

## SELEÇÃO E DESENHO DE gRNA PARA EDIÇÃO DE GENES RELACIONADOS A DOENÇA DE CITROS

Amanda de Carvalho Bernardi<sup>1,2\*</sup>

Marco Aurélio Takita<sup>1,2</sup>

### RESUMO

CRISPR (“Clustered Regularmente Interspaced Palindromic Repeats”) foram descobertos como parte de um sistema imunológico bacteriano, mas recentemente passaram a ser utilizados como uma poderosa ferramenta de biologia molecular, uma vez que esta pode ser facilmente personalizada para gerar mutações em um determinado local desejado do genoma. O sistema é constituído principalmente por 2 componentes: a endonuclease Cas9 e um RNA guia, que direciona a clivagem do DNA. A escolha da sequência-alvo e o desenho do RNA guia, são etapas preliminares e cruciais para trabalhos de edição de genoma e envolvem uma série de análises, que podem determinar se o sistema vai ou não funcionar corretamente. Este estudo mostra passo-a-passo a seleção de sequências-alvo e o desenho de RNAs guia para edição de genes possivelmente relacionados a tolerância ao cancro, doença de extrema importância para a citricultura. Este trabalho gera recursos para o desenvolvimento de estudos funcionais de citros, o que seria de grande importância para a cultura, além de servir como guia para o desenho de RNAs guia para outros trabalhos, contribuindo para o desenvolvimento da técnica CRISPR/CAS9 no Brasil.

Palavras-Chave: CRISPR, Cas9, Terpeno, CTV, Cancro Cítrico.

---

<sup>1</sup>Universidade Federal de São Carlos, Centro de Ciências Agrárias. Rodovia Anhanguera, Km 174 - Zona Rural, Araras - SP, 13604-900

<sup>2</sup>Instituto Agronômico de Campinas, Centro de Citricultura Sylvio Moreira. Rodovia Anhanguera, 1460 – Cascalho, Cordeirópolis - SP, 13490-000

\*Autor para correspondência – [bernardi.amanda@hotmail.com](mailto:bernardi.amanda@hotmail.com)

Recebido em: 30/01/2017 Aceito para publicação em: 22/06/2017

## ABSTRACT

CRISPR (Clustered Regularly Interspaced Palindromic Repeats) were discovered as part of a bacterial immune system but have recently come to be used as a powerful biomolecular tool since it can be easily customized to generate mutations at a particular desired location in the genome. The system consists mainly of 2 components: the Cas9 endonuclease and a guide RNA, which directs DNA cleavage. The choice of target sequence and guide RNA design are preliminary and crucial steps for genome editing work and involve a series of analyzes that can determine whether or not the system will function correctly. This study shows step-by-step the selection of target sequences and the design of RNAs guide for edition of genes possibly related to citrus canker tolerance a diseases of extreme importance for citriculture. This work generates resources for the development of functional studies of citrus, which would be of great importance for the culture, besides working as guide for the design of RNAs guide in other works, contributing to the development of CRISPR/Cas9 technique in Brazil.

Keywords: CRISPR, CAS9, Terpene, CTV, Citrus Canker.

## INTRODUÇÃO

Os CRISPR (“Clustered Regularly Interspaced Short Palindromic Repeats”), como o próprio nome diz, são aglomerados de sequências de ácidos nucleicos palindrômicas repetitivas e intercaladas por outras sequências, chamadas de espaçadores. Essa estrutura foi primeiramente identificada a partir de estudos básicos do gene *iap* de *Escherichia coli* (ISHINO et al., 1987) e posteriormente foi descrita como parte de um mecanismo de defesa contra ácidos nucléicos invasores (revisado por HORVATH e BARRANGOU, 2010). Entretanto a partir de 2013, o nome “CRISPR” ganhou visibilidade notória, quando grupos de pesquisas estabeleceram a partir desse mecanismo de defesa, uma técnica que se tornou uma das ferramentas de biologia molecular mais promissoras da atualidade: a edição de genomas via CRISPR/Cas9 (JINEK et al., 2012, CONG et al., 2013, MALI et al., 2013).

