



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



FACULTAD DE
CIENCIAS

UDELAR fcienc.edu.uy

Generación de herramientas moleculares para la optimización de metodologías de remplazo alélico en soja

Tesina de Grado - Licenciatura en Ciencias Biológicas 2020

Laboratorio de Biología Molecular Vegetal

María Pía Coronel

Tutora: Andrea Luciana Fleitas

Co-tutora: Sabina Vidal



Contenidos

Agradecimientos	3
Abreviaciones.....	4
Resumen	7
Introducción.....	8
Mecanismos de reparación del ADN	9
Edición del genoma utilizando nucleasas programables	13
Zinc Finger Nucleases	14
Transcription Activation-like Efector Nucleases	14
Meganucleasas.....	15
Sistema CRISPR/Cas9.....	16
Proteína Cas9	18
Aplicación del sistema CRISPR/Cas9 en plantas	19
Sistemas CRISPR-Cas9 libres de ADN.....	22
Reemplazo alélico y <i>gene-targeting</i> en plantas.....	24
Propuesta FMV: Optimización de técnicas de edición génica utilizando un modelo de resistencia a herbicida.....	27
Mecanismos de resistencia al herbicida glifosato	27
Objetivos	29
Materiales y métodos	30
1. Extracción de ADN genómico	30
2. Extracción de ARN y síntesis de ADNc	30
3. Clonado de CDS <i>EPSPS1</i> en vector de entrada pENTR2B.....	31
4. Generación de mutantes TIPS y TIPA sobre el vector de entrada.....	32
5. Subclonado en vector binario pUB-Dest.....	33
6. Generación de células quimiocompetentes TOP 10 de <i>E. coli</i>	34
7. Transformación de células quimiocompetentes de <i>E. coli</i>	34
8. Generación de células electrocompetentes de <i>Agrobacterium tumefaciens</i> C58C1	34
9. Transformación de células electrocompetentes de <i>A. tumefaciens</i>	35
10. Chequeo de transformantes de <i>A. tumefaciens</i> con el vector binario.....	35
11. Material Vegetal	36
12. Transformación de <i>Arabidopsis thaliana</i>	36
13. Diseño de adaptadores y generación de construcciones para síntesis de sgARNs	37

14.	Síntesis de sgARN por IVT	38
15.	Expresión y purificación de la proteína SpCas9-MBP-GFP	39
16.	SDS-PAGE.....	40
17.	Validación <i>in vitro</i> de los sgARNs.....	41
18.	Generación de vector de remplazo	42
19.	Minipreparación casera (miniprep)	43
20.	Minipreparación de kit Qiagen (QIAprep Spin Miniprep Kit Cat. 27106)	44
21.	Purificación de producto PCR de kit Qiagen (QIAquick PCR Purification Kit Cat. 28104) 44	
22.	Purificación de gel de kit Qiagen (QIAquick Gel Extraction Kit Cat. 28706)	45
23.	Electroforesis en gel de agarosa	45
	Resultados y discusión	46
	Identificación y análisis de genes <i>EPSPS</i> de soja.....	46
	Generación y caracterización de mutantes TIPS y TIPA de la <i>EPSPS1</i> de soja	50
	Herramientas moleculares para edición del genoma de soja.....	60
	Conclusiones y perspectivas.....	77
	Bibliografía	80
	Anexo	87
	Secuencia genómica del gen <i>EPSPS1</i> de <i>Glycine max</i> y con respectivos <i>primers</i> utilizados ...	88

Agradecimientos

En primera instancia quiero agradecer a Lu y a Sabi por haberme incluido en el equipo de BMV, siempre me acuerdo de aquella vez que fui para conocerlas al laboratorio, bien a principios del 2019, y entre risas y mucho nerviosismo (mío, obvio) me hicieron sentir parte desde el arranque. También quiero agradecer al resto de los miembros de BMV; Alex, Ceci, Joel, Juan, Marcel, Clau, Adrián y María por haber sido mis compañeros todo este tiempo y por haber estado dispuestos a ayudarme y sobretodo acompañarme en este proceso de aprendizaje y crecimiento personal.

Lu, no sé cómo agradecerle todo lo que has hecho por mí y todo lo que me has enseñado. Me emociono un poco porque, a mi entender, fuiste y seguís siendo una tutora y compañera increíble. Sin duda fue un camino largo y voy a recordar con mucho cariño la paciencia que siempre tuviste conmigo y las ganas de motivarme y hacerme enamorar de este trabajo y de lo que implica hacer ciencia. Vos sabes más que nadie que a veces me cuesta confiar plenamente en mis capacidades y gracias a vos puedo decir que estoy muy orgullosa de este trabajo de tesis y que estoy aún más entusiasmada por lo que se viene. Gracias por haberme hecho compañía durante todo este proceso, no lo cambiaría por nada.

Sabi, también quiero agradecerle particularmente por haberme guiado y por hacerme ver mi valor como “científica en construcción”. Siempre has estado atenta a mí y a mi trabajo y me parece que has enriquecido muchísimo esta experiencia. Gracias por haber estado en los momentos difíciles y por, junto con Lu, haberme dado un montón de oportunidades.

Por otro lado quiero agradecer a Manuel Sanguinetti y Ana Ramón, además de todo el equipo de su laboratorio y la sección, por haberme acercado por primera vez en la carrera a lo que es el trabajo científico y por haberme hecho ver que la biología molecular es a lo que me quiero dedicar. Sin dudas haber trabajado con ustedes durante el PAIE me generó aún más entusiasmo por esta carrera. También aprecio mucho el acompañamiento que me dieron durante esa etapa.

Casi terminando quiero agradecer a mi familia, a mis padres y mi hermano por haberme “aguantado” durante estos cuatro años de licenciatura y sobretodo en esta última etapa, que ustedes más que nadie saben el esfuerzo y la motivación por detrás. Gracias por estar durante los momentos más difíciles y haberme guiado toda la vida. También agradezco la libertad que me han dado para elegir mi camino.

Finalmente quiero agradecer a mis amigos, a “La Onda” (Diego, Lucas, Nicasio y Agustín), a Bruno, Cami, Pablo y a Ome porque son los mejores compañeros de facultad (y de la vida) que uno podría desear y por haberme hecho disfrutar de semejante etapa para nuestras vidas y por sobre todo por estar ahí para mí, siempre. También agradezco a mis amigas de siempre Lupe, Luli, Anita, Marga, Pocha, Paz y Jose por acompañarme y estar a mi lado hace tantos años. Gracias por haber sido pacientes durante mis ausencias. Les tengo muchísimo cariño y siempre voy a agradecer su apoyo. Por último, gracias Lupe especialmente por colaborar conmigo con el diseño de la caratula para este trabajo y compartir tu talento.

Este trabajo se lo quiero dedicar especialmente a Iris, que desde que tengo memoria apoyó mis ganas de hacer ciencia y por todo lo que hizo por mí y por mi vida.

Abreviaciones

APS	<i><u>A</u>mmonium <u>p</u>ersulfate</i>
BER	<i><u>B</u>ase <u>E</u>xcision <u>R</u>epair</i>
BrET	<i><u>B</u>romuro de <u>e</u>tidio</i>
Cas9	<i><u>C</u>RISPR <u>a</u>sociated <u>9</u></i>
cADN	<i><u>C</u>omplementary <u>A</u>DN</i>
CDS	<i><u>C</u>oding <u>S</u>equence</i>
CRISPR	<i><u>C</u>lustered <u>R</u>egularly <u>I</u>nterspaced <u>S</u>hort <u>P</u>alindromic <u>R</u>epeats</i>
cr-ARN	<i><u>C</u>RISPR <u>R</u>NA</i>
CV	<i><u>C</u>olumn <u>v</u>olumes</i>
DDR	<i><u>D</u>NA <u>D</u>amage <u>R</u>esponse</i>
DEPC	<i><u>D</u>ietilpiro<u>c</u>arbonato</i>
DNA	<i><u>D</u>eoxyribo<u>n</u>ucleic <u>a</u>cid</i>
DSB	<i><u>D</u>ouble <u>S</u>trand <u>B</u>reak</i>
DTT	<i><u>D</u>itio<u>t</u>reitol</i>
EDTA	<i><u>E</u>thylene<u>d</u>iamine<u>t</u>etra<u>a</u>cetic <u>a</u>cid</i>
EPSP	5-enolpiruvatoshikimato-3-fosfato
EPSPS	5-enolpiruvatoshikimato-3-fosfato sintasa
EtOH	Ethanol
FNU	<i><u>F</u>racción <u>N</u>o <u>U</u>nida</i>
Fw	<i><u>F</u>or<u>w</u>ard</i>
GFP	<i><u>G</u>reen <u>f</u>luorescent <u>p</u>rotein</i>
GT	<i><u>G</u>ene <u>T</u>argeting</i>
His-Tag	Cola de polihistidina
HR	<i><u>H</u>omologous <u>R</u>ecombination</i>
IEX	<i><u>I</u>on <u>E</u>xchange <u>C</u>hromatography</i>

IMAC	<i><u>I</u>mmobilized <u>M</u>etal <u>A</u>ffinity <u>C</u>hromatography</i>
IPTG	<i><u>I</u>sopropyl <u>β</u>-<u>D</u>-1-<u>thi</u>ogalactopyranoside</i>
IVT	<i><u>I</u>n <u>V</u>itro <u>T</u>ranscription</i>
KI	<i><u>K</u>nock <u>I</u>n</i>
KO	<i><u>K</u>nock <u>O</u>ut</i>
LB	<i><u>L</u>eft <u>B</u>order</i>
MBP	<i><u>M</u>altose <u>B</u>inding <u>P</u>rotein</i>
MGAP	<i><u>M</u>inisterio de <u>G</u>anadería <u>A</u>gricultura y <u>P</u>esca</i>
MMR	<i><u>M</u>ismatch <u>R</u>epair</i>
MOPS	<i>3-(<u>N</u>-<u>m</u>orpholino)<u>p</u>ropane<u>s</u>ulfonic acid</i>
MPM	<i><u>M</u>arcador de <u>P</u>eso <u>M</u>olecular</i>
mARN-Cas9	<i><u>M</u>essenger <u>ARN</u> of <u>Cas9</u></i>
N ₂	<i><u>N</u>itrógeno líquido</i>
NaAc	<i><u>S</u>odium acetate</i>
NER	<i><u>N</u>ucleotide <u>E</u>xcision <u>R</u>epair</i>
NHEJ	<i><u>N</u>on <u>H</u>omologous <u>E</u>nd <u>J</u>oining</i>
NLS	<i><u>N</u>uclear <u>L</u>ocalization <u>S</u>ignal</i>
NUC	<i><u>N</u>uclease Lobe</i>
ON	<i><u>O</u>ver<u>n</u>ight</i>
PAM	<i><u>P</u>rotospacer <u>A</u>djacent <u>M</u>otif</i>
PCR	<i><u>P</u>olymerase <u>C</u>hain <u>R</u>eaction</i>
PEG	<i><u>P</u>oly<u>e</u>thylene <u>g</u>lycol</i>
PEP	<i><u>P</u>hospho<u>e</u>no<u>l</u>pyruvate</i>
RB	<i><u>R</u>ight <u>B</u>order</i>
REC	<i><u>R</u>ecognition Lobe</i>
RGEN	<i><u>R</u>NA <u>G</u>uided <u>E</u>ndo<u>n</u>ucleases</i>
ARN	<i><u>R</u>ibo<u>n</u>ucleic <u>A</u>cid</i>

ARNi	<i>ARN Interference</i>
RNP	<i>Ribonucleoprotein</i>
RT	<i>Room Temperature</i>
Rv	<i>Reverse</i>
RVDs	<i>Repeat Variable Diresidues</i>
S3P	<i>Shikimate-3-Phosphate</i>
SDSA	<i>Synthesis-Dependent Strand Annealing</i>
SDS-PAGE	<i>Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis</i>
sgARN	<i>Single Guide RNA</i>
SSA	<i>Single strand annealing</i>
SSB	<i>Single Strand Break</i>
TAd	<i>Target Adaptor</i>
TALE	<i>Transcription Activator-like Effector</i>
TALEN	<i>Transcription Activation-like Effector Nuclease</i>
TCEP	<i>Tris (2-carboxyethyl) phosphine</i>
T-DNA	<i>Transfer DNA</i>
TEMED	<i>Tetramethylethylenediamine</i>
TIPA	<i>Thr102Ile + Pro106-Ala</i>
TIPS	<i>Thr102Ile + Pro106-Ser</i>
tracrARN	<i>Trans-activating CRISPR ARN</i>
UV	<i>Ultravioleta</i>
WT	<i>Wild Type</i>
X-gal	<i>5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside</i>
ZFN	<i>Zinc-finger Nucleases</i>
ZFP	<i>Zinc-finger Protein</i>

Resumen

La soja es el principal cultivo de exportación de Uruguay, además de ocupar la mayor superficie de siembra del país. Sin embargo, la productividad de este cultivo es sensiblemente menor a la reportada para la región y para los Estados Unidos, lo cual en parte se debe a factores de estrés biótico y abiótico. Algunos enfoques utilizados para mejorar caracteres de rendimiento y calidad en este cultivo se basan en estrategias de ingeniería genética para sobre-expresar o silenciar genes candidatos utilizando la transgénesis. Más recientemente se han desarrollado metodologías de edición génica, que permiten modificar genes específicos de manera precisa, siendo el sistema CRISPR/Cas9 la tecnología más utilizada y robusta. En algunas especies se ha podido realizar la modificación del genoma utilizando metodologías “libres de ADN”, en las que la proteína Cas9 se incorpora bajo la forma de proteína o como su ARN mensajero codificante. Esta estrategia garantiza la no incorporación de ADN foráneo y reduce los efectos fuera del sitio objetivo (*off-target*).

El objetivo general de este trabajo es optimizar metodologías de edición génica en soja (*Glycine max*) con el fin de generar plantas no transgénicas con modificaciones sitio-específicas utilizando el sistema CRISPR/Cas9. En particular, el trabajo propone optimizar metodologías de remplazo alélico. De manera de poner a punto estas metodologías sobre un modelo que proporcione un fenotipo fácil de identificar, el abordaje experimental propuesto consiste en remplazar un gen que codifica para una 5-enolpiruvato-shikimato-3-fosfato sintasa (EPSPS) por una versión de la enzima que sea insensible al herbicida glifosato.

Como resultado de este trabajo se generaron variantes de la EPSPS1 de soja: Thr102-Ile + Pro106-Ser (TIPS) y Thr102-Ala + Pro106-Ser (TIPA). Las variantes mutantes y la wild type (WT) fueron clonadas en un vector para la expresión constitutiva en plantas y transformadas en *Arabidopsis thaliana*. Paralelamente, se generaron vectores específicos para modificar el genoma de soja mediante remplazo alélico del gen *EPSPS1* por la versión conteniendo las mutaciones TIPS o TIPA. A su vez, se diseñaron *Single Guide RNAs* (sgARNs) para dirigir el corte de Cas9 a una región próxima al sitio donde se busca introducir las mutaciones. Se purificó Cas9 recombinante que se utilizó para evaluar la actividad de los sgARNs diseñados mediante ensayos de corte *in vitro*, utilizando complejos ribonucleoproteicos sobre la secuencia de *EPSPS1*.

En el futuro, se determinará si las mutaciones TIPS y TIPA confieren efectivamente resistencia al glifosato en *A. thaliana*. Utilizando todas las herramientas generadas, se apuntará a dilucidar cuál es la forma más eficiente de entrega de la proteína Cas9 y los sgARNs y qué tejido es más adecuado para realizar el remplazo alélico en soja.

Introducción

La soja (*Glycine max*) es una leguminosa cultivada principalmente en América del Norte y en América del Sur, donde se da el 40 % y el 45 % de la producción mundial, respectivamente. Este cultivo es una fuente de alimento rico en proteínas para personas y animales, siendo clave en la industria alimenticia en el mundo (FAO, 2012). En Uruguay la soja es de los principales cultivos, ocupando 966.000 hectáreas en la zafra de 2018-2019 y 910.000 hectáreas cultivadas en la zafra 2019-2020 (*Anuario Estadístico Agropecuario 2019, MGAP*). En cuanto a la producción, en el período 2018-2019 se cosecharon 2.8 millones de toneladas de soja según datos del Ministerio de Ganadería Agricultura y Pesca (MGAP) (*Anuario Estadístico Agropecuario 2019, MGAP*). Su exportación representó en dicho período (2018-2019) una ganancia cercana a los US\$ 903 millones. Sin embargo, la productividad de este cultivo es sensiblemente menor a la reportada para la región y para los Estados Unidos ya que es afectado por factores bióticos y abióticos. La dificultad para mejorar caracteres complejos y la estrecha base genética con la que cuentan los programas de mejoramiento hacen que la disponibilidad de cultivares con rasgos de tolerancia a diversos tipos de estrés sea escasa (Waqas et al., 2019).

El mejoramiento biotecnológico convencional es un proceso laborioso y lento para lograr resultados deseables (Kanchiswamy, 2016). Algunos enfoques utilizados actualmente para mejorar caracteres de rendimiento y calidad en este cultivo se basan en estrategias de ingeniería genética para sobreexpresar o silenciar genes candidatos utilizando la transgénesis. Si bien estas metodologías han permitido grandes avances, también presentan algunas desventajas. Una de ellas tiene que ver con el carácter aleatorio que tiene la inserción del ADN en el genoma vegetal cuando se emplean las metodologías clásicas de transformación de plantas (*Agrobacterium tumefaciens* o por biolística) (Raji et al., 2018). La inserción de un fragmento de ADN en el genoma de la planta puede eventualmente causar la disrupción de genes endógenos o el silenciamiento del transgen, dependiendo del sitio y el patrón de inserción en el genoma. Por otro lado, las metodologías de silenciamiento génico basadas en la interferencia del ARN (ARNi) pueden resultar en el silenciamiento no deseado de familias génicas enteras (Sledz & Williams, 2005). Por otra parte, los costos que insumen el desarrollo y la evaluación de un producto transgénico previo a su liberación son muy elevados (aproximadamente US\$ 10 millones por evento transgénico), lo cual deriva en que los desarrollos biotecnológicos que involucran la producción de cultivos transgénicos, están liderados casi exclusivamente por compañías multinacionales (Schmidt, 2005).

En la actualidad, las proyecciones de crecimiento poblacional mundial generan una gran demanda de productos alimenticios. El impacto del cambio climático y la emergencia de nuevos patógenos vegetales hacen imperiosa la búsqueda de nuevas herramientas de mejoramiento vegetal (*Kanchiswamy, 2016*). Recientemente se ha instalado el uso de metodologías de edición génica, que permiten modificar genes específicos de manera muy precisa, superando algunas flaquezas de las metodologías antes mencionadas (*Kanchiswamy, 2016*). Al utilizar estas metodologías, los tiempos para la generación de los organismos modificados se aceleran y las modificaciones generadas pueden ser indistinguibles de la variación natural, lo cual es de esperar que facilite significativamente los procesos de evaluación para la liberación comercial de las nuevas variedades editadas.

Todas las metodologías de edición génica disponibles en la actualidad, se basan en la generación de un daño en secuencias específicas del ADN, que luego es reparado por los sistemas celulares de reparación del ADN (*Voytas, 2013*). Durante el proceso de reparación del ADN es frecuente que ocurran errores, en particular en ausencia de una molécula con homología con la región dañada. Los errores incorporados pueden resultar en la sustitución de un nucleótido por otro, o en pequeñas inserciones o deleciones (*indels*) que pueden ocasionar el corrimiento del marco de lectura de genes y originar proteínas truncadas o aberrantes, no funcionales (*Voytas, 2013*).

Utilizando metodologías de edición génica, también es posible modificar la expresión de genes alterando las secuencias promotoras (*Elison et al., 2017*), o bien realizar estudios funcionales sobre genes no codificantes, como aquellos que controlan la expresión de pequeños ARNs (*Zhou et al., 2017*). Otra posibilidad que ofrece el uso de estas tecnologías tiene que ver con la inserción sitio-específica de una secuencia (*knock-in*), o el remplazo de una secuencia por otra (*gene targeting/gene replacement*). Estos fenómenos se pueden dar cuando, junto con la maquinaria que genera el daño en el genoma, se introduce una molécula de ADN que es utilizada como molde para reparar el ADN genómico (*Voytas, 2013*).

En las siguientes secciones se comentan los mecanismos moleculares de daño y reparación del ADN y los fundamentos de las metodologías de edición génica más utilizadas en la actualidad.

Mecanismos de reparación del ADN

Todos los organismos están constantemente expuestos a factores ambientales estresantes y productos genotóxicos que producen, directa o indirectamente, daño en el ADN (*Spampinato, 2017*). Debido a esto, la integridad genómica de cada organismo se ve permanentemente desafiada por factores endógenos y exógenos que dañan el ADN. En este

sentido, es crucial para los organismos desarrollar mecanismos eficientes para reparar los daños causados. Los procesos de reparación del ADN se han caracterizado en sistemas de modelo bacterianos, fúngicos y de mamíferos. La descripción de los procesos de reparación en plantas se inició de forma relativamente reciente (*Manova & Gruszka, 2015*).

Las células responden al daño del ADN promoviendo las vías DDR (DNA Damage Response, respuesta al daño en el ADN). Éstas implican la activación de varios mecanismos que conducen a la reparación del ADN, al arresto del ciclo celular o a la muerte celular programada. Las vías DDR en plantas también incluyen endoreduplicación, proceso que implica la replicación del ADN nuclear sin necesidad de realizar citocinesis (*Spampinato, 2017*).

Las lesiones de ADN se dividen en dos categorías principales: de simple y de doble cadena (*Manova & Gruszka, 2015*). La primera categoría está compuesta por lesiones que perturban solo una cadena del ADN, como oxidación o alquilación de bases, pérdida de bases, aductos de ADN, enlaces cruzados intracatenarios, fotoproductos de ADN y cortes de ADN de cadena sencilla (Single Strand Break, SSB) (*Manova & Gruszka, 2015*). En la segunda categoría se incluyen lesiones que afectan a ambas cadenas de ADN, tales como *cross-link* y cortes de ADN de doble cadena (Double Strand Break, DSB), siendo este último el tipo de lesión más drástica en el ADN del genoma eucariota (*Manova & Gruszka, 2015*).

Los daños en el ADN pueden repararse por varias vías. En plantas, se pueden encontrar: la fotorreactivación, la reparación por escisión de base (Base Excision Repair, BER), la reparación por escisión de nucleótidos (Nucleotide Excision Repair, NER), la reparación por *mismatch* (Mismatch Repair, MMR), por recombinación homóloga (Homologous Recombination, HR) y por recombinación no homóloga o unión de extremos no homólogos (Non Homologous End Joining, NHEJ) (*Manova & Gruszka, 2015*). Estas vías son activas a lo largo de diferentes etapas del ciclo celular y del desarrollo (*Manova & Gruszka, 2015*). Las vías BER, NER y MMR actúan sobre daños causados sobre una de las hebras, mientras que la HR y la NHEJ reparan daños que afectan ambas hebras del ADN (*Manova & Gruszka, 2015*).

La fotorreactivación es la reversión de los dímeros de pirimidina inducidos por la luz ultravioleta (UV). Este mecanismo está mediado por flavoenzimas activadas por la luz llamadas fotoliasas (*Spampinato, 2017*). La fotorreactivación es el mecanismo más efectivo utilizado por las plantas para reducir rápidamente los efectos negativos de la exposición a la luz solar (*Manova & Gruszka, 2015*).

La reparación por BER reconoce y repara lesiones que incluyen bases desaminadas, oxidadas y alquiladas, sitios abásicos (apurínicos o apirimidínicos, AP) y SSBs. Tras el reconocimiento del sitio afectado y la eliminación de la base dañada por las ADN glicosilasas y

endonucleasas AP, la región eliminada es sintetizada por la ADN polimerasa y ligada (Spampinato, 2017).

