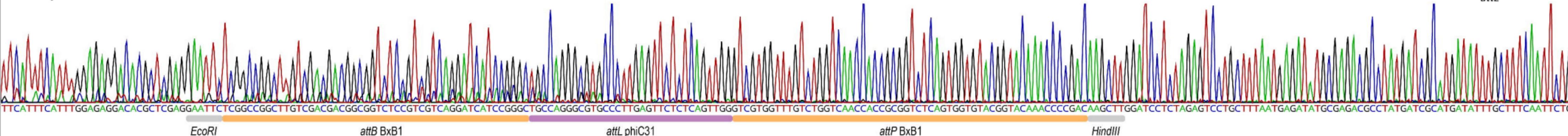


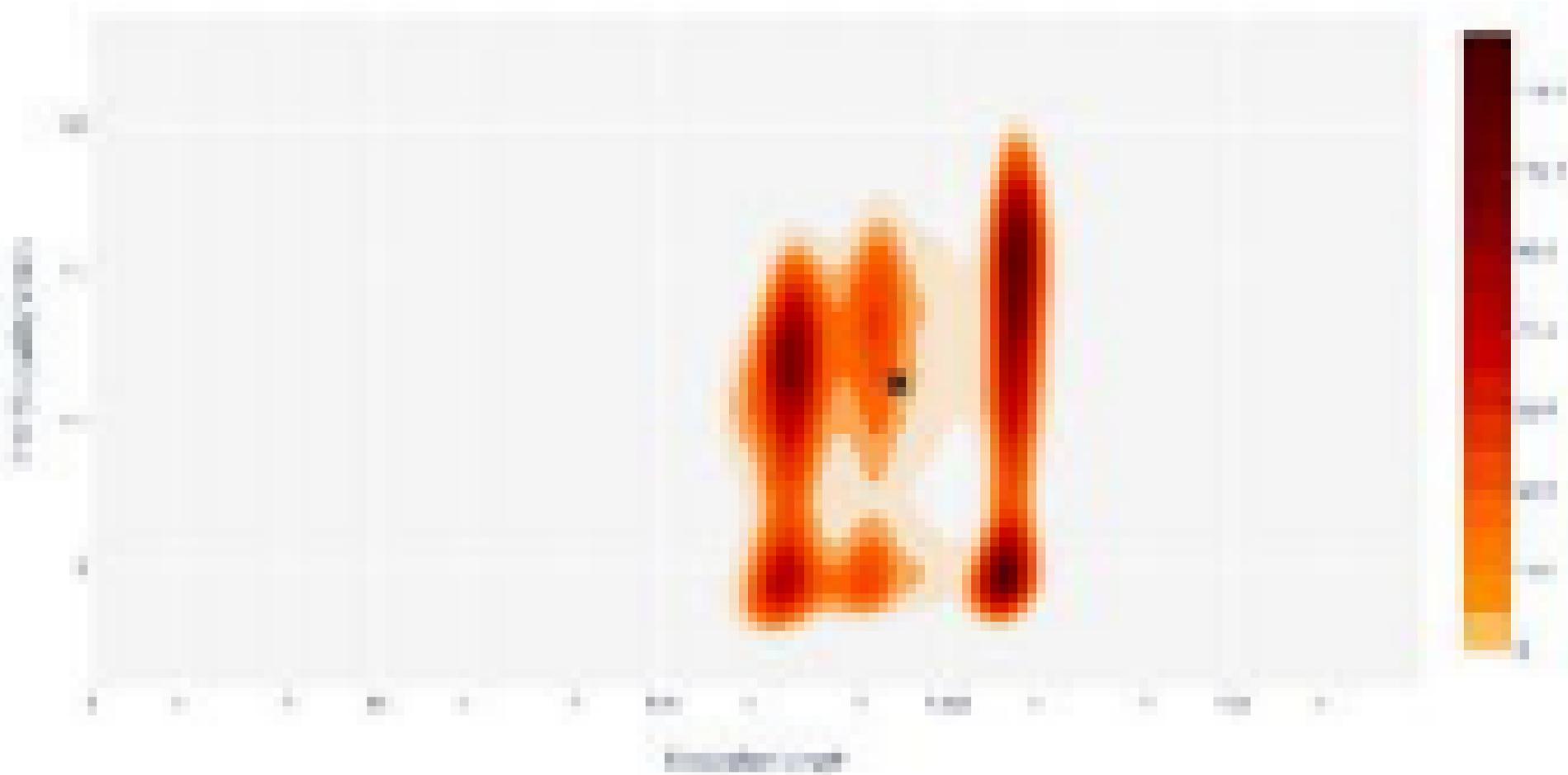
E

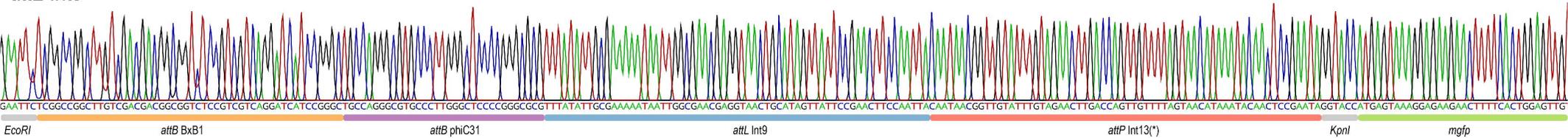
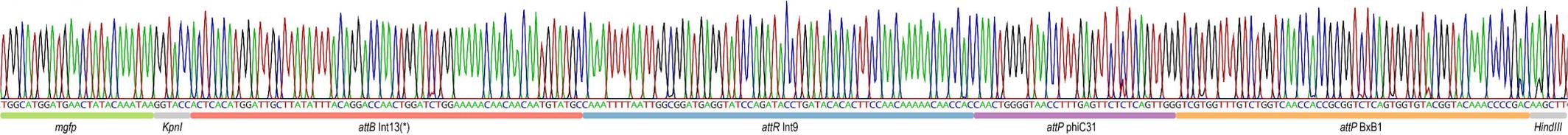
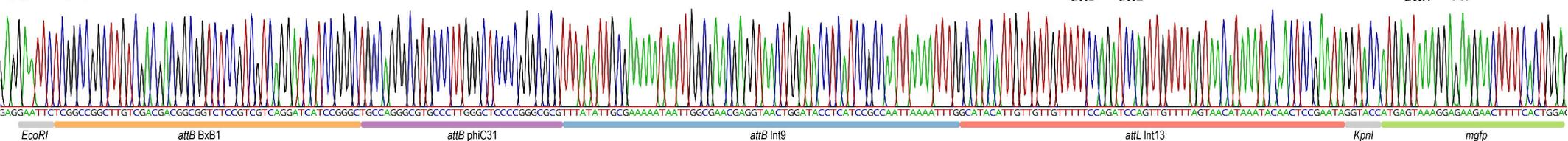


A*attL* BxB1**B***attR* BxB1**C***attR* BxB1

attL phiC31





A*attL* Int9*attR* Int9**B***attL* Int13*attR* Int13