Uma das razões para que trabalhos utilizando essa técnica tenham emergido tão rapidamente é a versatilidade da ferramenta, uma vez que a mesma pode ser facilmente customizada para gerar mutações em um sítio específico do genoma, além

de substituição gênica, entre outras aplicações. O sistema utilizado baseia-se no mecanismo de defesa tipo II, e conta com um RNA guia único (sgRNA), que contém uma sequência alvo, responsável por direcionar a clivagem da sequência de interesse pela endonuclease Cas9 (JINEK et al., 2012). A edição do genoma se dá através da participação do sistema natural de reparo na célula, o qual utiliza primordialmente dois mecanismos, a ligação de extremidade não homóloga (NHEJ) e a recombinação homóloga (HR) (CHAPMAN et al., 2012, LIU et al., 2014), que são passíveis de erro, aleatórios e induzidos, respectivamente.

Tanto a endonuclease Cas9 quanto o RNA guia (gRNA), principais componentes do sistema, podem ser obtidos por vias acadêmicas, comerciais ou por síntese. A síntese do gRNA pode se dar via transcrição *in vitro* ou através da produção de vetores para sua expressão, sendo que essa última envolve a construção de um cassete contendo a. um promotor da RNA polimerase III para pequenos RNAs, b. o gRNA, que por sua vez é composto pela sequência-guia (~20 nt), seguido por uma sequência específica de aproximadamente 80 nt, c. um terminador (normalmente “TTTTTT”) (PEREIRA, 2016). A escolha da sequência-guia constitui etapa crucial para o sucesso de experimentos de edição de genoma, e existem dezenas de programas “online” disponíveis que utilizam critérios básicos para obtê-las de maneira simplificada (Ex. CRISPR design, E-CRISPR, MultiCRISPR, CRISPR-P). Entretanto, muitos pesquisadores preferem ser mais cautelosos e fazer a seleção e desenho manualmente, permitindo assim que haja um controle minucioso dos parâmetros de seleção afim de otimizar a ação do sistema.

O cancro cítrico, é uma das doenças que mais ameaçam a citricultura mundial. Causada prioritariamente pela bactéria *Xanthomonas citri* subsp. *citri*, afeta todas as espécies e variedades de citros de importância comercial, causando desfolha de plantas, lesões em frutos, queda prematura de frutos e restrição da produção para áreas livres da doença (FUNDECITRUS, 2016b). Tendo isso em vista, este trabalho tem como objetivo descrever passo a passo os critérios de seleção para o desenho manual de gRNA baseados na expressão do vetor pChimera (SCHIML, FAUSER e PUCHTA, 2014) com alvo em genes possivelmente relacionados a tolerância a essa doença, e comparar os resultados com os obtidos através do programa CRISPR-P, um dos mais utilizados para edição de genoma de plantas. Este trabalho poderá ser usado como um

guia prático para o desenho gRNA de interesse em outros estudos, contribuindo assim para expansão da aplicação desta tecnologia no Brasil.

## **METODOLOGIA**

O primeiro passo para o desenho do gRNA consiste na seleção dos genes-alvo de interesse, e deve se basear em estudos anteriores que indiquem a necessidade ou possibilidade de obtenção de resultados positivos com a edição do mesmo. É imprescindível que a sequência do gene de interesse esteja disponível, o que normalmente acontece nos principais bancos de dados “online”. Neste trabalho os genes alvos selecionados se basearam em estudos anteriores sobre cancro cítrico (RODRIGUEZ et al., 2011a, 2011b) e as sequências desses genes foram obtidas no genoma disponível de *Citrus sinensis* (<http://citrus.hzau.edu.cn/orange/>).