En la vía MMR, los desajustes producidos por nucleótidos insertados o eliminados accidentalmente durante la replicación son reparados tomando como molde la hebra parental. Las proteínas Mut son fundamentales para reconocer las bases mal pareadas y dirigir a la maquinaria de reparación hacia la región (Manova & Gruszka, 2015).

La reparación NER participa en la reparación de daños de mayor gravedad, que distorsionan la doble hélice del ADN, como la formación de dímeros de pirimidina (dímeros de timina y 6,4-fotoproductos). El reconocimiento del sitio dañado lleva a la remoción de un tramo de una de las hebras del ADN, que es vuelto a sintetizar por la ADN polimerasa y ligado (Spampinato, 2017).

Cuando ocurren daños más serios sobre el ADN, por ejemplo, cortes en ambas cadenas (DSBs), se ponen en marcha los mecanismos de reparación por HR o NHEJ; la elección de la vía dependerá de la fase del ciclo celular. En la HR, los extremos dañados del ADN se corrigen en base a regiones con homología, mientras que en la vía NHEJ la información de la secuencia no juega un papel importante en la reparación (Manova & Gruszka, 2015).

La vía HR en plantas es favorecida durante la fase S/G2 del ciclo celular, cuando las cromátidas hermanas están disponibles para ser utilizadas como moldes homólogos (Mao et al., 2008). Los dos mecanismos principales dependientes de HR para corregir DSBs son 1) la hibridación de una hebra única (SSA: *Single strand annealing*) y 2) la hibridación dependiente de la síntesis de una hebra (SDSA: *synthesis-dependent strand annealing*) (Manova & Gruszka, 2015). En ambas vías se generan extremos 3'-protruyentes por acción de exonucleasas luego de la producción de un DSB. En el caso de la vía SSA, se generan extremos protruyentes complementarios entre sí en ambos extremos del ADN (Figura 1). Estas secuencias permiten la hibridación de los extremos autocomplementarios, hecho que es seguido de la extensión de las cadenas por la ADN polimerasa y ligación. Es importante remarcar que este tipo de reparación no requiere de otra molécula de ADN de doble cadena (dsADN: *double stranded ADN*) homólogo, sino que utiliza la homología dentro de sí misma (Manova & Gruszka, 2015). En la vía SDSA, en cambio, uno de los extremos 3'-protruyentes "invade" una doble hebra homóloga formando una estructura de tipo D-loop (*displacement loop*) (Roth et al., 2012). La cadena invasora, toma como molde la cadena complementaria en la otra molécula para repararse; este ADN sintetizado es complementario al extremo 3'-protruyente remanente, por lo que luego de la síntesis, ambas cadenas logran rehibridarse. Los gaps en el ADN terminan de sintetizarse y las cadenas se ligan (Roth et al., 2012).

Por otro lado, la vía NHEJ se encuentra activa durante todo el ciclo celular, incrementando su actividad durante las fases G1 y G2, cuando no existen moléculas de ADN con homología que puedan funcionar como moldes para la reparación (Mao et al., 2008). En esta vía, regiones de microhomología presentes en ambos extremos libres de la cadena dañada guían la reparación (Figura 1). Cuando los extremos protruyentes son totalmente complementarios, es posible que ocurra la reparación precisa del DSB, pero si la complementariedad es imprecisa, ocurre la pérdida o adición de nucleótidos. Varios factores son relevantes para el procesamiento de los extremos, la remoción de bases mal apareadas y la reparación de los gaps. Algunos de ellos son específicos de la vía NHEJ, como las ADN polimerasas λ y el complejo ligasa IV (Mao et al., 2008; Bennardo et al., 2008).

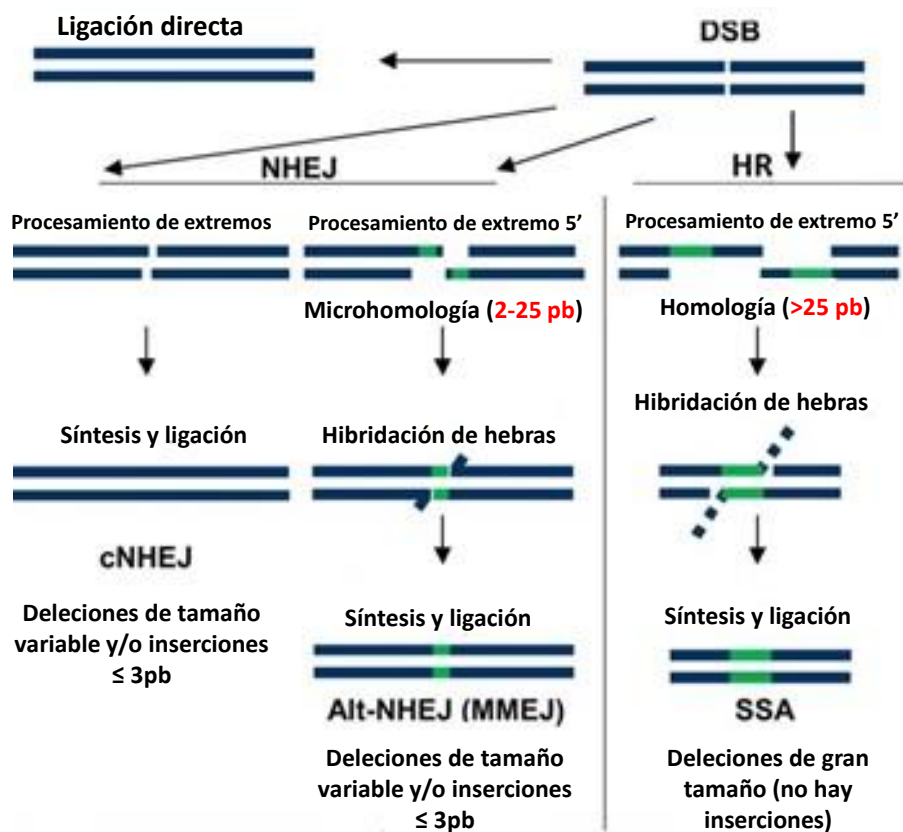


Figura 1. Esquema de las vías de reparación de DSB y sus potenciales resultados. Se muestra la vía de reparación NHEJ (*Non Homologous End Joining*) con sus dos mecanismos posibles; canónica y alternativa (cNHEJ y aNHEJ o MMEJ). La vía cNHEJ media la ligación directa de las hebras ADN dañadas y a su vez, no requiere un molde de homología para reparar las lesiones. La vía MMEJ (*Microhomology-mediated end joining*) hace uso de secuencias microhomólogas durante la alineación de los extremos dañados antes de hibridarlos, lo que resulta en deleciones que flanquean el DSB. Por otra parte, se muestra la vía HR (*Homologous Recombination*), particularmente la vía SSA (*Single Strand Annealing*) en la que se generan extremos protruyentes complementarios entre sí en ambos extremos del ADN y se permite la hibridación de dichos extremos autocomplementarios. **Imagen adaptada de Vu et al., 2014 (doi: 10.1105/tpc.114.126607).**

Las vías de reparación del ADN son de gran relevancia cuando se busca realizar mejoramiento genético mediante edición génica, ya que estas nuevas metodologías de mejoramiento se basan en la generación de daños en el ADN que son reparados por las maquinarias endógenas. Debido a que en general los daños son generados sobre ambas cadenas del ADN, las vías HR y NHEJ son las más relevantes en este aspecto.

Edición del genoma utilizando nucleasas programables

La modificación precisa del genoma sobre *loci* específicos es de gran interés para la biotecnología vegetal, al igual que para otras áreas de la biotecnología, aplicadas a la salud humana y animal (Gaj et al., 2016). Los sistemas convencionales de modificación sitio-específica basados en HR se han utilizado para realizar modificación de genes en algunos sistemas eucariotas, como la levadura. Sin embargo, con excepción de estos organismos, la eficiencia de HR en la mayor parte de las células eucariotas es extremadamente baja (1 en 10^6 a 1 en 10^7 ; Kim & Kim, 2014). En plantas, el musgo *Physcomitrella patens* es una herramienta poderosa para realizar estudios de genética reversa ya que su tasa de HR es tan alta como la de levadura (90 %; Schaefer & Zryd, 1997).

Con el surgimiento de las nucleasas programables, se abrió una posibilidad hasta el momento inexistente: la de modificar de manera precisa y dirigida genes de interés, sea para producir mutantes de pérdida de función (KOs, *Knockouts*): o, en menor frecuencia, para producir modificaciones específicas (*gene targeting*). Las nucleasas programables han sido utilizadas para manipular genes en una gran variedad de organismos y de tipos celulares con fines aplicados y básicos (Yu et al., 2014).

Los sistemas de edición génica pueden clasificarse en dos grandes grupos en base a la naturaleza de la molécula que participa en el reconocimiento de la secuencia blanco en el ADN. Por un lado, están los sistemas basados en proteínas quiméricas, sintéticas, capaces de reconocer y efectuar un corte en una secuencia blanco. Por otro lado, están los sistemas basados en complejos ARN-proteínas (RGENs, *RNA Guided Endonucleases*), en los cuales la molécula de ARN cumple la función de reconocimiento específico de la secuencia blanco (Khan, 2019).

Dentro de los sistemas basados en proteínas, encontramos las nucleasas de dedos de zinc (ZFNs), las nucleasas con efectores de tipo activadores de transcripción (*Transcription Activation-like Effector Nuclease*, TALENs) y las meganucleasas (Khan, 2019). Las ZFNs y las TALENs utilizan dominios proteicos de unión al ADN, presentes en factores de transcripción de plantas o de bacterias fitopatógenas, combinados de forma tal que la proteína sintética puede ser programada para reconocer una determinada secuencia específica en el ADN (Khan, 2019).

Estos dominios de reconocimiento están asociados a dominios catalíticos de endonucleasas que producen DSBs en el ADN blanco, como la enzima *FokI* de bacterias (Yu *et al.*, 2014). La enzima *FokI* está compuesta por un dominio nucleasa y un dominio de reconocimiento y para poder realizar cortes sobre el ADN es necesario que ocurra dimerización de la misma (Bitinaite *et al.*, 1998).

Dentro de las metodologías de edición génica, el sistema CRISPR/Cas9 es la tecnología más utilizada y robusta (Khan, 2019). Estas metodologías se basan en la utilización de ciertos componentes presentes en los *loci* CRISPR (Clustered Regular Interspaced Short Palindromic Repeats), los cuales cumplen una función en la inmunidad adaptativa de bacterias y arqueas frente a ácidos nucleicos invasores. En el sistema CRISPR, el reconocimiento del ADN está mediado por una pequeña molécula de ARN que forma un complejo ribonucleoproteico con la nucleasa Cas (*CRISPR Associated nuclease*) (Khan, 2019).

Las principales características de los distintos sistemas de edición génica se describen a continuación.

Zinc Finger Nucleases

Las ZFN son proteínas sintéticas que presentan una estructura modular compuesta por dos dominios: un dominio de unión al ADN de dedos de zinc (ZFP) y un dominio nucleasa, derivado de la enzima de restricción *FokI* (Kim & Kim, 2014) (Figura 2.A). Debido a que la enzima *FokI* actúa de manera dimérica, son necesarios dos monómeros de ZFP para crear una nucleasa activa. La especificidad de secuencia de las ZFN está determinada por los monómeros ZFP, que consisten en conjuntos en tándem de dedos de zinc C₂H₂. Cada dedo de zinc reconoce 3 pb de la secuencia del ADN y, por lo tanto, con 3 a 6 dedos de zinc se puede generar enzimas quiméricas capaces de identificar una secuencia de ADN de entre 9 y 18 pb (Figura 2.A). Por otra parte, la especificidad de unión al ADN de las ZFN puede modificarse por mutagénesis para generar nuevas versiones de nucleasas programables (Kim & Kim, 2014).

Varios grupos han logrado de forma exitosa realizar edición génica en plantas utilizando ZFNs, tanto para realizar KOs como para realizar *gene targeting*. En soja, Curtin *et al.* (2011) generaron mutaciones independientes en los genes parálogos DCL4a y DCL4b (*dicer-like*). Recientemente, Bonawitz *et al.* (2019) lograron integrar múltiples transgenes en el locus *FAD2-1a*.

Transcription Activation-like Effector Nucleases

Al igual que las ZFNs, las TALEN contienen el dominio de nucleasa *FokI*, pero a diferencia de las primeras, utilizan una clase diferente de dominios de unión al ADN. Estos

dominios derivan de proteínas presentes en bacterias del género *Xanthomonas* spp., que actúan como factores de virulencia (efectores) en plantas, activando genes vegetales relevantes para la patogenicidad de la bacteria. Estos factores, conocidos como efectores activadores de transcripción (TALE) (Kim & Kim, 2014), están compuestos de repetidos de 33–35 aminoácidos, cada uno de los cuales reconoce un único par de bases. La especificidad de nucleótidos de cada dominio repetido está determinada por dos aminoácidos en las posiciones 12 y 13, denominados RVDs (Repeat Variable Diresidues). Este reconocimiento permite mayor versatilidad que la que presentan las ZFNs en cuanto a las secuencias reconocidas (Figura 2.B; Kim & Kim, 2014).

Las TALENs han sido utilizadas en soja para editar simultáneamente dos genes codificantes para enzimas fitoeno desaturasa (*GmPDS11* y *GmPDS18*; Du et al., 2016), así como enzimas de la familia *dicer-like* (DCL2; Curtin et al., 2018). Finalmente, con esta tecnología se han generado variedades de soja con mayor contenido de ácidos grasos insaturados en su aceite, mediante mutaciones en genes codificantes para las enzimas FAD2 -1a y FAD2 -1b (Haun et al., 2014).

Meganucleasas

Las meganucleasas son proteínas de gran tamaño que reconocen largas secuencias de ADN (18-40 pb). También funcionan de forma dimérica, pero debido a que el reconocimiento del ADN está dado por un único sitio activo dentro de la nucleasa, es complejo generar variantes que puedan reconocer nuevas secuencias por lo que su versatilidad es significativamente menor a la que presentan los dos sistemas anteriormente descritos (Jurica & Stoddard, 1999).

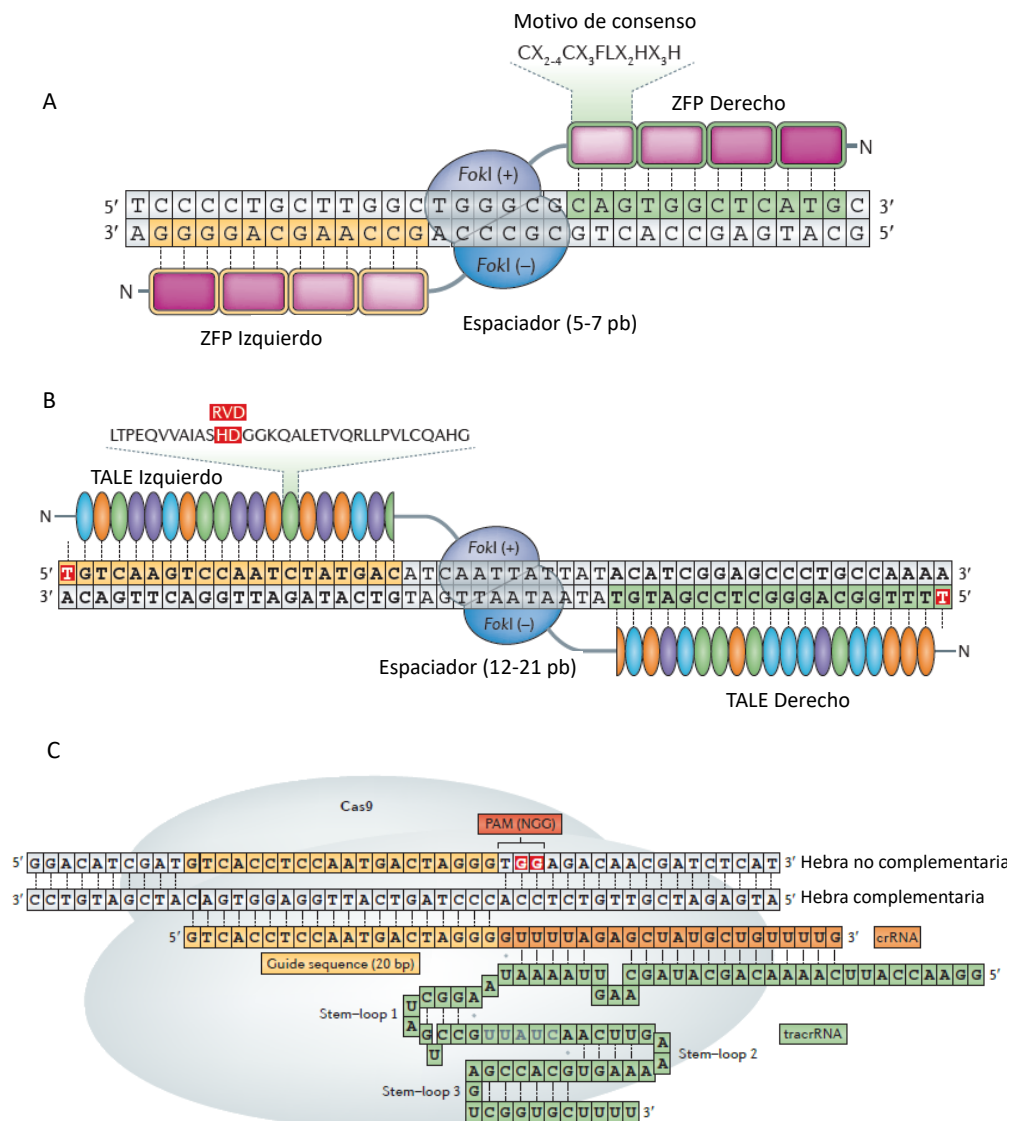


Figura 2. Nucleasas empleadas para edición génica. **A)** Esquema de la estructura de una nucleasa de dedo de zinc (ZFN). Cada ZFN está compuesto por una proteína de dedo de zinc (ZFP) en el extremo amino y el dominio nucleasa FokI en el extremo carboxilo. Se muestra el motivo aminoacídico consenso de los ZFN. **B)** Esquema de la estructura de una nucleasa TALE (TALENs). Cada TALEN está compuesta de dominios del unión al ADN provenientes de efectores activadores de transcripción (TALE) en el extremo amino y el dominio nucleasa FokI en el extremo carboxilo. Cada repetición de TALE está compuesta por 33–35 aminoácidos y reconoce un solo par de bases a través de los aminoácidos en las posiciones 12 y 13 gracias al motivo RVD. **C)** Esquema del sistema CRISPR/Cas9. La proteína Cas9 se asocia al ARN CRISPR (crARN) y a un crARN transactivador (tracrARN) que guían a la proteína hacia la secuencia blanco complementaria al crARN. El motivo adyacente a protoespaciador (PAM), localizado en la hebra no complementaria, es reconocido por la proteína Cas9 y es fundamental para que ésta produzca el corte en la secuencia. *Imagen adaptada de Kim & Kim, 2014 (doi: 10.1038/nrg3686).*

Sistema CRISPR/Cas9

Como se mencionó anteriormente, en los sistemas de tipo RGENs, una proteína de tipo nucleasa es guiada hacia su sitio de reconocimiento por pequeños ARNs. En bacterias y arqueas, los sistemas de escisión de ADN guiados por ARN proporcionan inmunidad adaptativa contra fagos o plásmidos invasores. Estos organismos a menudo capturan pequeños fragmentos de ADN (de aproximadamente 20 pb) del fago o del plásmido invasor e insertan

estas secuencias (denominadas protoespaciadores) en su propio genoma para formar una matriz CRISPR (Kim & Kim, 2014).

Los sistemas CRISPR/Cas se clasifican en 3 grupos: de tipo I, II y III (Figura 3). En los sistemas CRISPR de tipo II se encuentran diversas proteínas; Cas9 (Csn1), Cas12a (Cpf1), Cas13a (C2c2) y Cas13b (C2c6), siendo las más utilizadas la Cas9 y Cas12a (Makarova et al., 2018).

La proteína Cas9 (*CRISPR associated 9*) se asocia a dos ARNs no codificantes: el crispr ARN (crARN) y el ARN transactivador (*transactivating crispr RNA*, tracrARN). Ambas moléculas de ARN se sintetizan a partir del locus CRISPR. Los crARN se transcriben como pre-crARN y se procesan para generar pequeños ARNs con homología de secuencia a regiones específicas de ácidos nucleicos provenientes de infecciones previas. Por su parte, el tracrARN también se transcribe desde el locus CRISPR y contribuye al procesamiento del pre-crARN (Figura 3). La proteína Cas12a se asocia únicamente a un crARN y la generación de estos crARN maduros es independiente de un tracrARN (Yang et al., 2019). Por otra parte, en los sistemas CRISPR que no son de tipo II los *loci* CRISPR contienen múltiples proteínas Cas, que forman complejos con el crARN (complejo CASCADE para tipo I y complejos Cmr o Csm RAMP para tipo III) para facilitar el reconocimiento y destrucción de ácidos nucleicos blanco (Hsu et al., 2014).

En el sistema CRISPR/Cas9, el crARN cumple la función de reconocimiento por homología de la secuencia blanco, mientras que el tracrARN hibrida con crARN dándole estabilidad al dúplex crARN:tracrARN y permitiendo su acoplamiento con la nucleasa Cas9 para dirigir la escisión de secuencias blanco en el ADN (Figura 2.C). Un elemento fundamental para que ocurra un corte específico en el ADN, es la presencia en la secuencia blanco de un motivo adyacente al protoespaciador (*Protospacer Adjacent Motif*, PAM). Esta secuencia, que en el caso del sistema CRISPR/Cas9 de *Streptococcus pyogenes* suele ser un 5'-NGG-3' a continuación de la secuencia blanco, es reconocida directamente por la proteína Cas9. Proteínas Cas provenientes de otros organismos pueden reconocer secuencias PAM diferentes (Nishimasu et al., 2014). Cas12a reconoce principalmente PAMs ricos en T (TTTN) que preceden a la secuencia blanco y corta el ADN en un patrón escalonado generando extremos cohesivos y dejando *overhangs* de 4-5 nucleótidos (Yang et al., 2019).