A partir da sequência de cada gene, foi feita a seleção das sequências-alvo candidatas, contendo 20 nucleotídeos únicos com conteúdo “GC” próximo a 50% (Tolera-se 40%-60%) posicionados exatamente adjacentes a três nucleotídeos específicos chamados de PAM, que varia de acordo com a nucleasse utilizada, e no caso deste trabalho é “NGG” (PEREIRA, 2016). Essas candidatas deverão ser avaliadas quanto a formação de estruturas secundárias afim de garantir sua condição ótima para complementariedade com o alvo e quanto a sua especificidade para impedir que ela se ligue a outras regiões no genoma que podem causar ruídos na interpretação das análises fenotípicas.

Antes de se avaliar a estrutura da sequência-guia em si, quando se trata de nocaute de genes codificadores de proteínas, é necessário levar em consideração a posição da sequência em relação ao gene alvo. A sequência-alvo deve preferencialmente estar localizada em “exons” localizados próximos ao códon de iniciação, isso porque as mutações causadas por esse sistema (chamadas de “indels”) normalmente alteram a fase de leitura, assim quanto mais próximas do início da sequência, maior a probabilidade de que as alterações levem à perda da função gênica.

Para avaliar a formação de estruturas secundárias foi utilizado o programa “Gene Runner”, e foram descartadas todas as candidatas que apresentaram estruturas secundárias em temperaturas de “melting” superior a 0°C. Além disso também foram

excluídas todas as candidatas que possuíam 4 nucleotídeos idênticos seguidos (“AAAA”, “TTTT”, “CCCC”, “GGGG”). Para avaliar os possíveis “off-targets”, ou regiões do genoma alvo em que a sequência poderia se ligar não especificamente, foi feito um alinhamento local utilizando-se do Blast (ALTSCHUL et al., 1990), que consiste em uma ferramenta que encontra regiões de similaridade entre sequências biológicas. Todas as candidatas que apresentavam alinhamento significativo de mais de 16 nucleotídeos seguidos com regiões “off-target” foram excluídas.

Outro critério interessante de ser avaliado nas candidatas a sequências-guia são a presença de sítios de restrição, que pode ser vantajosa ou desvantajosa dependendo do contexto. Se a sequência possuir sítios de restrição para enzimas que serão utilizadas para a clonagens futuras, isso inviabilizaria o processo, por outro lado se esse não for o caso, a presença de um sítio de restrição na sequência de nucleotídeos da região onde ocorrerá a mutação, pode ser um aliado futuramente, possibilitando que se faça uma varredura das células modificadas caso a mutação gerada leve à perda deste sítio. A avaliação da presença dos sítios de restrição é feita no programa “online” NEBcutter (<http://nc2.neb.com/NEBcutter2/>).

## **RESULTADOS E DISCUSSÕES**

### **Seleção de genes de interesse**

Estudos anteriores revelaram que a supressão da produção de limoneno através da expressão anti-senso conduz a uma resistência em plantas cítricas a vários patógenos tais como *Xanthomonas citri* (RODRIGUEZ et al., 2011a, 2011b). O limoneno é um terpeno responsável por aproximadamente 90% de toda a composição de óleo essencial de citros e *Xanthomonas citri* é uma bactéria responsável pelo cancro cítrico. Por essa razão, genes relacionados a síntese desse terpeno foram selecionados como alvo, para avaliar futuramente se um fenótipo com supressão do mesmo apresentaria tolerância a essa doença. Os possíveis genes codificantes de terpeno sintases em *C. sinensis*, foram selecionados a partir de uma busca no genoma da espécie, disponível em <http://citrus.hzau.edu.cn/orange/index.php>, utilizando-se como palavra-chave “limonene”. Foram identificados nessa pesquisa 62 genes, sendo selecionados quatro para esse trabalho (Tabela 1).