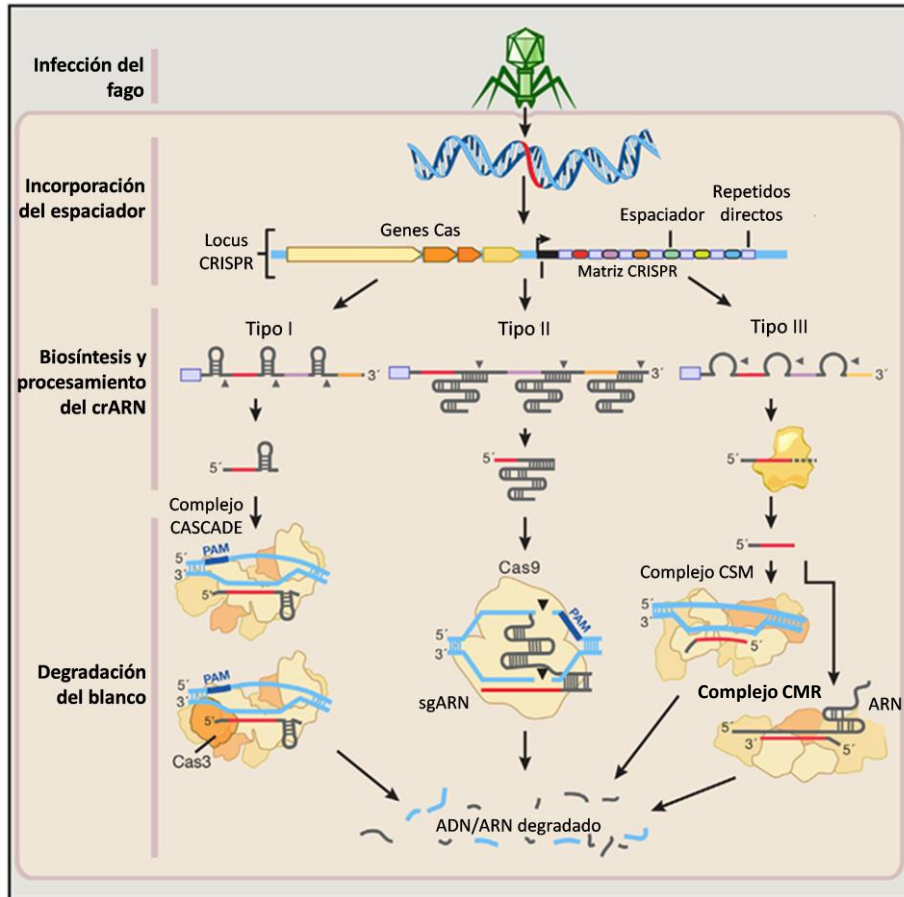


Figura 3. Esquema de la invasión celular por elementos genéticos foráneos de bacteriófagos o plásmidos. En una primera etapa, ocurre la infección del fago y en un segundo paso (*incorporación del espaciador*) enzimas asociadas a Cas obtienen espaciadores exógenos y los insertan en el locus CRISPR. La matriz CRISPR se compone por una secuencia líder rica en AT seguida de repeticiones cortas, de longitud variable, entre las cuales se hallan los protoespaciadores foráneos incorporados. A partir de la matriz CRISPR ocurre la transcripción de un pre-crARN no codificante que es madurado enzimáticamente mediante mecanismos característicos de cada sistema CRISPR (sea I, II o III). Para el tercer paso (biosíntesis y procesamiento del crARN), en los sistemas de tipo I y III, el crARN es clivado entre los repetidos por acción de ribonucleasas, mientras que en el sistema de tipo II el tracrARN se hibrida con los repetidos directos, formando un dúplex que es procesado por la ARNsa III. Los crARNs madurados en el sistema I y III son cargados en complejos proteicos, en tanto que en el sistema de tipo II el crARN-tracrARN forman un complejo con la proteína Cas9. En todos los casos, los complejos ribonucleoproteicos (RNPs) producen el corte de los ácidos nucleicos blanco. **Imagen adaptada de Hsu et al., 2014 (doi: 10.1016/j.cell.2014.05.010).**

Proteína Cas9

Los estudios cristalográficos de la proteína Cas9 demostraron que la dicha proteína consta de dos lóbulos: un lóbulo de reconocimiento (REC) y un lóbulo con función nucleasa (NUC). El lóbulo REC se puede dividir en tres regiones, una “hélice-puente”, un dominio REC1 y un dominio REC2. El lóbulo NUC consta de los dominios RuvC, HNH y de interacción con el PAM (PI) (Nishimasu et al., 2014). Los dominios nucleasa RuvC y HNH son los responsables de cortar respectivamente las hebras del ADN directa y complementaria respecto a la secuencia blanco de 20 nucleótidos. Ambos dominios poseen sitios de unión a metal, lo que determina que la actividad de la proteína sea dependiente de Mg^{2+} (Nishimasu et al., 2014). El

heteroduplex ARN:ADN blanco, negativamente cargado, se acomoda en un surco con densidad de carga positiva entre los lóbulos REC y NUC. El dominio PI es el principal determinante de la especificidad por el PAM y es fundamental para la actividad enzimática (Figura 4, *Nishimasu et al., 2014*). La estructura modular de la proteína facilita la modificación de una o más regiones de forma independiente lo que permite la generación de proteínas Cas con diferentes propiedades. Por ejemplo, la variante dCas9 (*dead Cas9*) contiene las mutaciones D10A y H840A en los dominios RuvC I y HNH que inactivan la actividad endonucleasa por completo. Esta variante inactiva mantiene su capacidad de unión al ADN y puede adosarse a dominios activadores o represores de la transcripción para modificar la expresión génica (*Qi et al., 2013; Chen & Qi, 2017*) o a proteínas fluorescentes para estudiar la dinámica del ADN (*Chen et al., 2013*), entre otras aplicaciones. Las mutaciones simples D10A o H840A permite la generación de nickasas, que sólo cortan una de las hebras del ADN (la hebra blanco o la complementaria, respectivamente). Estas variantes han sido utilizadas en algunos sistemas eucariotas para promover la HR sobre la vía NHEJ (*Atsushi et al., 2017*).

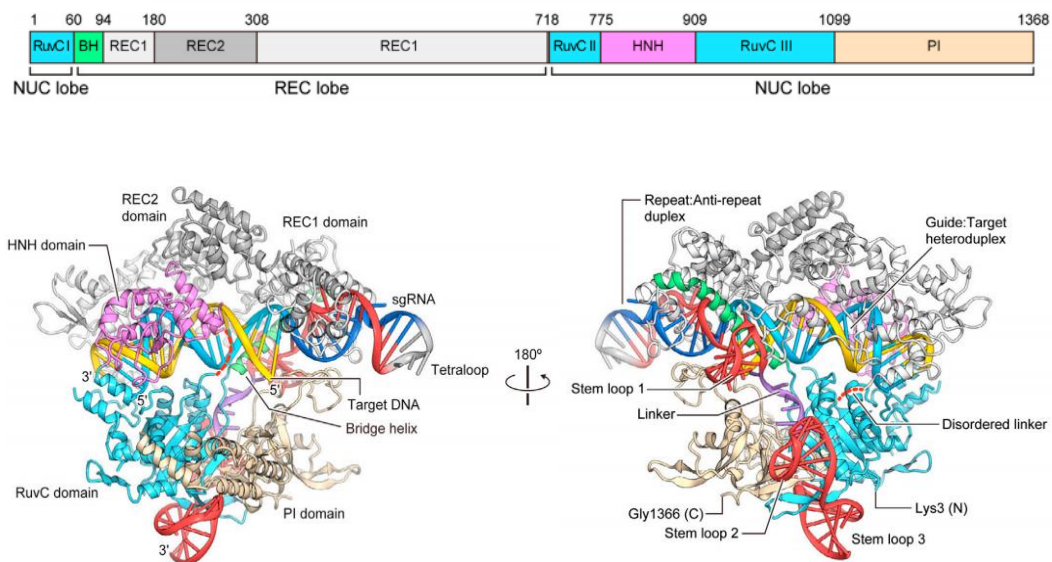


Figura 4. Estructura del complejo ternario Cas9-sgARN-ADN. **A)** Organización de los dominios de Cas9 de *S. pyogenes*. Se muestran los lóbulos REC y NUC con sus respectivos dominios. **B)** Representación en cinta del complejo Cas9-sgARN-ADN. Se muestran los dominios de Cas9 en complejo con el sgARN y el ADN blanco. **Imagen tomada de Nishimasu et al., 2014 (doi: 10.1016/j.cell.2014.02.001).**

Aplicación del sistema CRISPR/Cas9 en plantas

El sistema CRISPR/Cas9 posee ciertas ventajas sobre las herramientas como las ZFNs o las TALENs. Por un lado, en las nucleasas en las que el reconocimiento de las secuencias blanco está mediado exclusivamente por proteínas, el buen funcionamiento radica en la correcta

expresión de proteínas de gran tamaño. A pesar de que existen algunas soluciones comerciales para la generación de construcciones, tanto el ensamblado de las construcciones como la expresión de los productos proteicos dentro de los organismos pueden ser problemáticos (*Liu & Fan, 2014*).

El sistema CRISPR/Cas9 ofrece una forma más simple para obtener especificidad y presenta igual o mayor eficiencia de corte que las ZFNs o las TALENs. Para el uso del sistema en el laboratorio, el crARN y el tracrARN fueron modificados para formar una única molécula de ARN híbrida, denominada sgARN (*single guide RNA*), lo que facilita la generación de las construcciones para realizar edición génica (*Nishimasu et al., 2014*).

En plantas, la estrategia más común para realizar edición génica es a través de la transformación de plantas con construcciones génicas que expresen tanto los sgARNs, como la Cas9. Estas construcciones pueden tener como blanco de edición un único locus o múltiples loci. Asimismo, las construcciones son relativamente pequeñas, debido a que la expresión de los sgARNs en general está bajo el control de promotores para la ARN polimerasa III (por ejemplo, los promotores U3 o U6 de pequeños ARNs nucleares), los cuales son pequeños, como también lo son sus terminadores (tramo corto de timidinas; *Ma et al., 2016*). Existen metodologías que simplifican la generación de construcciones, como el clonado por Golden Gate, que permiten el ensamblado de múltiples cassettes para la expresión de sgARNs de forma sencilla (*Ma et al., 2016*), y también estrategias para el ensamblado de cassettes multiplex en las que varios sgARNs son expresados a partir de un único promotor como un ARN policistrónico que luego es procesado por maquinarias endógenas (*Xie et al., 2015*).

En el trabajo con plantas, muchas veces un cuello de botella es la transformación y la regeneración de plantas transgénicas. En el caso de la soja, los dos métodos más utilizados para la transformación son la transformación de cotiledones mediada por *Agrobacterium tumefaciens* (*Li et al., 2017*) y el bombardeo de partículas (biolística) sobre embriones somáticos (*Finer, 2016*). La transformación mediada por *Agrobacterium* tiene algunas ventajas sobre los métodos de biolística. Por ejemplo, a diferencia de los métodos físicos de transformación, el tamaño de las construcciones génicas no es un factor relevante en la eficiencia de transformación con *Agrobacterium* (*Bao et al., 2020*). A su vez, éste método suele resultar en la inserción en el genoma de un número menor de copias del transgen en comparación con los métodos basados en biolística. Sin embargo, estos últimos presentan niveles mayores de eficiencia de transformación de soja que los obtenidos con *Agrobacterium* (*Bao et al., 2020*). A su vez, ambas metodologías tienen la desventaja de ser fuertemente dependientes de la variedad; en el caso de la transformación con *Agrobacterium* es necesario utilizar variedades que sean susceptibles a este fitopatógeno, mientras que en el caso de la

biolística, es necesario que los genotipos de soja sean eficientes en la producción de embriones somáticos (*Finer, 2016*). Esos requisitos hacen que muchas veces los genotipos transformados no corresponden a cultivares de elite, sino que son líneas de referencia que luego deben ingresar a los programas de mejoramiento genético clásico a través de cruzamientos (*Bao et al., 2019*).

Una vez que se obtienen las plantas transgénicas, los módulos para la expresión de la proteína Cas9 y de los sgARNs deben expresarse correctamente para producir la modificación deseada. Tanto en la transformación mediada por *Agrobacterium* como con la biolística, la inserción de las construcciones en el genoma de las plantas es al azar. La posición donde ocurra la inserción afectará directamente los niveles de expresión de Cas9 y de los sgARNs y, por lo tanto, su eficiencia. Los promotores y terminadores utilizados para la expresión de los componentes del sistema afectarán de igual manera y también lo hará el uso de codones que tenga el organismo, en particular en el caso de la expresión de la proteína. Todos estos aspectos son puntos donde hay espacio para la optimización (*Ma et al., 2016*).

Una vez generado el corte por la nucleasa Cas9, el DSB puede ser reparado mediante NHEJ o por HR. Los mecanismos de reparación ilegítima dan lugar a inserciones o deleciones que pueden ocasionar cambios en el marco de lectura de proteínas o modificaciones en secuencias promotoras, pudiendo producir la eliminación de una proteína o cambios en la expresión génica. La reparación por HR puede ocurrir cuando se incluye un molde con homología con la región dañada; esto puede ser utilizado para integrar exactamente la modificación que uno desea en el lugar que uno desea (*Kim & Kim, 2014*). En plantas, una vez que se produce la modificación esperada, la construcción puede ser eliminada del genoma por retrocruza con un individuo no transgénico o por autofecundación y segregación, seleccionando las plantas que contienen la modificación y carecen de la construcción (*Bao et al., 2019*).

El sistema de edición génica basado en CRISPR/Cas9 tiene muchas aplicaciones para estudios funcionales de genes en plantas ya que proporciona una herramienta robusta para la mejora genética de rasgos importantes en los cultivos, como el rendimiento, la arquitectura, el uso de nutrientes, la resistencia a enfermedades y la adaptación al estrés (*Bao et al., 2019*). El sistema simplifica el estudio individual de genes pertenecientes a familias multigénicas, la eliminación simultánea de genes parálogos o con elevada identidad de secuencia y la eliminación de manera independiente de genes ligados (*Bao et al., 2019*). En la Tabla 1 se resumen algunas características que han sido modificadas en soja.

Tabla 1. Edición génica en soja utilizando nucleasas programables.

Gen editado	Estrategia de transformación	Explante transformado	Regeneración de individuo editado	Referencia
<i>ZFN</i>				
<i>FAD2-1^a</i>	Biolística	Embriones somáticos	Sí	<i>Bonawitz et al., 2019</i>
<i>DCL4a/DCL4b</i>	<i>Agrobacterium</i>	Cotiledones	Sí	<i>Curtin et al., 2011</i>
<i>DCL1a/DCL1b</i>	<i>Agrobacterium</i>	Cotiledones	Sí	<i>Curtin et al., 2016</i>
<i>TALENs</i>				
<i>GmDcl2a+</i> <i>GmDcl2b</i>	<i>Agrobacterium</i>	Cotiledones	Sí	<i>Curtin et al., 2018</i>
<i>FAD2-1a+</i> <i>FAD2-1b</i>	<i>Agrobacterium</i>	Cotiledones	Sí	<i>Haun et al., 2014</i>
<i>CRISPR</i>				
<i>CPR5</i>	Biolística	Embriones somáticos	Sí	<i>Campbell et al., 2019</i>
<i>FAD2-2</i>	<i>Agrobacterium</i>	Cotiledones	Sí	<i>Amin et al., 2019</i>
<i>GmSPL9</i>	<i>Agrobacterium</i>	Cotiledones	Sí	<i>Bao et al., 2019</i>
<i>GmLHY</i>	<i>Agrobacterium</i>	Cotiledones	Sí	<i>Cheng et al., 2019</i>
<i>PDS11+PDS18</i>	<i>Agrobacterium</i>	Cotiledones	Sí	<i>Du et al., 2016</i>
<i>FT2a</i>	<i>Agrobacterium</i>	Cotiledones	Sí	<i>Cai et al., 2018</i>
<i>Glyma06g14180,</i> <i>Glyma08g02290,</i> <i>Glyma12g37050</i> <i>Glyma07g14530/</i> <i>Glyma01g38150 +</i> <i>Glyma11g07220/</i> <i>Glyma01g38150</i> <i>and</i> <i>Glyma11g07220</i>	PEG	Protoplastos	No	<i>Sun et al., 2015</i>
<i>Rfg1</i>	<i>Agrobacterium</i>	Hairy root	No	<i>Jacobs et al., 2015</i>
<i>Rfg1</i>	<i>Agrobacterium</i>	Hairy root	No	<i>Fan et al., 2017</i>

Notas: Dcl: Dicer-like. FAD: Fatty acid desaturase. SPL: Squamosa promoter binding protein-like. LHY: Late elongated hypocotyl. PDS: Phytoene desaturase FT: Flowering locus T. CPR: Constitutive expression of PR genes. Rfg1: Factor de nodulación.

Sistemas CRISPR-Cas9 libres de ADN

Recientemente han surgido metodologías de CRISPR/Cas9 libres de ADN, con origen en la edición de líneas celulares animales o embriones (*Metje-Sprink et al., 2019*). En estas metodologías, la introducción de los sgARNs y la proteína Cas9 en las células ocurre bajo la forma de ARN (mARN-Cas9 y sgARN) o mediante complejos ribonucleoproteicos Cas9-sgARNs pre-ensamblados (RNPs; *Woo et al., 2015*). Estas estrategias evitan la incorporación de ADN foráneo al genoma (*Metje-Sprink et al., 2019*) y por lo tanto, tienen la ventaja de que las plantas generadas no son transgénicas; en ninguna etapa del proceso de edición hay intermediarios transgénicos. Este sistema es especialmente útil para especies que se propagan vegetativamente o poseen ciclos de vida muy largos, donde el cruzamiento no es posible o consume mucho tiempo. Asimismo, las metodologías libres de ADN tiene la ventaja adicional

de que al estar Cas9 expresada en forma transitoria, la posibilidad de que se generen mutaciones en sitios no deseados (*off-target*), se reduce considerablemente (*Woo et al., 2015*). Otra ventaja adicional es que no es necesario realizar la optimización de las secuencias promotoras o codificantes según la especie con la que se trabaje (*Metje-Sprink et al., 2019*).

Por otra parte, la principal dificultad de las metodologías libres de ADN está en desarrollar estrategias eficientes para realizar la entrega de los RNPs o del mARN-Cas9 en las células de la planta y la posterior regeneración de las plantas a partir de las células editadas (*Metje-Sprink et al., 2019*).

A través de metodologías libres de ADN se han logrado editar diferentes especies vegetales, utilizado como estrategia de entrega la introducción en protoplastos mediada por polietilenglicol o la biolística sobre embriones somáticos, aunque no en todos los casos se han regenerado organismos editados (*Kim et al., 2017*). En la Tabla 2 se resumen los ejemplos de edición génica libres de ADN reportados hasta la fecha. La principal desventaja que tienen estas metodologías es que al no generar organismos transgénicos la eficiencia es menor en comparación con otros métodos clásicos. Además, como las plantas se regeneran en ausencia de selección, solo es posible identificar los organismos editados mediante secuenciación o con técnicas complejas y tediosas de genotipado (*Metje-Sprink et al., 2019*).

Tabla 2. Uso del sistema CRISPR/Cas libre de ADN.

Gen editado	Organismo	Cas9+sgARN	Estrategia de transformación	Explantado transformado	Regeneración de individuo editado	Referencia
<i>BRI1/AOC/PHY</i>	<i>A. thaliana</i>	Cas9 RNP	PEG	Protoplastos	Sí	<i>Woo et al., 2015</i>
<i>BIN2</i>	Lechuga (<i>L. sativa</i>)	Cas9 RNP	PEG	Protoplastos	Sí	<i>Woo et al., 2015</i>
<i>P450/DWD1</i>	Arroz (<i>O. sativa</i>)	Cas9 RNP	PEG	Protoplastos	Sí	<i>Woo et al., 2015</i>
<i>MLO7</i>	Uva (<i>V. vinifera</i>)	Cas9 RNP	PEG	Protoplastos	Sí	<i>Malnoy et al., 2016</i>
<i>DIPM1/DIPM2/DIPM4</i>	Manzana (<i>M. domestica</i>)	Cas9 RNP	PEG	Protoplastos	Sí	<i>Malnoy et al., 2016</i>
<i>NR</i>	Petunia (<i>P. hybrid</i>)	Cas9 RNP	PEG	Protoplastos	Sí	<i>Subburaj et al., 2016</i>
<i>LIG/ALS2/MS2/MS45</i>	Maíz (<i>Z. mays</i>)	Cas9 RNP	Biolística	Embriones inmaduros	Sí	<i>Svitashev et al., 2016</i>
<i>TaGW2</i>	Trigo (<i>T. aestivum</i>)	Cas9 RNP	Biolística	Callos	Sí	<i>Liang et al., 2016</i>
<i>TaGW2</i>	Trigo (<i>T. aestivum</i>)	Cas9 mARN + SgARN	Biolística	Callos	Sí	<i>Zhang et al., 2016</i>
<i>GT4</i>	Papa (<i>S. tuberosum</i>)	Cas9 RNP (crARN+tracrARN)	PEG	Protoplastos	Sí	<i>Anderson et al., 2018</i>
<i>ALS2</i>	<i>N. bethamiana</i>	TALEN mARN	PEG	Protoplastos	No	<i>Stoddart et al., 2016</i>
<i>FAD2</i>	Soja (<i>G. max</i>)	Cpf1 RNP	PEG	Protoplastos	No	<i>Kim et al., 2017</i>

Notas: BRI: Brassinosteroid insensitive. PHY: Phytochrome. AOC: Allene oxide cyclase. BIN: BR-Specific Hormone-Insensitive. DWD: DDB1 binding WD40 motif. DPIM1/2: DspA/E-interacting proteins of *Malus domestica*. GW: Grain width and weight. NR: Nitrate reductase. LIG: Liguleless. ALS: Acetolactate synthase. MLO: Mildew Locus O. FAD-2: Delta(12)-fatty-acid desaturase.

Remplazo alélico y *gene-targeting* en plantas

El *gene-targeting* GT consiste en introducir fragmentos de ADN de interés y promover la recombinación homóloga para modificar secuencias endógenas. Esta técnica puede utilizarse para producir KOs de genes o exones específicos, generar inserciones y/o realizar remplazos de genes o alelos, conteniendo mutaciones puntuales (Voytas, 2013). Una vez que el fragmento de ADN que se quiere incorporar fue introducido dentro de la célula, se da el intercambio de información genética entre el ADN molde y el locus nativo, utilizando las vías de reparación del ADN basadas en HR. De esa manera, es posible alterar específicamente un locus a través de la utilización de un ADN molde que incluya las modificaciones genéticas buscadas (Voytas, 2013). Si bien existen algunas líneas celulares eucariotas en las que la HR ocurre eficientemente, en la mayoría de las células somáticas el ADN se integra en sitios aleatorios en el genoma a través de la vía NHEJ (Voytas, 2013).

En este sentido, muchas de las investigaciones se han abocado a la búsqueda de estrategias para aumentar la frecuencia de HR. Al generar DSBs en el sitio donde se desea que ocurra la incorporación de ADN, es posible aumentar considerablemente la frecuencia de HR, por lo que esta solución ha sido la más utilizada. También se ha visto que el silenciamiento de

moléculas de la vía NHEJ, como KU70, KU80 o la ligasa IV y la inhibición química de la ligasa IV producen cierta mejora (Voytas, 2013).

Como se mencionó anteriormente, existen distintas herramientas que pueden ser utilizadas para provocar DSBs en el genoma. Las meganucleasas han resultado poco versátiles para reconocer secuencias deseadas debido a que es difícil manipular los sitios de unión de las nucleasas sin que esto resulte en la pérdida de su actividad (Puchta & Fauser, 2013). El uso de ZFN para GT ha sido más exitoso. Se ha logrado realizar GT en tabaco, maíz, *Arabidopsis thaliana*, entre otros organismos (Puchta & Fauser, 2013). Los TALENs, también han sido empleados para realizar GT en arroz, tabaco, *Arabidopsis*, entre otros (Li et al., 2012; Zhang et al., 2013; Shan et al., 2018), en algunos casos con frecuencias muy altas (14 % en protoplastos y 4 % en callos de tabaco regenerados sin selección; Zhang et al., 2013).

El sistema CRISPR/Cas ofrece una simplicidad y multiplexibilidad para generar DSBs incomparables. Este sistema ha sido el más utilizado para realizar GT en plantas (Sanagala et al., 2017). En la Tabla 3 se resumen ejemplos de GT y se indican las particularidades experimentales.

Además de la baja tasa de HR, otro factor que limita la eficiencia de GT es la entrega en abundancia de moldes con homología para la reparación. Muchos ejemplos disponibles realizan la entrega del molde para la reparación del ADN utilizando vectores virales replicativos ectópicos de alto número de copias, basados en geminivirus (Baltes, 2014; Gil-Humanes, 2017). La desventaja de esta aproximación es que al encontrarse en grandes cantidades, fragmentos de ADN plasmídico pueden también ser incorporados dentro del genoma de la planta (Sanagala et al., 2017). También se han desarrollado estrategias de entrega del ADN molde mediante el uso de vectores binarios para *Agrobacterium*. Una estrategia consiste en llevar adelante un doble evento de transformación (Endo, 2016), generando en primer lugar, un explante transgénico conteniendo la construcción para la expresión de Cas9 y los sgARN, seguida de una segunda instancia de transformación donde se introduce el molde reparador (Endo, 2016). Otra estrategia consiste en transformar la planta con una construcción que contenga los módulos de expresión para Cas9 y para los sgARN y además un molde con homología para la reparación. Este molde debe estar flanqueado por las secuencias de reconocimiento de la proteína Cas9, de manera que se libere un fragmento de ADN lineal tras el corte por la proteína (Fauser, 2012; Shan et al., 2018). Estas estrategias presentan el inconveniente de que cabe la posibilidad de que ocurra la integración de ADN foráneo en el genoma (Sanagala et al., 2017).