Tabela 1 – Sequências candidatas para desenho de RNA-guia com alvo em terpeno sintase.

	a. presença de 4 nucleotídeos idênticos seguidos	b. % de GC	c. formação de estruturas secundárias em Tm > 0°C	d. alinhamento local com mais com regiões "off-target"	e. presença de sítios de restrição de enzimas da clonagem	f. presença de sítio de restrição único	viável
<b>1. Cs3g04170</b>							
AGACATTTGAA AATTTATAG	não	20%	-	-	-	-	x
TGCTTTATTTGT TGCAAGAG	não	35%	-	-	-	-	x
TAAAGCAGCCA TCAGAATCA	não	40%	não	sim	-	-	x
CATGATTCTGAT GGCTGCTT	não	45%	Hairpin a 6,8°C	-	-	-	x
TTATTTTTGGCC ATGATTCT	sim	-	-	-	-	-	x
CTGGCTTATTTT TGGCCATG	sim	-	-	-	-	-	x
GCTGACAAGGC ATTGGACTG	não	55%	não	sim	-	-	x
ATGATAATTTGA CAGTTGAT	não	25%	-	-	-	-	x
AGTTTGCTGAT CTCCTATCA	não	40%	Hairpin a 36,9°C	-	-	-	x
TGGTAGTTTGC TGATCTCCT	não	45%	não	não	não	não	sim
<b>2. Cs3g04190</b>							
TGAGGGATTAA TGCAAGAAG	não	40%	não	sim	-	-	X
TTGCATTAATCC CTCAACCT	não	40%	Hairpin a 6°C	-	-	-	X
AACCAAGGTTG AGGGATTAA	não	40%	não	sim	-	-	X
ACCTTGGTTAC CTCTGTAAA	não	40%	não	não	não	sim	sim
AGACATTTGAA ACCATTAC	não	30%	-	-	-	-	X
CTGGCTTATTTT TTGCCATG	sim	-	-	-	-	-	x
ATTGGACTGGC TTATTTTTT	sim	-	-	-	-	-	x
GCTGACAAGGC ATTGGACTG	não	55%	não	sim	-	-	x
ATTATAATTTGA CAGTTGAT	não	20%	-	-	-	-	x
AGTTTGCTGAT CTCCTATCA	não	40%	Hairpin a 36,9°C	-	-	-	x
<b>3. Cs3g04260</b>							
GAAGATGTGCC ATCCAGTGA	não	50%	não	-	-	-	x
TGAAGATGTGC CATCCAGTG	não	50%	não	-	-	-	x
ATGCACTTATTG TGACTATT	não	30%	-	-	-	-	x
ACAATAAGTGC ATACCTTTC	não	35%	-	-	-	-	x
TTTGTACCAGA AAGGTATGC	não	40%	Hairpin a 40,8°C	-	-	-	x
TACAAATCCGA TCATTGAGA	não	35%	-	-	-	-	x
GTAATCCAGAT TTAGTTCAC	não	35%	-	-	-	-	x
TCGTCAGATGA GATACAGAG	não	45%	não	não	não	não	sim
CGTCAGATGAG ATACAGAGA	não	45%	não	não	não	não	sim
GTCAGATGAGA TACAGAGAG	não	45%	não	não	não	não	sim

	a. presença de 4 nucleotídeos idênticos seguidos	b. % de GC	c. formação de estruturas secundárias em $T_m > 0^\circ\text{C}$	d. alinhamento local com regiões "off-target"	e. presença de sítios de restrição de enzimas da clonagem	f. presença de sítio de restrição único	Viável
4. Cs3g04410							
GCCTCGTATT ACAGCTTTGA	não	45%	não	sim	-	-	x
CTTTGAAGGA GAAAGCATCA	não	40%	Hairpin a $45.1^\circ\text{C}$	-	-	-	x
TGAAGGAGAA AGCATCATGG	não	45%	não	sim	-	-	x
CTCCTCCATG ATGCTTTTCTC	não	50%	não	sim	-	-	x
GAATTGCCAG GCCTCCTCCA	não	60%	não	sim	-	-	x
ATGATCAGCA AGAGCAAGCA	não	45%	não	sim	-	-	x
GATCAGCAAG AGCAAGCAAG	não	50%	não	sim	-	-	x
GAACAAGCAA AGCGCGCAC	não	57%	não	não	não	sim	sim
GCTCCAGTGC GCGCTTTGCT	não	65%	-	-	-	-	x
AGCTCCAGTG CGCGCTTTGC	não	65%	-	-	-	-	x
GCCAACGCAC ATACCTTTCT <sup>1</sup>	não	50%	não	sim	não	não	x

<sup>1</sup> Sequência obtida através do programa CRISPR-P.