En mamíferos las estrategias para la entrega del ADN molde han sido más exitosas que en plantas (Li et al., 2016). Existen ejemplos donde el molde se entrega como plásmidos (Aida

et al., 2015), dsADN (*double stranded ADN*, Shy et al., 2016), lsADN (*long single stranded ADN*; Quadros et al., 2017) o ssADN (*short single stranded ADN*; Igoucheva et al., 2001). La biotecnología vegetal tiene mucho para explorar en estos aspectos.

Tabla 3. Gene targeting y remplazo alélico en plantas.

Gen editado	Organismo	Cas9+SgARN	ADN homólogo	Explante transformado	Referencia
ALS	Arroz (<i>O. sativa</i>)	Vector vía <i>Agrobacterium</i>	Vector vía <i>Agrobacterium</i> (transformación secuencial)	Callos embriogénicos	Endo et al., 2016
EPSPS	Arroz (<i>O. sativa</i>)	Vector vía biolística	Plásmido circular vía biolística (transformación simultánea)	Callos embriogénicos	Li et al., 2016
ACT1/GST	Arroz (<i>O. sativa</i>)	Vector vía <i>Agrobacterium</i>	Geminivirus vía <i>Agrobacterium</i> (transformación simultánea)	Callos embriogénicos	Wang et al., 2017
ALS2	Maíz (<i>Z. mais</i>)	Vector vía biolística	dsADN o ssADN oligo vía biolística (transformación simultánea)	Embriones inmaduros	Svitashev et al., 2015
ANT1	Tomate (<i>S. lycopersicum</i>)	Vector vía <i>Agrobacterium</i>	Geminivirus vía <i>Agrobacterium</i> (transformación simultánea)	Cotiledones	Čermák et al., 2015
GFP (promotor)	Trigo (<i>T. aestivum</i>)	Vector vía PEG	Geminivirus vía PEG (transformación simultánea)	Protoplastos	Gil-Humanes et al., 2017
GFP (promotor)	Trigo (<i>T. aestivum</i>)	Vector vía biolística	Geminivirus vía biolística (transformación simultánea)	Escutelo	Gil-Humanes et al., 2017
At1g53430	<i>A. thaliana</i>	Vector vía <i>Agrobacterium</i>	Mismo vector	Flores	Shan et al., 2018
DD20/DD43/ALS1	Soja (<i>G. max</i>)	Vector vía biolística	dsADN vía biolística (transformación simultánea)	Embriones somáticos	Li et al., 2015
ARGOS8 (promotor)	Maíz (<i>Z. mais</i>)	Vector vía biolística	Plásmido circular (transformación simultánea)	Embriones inmaduros	Shi et al., 2018
EPSPS	Casava (<i>M. esculenta</i>)	Vector vía <i>Agrobacterium</i>	Geminivirus vía <i>Agrobacterium</i> (simultánea)	Callos embriogénicos	Hummel et al., 2018

Notas: ANT: Anthocyanin mutant. ALS: Acetolactate synthase. ACT: Actin. GST: Glutathion-S-transferase. EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase. GFP: green fluorescent protein. ARGOS8: negative regulator of ethylene responses. At1g53430: Leucine-rich repeat transmembrane protein kinase. DD20/DD43: two genomic sites in chromosome 4 of *Glycine max*.

Propuesta FMV: Optimización de técnicas de edición génica utilizando un modelo de resistencia a herbicida

Este trabajo de tesis se enmarca en un proyecto financiado por el Fondo María Viñas de la Agencia Nacional de Investigación e Innovación. Su objetivo es optimizar metodologías de remplazo alélico utilizando complejos ribonucleoproteicos pre-ensamblados o utilizando ARN codificante para la proteína Cas9 y, a su vez, desarrollar técnicas de regeneración novedosas de plantas de soja editadas. La estrategia experimental plantea utilizar como blanco de edición, un gen codificante para la 5-enolpiruvato-shikimate-3-fosfato (EPSP) sintasa (EPSPS), con el propósito de conferir resistencia al herbicida glifosato. Esta estrategia simplifica la detección de las plantas editadas ya que confiere un marcador fenotípico fácil de detectar. Tiene la ventaja adicional de que los mecanismos de resistencia al glifosato están bien caracterizados (*Sammons & Gaines, 2014*), lo que sustenta en buena medida la estrategia planteada. Con el fin de brindar mayor comprensión sobre la estrategia experimental, en la siguiente sección se comentan los mecanismos de resistencia al herbicida glifosato.

Mecanismos de resistencia al herbicida glifosato

El herbicida glifosato es un derivado de glicina modificado con N-fosfonometilo que es capaz de formar un complejo inactivo estable con la enzima EPSPS (*Sammons & Gaines, 2014*). La EPSPS participa en la biosíntesis de los aminoácidos aromáticos como la fenilalanina, la tirosina y el triptófano a través de la vía del Shikimate en bacterias, hongos y plantas (*Pollegioni et al., 2011*). En la vía del shikimate, ubicada a nivel cloroplástico, la EPSPS convierte el fosfoenolpiruvato (PEP) y shikimate-3-fosfato (S3P) en 5-enolpyruvylshikimate-3-fosfato (EPSP) y fosfato inorgánico. El herbicida actúa como análogo de estado de transición competitivo con el fosfoenolpiruvato (PEP), al formar un complejo ternario estable no covalente con el complejo enzima-S3P-glifosato. El glifosato causa la muerte de las plantas al privarlas de aminoácidos aromáticos (*Firdous et al., 2018*).

La primera estructura cristalográfica de EPSPS que se determinó fue la de la proteína de *E. coli* en estado libre por un grupo de investigación de Monsanto en 1991. La estructura mostró tener un plegamiento proteico único (barril α/β invertido) con dos dominios globulares, cada uno compuesto por tres unidades de plegamiento idénticas, conectadas entre sí por una región bisagra. Estudios estructurales posteriores en complejo con S3P y glifosato demostraron que los dos dominios globulares de la enzima se acercan entre sí tras la unión del S3P, sufriendo un cambio conformacional abierto-cerrado. Estos estudios también demostraron que el PEP y el glifosato se unen al mismo sitio de la enzima a través de interacciones similares (*Pollegioni et al., 2011*).

Existen diversos mecanismos para conferir resistencia al glifosato, entre ellos: *target-site resistance* (cambios aminoacídicos en la enzima blanco que afectan la interacción con el herbicida), metabólicos (modificaciones químicas del herbicida, por conjugación o degradación) y de exclusión (exclusión física, a través de reforzamiento de la cutícula, o exclusión fisiológica, fundamentalmente secuestro en la vacuola mediado por transportadores) (Sammons & Gaines, 2014). El aumento del número de copias del gen *EPSPS* es otro mecanismo que confiere resistencia al glifosato (Sammons & Gaines, 2014).

En este trabajo nos interesa en particular usar los mecanismos de *target-site resistance*. Las mutaciones más frecuentes encontradas naturalmente involucran a la Pro106 que en general es sustituida por los aminoácidos Ala, Leu, Ser o Thr. Estas mutaciones se han encontrado naturalmente en las especies *Lolium multiflorum*, *Eleusine indica* y *Amaranthus tuberculatus* (Sammons & Gaines, 2014). La Pro106 no está directamente involucrada en interacciones moleculares con el glifosato o el sustrato PEP, pero el cambio aminoacídico provoca un cambio estructural en el sitio activo y desplaza otros aminoácidos hacia el inhibidor, reduciendo el espacio disponible (Sammons & Gaines, 2014). Asimismo, la mutación Pro106-Ser produce un aumento en la constante de inhibición (Sammons & Gaines, 2014).

La generación de mutaciones simultáneas confiere niveles de resistencia mayor y a la vez permiten que la proteína mantenga una afinidad suficiente por el sustrato PEP, aventajando la mutación simple Pro106-Ser. Entre ellas se encuentra la doble mutación Thr102Ile + Pro106Ser (TIPS) en el sitio de unión al PEP; esta variante es utilizada en cultivos transgénicos de maíz (Roundup Ready) y se ha reportado su aparición naturalmente en poblaciones de *E. indica*. De igual forma se encontró que la mutación Thr102Ile + Pro106-Ala (TIPA) confiere resistencia al herbicida (Healy-Fried et al., 2007).

Teniendo en cuenta el conocimiento disponible, el objetivo general del proyecto es el de generar una metodología eficiente para realizar GT en soja, utilizando como marcador fenotípico del éxito de la técnica la sustitución de uno de los genes *EPSPS* (*EPSPS1*) de soja por sus variantes TIPS o TIPA. En este trabajo se describe la generación y caracterización de las variantes TIPS y TIPA de la *EPSPS* de soja (de manera de determinar cuál de ellas incorporar) y la generación de otras herramientas moleculares que serán utilizadas en etapas posteriores para desarrollar un protocolo de remplazo alélico para este cultivo.

Objetivos

El objetivo general de este trabajo es optimizar metodologías de edición génica en soja (*Glycine max*) con el fin de generar plantas no transgénicas con modificaciones sitio-específicas utilizando el sistema CRISPR/Cas9. En particular, el trabajo propone optimizar metodologías de remplazo alélico. De manera de poner a punto estas metodologías sobre un modelo que proporcione un fenotipo fácil de identificar, se propone utilizar como secuencia genómica blanco un gen que codifica para una 5-enolpiruvatoshikimato-3-fosfato sintasa (EPSPS), enzima involucrada en la resistencia al herbicida glifosato.

El trabajo parte de la hipótesis de que si se reemplaza un gen de la EPSPS nativo por una versión que contenga mutaciones que confieran insensibilidad por el herbicida, las plantas de soja que se regeneren exhibirán un fenotipo de resistencia al mismo. De cumplirse esta premisa, se estará generando un fenotipo fácilmente seleccionable, que podrá ser utilizado para optimizar las metodologías de remplazo alélico para el cultivo de soja. La EPSPS es una enzima que está muy conservada a nivel de su secuencia aminoacídica y existen variantes mutantes de genes *EPSPS* que están bien caracterizadas en diferentes organismos (*Lolium multiforum*, *Eleusine indica* y *Amaranthus tuberculatus*) y que son insensibles a glifosato. Teniendo en cuenta estos aspectos, este trabajo tiene los siguientes objetivos específicos:

- 1- Introducir las mutaciones TIPS (Pro106-Ser + Thr102-Ile) o TIPA (Pro106-Ser + Thr102-Ala) generadas en la *EPSPS1* de soja en *Arabidopsis thaliana*.
- 2- Diseñar y validar ARNs guía que puedan ser utilizados para editar el genoma de soja.
- 3- Generar un vector de remplazo que contenga las mutaciones TIPS o TIPA para su uso durante los ensayos de remplazo alélico.

Materiales y métodos

1. Extracción de ADN genómico

En un tubo de tipo eppendorf, se colocaron de 1 a 3 discos de hoja de soja (*Glycine max*, cultivar Jack), junto con cinco partículas de zirconio de (2 mm). Luego de sumergir las muestras en nitrógeno líquido (N₂), se procedió rápidamente a moler el material vegetal utilizando un homogeneizador de tejidos de tipo beadbeater (Biospec) durante 1 minuto. A cada muestra se le agregaron 700 µL del buffer 2X CTAB (CTAB 2 % w/v, NaCl 1.4 M, EDTA 20 mM, Tris-HCl 100 mM pH 7.5, β-mercaptoetanol 1 %). Se incubó durante 30 minutos a 65 °C mezclando suavemente cada 10 minutos. Se agregaron 600 µL de cloroformo:alcohol isoamílico 24:1. Se centrifugó por 10 minutos a 8000 x g a temperatura ambiente (RT: *room temperature*). Se tomó la fase superior acuosa y se transfirió a un tubo de microcentrífuga nuevo. Se agregaron 5 µL de ARNsa (20 mg/mL) y se incubó 10 minutos a RT. Se agregaron nuevamente 600 µL de cloroformo:alcohol isoamílico 24:1 y se centrifugó por 10 minutos a 8000 x g a RT. Se transfirió la fase acuosa a otro tubo, al cual se le agregó volumen igual de isopropanol. Se dejó a -20 °C de 1 hora y 30 minutos a toda la noche (ON: *over night*). Se centrifugó a 16000 x g durante 20 minutos. Se lavó el *pellet* con EtOH 70 %, se dejó secar y se resuspendió en H₂O-miliQ.

2. Extracción de ARN y síntesis de ADNc

Para la extracción de ARN total, se utilizó el kit *PureLink™ Plant ARN Reagent* de Invitrogen (Cat. 12322-012).

Se pulverizó 100 mg de material vegetal (hojas de *Glycine max* de la variedad Jack) con N₂ líquido y se añadieron 500 µL de *PureLink Plant ARN Reagent* frío. Se agregaron 100 µL de NaCl 5 M y 300 µL de cloroformo. Se centrifugó la muestra a 12000 x g por 10 minutos a 4 °C para separar las fases. Se transfirió la fase superior (acuosa) a un nuevo tubo y se añadió 1 volumen de isopropanol. Se incubó el tubo a RT por 10 minutos y se centrifugó a 12000 x g por 30 minutos a 4 °C. Se descartó el sobrenadante y el *pellet* se lavó con EtOH 75 %, se secó y se resuspendió en H₂O-MiliQ. El ARN se almacenó a -80 °C hasta su uso. El H₂O-MiliQ y la solución de NaCl 5M fueron tratadas previamente con dietilpirocarbonato (DEPC) a una concentración final de 0.1 % v/v ON y autoclavadas durante 40 minutos antes de su uso.

La calidad del ARN se evaluó mediante electroforesis desnaturalizante en gel de agarosa. Para ello, se resuspendió aproximadamente 1 µg de ARN total en una mezcla de formaldehído 2.2 M, formamida 50 % v/v, 1X buffer MOPS (ácido 3-(n-morfolino)-propanosulfónico 0.2 M pH 7, acetato de sodio [NaAc] 0.05 M y EDTA 0.01 M) con bromuro de

etidio (BrET) 70 ng/μL. Las muestras se desnaturalizaron a 55 °C durante 10 minutos y se separaron en geles de agarosa desnaturalizante conteniendo formaldehído 2.2 M y 1X MOPS como buffer para la electroforesis. El gel se corrió a 8 V/cm durante 30 minutos.

Previamente a la síntesis de ADNc, aproximadamente 2 μg de ARN fueron tratados con la enzima ADNsa I (Thermo, Cat. EN0521). Para ello se mezcló el ARN con 2 U de enzima en buffer de reacción con MgCl₂ 1X en un volumen de 10 μL. Se incubó la muestra a 37 °C por 30 minutos, se añadió 1 μL de EDTA 50 mM y se incubó a 65 °C por 10 minutos. La síntesis del ADNc se llevó a cabo utilizando el kit *RevertAid Reverse Transcriptase* (Thermo, Cat. K1691). Dos μg de ARN total se mezclaron con 0.5 μL del cebador oligodT (100 μM), 2 μL de dNTPs (10 mM), 4 μL de Buffer 5X, 0.5 μL de inhibidor de ARNsa y 1 μL de transcriptasa reversa en un volumen final de 20 μL. Se incubó por 60 minutos a 42 °C y luego 10 minutos a 70 °C para inactivar la enzima.

3. Clonado de CDS *EPSPS1* en vector de entrada pENTR2B

La secuencia codificante del gen *EPSPS1* (Glyma.01G139600.1) fue amplificada a partir de ADNc de la variedad Jack de soja utilizando la polimerasa Phusion HF (Thermo, Cat. F530L). La reacción se preparó de acuerdo a las especificaciones sugeridas por el proveedor en un volumen de 20 μL: buffer Phusion HF 1X, 200 μM dNTPs, 0.5 μM de cada uno de los cebadores (*primers*) directo (forward, Fw) 268 y reverso (reverse, Rv) 269 (Tabla 4), 0.02 U de polimerasa y 1 μL de ADNc como molde.

Se añadieron sitios de restricción para las enzimas *Sall* y *NotI* llevando a cabo una segunda amplificación por PCR, utilizando como molde el producto de la reacción anterior y los *primers* 270 y 271, en las mismas condiciones de reacción.

En ambos casos el programa de ciclado utilizado fue: desnaturalización inicial: 98 °C, 3 minutos; 35 ciclos de: desnaturalización: 98 °C, 10 segundos, hibridación: 59 °C, 30 segundos, extensión: 72 °C, 1 minuto 20 segundos; extensión final: 72 °C 10 minutos.

Tabla 4. *Primers* utilizados para generar las construcciones de entrada

Nº de <i>Primer</i>	Nombre	Secuencia
268	FwGmEPSPS1	5'-CAGACCAAAGTGCAGAGAACAAC-3'
269	RvGmEPSPS1	5'-CCACATATGAGTAGAGCCAAGTCG-3'
270	FwGmEPSPS1- <i>Sall</i>	5'-TTTGTCTGACATGGCCCAAGTGAGCAGAG-3'
271	ReGmEPSPS1- <i>NotI</i>	5'-TGCGGCCGCTTAGTGCTTTGTAAACCTC-3'
272	FwTIPS	5'- TTGGAAATGCTGGTAtTGCAATGCGTtCTTTGACAGCAGCTG-3'
273	RvTIPS	5'- CAGCTGCTGTCAAAGaACGCATTGCAaTACCAGCATTCCAA-3'
274	FwTIPA	5'- TTGGAAATGCTGGTAtTGCAATGCGTgCTTTGACAGCAGCTG-3'

275	RvTIPA	5'- CAGCTGCTGTCAAAGcACGCATTGCAaTACCAGCATTTC CAA -3'
296	FwTIPsb	5'-CTGGTAtTGCAATGCGTtCTTTGACAGCAGCTG-3'
297	RvTIPsb	5'-GTCAAAGaACGCATTGCAaTACCAGCATTTC CC -3'
298	FwTIPAb	5'-CTGGTAtTGCAATGCGTgCTTTGACAGCAGCTG-3'
299	RvTIPAb	5'-GTCAAAGcACGCATTGCAaTACCAGCATTTC CC -3'

*La secuencia subrayada indica el sitio de restricción introducido.

El producto de PCR obtenido fue purificado utilizando columnas de purificación comerciales (ver punto 21).

Tanto el vector pENTR2B como el producto de PCR fueron digeridos con las enzimas *Sall* y *NotI*. Para ello, 4 µg de plásmido o del molde de PCR se incubaron con 20 U de *Sall* y 10 U de *NotI* en buffer NEB 3.1 por 2 horas a 37 °C. El inserto digerido fue purificado utilizando columnas de purificación comerciales (ver punto 21). Por otra parte, el vector digerido fue corrido en un gel de agarosa 1 % (ver punto 23) y la banda correspondiente al vector carente del gen letal *ccdB* fue escindida del gel y purificada por kit de extracción comercial (ver punto 21). El vector digerido fue desfosforilado utilizando la enzima FastAP (Thermo, Cat. EF0651). Para ello, 1 µg de vector fue incubado con 1 U de fosfatasa alcalina termosensible en buffer 1X durante 10 minutos a 37 °C. La enzima fue inactivada a 75 °C por 5 minutos.

El vector y el inserto fueron ligados entre sí usando la enzima T4 ADN ligasa (Thermo, Cat. EL0019). Para ello, se mezclaron 25 ng de pENTR2B con 6 ng de inserto digerido (relación 5-1) y 1 U de ligasa T4 en buffer T4 ligase 1X. La mezcla de reacción se incubó a RT ON. La ligasa fue inactivada incubando a 65 °C durante 10 minutos.

Se transformaron células quimiocompetentes TOP 10 (ver punto 7) y se plaquearon en medio LB-agar con agregado de kanamicina a una concentración final de 50 µg/mL. A partir de las colonias obtenidas, se realizaron minipreparaciones de plásmidos (ver punto 19), las cuales fueron analizadas mediante restricción con *NcoI* para verificar la presencia del inserto. Para ello, se incubaron 6 µL de plásmido (aproximadamente 500 ng de plásmido) con 2 U de enzima en buffer Tango 1X a 37 °C durante 1 hora. Las digestiones fueron visualizadas mediante una electroforesis en gel de agarosa 1 %. Los clones pENTR2B:EPSPS1_{CDS} en los que se observó el patrón de restricción correcto fueron confirmados por secuenciación (Institut Pasteur Montevideo).

4. Generación de mutantes TIPS y TIPA sobre el vector de entrada

Los mutantes puntuales fueron generados mediante mutagénesis QuickChange (*Liu & Naismith, 2008*) utilizando como molde la construcción pENTR2B:EPSPS1_{CDS}. Se utilizó la polimerasa Phusion HF de Thermo en las condiciones sugeridas por el proveedor. Para realizar

la mutagénesis dirigida de TIPS y TIPA se emplearon las siguientes estrategias: 1) utilización de *primers* perfectamente complementarios (272 y 273 para TIPS o 274 y 275 para TIPA); y 2) utilización de *primers* parcialmente complementarios (*primers* 296 y 297 para TIPS o 298 y 299 para TIPA). Los *primers* utilizados se resumen en la Tabla 4.

El programa utilizado para la PCR fue: desnaturalización inicial: 98 °C, 30 segundos; 35 ciclos de: desnaturalización: 98 °C, 10 segundos, hibridación/extensión: 72 °C, 2 minutos; extensión final: 72 °C, 10 minutos. Se utilizó la enzima de restricción *DpnI* (10 U) para digerir el ADN metilado ON (de las hebras parentales) y se inactivó la enzima a 80 °C por 20 minutos. Se verificó la PCR por electroforesis en gel de agarosa 1 % (ver punto 23), se transformaron células quimiocompetentes (ver punto 7) y se plaquearon en medio LB-agar con agregado de kanamicina a una concentración final de 50 µg/mL.

Se realizó una minipreparación casera (ver punto 19) y el ADN plasmídico se separó en un gel de agarosa 1 % para verificar que el tamaño fuera el correcto. Las construcciones finales pENTR2B:EPSPS1_{CDS} conteniendo las mutaciones TIPS y TIPA fueron confirmadas por secuenciación.

5. Subclonado en vector binario pUB-Dest

Las variantes mutantes TIPS y TIPA así como la versión wild type (WT) de la EPSPS1 fueron subclonados en un vector binario para transformación vegetal, pUB-Dest (*Grefen et al., 2010*) mediante recombinación con la clonasa LR, utilizando el sistema *Gateway* (Invitrogen, Cat 11791020). El vector pUB-Dest permite la expresión constitutiva de los genes de interés bajo el control de un promotor de poliubiquitina de *Arabidopsis*.

Para generar la construcción del vector destino, se realizó una recombinación de los vectores de entrada conteniendo las variantes salvaje (WT: *wild type*), TIPS o TIPA con el vector binario pUB-Dest. Para ello, se incubaron ON 150 ng de vector pUB-Dest y 100 ng de los vectores pENTR2B:EPSPS1_{CDS}-WT, TIPS o TIPA linealizados con la enzima *BsrBI*, con la LR Clonase 1X en un volumen final de 10 µL. Para inactivar la enzima, la mezcla de reacción fue incubada con 1 µL de proteinasa K (2 µg/µL) a 37 °C durante 10 minutos.

Se transformaron células quimiocompetentes (ver punto 7) y se seleccionaron colonias resistentes al medio selectivo (espectinomicina 50 µg/µL) para realizar minipreparaciones de plásmidos. Se analizó la incorporación de la EPSPS1 mediante digestión con la enzima *EcoRI*. Las construcciones finales pUB-Dest:EPSPS1_{CDS}-WT, TIPS o TIPA fueron confirmadas mediante secuenciación.

6. Generación de células quimiocompetentes TOP 10 de *E. coli*

Se inocularon 250 mL de medio SOB (extracto de levadura 5 g/L, triptona 20 g/L, NaCl 10 mM, KCl 2.5 mM, MgSO₄ 20 mM) con 100 µL de un precultivo de células TOP 10 crecidas ON a 20 °C. El cultivo se incubó a 20 °C con agitación hasta que alcanzó una densidad óptica a 600 nm (OD_{600nm}) de 0.3 (aproximadamente 16 horas). Se centrifugó a 3000 x g y a 4 °C durante 10 minutos en un tubo de centrífuga de fondo plano. Se resuspendieron las células en 80 mL de buffer CCM80 frío (acetato de potasio 10 mM pH 6.4, CaCl₂ 80 mM, MnCl₂ 20 mM, MgCl₂ 10 mM, glicerol 10 % v/v). Se incubaron las células en hielo por 20 minutos. Se centrifugó nuevamente en iguales condiciones y se resuspendió en 2 mL de CCM80 frío. Se incubó en hielo por 20 minutos adicionales. Se alicuotaron las células en tubos de microcentrífuga previamente enfriados, se congelaron en N₂ líquido y se almacenaron a -80 °C. Para el cálculo de la eficiencia de transformación, se transformó 10 pg del vector pUC (10 pg/µL). Las células se plaquearon en LB-agar con ampicilina a una concentración final de 100 µg/mL. Al día siguiente, se calculó la eficiencia (número de colonias x factor de dilución x 10⁵/µg de ADN).

7. Transformación de células quimiocompetentes de *E. coli*

En todos los casos se utilizó la cepa TOP 10 de *E. coli*. Se colocó una alícuota de 50 µL de células quimiocompetentes en hielo para descongelar (5 minutos) y se agregó hasta un 10 % del volumen del producto a transformar (5 µL). Se incubó en hielo por 30 minutos y se realizó un *heatshock* por 45 segundos a 42 °C. Se incubó nuevamente en hielo por dos minutos, se agregó el medio de elección (LB, SOC, TB) y se recuperaron las células a 37 °C por 1 hora en agitación. Se plaquearon las células en LB-agar con agregado de antibiótico acorde a la resistencia conferida por el plásmido transformado.

8. Generación de células electrocompetentes de *Agrobacterium tumefaciens* C58C1

Se inocularon 500 mL de medio LB con 5 mL de un cultivo fresco de *A. tumefaciens* de la cepa C58C1 y el cultivo se dejó crecer a 28 °C con agitación ON. Se mantuvieron las células en hielo por 20 minutos, se transfirieron a un tubo de centrífuga de 500 mL (previamente enfriado) y se centrifugó por 15 minutos a 4000 x g, a 4 °C. Se descartó el sobrenadante y las células se resuspendieron en 500 mL de glicerol 10 % (glicerol 10 % v/v en H₂O-MiliQ) estéril frío. Se centrifugó nuevamente por 15 minutos a 4000 x g, a 4 °C y se descartó el sobrenadante. Se resuspendió el *pellet* en 250 mL de glicerol 10 % frío y se repitió la centrifugación. Se descartó el sobrenadante y se repitió este paso dos veces. Se alicuotaron las

células en tubos de microcentrífuga previamente enfriados, se congelaron en N₂ líquido y se almacenaron a -80 °C.

9. Transformación de células electrocompetentes de *A. tumefaciens*

En todos los casos se utilizó la cepa C58C1 de *A. tumefaciens*. Se esterilizaron celdas de electroporación de 2 mm con EtOH 70 % y se lavaron con H₂O-MiliQ estéril. Se colocó una alícuota de 50 µL de células electrocompetentes en hielo para descongelar (cinco minutos) y se agregó un volumen no mayor a 5 % del volumen de la alícuota a transformar. Se colocó las células en las cubas estériles. Se utilizó un electroporador Gene Pulser XCell (Biorad) y un programa predefinido para *A. tumefaciens* (voltaje: 2.5 kV, campo: 12.5 kV/cm, capacitancia: 25 µF, resistor: 200 Ω, tiempo constante: 3.8 ms). Se recuperaron las células en medio LB a 28 °C por 2 horas. Se plaquearon las células en LB-agar con agregado de antibiótico para la selección de la cepa C58C1 (rifampicina 100 µg/mL) y para la construcción (espectinomicina 100 µg/mL).

10. Chequeo de transformantes de *A. tumefaciens* con el vector binario

Las bacterias transformadas con las construcciones pUB-Dest:EPSPS1_{CDS}-WT, TIPS o TIPA fueron chequeadas por PCR de colonia. Para cada construcción, tres colonias aisladas fueron picadas y resuspendidas en 20 µL de H₂O-MiliQ estéril. Los tubos se incubaron a 95 °C por 5 minutos.

Para verificar la presencia del vector en las bacterias se llevó a cabo una PCR utilizando el protocolo de DFS-Taq ADN polimerasa (Bioron). La reacción se preparó de acuerdo a las especificaciones sugeridas por el proveedor en un volumen de 20 µL utilizando buffer *yellow* 1X, dNTPs 200 µM, *primers* Fw (305 o 306 o 307) y RvUbiseq (Tabla 5) cada uno a 0.5 µM, 1.25 U de enzima Taq polimerasa y 1 µL del lisado bacteriano. Se realizaron controles negativos (en iguales condiciones sin molde) y positivos (mismas condiciones pero utilizando como molde 5 ng de ADN plasmídico de las construcciones pUB-Dest:EPSPS1_{CDS}-WT, TIPS o TIPA). El programa de ciclado utilizado fue el siguiente: desnaturalización inicial: 94 °C, 2 minutos; 30 ciclos de: desnaturalización: 94 °C, 10 segundos, hibridación: 50 °C, 20 segundos, extensión: 72 °C, 1 minutos; extensión final: 72 °C, 10 minutos.

Tabla 5. *Primers* utilizados para chequear transformantes de *Agrobacterium* por PCR colonia

N° de Primer	Nombre	Secuencia
305	FwEPSPS1wtcheck	5' – CTTGGAAATGCTGGTACTGCAATGCGTC – 3'
306	FwEPSPS1TIPSccheck	5' – CTTGGAAATGCTGGTATTGCAATGCGTT – 3'
307	FwEPSPSTIPAccheck	5' – CTTGGAAATGCTGGTATTGCAATGCGTG – 3'
-	RvUbiseq	5' –AGAGATAGATTTGTAGAGAG– 3'

Los productos de PCR fueron separados en un gel de agarosa 1 %. Se hicieron precultivos líquidos de aquellas colonias positivas en medio LB con agregado de rifampicina (100 µg/mL) y espectinomicina (100 µg/mL). Se agregó 24 % v/v de glicerol para almacenar las bacterias a -80 °C.

11. Material Vegetal

Para la obtención de ADN genómico y ARN de soja se utilizó la variedad Jack de *Glycine max* (L.) Merrill en estadio V1. Las plantas de soja se cultivaron en cámara de crecimiento bajo lámparas de halogenuro metálico y lámparas de vapor de sodio a alta presión con una intensidad lumínica de 800 µmol fotones m⁻² seg⁻¹ y un fotoperiodo 16 horas luz/8 horas oscuridad, a 25 °C y una humedad relativa de 60-40 %. Para su crecimiento se utilizaron macetas de 9 cm de diámetro por 8 cm de altura. Se utilizó una mezcla de partes iguales de turba y perlita como sustrato. Las plantas fueron fertilizadas con medio hakaphos rojo (Compo Expert) en una concentración de 2 g/L una vez a la semana.

Para la obtención de líneas sobreexpresantes de la EPSPS1 de soja y de sus variantes TIPS y TIPA se usó *Arabidopsis thaliana* ecotipo Col-0. Las plantas fueron germinadas *in vitro* luego transferidas a cámara de crecimiento. Se esterilizaron las semillas sumergiéndolas en EtOH 70 % por un máximo de 5 minutos, seguido de una solución de hipoclorito de sodio 25 g/L y Tritón X-100 0.1 % durante 15 minutos en rotación. Pasando este tiempo, se realizaron al menos 4 lavados con H₂O-MiliQ estéril. Las semillas se sembraron sobre placas de medio MS (*Murashige y Skoog, 1962*) con la mitad de concentración (½ MS: 2.1 g/L medio MS, pH 5.7, 0.5 % agar) y se dejaron a 4 °C por 3-5 días para su estratificación. Posteriormente, las placas se transfirieron a cámara de crecimiento *in vitro* bajo lámparas de LED con una intensidad lumínica de 100 µmol fotones m⁻² seg⁻¹, un fotoperiodo de 16 horas luz/8 horas oscuridad a 22 °C entre una y dos semanas. Las plantas germinadas fueron trasplantadas a turba y se cultivaron en cámara de crecimiento con 150 µmol fotones m⁻² seg⁻¹ de intensidad lumínica y un fotoperiodo de 16 horas luz/8 horas oscuridad a 25 °C. Como solución de riego se utilizó agua destilada, suplementando con medio ½ MS una vez a la semana.

12. Transformación de *Arabidopsis thaliana*

Para cada construcción se transformaron al menos 5 macetas con 4 plantas de *A. thaliana* en cada una. Se crecieron las plantas hasta alcanzar la etapa de floración. Se cortaron las inflorescencias primarias para favorecer el surgimiento de nuevas inflorescencias. Cuando las plantas desarrollaron un buen número de botones florales, se realizó la transformación por inmersión floral (*Clough y Bent, 1998*). Para ello, se preparó un cultivo líquido de las cepas

Agrobacterium transformadas con cada construcción en diferentes matraces conteniendo 500 mL de medio YEP con agregado de 100 µg/mL rifampicina y 100 µg/mL espectinomicina y se incubó a 28 °C, ON, con agitación a 200 rpm hasta alcanzar una DO₆₀₀ de aproximadamente 2. Las bacterias se cosecharon por centrifugación a 2000 x g durante 20 minutos y se resuspendieron en 500 mL de medio ½ MS, sacarosa 5 % y el detergente silwett-L77 0.05 %. Las plantas fueron sumergidas en la solución por unos segundos y fueron cubiertas con un nylon por 24 horas para mantener la humedad y favorecer la transformación. Posteriormente, se repitió el procedimiento para transformar nuevas inflorescencias de forma individual. Para estas inoculaciones, se depositó la suspensión bacteriana detallada anteriormente sobre cada botón floral (Martinez et al., 2012). De igual forma, se cubrió las plantas con nylon hasta el día siguiente.

13. Diseño de adaptadores y generación de construcciones para síntesis de sgARNs

Se generaron sgARNs dirigidos hacia el exón 2 del gen *EPSPS1*. Las secuencias blanco fueron seleccionadas utilizando el recurso web ccTop (Stemmer et al., 2015). Los oligos Fw y Rv necesarios para la producción de los adaptadores (*target adaptors*, TAd) fueron adquiridos de la empresa Macrogen.

Los adaptadores fueron clonados en el vector pEn-Chimera T7 Expression (Fauser et al., 2014). Este vector permite el clonado del sgARN hacia el 3' del promotor T7 lo que permite la síntesis del sgARN por IVT utilizando la T7 ARN polimerasa. El clonado de los adaptadores en el vector pEN-Chimera T7 Expression se realizó mediante restricción, siendo la enzima *BbsI* la empleada para este fin.

A las secuencias predichas por el programa ccTop se les agregaron secuencias protruyentes (*overhangs*), compatibles con los generados en el vector tras el corte con *BbsI*. Esto es: Fw: 5'-GAGAGG + protoespaciador-3' y Rv: 5'- AAAC + protoespaciador + CC-3'. Se realizó la hibridación de los *primers* colocando en un tubo de microcentrífuga 46 µL de H₂O-miliQ y 2 µL de cada *stock* de oligos a 100 µM de concentración (4 µM final de cada uno, ver *Tabla 6*). Dicho tubo se incubó 5 minutos a 95 °C y posteriormente 20 minutos a RT. Paralelamente, se realizó una digestión del vector pEN-Chimera con *BbsI* en una reacción conteniendo 2 µg de vector, buffer cutsmart 1X y 10 U de *BbsI* en un volumen final de 20 µL. Dicha digestión se incubó a 37 °C, ON. El vector digerido fue purificado utilizando columnas de purificación comerciales (ver punto 21).

Tabla 6. Primers para generación de adaptadores

N° de Primer	Nombre	Secuencia	Aplicación
280	T1EPSPSFw	5' – TAGGGATTGTTTCCCCTAGTA – 3'	Generación de adaptador TAd1
281	T1EPSPSRv	5' – AAATACTAGTGGGAAACAATC – 3'	Generación de adaptador TAd1
282	T4EPSPSFw	5' – TAGGAGACAACCTTGTGTATAGTG – 3'	Generación de adaptador TAd4
283	T4EPSPSFw	5' – AAACCACTATACAACAAGTTGTCT – 3'	Generación de adaptador TAd4
284	T11EPSPSFw	5' – TAGGCGTCCTTTGACAGCAGCTG – 3'	Generación de adaptador TAd11
285	T11EPSPSFw	5' – AAACCAGCTGCTGTCAAAGGACG – 3'	Generación de adaptador TAd11
286	T15EPSPSFw	5' – TAGGACAGCAGCTGTGGTTGCTGC – 3'	Generación de adaptador TAd15
287	T15EPSPSFw	5' – AAACGCAGCAACCACAGCTGCTGT – 3'	Generación de adaptador TAd15

Se realizó la ligación de los adaptadores con el vector digerido en una reacción conteniendo 10 ng del vector pEN-Chimera, 3 µL de adaptador (TAd1, TAd4, TAd11, TAd15), buffer ligasa 1X y 1 U de ligasa T4 (Thermo Cat. EL0011) en un volumen final de 10 µL. El tubo se incubó a RT por una hora. Se inactivó la ligasa a 65 °C por 10 minutos. La mezcla de ligación se transformó en células quimiocompetentes de *E. coli* (ver punto 7) y se plaquearon en LB-agar con ampicilina a una concentración final de 100 µg/mL. Se realizaron minipreparaciones de plásmidos de las colonias resistentes a ampicilina (ver punto 19) y se corroboró la correcta incorporación del adaptador mediante restricción con *Bbs*I. Para este análisis, se mezclaron aproximadamente 500 ng de plásmido con 1 U de *Bbs*I en buffer cutsmart 1X y se incubó por una hora a 37 °C. Se confirmó la presencia del adaptador en las construcciones finales pEN-Chimera:TAd1, TAd4 y TAd11 por secuenciación.

14. Síntesis de sgARN por IVT

La transcripción *in vitro* (IVT: *in vitro* transcription) de los sgARNs se realizó utilizando el kit *TranscriptAid T7 High Yield Transcription Kit* (Thermo Cat. K0441) siguiendo las recomendaciones del proveedor. Previo a la síntesis de ARN, las construcciones pEN-Chimera:TAd1, TAd4 y TAd11 fueron linealizadas utilizando la enzima *Eco*RI en una reacción conteniendo 1 µg de ADN y 5 U de enzima en buffer 1X *Eco*RI e incubando la mezcla de reacción a 37 °C, ON. El vector digerido fue purificado por columna (ver punto 22). Para la reacción de IVT, 1 µg de ADN lineal fue incubado con NTPs (10 mM de cada NTP) y 2 µL de mix de enzima *TranscriptAid* en buffer *TranscriptAid Reaction* 1X a 37 °C durante 2 horas. Pasado este tiempo, la mezcla se incubó con *ADNsa* I (Thermo, Cat. EN0521) a 37 °C durante 15 minutos. El ARN sintetizado se purificó utilizando el kit *ARNeasy MinElute Cleanup* (Quiagen, Cat. 74204) según el protocolo provisto por el proveedor. Se cuantificó el ARN por absorbancia a 260 nm utilizando un *NanoDrop* (Thermo).

15. Expresión y purificación de la proteína SpCas9-MBP-GFP

La proteína fue expresada a partir del plásmido pMJ922 obtenido a partir del repositorio AddGene (<http://www.addgene.org/>). Dicho vector permite la expresión de la proteína SpCas9 conteniendo secuencias de localización nuclear como fusión N-terminal a la proteína de unión a maltosa (MBP) junto a una cola de histidinas (HisTag) y una fusión C-terminal a GFP. La MBP está unida por un *linker* que contiene una secuencia de corte para la proteasa TEV (*tobacco etch virus protease*). Por otro lado, la HisTag permite la purificación por cromatografía de afinidad por metales inmovilizados (IMAC), utilizando níquel en este caso.

El plásmido pMJ922 fue transformado en la cepa Rosetta DE3 de *E. coli*. Colonias individuales fueron utilizadas para estriar a completitud placas de LB sólido con ampicilina 100 µg/mL. Estas placas fueron incubadas ON a 37 °C. La placa fue raspada utilizando una espátula estéril y las bacterias fueron resuspendidas en 5 mL de medio LB-líquido. Esta suspensión se utilizó para inocular 1 L de medio TB (extracto de levadura 24 g/L, triptona 20 g/L, glicerol 4 mL/L, buffer fosfato KH₂PO₄ 17 mM, K₂HPO₄ 72 mM) con agregado de ampicilina 200 µg/mL. Se creció el cultivo a 37 °C y a 200 rpm hasta alcanzar una DO₆₀₀ de 0.6. En este punto, se enfrió el cultivo durante al menos 30 minutos a 4 °C. La expresión de la proteína recombinante se indujo con el agregado de isopropil-β-D-1-tiogalactopiranosido (IPTG) 200 µM y se reforzó el cultivo con 200 µg/mL de ampicilina adicional. Se realizó la inducción por 16 horas a 18 °C y en agitación (180 rpm). Se cosecharon las células mediante centrifugación a 3000 x g durante 20 minutos y se guardaron los *pellets* a -80 °C hasta su procesamiento.

Seis gramos de *pellet* bacteriano se resuspendieron en 30 mL de buffer A IMAC (tris 50 mM pH 7.5, NaCl 500 mM, MgCl₂ 5 mM; 5 mL por gramo de *pellet*). La suspensión bacteriana fue sometida a tres ciclos de congelado/descongelado a -80 °C. Se transfirió el líquido a un vaso de bohemia y se agregó ditioneitol (DTT) 1 mM, lisozima 0.1 mg/mL, MgCl₂ 2 mM, una tableta de inhibidores de proteasa (Roche, Cat. 05892970001) y 1 µL de benzonasa (25 kU/µL, Pierce, Cat. 88701). Se incubó en agitación muy suave por 30 minutos. Posteriormente se realizaron 4 ciclos de sonicado utilizando pulsos de 40 % de amplitud durante 3 minutos con pausas de 2 minutos entre sí. Para realizar el clarificado, se centrifugó por 15 minutos a 16600 x g y a 4 °C, se transfirió el sobrenadante a un tubo nuevo y se centrifugó nuevamente durante 45 minutos a 16600 x g y a 4 °C.

La fracción soluble se sembró en una columna IMAC HiTrap FastFlow de 5 mL (GE, Cat. 17092104), previamente lavada con 10 volúmenes de columna (CV: *column volumes*) de H₂O-MiliQ y equilibrada en 10 CV de buffer A IMAC. La columna se lavó con 25-30 CV del buffer A IMAC. Las proteínas unidas a la resina fueron eluidas en fracciones de 0.5 mL en buffer B IMAC (tris 50 mM pH 7.5, NaCl 250 mM, MgCl₂ 5 mM, imidazol 400 mM).

Las fracciones conteniendo proteína fueron agrupadas y se realizó un cambio de buffer en columna de desalado PD-10 (GE, Cat. 17-0851-01) equilibrada con buffer Sp A (tris 50 mM, NaCl 100 mM, MgCl₂ 5 mM). Se sembraron 2.5 mL de muestra, se permitió la entrada del volumen a la columna y se agregó 3.5 mL de buffer Sp A para eluir. Tras la elución, se añadió DTT en una concentración de 1 mM final.

El segundo paso de la purificación consistió en una cromatografía de intercambio iónico utilizando una columna SpHp HighTrap (GE, Cat. GE29-0513-24) de intercambio catiónico. La columna fue equilibrada en buffer Sp A, se sembró la proteína y se lavó con 25 CV de buffer Sp A. Se eluyó la proteína utilizando buffer Sp B (tris 50 mM, NaCl 1 M, MgCl₂ 5 mM), recogiendo en fracciones de 0.5 mL.

Finalmente, se utilizó la columna PD-10 para transferir la proteína hacia el buffer de almacenamiento (HEPES pH 7.5 30 mM, NaCl 200 mM, MgCl₂ 2 mM, tris (2-carboxietil) fosfina [TCEP] 1 mM).

La concentración de proteína se determinó por medida de absorbancia a 280 nm utilizando un coeficiente de absortividad molar predicho de 211920 M⁻¹ cm⁻¹ (<https://web.expasy.org/protparam/>). La proteína fue almacenada a una concentración de 1 mg/mL a -80 °C tras congelamiento en N₂ líquido.

16. SDS-PAGE

La purificación de la SpCas9 se analizó mediante electroforesis desnaturizante en gel de poliacrilamida. Se preparó un gel inferior (separador) de 5 mL con una relación acrilamida/bisacrilamida 19:1 y un porcentaje de 8 % (1.25 mL Tris 1.5 M pH 8.8 con agregado de SDS 0.1 %, 1 mL 40 % acril-bisacril mix, 50 µL de persulfato de amonio [APS] 10 % y 2 µL de tetrametiletilendiamina [TEMED]). A su vez, se preparó un gel superior (concentrador) de 2 mL con una relación acrilamida/bisacrilamida 19:1 y un porcentaje de 5 % (0.250 mL buffer tris 1M (pH 6.8) con agregado de SDS 0.1 %, 0.250 mL 40 % acril-bisacril mix, 20 µL de APS y 2 µL de TEMED). Las fracciones colectadas fueron sembradas con un buffer de carga 4X (tris-HCl 200 mM pH 6.8, DTT 400 mM, SDS 8 %, azul bromofenol 0.4 %, glicerol 40 %) junto con un marcador de peso molecular (NEB, Cat. P77045). La corrida electroforética fue realizada en buffer de electrodo (tris 25 mM, glicina 192 mM, SDS 0.1 %, pH 8.3) a 10 mA hasta que las proteínas ingresaron al gel separador y a 20 mA hasta finalizar la corrida. Al finalizar la corrida, se lavó el gel con agua, se tiñó con solución de Coomassie (Coomassie brilliant blue R250 0.025 %, etanol 40 % v/v, ácido acético 7 % v/v) y se destiñó con solución de desteñido (etanol 40 % v/v, ácido acético 7 % v/v).

17. Validación *in vitro* de los sgARNs

Se prepararon reacciones para ensayos de corte de Cas9 utilizando como blanco productos de PCR de los genes *EPSPS1* (Glyma.01G139600.1) y *EPSPS3* (Glyma.03G027400.1) de soja (variedad Jack). Las secuencias correspondientes a cada gen fueron amplificadas utilizando la polimerasa Phusion HF. Las reacciones se prepararon de acuerdo a las especificaciones sugeridas por el proveedor en un volumen de 20 μ L: buffer Phusion HF 1X, 200 μ M dNTPs, 0.5 μ M de cada uno de los *primers*, 0.02 U de polimerasa y 100 ng de ADN genómico. Los *primers* utilizados fueron el 276 y 277 para *EPSPS1* y 278 y 279 para *EPSPS3* (Tabla 7). El programa utilizado fue: desnaturalización inicial: 98 °C, 30 segundos; 35 ciclos de: desnaturalización: 98 °C, 10 segundos, hibridación: 65 °C para *EPSPS1* o 66 °C para *EPSPS3*, 30 segundos, extensión: 72 °C, 1 minuto 30 segundos; extensión final: 72 °C, 10 minutos.