### Seleção e Avaliação de Sequências-Alvo

Foram selecionadas as 10 sequências candidatas mais próximas ao promotor de cada um dos genes, utilizando como critério a busca por "GG" no gene a partir da posição 21 (20 nucleotídeos de sequência + 1 correspondente ao "N" do "NGG"). As sequências candidatas foram avaliadas quando aos critérios eliminatórios na seguinte ordem: a. Presença de 4 nucleotídeos idênticos seguidos; b. Porcentagem de GC; c. Formação de estruturas secundárias em  $T_m > 0^\circ\text{C}$  (Avaliado pela ferramenta "oligo analyzer" do GeneRunner), d. Alinhamento local com mais de 16 nucleotídeos seguidos com regiões "off-target" (Avaliado pelo Blast), e. Presença de sítios de restrição de enzimas que serão utilizadas posteriormente para clonagem nos vetores escolhidos (*BpiI*, *AvrII*, *NheI*, *XbaI* e *SacI*). As candidatas que não atendiam algum dos critérios de avaliação foram descartadas sem passar pelas avaliações seguintes. Apesar de não eliminatório, também foi avaliado o critério de f. presença de sítio de restrição (avaliado pelo *software* NEBcutter) único para facilitar a varredura de mutações.

Em outros casos onde não é possível excluir as sequências, pode-se levar em consideração quais são as regiões de promiscuidade e alta fidelidade da sequência-

guia (HSU et al., 2013; PATTAANAYAK et al., 2013). Estudos identificaram a região promíscua como os nove primeiros nucleotídeos mais próximos à extremidade 5', uma vez que essa região tolera mal pareamento com o DNA, enquanto a região de alta fidelidade foi chamada de "seed" e necessita de complementariedade total. Desta forma, não é necessária a exclusão de sequências-guias que apresentem alinhamento "off-target" porém na região "seed".

Os resultados das análises das sequências candidatas encontram-se na tabela 1, e a partir deles foi possível chegar a 6 sequências propícias para edição de genes de limoneno sintase, sendo elas: "TGGTAGTTTGCTGATCTCCT" (Cs3g04170), "ACCTTGTTACCTCTGTAAA" (Cs3g04190), "TCGTCAGATGAGATACAGAG", "CGTCAGATGAGATACAGAGA", "GTCAGATGAGATACAGAGAG" (Cs3g04260), e "GAACAAGCAAAGCGCGCAC" (Cs3g04410).

A fim de fazer uma comparação entre os resultados obtidos manualmente e os resultados obtidos através de um dos programas mais utilizados para seleção de espaçadores em plantas, uma busca foi feita utilizando como alvo os mesmos quatro genes possivelmente relacionados à síntese de limoneno. Infelizmente, apenas um desses genes foi encontrado no banco de dados de CRISPR-P (Cs3g04410), e conseqüentemente foi o único passível desta análise que resultou em diversos possíveis espaçadores. Essa ausência dos outros três genes pode ser devido a divergência do genoma base utilizado, que para esse programa encontra-se desatualizado em uma versão.

O programa pontua as sequências candidatas e atribui uma nota de acordo com seus critérios de avaliação, e nesse caso, a sequência "GCCAACGCACATACCTTTCT" foi a única que obteve nota máxima e então foi avaliada segundo os mesmos critérios das sequências candidatas escolhidas manualmente para fins de comparação. Os resultados dessa avaliação também se encontram na tabela 1 e através deles é possível notar que a sequência atende a todos os requisitos exceto a presença de "off-targets", que uma vez analisadas apresentou 94% de identidade com um possível fator de transcrição "RF2b". Além disso quando comparamos a posição dessa sequência em relação ao gene, verificamos que ela está localizada ao final do gene, ou seja distante



do “códon” iniciador, o que não é desejado quando se trata de edição de genes codificadores de proteínas.

## CONCLUSÃO

A seleção de sequências alvo é uma etapa crucial para o sucesso de experimentos de edição de genomas, e apesar de poder contar com a ajuda de *softwares*, pode ser realizada manualmente afim de garantir um controle minucioso das etapas de seleção. Além disso a seleção manual permite ao pesquisador selecionar sequências mais favoráveis às individualidades de cada experimento como por exemplo a ausência de sítios de restrição de enzimas que serão posteriormente utilizadas no processo de clonagem do RNA-guia.

Neste trabalho foram obtidas sequências guias para a edição de genoma de 4 genes relacionados à síntese de limoneno, e serão utilizadas para customizar RNAs-guia, otimizando assim as chances da geração de uma mutação funcional e consequentemente do sucesso do experimento. Os critérios listados mostraram-se eficientes na seleção de sequências-alvo quando comparados ao resultado obtido “online”. Desta forma, este trabalho representa uma contribuição para a expansão da técnica CRISPR/CAS9, uma vez que pode ser utilizado para orientar pesquisadores em uma análise rápida e pouco trabalhosa, gerando uma economia de tempo e dinheiro futuros.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALTSCHUL, S.F. et. Al. Basic local alignment search tool. **Journal of Molecular Biology**. v.215. p.403-410, 1990.

CHAPMAN, J. R., TAYLOR, M. R. G., BOULTON, S. J. Playing the end game: DNA double-strand break repair pathway choice. **Molecular Cell**. v.47, n.4, p.497–510, 2012.

CONG, L., et.al. Multiplex genome engineering using CRISPR/Cas systems. **Science**. v.339, p.819–823, 2013.

HORVATH, P., BARRANGOU, R. CRISPR/Cas, the immune system of bacteria and archaea. **Science**. 327(5962), 167-70, 2010.

HSU, P.D. DNA targeting specificity of RNA-guided CAS9 nuclease. **Nature Biotechnology**. V. 31. P. 827-832, 2013.

ISHINO, Y et. al. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. **Journal of Bacteriology** v.69, n.12, p.5429-5433, 1987.

JINEK, M., et. al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. **Science**. 337, 816–821, 2012.

LIU, Y.; HEYING, E.; TANUMIHARDJO, S. A. History, global distribution, and nutritional importance of citrus fruits. **Comprehensive Reviews in Food Science and Food Safety**. v.11, n.6, p. 530-545, 2012.

MALI, P., ESVELT, K. M., CHURCH, G. M. Cas9 as a versatile tool for engineering biology. **Nature Methods**. v.10, p.957–963, 2013.

PATTANAYAK, V. et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed CAS9 nuclease specificity. **Nature Biotechnology**. v.31. p.839-843,2013.

PEREIRA, T. C. Introdução à Técnica de CRISPR. Ribeirão Preto, SP: Sociedade Brasileira de Genética, 250 p, 2016.

RODRÍGUEZ, A., SAN ANDRÉS, V, CERVERA, M., REDONDO, A., ALQUÉZAR, B., SHIMADA, T., GADEA, J., RODRIGO, M.J., O ZACARÍAS, L., PALOU, L., LÓPEZ, M. M., CASTAÑERA, P., PEÑA. Terpene down-regulation in orange reveals the role of fruit aromas in mediating interactions with insect herbivores and pathogens. **Plant Physiology** 156, 793–802, 2011a.

RODRIGUEZ, A.; ANDRÉS, V. A.; CERVERA, M.; REDONDO, A.; ALQUÉZAR, B.; SHIMADA, T.; GADEA, J.; RODRIGO, M.; ZACARÍAS, L.; PALOU, L.; LOPEZ, M. . ; CASTAÑERA, P.; PEÑA, L. The monoterpene limonene in orange peels attracts pests and microorganisms. **Plant Signaling & Behavior**. v.6, n.11, p.1820-1823, 2011b.

SCHIML S., FAUSER F. AND PUCHTA H. The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. **Plant Journal**. v.80, n.6, p.1139-1150, 2014.