Tabla 7. *Primers* utilizados para amplificar *EPSPS1* y *EPSPS3*

N° de Primer	Nombre	Secuencia	Aplicación
276	FwEPSPSint-Sall	5'-TTTGTGCGACGGGGCCTCAAATCTCGCATC-3'	Amplificación del gen <i>EPSPS1</i>
277	RvEPSPSint-NotI	5'-TGCGGCCGCTCAATTTCCACATCACCAAGAGC-3'	Amplificación del gen <i>EPSPS1</i>
278	EPSPSCh3Fw	5' – AGGAGTTTGATGCGTGGAGTGG – 3'	Amplificación del gen <i>EPSPS3</i>
279	EPSPSChRv	5' – AATTTCCACGTCCCAAGAGC – 3'	Amplificación del gen <i>EPSPS3</i>

Para los ensayos de corte, se incubó el producto de PCR (300 ng) junto con los complejos ribonucleoproteicos (RNPs) pre-ensamblados en buffer de reacción (HEPES 20 mM, KCl 150 mM, MgCl₂ 10mM, EDTA 0.1 mM, DTT 0.5 mM). Para ensamblar los RNPs, 4 μ g de Cas9 se incubaron junto con 1120 ng de sgARN durante 15 minutos a RT en buffer de reacción. Pasado este tiempo, se agregaron 300 ng de producto de PCR y la mezcla se incubó durante 2 horas a 37 °C. La reacción se detuvo incubando con 1 μ L de proteinasa K (2 μ g/ μ L) durante 10 minutos a 37 °C. Se analizó el corte del producto de PCR por electroforesis en gel de agarosa 1 %. Las reacciones de corte y los controles montados se resumen en la Tabla 8.

Tabla 8. Ensayos de corte *in vitro* con endonucleasa Cas9.

<u><i>EPSPS1</i></u>	Control negativo	Sin sgARN	sgARN1	sgARN4	sgARN11
Buffer 10X + DTT	5 μ L	5 μ L	5 μ L	5 μ L	5 μ L
H ₂ O-DEPC	40 μ L	26.7 μ L	26.2 μ L	26.2 μ L	25.8 μ L
Cas9 GFP (0.3 mg/ μ L)	-	13.3 μ L	13.3 μ L	13.3 μ L	13.3 μ L
sgARN (1120 ng)	-	-	0.5 μ L	0.5 μ L	0.9 μ L
PCR <i>EPSPS1</i> (300 ng)	5 μ L	5 μ L	5 μ L	5 μ L	5 μ L

<i>EPSPS 3</i>	Control negativo	Sin sgARN	sgARN1	sgARN4	sgARN11
Buffer 10X + DTT	5 µL	5 µL	5 µL	5 µL	5 µL
H ₂ O-DEPC	40 µL	26.7 µL	26.2 µL	26.2 µL	25.8 µL
Cas9 GFP (0.3 mg/µL)	-	13.3 µL	13.3 µL	13.3 µL	13.3 µL
sgARN (1120 ng)	-	-	0.5 µL	0.5 µL	0.9 µL
PCR <i>EPSPS3</i> (300 ng)	5 µL	5 µL	5 µL	5 µL	5 µL

18. Generación de vector de remplazo

Para generar la construcción para el remplazo alélico en soja, se realizó en una primer instancia el clonado de un fragmento genómico de 2423 pb del gen *EPSPS1* en el vector pCRII-Blunt TOPO (Invitrogen, Cat. K270020). Los sitios de corte para los sgARNs diseñados se encuentran aproximadamente en medio de este fragmento. El fragmento fue amplificado a partir de ADN genómico de soja (variedad Jack) utilizando las condiciones y *primers* descritos en el punto 17.

Se realizó el clonado en el vector pCRII-Blunt de acuerdo al protocolo recomendado por el proveedor. Se incubó 10 ng de pCRII-Blunt junto con el producto de PCR y solución salina (NaCl 1.2 M, MgCl₂ 0.06 M), durante 1 hora. Se transformaron células quimiocompetentes de *E. coli* (ver punto 7) y se plaqueó en LB-agar con agregado de kanamicina (50 µg/mL) y 5-bromo-4-cloro-3-indolil-β-D-galactopiranosido (X-gal). Se picaron colonias blancas para realizar minipreparaciones de plásmidos (ver punto 19). Se hicieron digestiones con la enzima *EcoRI* para verificar el clonado. Se incubó aproximadamente 500 ng de plásmido con 2 U de *EcoRI* en buffer *EcoRI* a 37 °C durante 1 hora. Los clones pCRII:EPSPS1_{genom} que presentaron el patrón de digestión adecuado fueron confirmados por secuenciación.

En un segundo paso, se introdujeron las mutaciones TIPS y TIPA en la construcción pCRII:EPSPS1_{genom}. Esto se realizó mediante mutagénesis Quickchange utilizando las mismas condiciones descritas en el punto 4. Se transformaron células quimiocompetentes de *E. coli* TOP 10 (ver punto 7) y se plaquearon en LB-Agar con agregado de kanamicina (50 µg/mL). Se realizaron minipreparaciones de plásmidos (ver punto 19) y el ADN plasmídico se corrió en un gel de agarosa 1 % para verificar que el tamaño fuera el correcto. Las construcciones finales pCRII:EPSPS1_{genom} TIPS y TIPA conteniendo las mutaciones fueron chequeadas por secuenciación.

En un tercer paso se introdujo una mutación silenciosa lisina-lisina (K-K) en el PAM asociado al TAd1. Para ello, se realizó una mutagénesis Quickchange utilizando la enzima

Phusion HF y los *primers* 300 y 301 (Tabla 9) y como molde las construcciones pCRII:EPSPS1_{genom}-TIPS o pCRII:EPSPS1_{genom}-TIPA. El programa utilizado para la PCR fue: desnaturalización inicial: 98 °C, 30 segundos; 40 ciclos de: desnaturalización: 98 °C, 10 segundos, hibridación: 65 °C, 30 segundos, extensión: 72 °C, 2 minutos 30 segundos; extensión final: 72 °C, 10 minutos. Se transformaron células quimiocompetentes de *E. coli* TOP 10 (ver punto 7) y se plaqueó en LB-Agar con agregado de kanamicina (50 µg/mL). Se realizó una minipreparación de plásmidos (ver punto 19) y el ADN se corrió en un gel de agarosa 1 % para verificar que el tamaño fuera el correcto. Las construcciones pCRII:EPSPS1_{genom}-TIPS-KK y pCRII:EPSPS1_{genom}-TIPA-KK fueron secuenciadas para confirmar la mutación.

En un cuarto paso, se introdujo en los bordes de la secuencia de homología la secuencia TA1+PAM mediante PCR. Los *primers* utilizados en este caso hibridan con regiones intrónicas hacia su 3' y presentan en su región 5' la secuencia TA1+PAM (5'-GGGATTGTTTCCCACTAGTAAGG-3') (Tabla 8). Se llevó a cabo una PCR utilizando la enzima Phusion HF siguiendo las recomendaciones del proveedor, utilizando los *primers* 302 y 303 y como moldes las construcciones pCRII:EPSPS1_{genom}-TIPS-KK o pCRII:EPSPS1_{genom}-TIPA-KK. El programa utilizado para la PCR fue: desnaturalización: 98 °C, 30 segundos; 35 ciclos de: desnaturalización: 98 °C, 10 segundos, hibridación/extensión 1 minuto; extensión final: 72 °C, 10 minutos.

Los productos de PCR obtenidos fueron ligados en el vector pCRII-Blunt TOPO. Se transformaron células quimiocompetentes de *E. coli* TOP 10 (ver punto 7) y se plaquearon en LB-agar con agregado de kanamicina (50 µg/mL) y X-gal. Los clones fueron chequeados con la enzima *EcoRI* para verificar la presencia del inserto y chequeados por secuenciación.

Tabla 9. *Primers* utilizados para generación del vector de remplazo

N° de Primer	Nombre	Secuencia	Aplicación
300	FwEPSPS1K-K	5'-CCCACTAGTAAaGAATCTAAAGATGAAATCAAT-3'	Generación de mutación en PAM
301	RvEPSPS1K-K	5'-TCTTTAGATTcTTACTAGTGGGAAACAATCCC-3'	Generación de mutación en PAM
302	RvEPSPS1+PAM	5'-GGGATTGTTTCCCACTAGTAAGG TTTGAACTGAGGTTTTGAAGGG-3'	Amplificación intrónica/TA1+PAM
303	FwEPSPS1+PAM	5'-GGGATTGTTTCCCACTAGTAAGG GTATTATTCTGTGAAATGAAGC-3'	Amplificación intrónica/TA1+PAM

19. Minipreparación casera (miniprep)

Se centrifugaron 3 mL de cultivo líquido bacteriano durante 2 minutos a 16600 x g en tubos de microcentrífuga. Se descartó el sobrenadante y se resuspendió el *pellet* en 300 µL de buffer P1 (tris-HCl 50 mM pH 8.0, EDTA 10 mM, ARNsA A 100 µg/mL). Se realizó la lisis alcalina

en 300 μ L de solución P2 (NaOH 200 mM, SDS 1 %). No se permitió que la lisis procediera por más de 5 minutos. Se neutralizó con 300 μ L de solución P3 (acetato de potasio 1.5 M, ácido acético glacial 86.25 mM). Se centrifugó a máxima velocidad (16600 x g) por 10 minutos a 20 °C. Se transfirió el sobrenadante con ADN plasmídico hacia un tubo de microcentrífuga nuevo y se agregaron 0.6 volúmenes de isopropanol. Se centrifugó nuevamente por 20 minutos a 16600 x g a 20 °C. Se descartó el sobrenadante y se lavó el *pellet* con 600 μ L de EtOH 70 %. Se secó el *pellet* a 70 °C por 2 minutos y se resuspendió el ADN plasmídico en 30 μ L de H₂O-MiliQ.

20. Minipreparación de kit Qiagen (QIAprep Spin Miniprep Kit Cat. 27106)

Se centrifugó de 1 a 5 mL de cultivo bacteriano líquido por 1 minuto a 16600 x g y a TA. Se resuspendió el *pellet* en 250 μ L de buffer P1 y se transfirió a un tubo de microcentrífuga. Se añadieron 250 μ L de solución P2 y se mezcló invirtiendo el tubo de cuatro a seis veces hasta que la solución fuera translúcida. No se permitió que la lisis procediera por más de 5 minutos. Se agregaron 350 μ L de solución N3, se mezcló invirtiendo el tubo. Se centrifugó por 10 minutos a 16600 x g a RT. Se transfirió el sobrenadante a una columna QIAprep 2.0, se centrifugó por 1 minuto y se descartó el *flow-through*. Para realizar un lavado a la columna, se añadieron 750 μ L de solución PE, se centrifugó por 1 minuto y se descartó el *flow-through*. Se centrifugó nuevamente por 2 minutos para secar completamente la columna. Se colocó la columna QIAprep 2.0 en un tubo de microcentrífuga nuevo y para eluir el ADN plasmídico se colocaron 50 μ L de H₂O-MiliQ. Se dejó reposar por 1 minuto y se centrifugó por 1 minuto. Se repitió este paso por segunda vez para maximizar la eficiencia de la elución.

21. Purificación de producto PCR de kit Qiagen (QIAquick PCR Purification Kit Cat. 28104)

Se añadieron 5 volúmenes de buffer PB por volumen de reacción de PCR y se mezcló para homogeneizar. Se colocó la mezcla sobre una columna QIAquick provista por el kit, se centrifugó por 1 minuto a TA y a 16600 x g y se descartó el líquido. Para lavar la columna, se añadieron 750 μ L de solución PE, se centrifugó por 1 minuto y se descartó el *flow-through*. Se centrifugó nuevamente por 2 minutos para secar completamente la columna. Se colocó la columna en un tubo de microcentrífuga nuevo y para eluir el ADN se agregó 50 μ L de H₂O-MiliQ. Se centrifugó por 1 minuto en las mismas condiciones. Se repitió este paso por segunda vez para maximizar la eficiencia de la elución.

22. Purificación de gel de kit Qiagen (QIAquick Gel Extraction Kit Cat. 28706)

Se escindió el fragmento de ADN de interés del gel de agarosa utilizando un bisturí. Se pesó la lámina de gel en un tubo de microcentrífuga y se añadieron 3 volúmenes de buffer QG por volumen de gel (100 mg de gel ~ 100 µL). Se incubó el tubo por 10 minutos a 50 °C hasta que el gel estuviera completamente disuelto. Se vortexeó cada dos o tres minutos para colaborar con la disolución.

Se incorporó un volumen de isopropanol y se mezcló. Se colocó una columna QIAquick en un tubo de recolección y para unir el ADN a la columna se añadió el volumen disuelto y se centrifugó por 1 minuto a 16600 x g y a TA. Para lavar la columna, se adicionaron 750 µL de solución PE y se centrifugó por 1 minuto en iguales condiciones. Se centrifugó nuevamente por 2 minutos para secar completamente la columna. Se colocó la columna en un tubo de microcentrífuga nuevo y para eluir el ADN se añadieron 50 µL de H₂O-MiliQ. Se centrifugó por 1 minuto en las mismas condiciones. Se repitió este paso por segunda vez para maximizar la eficiencia de la elución.

23. Electroforesis en gel de agarosa

Se preparó un gel de agarosa 1 % w/v utilizando buffer TAE 1X (tris 40 mM, acetato 20 mM, EDTA 1 mM), agarosa y BrET 0.05 µg/mL. Una vez solidificado el gel, se montó en una cuba electroforética con buffer TAE 1X y se corrieron las muestras junto con un marcador de peso molecular (ADN del fago lambda digerido con la enzima *Pst*I). Al finalizar la corrida, se observaron los resultados en un transiluminador.

Resultados y discusión

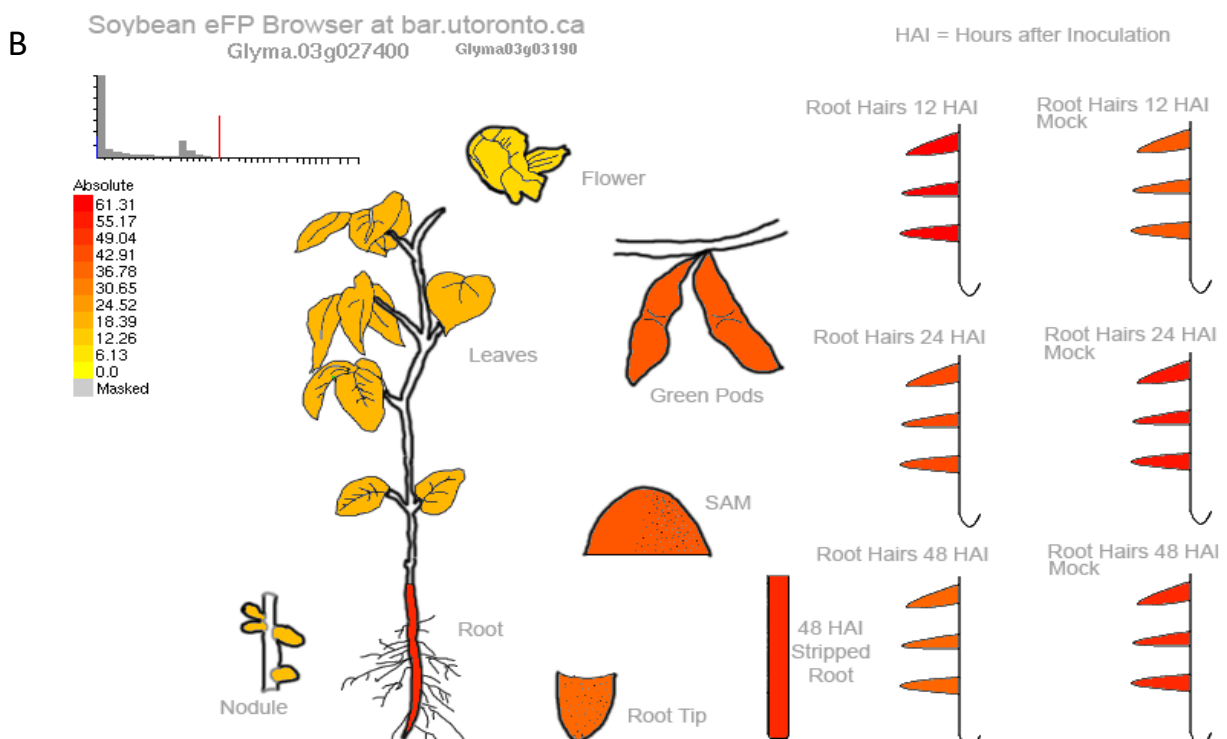
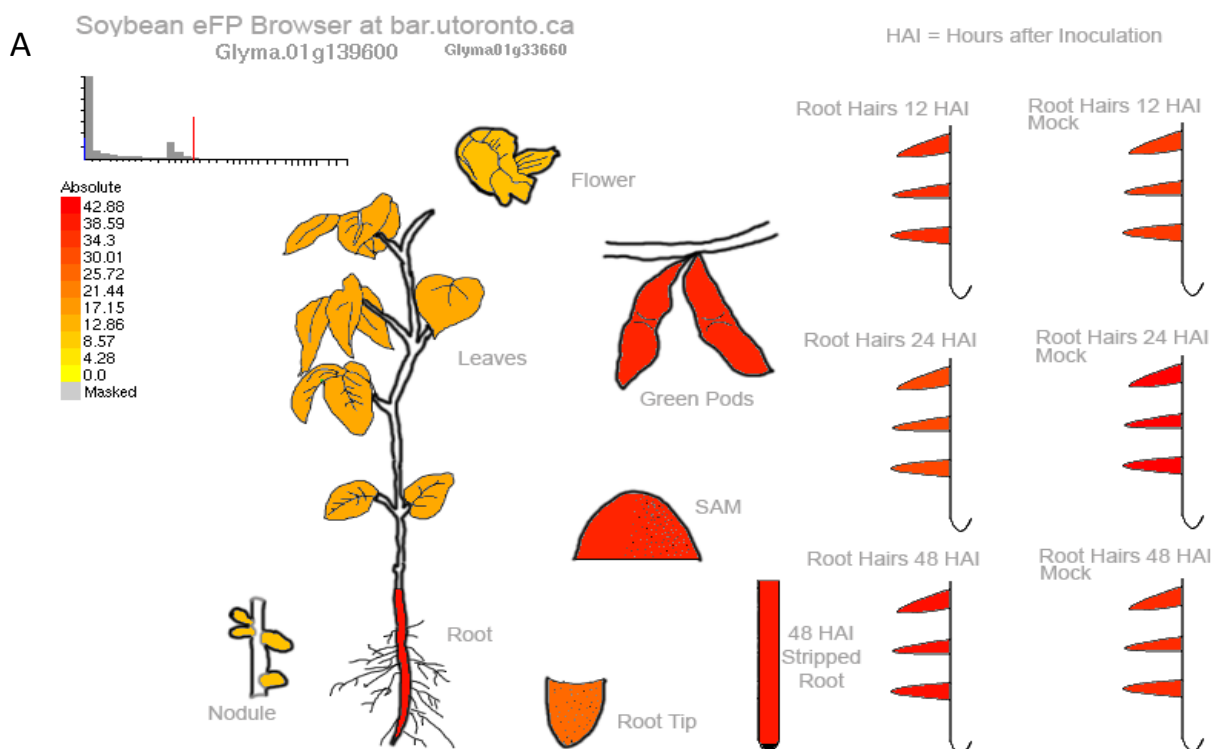
El objetivo de este proyecto es optimizar metodologías de edición génica en soja (*Glycine max*) con el fin de generar plantas con modificaciones sitio-específicas utilizando el sistema CRISPR/Cas9. En particular, este trabajo propone optimizar metodologías de remplazo alélico, no disponibles aún en la literatura para esta especie. Para este fin se requiere la utilización de herramientas moleculares para modificar el genoma de soja, así como diseñar e implementar una estrategia para sustituir una secuencia genómica por otra conteniendo los cambios deseados. Para ello, es necesario construir un vector de remplazo, optimizar una metodología para incorporar estos elementos genéticos en la planta en forma eficiente, así como procedimientos eficientes para seleccionar y regenerar los individuos editados.

La forma más conveniente de abordar la optimización de un protocolo es procurando obtener un fenotipo fácil de detectar. En ese sentido, en este trabajo se optó por generar una variante del gen de la 5-enolpiruvatoshikimato-3-fosfato (EPSP) sintasa (EPSPS), resistente al herbicida glifosato. Esta enzima es la sexta en la ruta metabólica del shikimato y es esencial para la síntesis de aminoácidos aromáticos. Las EPSPSs catalizan la transferencia del grupo enolpiruvato presente en el fosfoenolpiruvato (PEP) hacia la molécula de shikimato-3-fosfato (S3P), formando 5-enolpiruvato-3-shikimato fosfato (EPSP) y liberando fosfato inorgánico.

Las EPSPSs son proteínas formadas por dos dominios globulares similares. El sitio activo de la enzima se localiza en el surco inter-dominio (*Schonbrunn et al., 2001*), donde se unen los sustratos S3P y PEP. El herbicida glifosato es un inhibidor competitivo de la enzima EPSPS al unirse al sitio de unión del PEP (*Healy-Fried et al., 2007*). La unión de este compuesto a EPSPS puede interrumpir la síntesis de aminoácidos aromáticos, provocando la muerte de la planta (*Schonbrunn et al., 2000*). En plantas, esta proteína se localiza en los cloroplastos y en otros plástidos (*Garg et al., 2014*).

Identificación y análisis de genes EPSPS de soja

En soja existen dos genes codificantes para EPSPSs los cuales se encuentran en los cromosomas 1 (Glyma.01G139600.1) y 3 (Glyma.03G027400.1). Ambas copias comparten un 90 % de identidad de secuencia aminoacídica entre sí. El análisis *in silico* de la expresión de los genes *EPSPS1* y *EPSPS3*, utilizando la base de datos Soybean eFP de la Universidad de Toronto (bar.utoronto.ca), determinó que ambos genes poseen niveles de expresión similares en diferentes órganos y etapas del desarrollo (Figuras 5 y 6). Se escogió trabajar con el gen *EPSPS1* debido a que éste presenta un nivel de expresión en hojas levemente superior al de *EPSPS3*, siendo el glifosato un herbicida de acción foliar (Figura 5).



eFP by R. Patel. Drawings by R. Patel. Data for Root Hair series from Complete Transcriptome of Soybean Root Hair Cell, a Single Cell Model, and its alteration in response to *Bradyrhizobium japonicum* infection: Libault, M., Farmer, A., Brechnmacher, L et al. (2010). Plant Physiol. 152, 541-552. All other data from An Integrated Transcriptome Atlas of the Crop Model *Glycine max*, and its use in Comparative Analyses in Plants: Libault, M., et al. (2010) The Plant Journal 63, 86-99
 Data is mRNA-Seq derived data and all samples were sampled in triplicate.

Figura 5. Valores de expresión para los genes *EPSPS1* (Glyma.01G139600.1) y *EPSPS3* (Glyma.03G027400.1) en diferentes órganos de soja. Los valores fueron obtenidos a partir de datos de RNA-seq en la base de datos Soybean eFP Browser (<http://bar.utoronto.ca/efpsoybean/cgi-bin/efpWeb.cgi>). A la izquierda se observa una escala a color de valores absolutos de niveles de

expresión. Tanto para el gen *EPSPS1* (A) como para *EPSPS3* (B) se observan patrones similares de expresión en el organismo. *EPSPS1* posee mayor expresión en hojas.

Dataset: 7 developmental stages from data selection: GM_mRNASeq_SOYBEAN_GL-3
Showing 2 measure(s) of 2 gene(s) on selection: GM-1

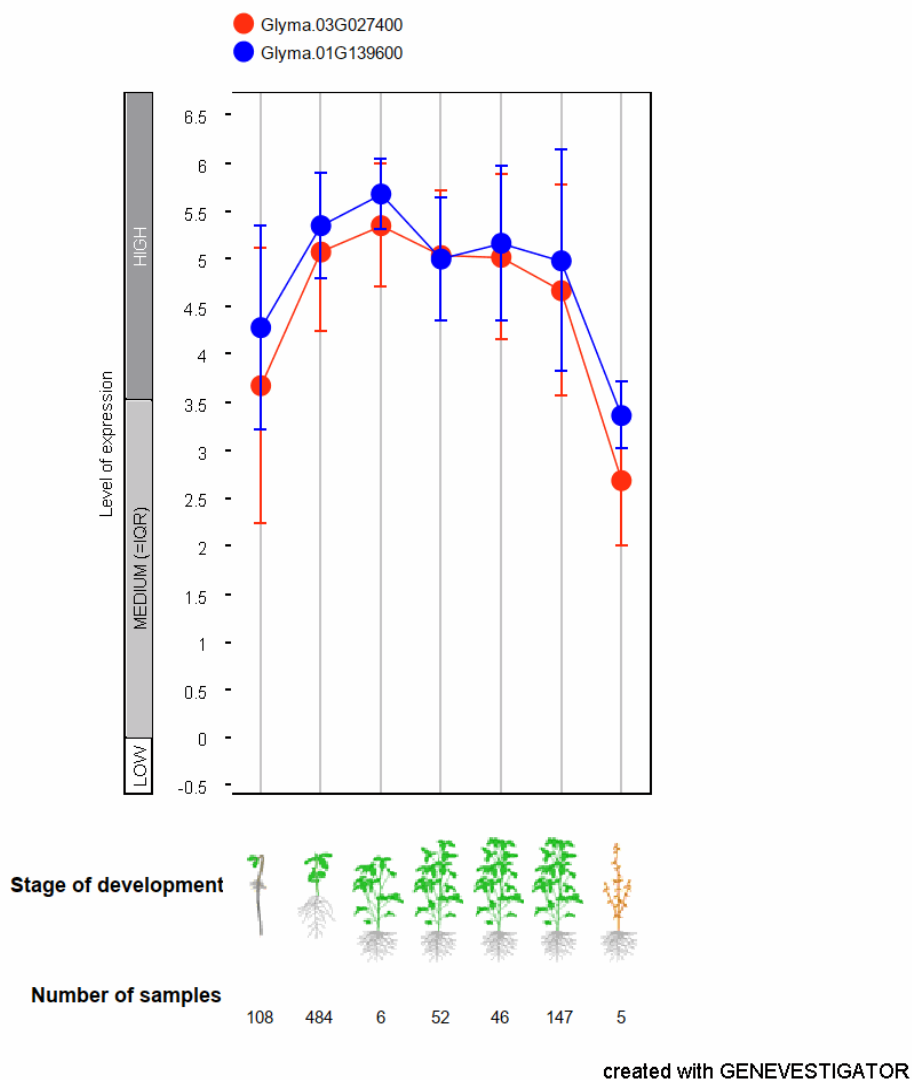


Figura 6. Valores de expresión para los genes *EPSPS1* (Glyma.01G139600.1) y *EPSPS3* (Glyma.03G027400.1) en diferentes estadios de desarrollo. Los valores fueron obtenidos a partir de datos de ARN-seq utilizando el recurso web Geneinvestigator (<https://geneinvestigator.com/gv/index.jsp>). Se observan patrones similares de expresión entre los genes en diferentes estadios de desarrollo.

Las EPSPSs poseen un grado de conservación grande (Figura 7), en especial en la región donde se ubica el sitio de unión al PEP (Figura 8) (Anexo). Las mutaciones TIPS (Thr102-Ile + Pro106-Ser) y TIPA (Thr102-Ile + Pro106-Ala) en esta región confieren insensibilidad al glifosato al mismo tiempo que permiten que la enzima mantenga inalterada la afinidad por el PEP (Healy-Fried et al., 2007). Nuestra estrategia de trabajo parte de la base de que al producir las sustituciones TIPS o TIPA en el gen *EPSPS1* de soja obtendremos plantas resistentes al

glifosato, conteniendo el gen *EPSPS1* funcional. Si bien las EPSPs son proteínas con alto grado de conservación, al momento no se han producido estas variantes en la EPSPs de soja. Por este motivo, en primera instancia se debe evaluar si estas mutaciones verdaderamente generan una variante de la EPSPs de soja que sea insensible a la inhibición por este herbicida. Con el fin de analizar *in vivo* si las mutaciones logran conferir insensibilidad al herbicida, se generaron plantas transgénicas de *Arabidopsis thaliana* que sobre-expresan las variantes mutantes TIPS y TIPA y la versión wild type (WT) de la EPSPs de soja.

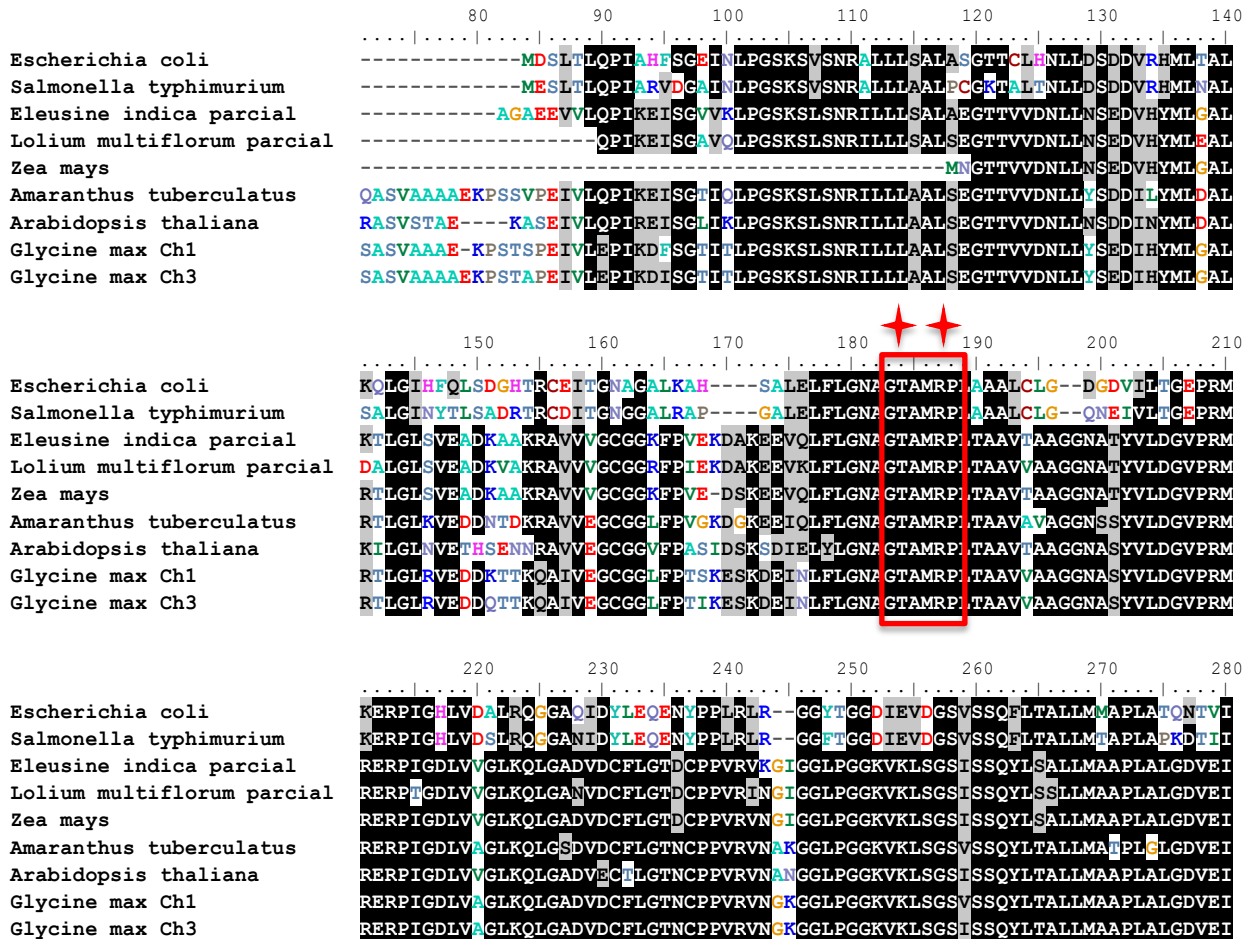


Figura 7. Alineamiento parcial de las EPSPs de soja y sus homólogos en diferentes especies (*G. max* Ch1 (Glyma.01G139600.1), *G. max* Ch3 (Glyma.03G027400.1), *A. thaliana* (AT1G48860.1), *Z. mays* (Zm00008a033811_T01), *E. indica* parcial (AAN63155.1), *L. multiflorum* parcial (AAZ79230.2), *A. tuberculatus* (ACV67278.1), *S. typhimurium* (AAA27028.1) y *E. coli* (NEY26641.1). En rojo se muestra la localización del sitio de unión al fosfoenolpiruvato y donde se hallan las mutaciones TIPS y TIPA. La conservación aminoacídica se encuentra sombreada con un *threshold* de 65. Con asteriscos se muestran los sitios de mutaciones T102 y P106 (numeración acorde a la EPSPs de *E. coli*).



Figura 8. Logo de la estructura primaria del sitio de unión al PEP y región flanqueante generado con el recurso weblogo (<https://weblogo.berkeley.edu/>). El logo muestra el elevado grado de conservación a nivel del sitio de unión al PEP, donde la identidad de secuencia alcanza el 96 %. En asteriscos se indican la Thr102 y la Pro106. Los números de acceso utilizados fueron NEY26641.1, AAA27028.1, AAN63155.1, AAZ79230.2, Zm00008a033811_T01, ACV67278.1, AT1G48860.1, Glyma.01G139600.1 y Glyma.03G027400.1.

Generación y caracterización de mutantes TIPS y TIPA de la *EPSPS1* de soja

Con el objetivo de obtener una variante funcional del gen *EPSPS1* de soja insensible al glifosato, se generaron las mutaciones TIPS y TIPA en la región codificante del gen y las variantes resultantes fueron introducidas en *Arabidopsis* para evaluar la funcionalidad de la enzima resultante.

1. Generación de construcciones para transformación de *A. thaliana*

Para la generación del vector de expresión del gen *EPSPS1* de soja, se generó ADNc a partir de ARN extraído de plantas de soja, siguiendo los protocolos descritos en Materiales y Métodos (ver punto 2). Se amplificó por PCR la región codificante (CDS) del gen *EPSPS1* (Glyma.01G139600.1), utilizando *primers* específicos para el inicio y la terminación de la CDS, la cual fue clonada en el vector pENTR2B entre los sitios para enzimas de restricción *Sall* y *NotI*. En este vector la CDS de *EPSPS1* queda flanqueada por los sitios attL1 y attL2, lo cual permite la posterior recombinación con vectores de destino que presenten los sitios attR1 y attR2, utilizando el sistema *Gateway*. El vector pENTR2B posee un gen de resistencia a kanamicina (neomicina fosfotransferasa II, *nptII*), antibiótico utilizado para seleccionar los transformantes (Figura 9).

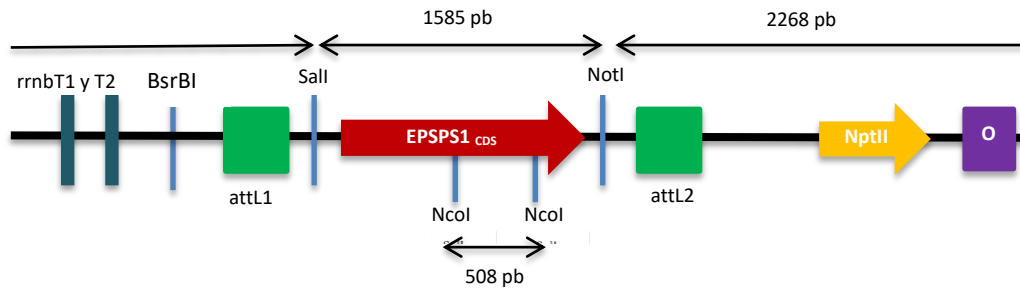


Figura 9. Esquema del vector de entrada pENTR2B:EPSPS1_{CDS}. Se muestran los sitios de recombinación attL1 y attL2, los sitios de restricción *SalI* y *NotI* (utilizados para clonar la CDS de la *EPSPS1* en el vector), *NcoI* (utilizado para chequear la presencia del inserto durante el clonado) y *BsrBI* (utilizado para linealizar el vector de entrada previo a la recombinación LR), el gen *nptII* (de resistencia a kanamicina), el sitio de origen de replicación (O) y la secuencia codificante del gen *EPSPS1* de soja. El vector contiene dos secuencias de terminación *rrnB* que impiden la transcripción basal del producto clonado en *E. coli*, evitando un posible efecto tóxico en las bacterias (Orosz *et al.*, 1991).

Para confirmar el clonado, los plásmidos purificados a partir de las colonias bacterianas resistentes a la kanamicina fueron digeridos con *NcoI* y analizados en geles de agarosa. Si bien el vector pENTR2B no contiene sitios de restricción para *NcoI*, esta enzima reconoce dos regiones dentro de la secuencia codificante de *EPSPS1*, en las posiciones 955 pb y 1463 pb. Por lo tanto, los tamaños esperados de la digestión de los clones positivos son de 508 pb y 3344 pb. Este fue el patrón obtenido en el caso de los clones 3, 5 y 6 (Figura 10). Estos clones fueron seleccionados y analizados mediante secuenciación (Figura 11).



Figura 10. Electroforesis en gel de agarosa 1 % de la digestión con *NcoI* del plásmido pENTR2B:EPSPS1_{CDS}. Los clones 3, 5 y 6 poseen el patrón de bandas esperado (3344 pb y 508 pb). MPM: marcador de peso molecular (ADN del fago lambda digerido con la enzima *PstI* [λ_{PstI}]).

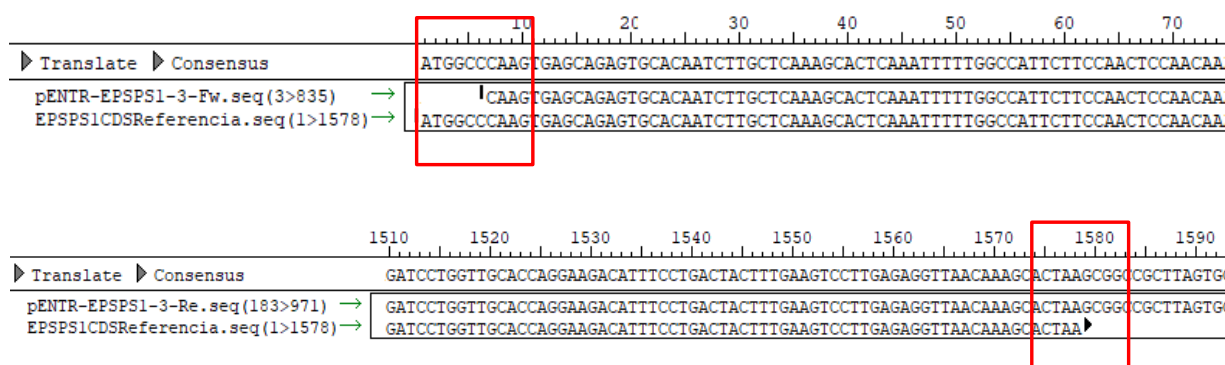


Figura 11. Resultados de secuenciación de los clones pENTR2B:EPSPS1_{CDS}. Mediante el recurso SeqMan de Lasergene se realizó un contig alineando los resultados de secuenciación con la secuencia de referencia de la CDS del gen *EPSPS1*. En la imagen superior se muestra el comienzo 5'-ATG (superior) y en la imagen inferior el final 3'- TAA (inferior) de la CDS de EPSPS1. El resultado de secuenciación no presentó diferencias con la secuencia de referencia.

Tomando como molde la construcción pENTR2B:EPSPS1_{CDS}, se introdujeron las mutaciones TIPS y TIPA mediante la metodología QuickChange, utilizando *primers* perfectamente complementarios y, posteriormente, parcialmente complementarios (Liu & Naismith, 2008).

La metodología de mutagénesis dirigida por PCR se basa en utilizar *primers* total o parcialmente complementarios entre sí, que hibridan con el plásmido sobre el que se quiere introducir la mutación. La técnica implica que durante los ciclos de PCR los *primers* se hibridan al molde de ADN replicando el plásmido en su totalidad e introduciendo la mutación deseada. La mezcla de cadenas resultantes incluye el ADN sintetizado por PCR conteniendo la mutación y el ADN parental metilado, que puede eliminarse por digestión con la enzima *DpnI*. El producto mutado posee extremos homólogos que permiten que el ADN adopte una conformación circular. Al ser transformado en *E. coli*, el daño es reparado por mecanismos de reparación propios de la bacteria, restaurando el plásmido circular (Liu & Naismith, 2008). Esta metodología es frecuentemente utilizada para introducir mutaciones puntuales en el ADN, realizar deleciones o realizar inserciones.

Por otro lado, las características de los *primers* (elevada complementariedad, gran número de pb) pueden derivar en la formación de dímeros de *primers* (Edelheit et al., 2009). Asimismo, al utilizar *primers* perfectamente complementarios, la amplificación por PCR es lineal: el ADN sintetizado *de novo* no puede ser utilizado como molde en las reacciones posteriores ya que no posee la secuencia complementaria para el otro *primer* (Figura 12). Esto puede llevar a un nivel de eficiencia menor del PCR.

En una primera instancia se utilizaron los *primers* 272 y 273 para introducir la mutación TIPS y 274 y 275 para introducir la mutación TIPA con el programa de PCR mencionado en el punto 4. Al no ser exitosa la amplificación, se buscó optimizar el protocolo de ciclado para los *primers* 272 y 273 realizando PCR en gradiente de temperaturas de hibridación: 47 °C, 52 °C, 57 °C, 62 °C, 67 °C, 72 °C. En ninguna de las condiciones se observó amplificación y la formación de dímeros de *primer* fue muy evidente (Figura 13). En el fundamento de esta técnica se recomienda que aunque no se observe amplificación por PCR se proceda a transformar *E. coli* igualmente (Edelheit et al., 2009). Sin embargo, en nuestra experiencia, no se obtuvieron resultados positivos en los casos en donde no se observó un producto de amplificación de ADN.

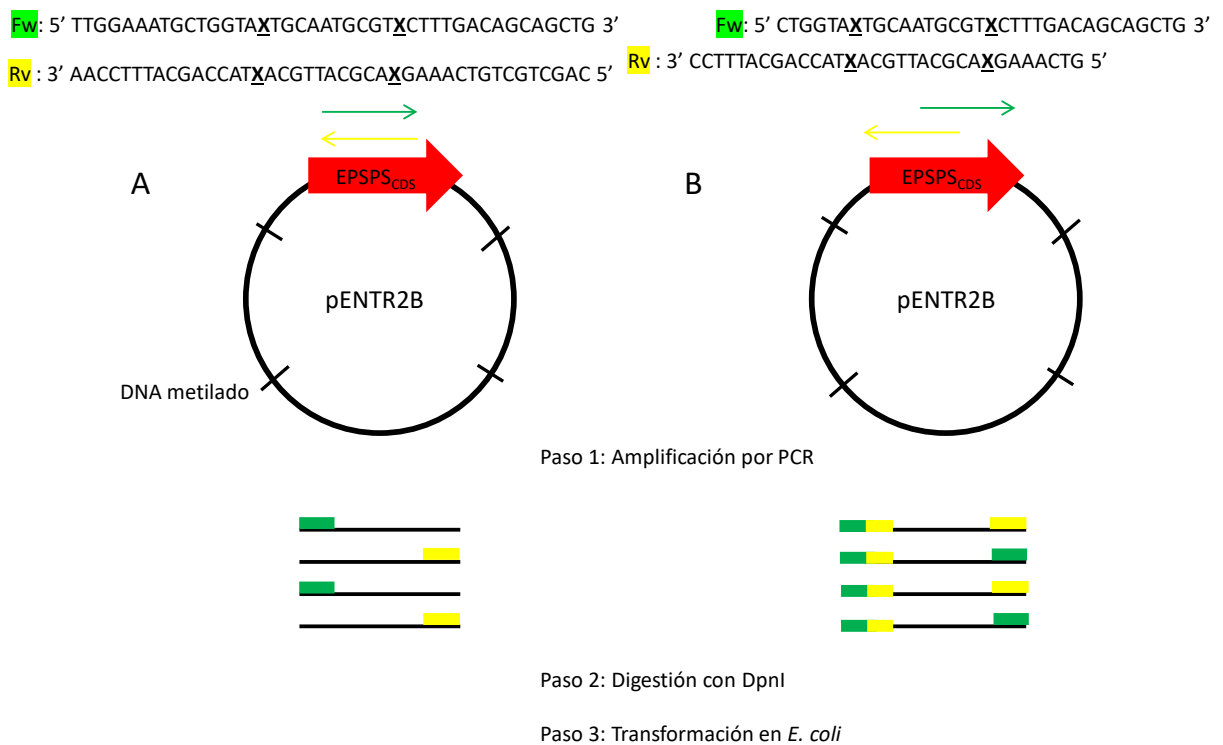


Figura 12. Esquema de metodología QuickChange utilizando *primers* perfectamente complementarios (**A**) y parcialmente complementarios (**B**). Se muestran los *primers* utilizados para cada estrategia con los sitios (X) donde se encuentran las mutaciones TIPS (mutación Thr102-Ile: ACT → ATT + mutación Pro106-Ser: CCT → TCT) o TIPA (mutación Thr102-Ile: ACT → ATT + mutación Pro106-Ala: CCT → GCT). La polimerasa amplifica en sentido 5' → 3' hasta alcanzar el extremo 5' del *primer*. En **A**, las cadenas amplificadas con el *primer* Fw solo poseen sitios de hibridación para el *primer* Fw y las amplificadas con el *primer* Rv poseen sitios de hibridación para el *primer* Rv. Como los productos de amplificación no pueden servir como molde para ambos *primers*, la amplificación es lineal. En **B**, como los *primers* son parcialmente complementarios, éstos mantienen complementariedad con la secuencia de ADN pero además poseen sitios de hibridación para cada *primer* complementario; por este motivo la amplificación es exponencial. Una vez culminado el PCR, el ADN metilado se elimina tratando la mezcla con la enzima *DpnI*.

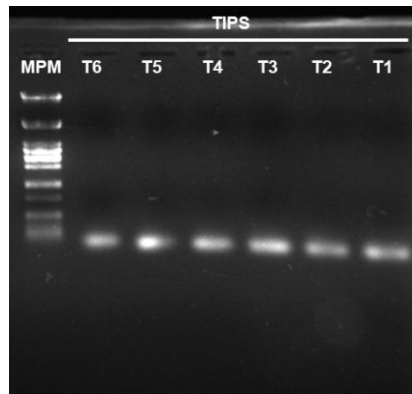


Figura 13. Electroforesis en gel de agarosa 1 % de la amplificación por PCR en gradiente (T6: 72°C, T5: 67°C, T4: 62°C, T3: 57°C, T2: 52°C, T1: 47°C) para mutación TIPS. Se observa formación de dímeros de primer de bajo peso molecular. MPM: marcador de peso molecular (λ_{PstI}).

En segundo lugar, se realizó una modificación de la metodología QuickChange haciendo uso de *primers* parcialmente complementarios entre sí (Xia et al., 2014; Figura 12). Utilizando esta estrategia se obtuvo el resultado mostrado en la Figura 14. Aunque no se observaron bandas limpias de amplificación, se transformaron células quimiocompetentes de *E. coli* y se obtuvieron colonias. Se realizaron minipreparaciones y los clones de tamaño correcto se chequearon por secuenciación (Figura 15).

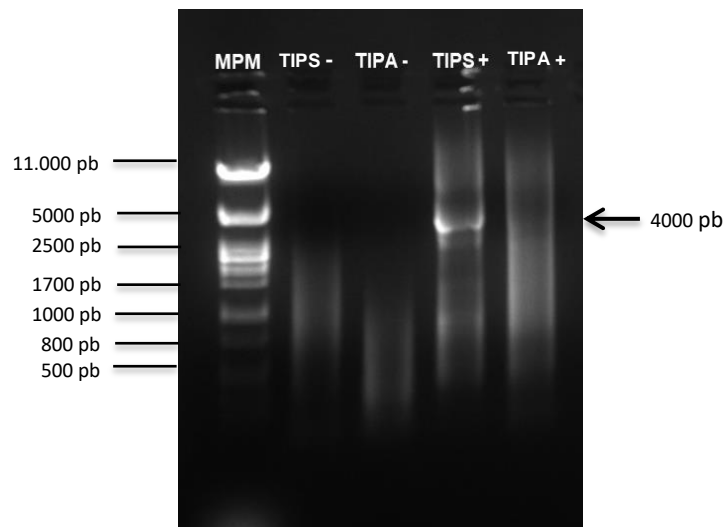


Figura 14. Electroforesis en gel de agarosa 1 % de la amplificación por PCR para TIPS (mutación Thr102-Ile: ACT \rightarrow ATT; mutación Pro106-Ser: CCT \rightarrow TCT) y TIPA (mutación Thr102-Ile: ACT \rightarrow ATT; mutación Pro106-Ala: CCT \rightarrow GCT) con *primers* parcialmente complementarios. Se muestran los controles negativos (TIPS-, TIPA-). En los controles negativos (TIPS-, TIPA-) no se observó amplificación, mientras que en los carriles 3 y 4 (TIPS+, TIPA+) se observó un *smear* de mayor peso molecular similar al vector completo (aproximadamente 4000 pb). MPM: marcador de peso molecular (λ_{PstI}).

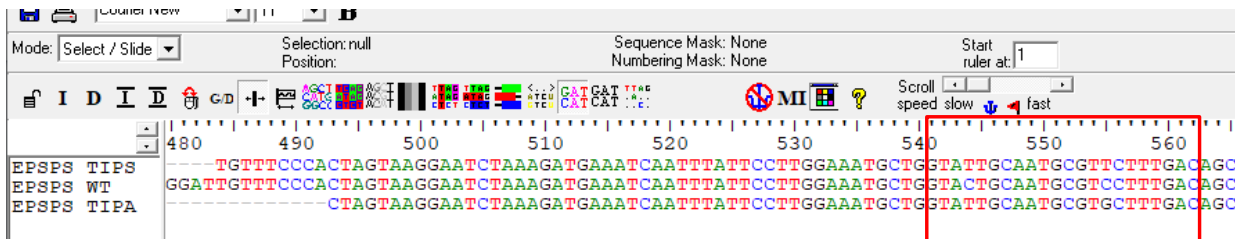
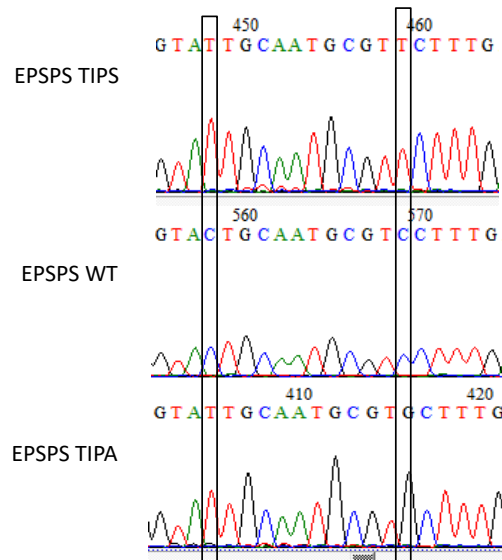


Figura 15. Resultados de secuenciación de los clones pENTR2B:EPSPS1_{CDS} WT, TIPS y TIPA. En el recuadro rojo se muestra los codones mutados. Mutación Thr102-Ile: ACT → ATT. Mutación Pro106-Ser: CCT → TCT. Mutación Pro106-Ala: CCT → GCT.

Una vez obtenido el vector de entrada con la CDS del gen *EPSPS1* y las variantes mutantes TIPS y TIPA, las secuencias se clonaron en un vector binario para la transformación de *Arabidopsis thaliana* mediada por *Agrobacterium tumefaciens*. Se utilizó la tecnología *Gateway* para incorporar, mediante recombinación LR, las secuencias presentes en los vectores pENTR2B en el vector binario. El sistema *Gateway* se basa en la recombinación de sitios attL1 y los attR mediada por la enzima LR Clonase. Gracias a la acción de esta enzima, los sitios attL1 y attL2 presentes en el vector de entrada (pENTR), recombinan con los sitios attR1 y attR2 presentes en el vector de destino (pDEST) (Reece-Hoyes y Walhout, 2018; Figura 16). En el presente en este trabajo, el vector de destino utilizado fue pUB-Dest (Grefen et al., 2010). Este vector permite la expresión del gen de interés bajo el control de un promotor constitutivo de ubiquitina 10 de *Arabidopsis* y el terminador CaMV35S (*Cauliflower Mosaic Virus*). El vector presenta los bordes izquierdo (LB) y derecho (RB) del T-ADN (ADN de transferencia), utilizados por *A. tumefaciens* para transferir esa región de ADN hacia el genoma de la célula vegetal. El vector presenta también un gen de resistencia a espectinomicina (SpeR, aminoglicosido adeniltransferasa) que permite seleccionar los transformantes en *E. coli* y *A. tumefaciens* y el

gen de resistencia a glufosinato de amonio (BastaR, fosfinotricina acetiltransferasa) que permite seleccionar las células vegetales transformadas (Figura 17).

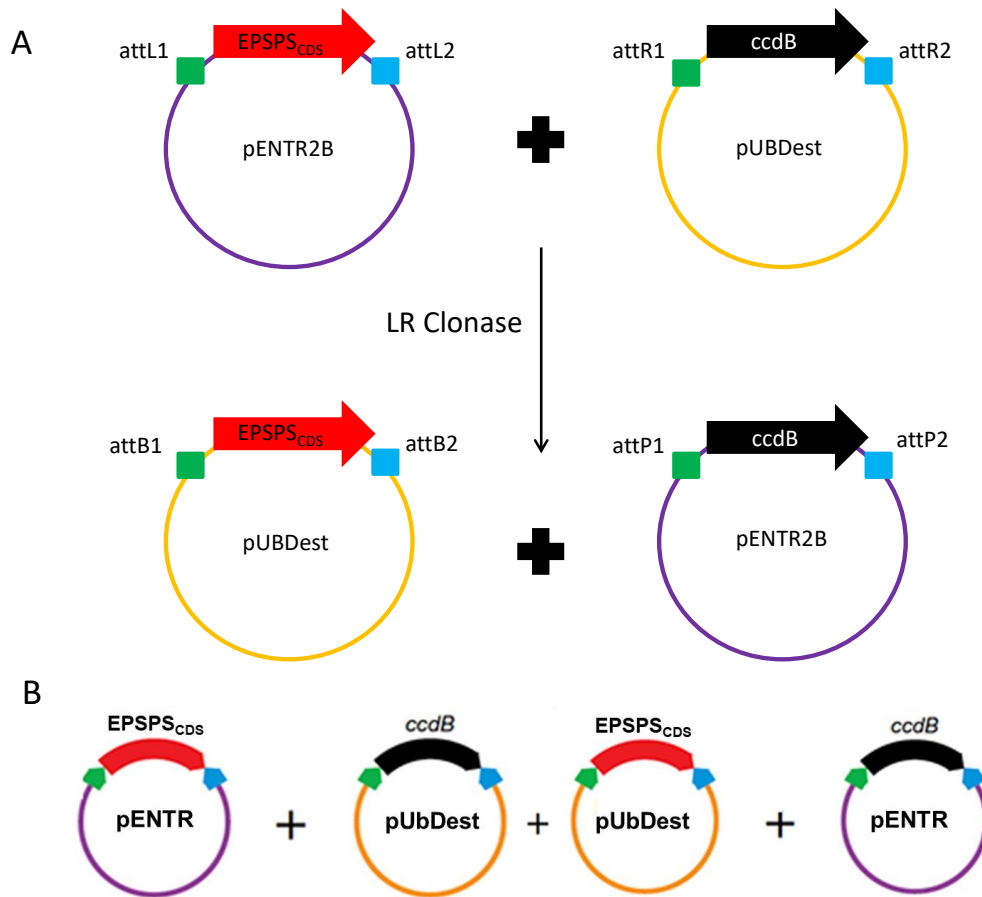


Figura 16. A) Mecanismo de clonado por *Gateway*. El sistema se basa en la presencia de un vector de entrada conteniendo una secuencia de interés flanqueada por sitios attL (1 y 2) que, mediante la enzima LR clonasa, serán recombinados con los sitios attR (1 y 2) del vector de destino, siendo transformados finalmente en sitios attB (vector final) y attP (producto saliente). La recombinación es un proceso unidireccional. El vector de destino posee un gen letal denominado *ccdB* que bloquea la actividad de una girasa y provoca la muerte de las bacterias transformadas que lo incorporen. **B)** Mezcla de recombinación una vez realizadas las reacciones LR. Se obtendrá una mezcla conteniendo pUB-Dest:EPSPS_{1CDS} y pENTR2B:ccdB, pero también podrían quedar pENTR2B:EPSPS_{1CDS} y pUB-Dest:ccdB. Las construcciones conteniendo *ccdB* producen la muerte de las bacterias. La construcción de interés (pUB-Dest:EPSPS_{1CDS}) se selecciona de las demás al conferir resistencia a espectinomicina.

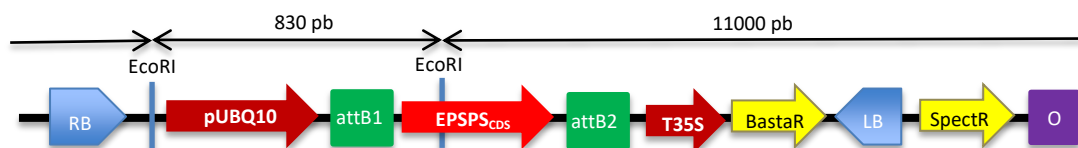
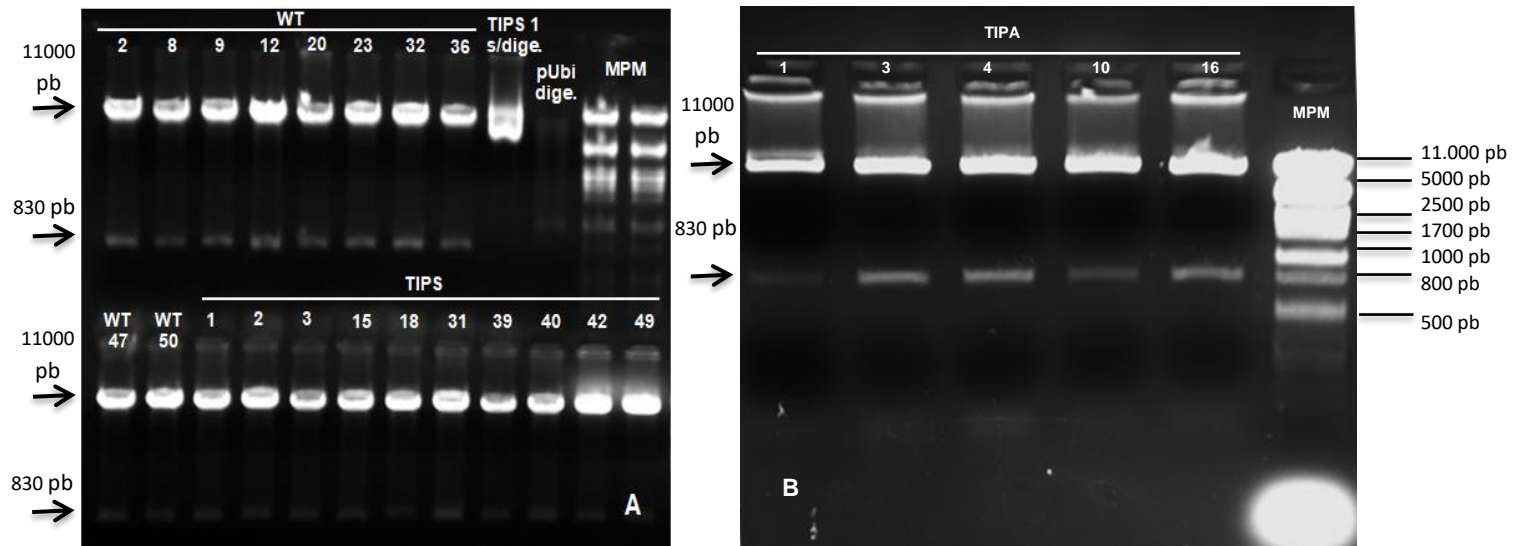


Figura 17. Esquema del vector binario pUB-Dest:EPSPS_{1CDS}. La secuencia codificante del gen EPSPS1 de soja queda bajo el control de un promotor pUBQ10 de *A. thaliana* y el terminador CaMV35S. Los bordes derecho (RB) e izquierdo (LB) permiten la transformación vegetal mediada por *A. tumefaciens*. Se muestran los sitios attB1 y attB2 obtenidos tras la recombinación LR, el gen SpectR (aminoglicosido

adenililtransferasa) de resistencia a espectinomicina en bacteria, el gen BastaR (fosfotricina acetiltransferasa) de resistencia a glufosinato de amonio en planta y el sitio de origen de replicación O.

En una primera instancia, la reacción de recombinación LR se incubó ON y posteriormente se transformaron células quimiocompetentes. Las células fueron plaqueadas en placas de Petri conteniendo LB-agar con el agregado de espectinomicina (ver punto 7). Debido a que no se obtuvieron colonias resistentes al medio selectivo, se optó por digerir el vector de entrada con la enzima de restricción *BsrBI* (Figura 9) para aumentar la eficiencia de recombinación. Se repitió la reacción en las mismas condiciones y se transformaron células quimiocompetentes (ver punto 7). En este caso sí hubo colonias.

Se realizaron minipreparaciones de plásmidos de las colonias obtenidas (ver punto 19) y se digirió el ADN plasmídico con *EcoRI* para verificar el clonado (Figura 16). Esta enzima presenta sitios de corte en la construcción, dentro de la CDS de la *EPSPS1* y en el promotor UBQ10, por lo que se esperaba obtener bandas de aproximadamente 11000 pb y 830 pb. Como se observa en la Figura 18, todos los clones analizados mostraron este patrón de bandas en electroforesis en geles de agarosa. Los clones WT 2, TIPS 2 y TIPA 1 fueron secuenciados para confirmar que contuvieran las variantes esperadas de la EPSPS (Figura 19).



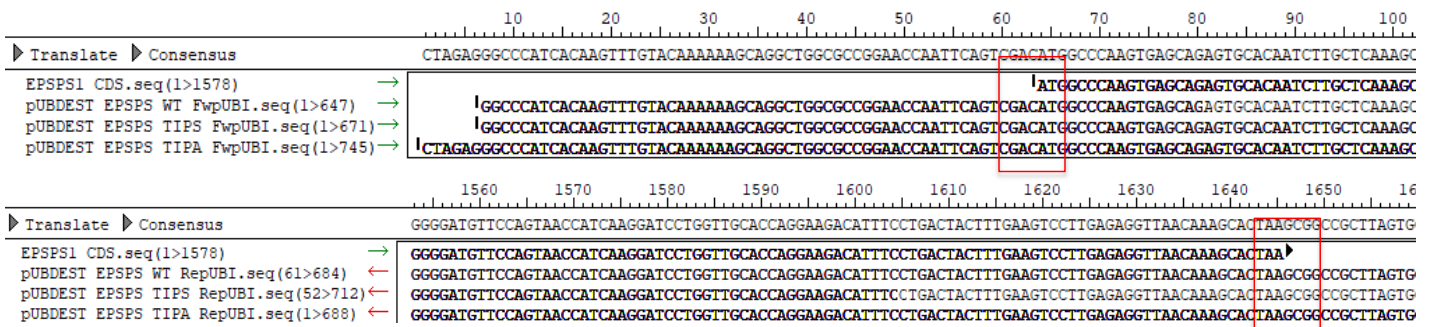


Figura 19. Resultados de secuenciación de los clones pUB-Dest:EPSPS1_{CDS} WT/TIPS/TIPA. Mediante el recurso SeqMan de Lasergene se realizó un contig alineando los resultados de secuenciación con la secuencia de referencia de la CDS del gen *EPSPS1*. En la imagen superior se muestra el comienzo 5'-ATG (superior) y en la imagen inferior el final 3'- TAA (inferior) de la CDS de EPSPS1. El resultado de secuenciación no presentó diferencias con la secuencia de referencia. Las diferentes versiones de la EPSPS clonadas en el vector p-UBDest no presentaron diferencias con la secuencia de referencia excepto las incorporadas por mutagénesis.

2. Transformación de *A. tumefaciens*

Los clones pUB-Dest:EPSPS1_{CDS} WT, TIPS y TIPA verificados por secuenciación fueron utilizados para transformar la cepa C58C1 de *A. tumefaciens*. Para ello se generaron células electrocompetentes (ver punto 8) y se transformaron las células con las construcciones de interés mediante electroporación (ver punto 9). Las células fueron seleccionadas en LB-agar con agregado espectinomicina 100 µg/mL y rifampicina 100 µg/mL. La resistencia a espectinomicina es conferida por el vector pUb-Dest mientras que la resistencia a rifampicina es conferida por una mutación cromosómica sobre el gen *rpoB* (subunidad β de la ARN polimerasa) en la cepa de *Agrobacterium*. Las colonias obtenidas fueron analizadas para la presencia del plásmido mediante PCR de colonia utilizando los *primers* Fw 305, 306 o 307 (Tabla 5), específicos para las variantes WT, TIPS o TIPA, respectivamente, en conjunto con un *primer* Rv (RvUbiseq) que hibrida en el terminador del vector (ver punto 10). Todas las colonias analizadas presentaron el plásmido transformado (Figura 20). Las colonias fueron almacenadas en glicerol a -80 °C hasta su uso.

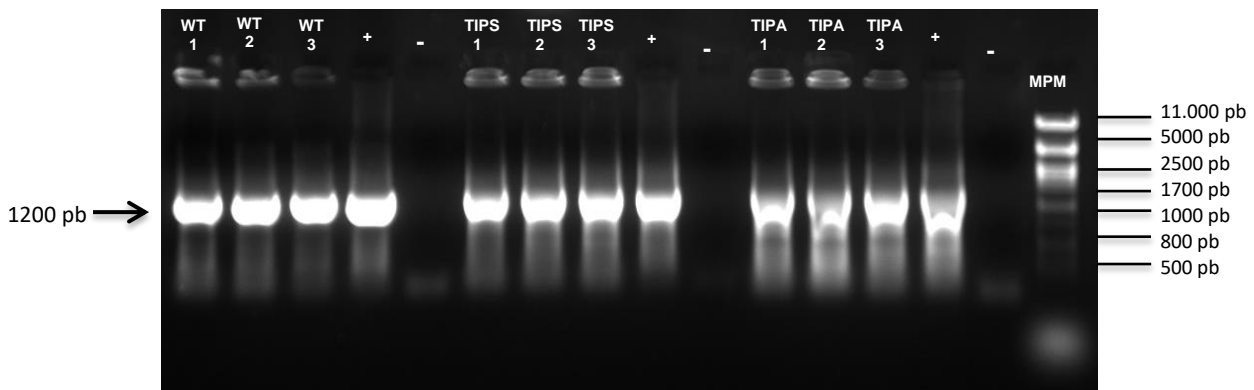


Figura 20. Electroforesis en gel de agarosa 1 % de amplificación por PCR para verificación de los clones pUB-Dest:EPSPS1_{CDS} WT/TIPS/TIPA en *A. tumefaciens*. Se observa amplificación en todos los carriles a los 1200 pb. MPM: marcador de peso molecular (λ_{PstI}).

3. Generación de plantas transgénicas de *A. thaliana*

Con el objetivo de evaluar la funcionalidad de las variantes WT y mutantes, de la enzima EPSPS1 de soja, se generaron plantas transgénicas de *Arabidopsis* que incorporaron las construcciones pUB-Dest:EPSPS1_{CDS} WT/TIPS/TIPA (ver punto 12). Para ello, se utilizó el método de inmersión floral de acuerdo al protocolo desarrollado por Clough y Bent (1998) y Martínez et al. (2012) (Figura 21).

Luego de la transformación, las semillas transformadas fueron seleccionadas en macetas con turba en presencia de glufosinato de amonio comercial (Basta Bayer) en una concentración de 30 mg/L, aplicado mediante pulverización cada dos días (Figura 21). Las plantas resistentes (generación T0), fueron trasplantadas a macetas individuales para generar semillas de la generación T1. Estas plantas serán analizadas en trabajos posteriores en cuanto a la presencia del transgen (PCR a partir de ADN genómico) y nivel de acumulación de transcripto para EPSPS mediante RT-PCR semicuantitativa. Utilizando plantas con grado de expresión similar, se evaluará la resistencia al herbicida glifosato.



Figura 21. A) Transformación de *A. thaliana* mediante inmersión floral utilizando las construcciones pUB-Dest:EPSPS1_{CDS} WT/TIPS/TIPA transformadas en *A. tumefaciens*. **B)** Germinación y selección de *A. thaliana* transgénica en turba con agregado de glufosinato de amonio 30 mg/L. Los círculos indican las plántulas resistentes a glufosinato de amonio.

Herramientas moleculares para edición del genoma de soja

Para promover el remplazo alélico en el genoma de una célula vegetal, se debe desencadenar un mecanismo de reparación del ADN vía HR. En la mayoría de los organismos eucariotas estudiados, la reparación vía HR en células no germinales se da en una frecuencia muy baja. Para estimularla, deben darse dos condiciones. Por un lado, la generación de un daño en el genoma, preferentemente cercano al sitio donde se quiere producir la modificación. Y por otro lado, la ocurrencia de un molde de ADN con homología de secuencia con la región donde ocurrió el daño. Como se mencionó en la Introducción, el sistema CRISPR/Cas9 permite generar un corte en el ADN en forma específica a través de la actividad de la nucleasa Cas9 guiada por sgARNs. En este sentido, en esta sección se describe la generación de los sgARNs, la producción de la proteína Cas9 recombinante y la generación de un vector de remplazo alélico. Estos componentes serán utilizados en el futuro para producir el remplazo de la copia WT del gen *EPSPS1* de soja por moldes portando las mutaciones TIPS o TIPA.

1. Generación de construcciones para síntesis de sgARN in vitro

Para generar modificaciones en el genoma de soja mediante edición génica es necesario producir sgARNs que tengan como blanco la región de interés del genoma. En este caso se escogió el exón 2 del gen *EPSPS1*. Los sgARNs están diseñados para cortar en el gen *EPSPS1*, muy cerca del sitio donde se quieren introducir las mutaciones TIPS y TIPA.

Los sgARNs se diseñaron utilizando el recurso CCTop-CRISPR/Cas9 *target online predictor* (<https://crispr.cos.uni-heidelberg.de/>). Este software permite obtener una predicción de posibles adaptadores y sus niveles de eficiencia para la secuencia ingresada, siendo en este caso la secuencia nucleotídica del gen *EPSPS1*. El programa permite seleccionar una serie de parámetros relevantes: la secuencia del PAM (*Protospacer Adjacent Motif*), la longitud de la región proximal al PAM (semilla), los *mismatches* tolerados y el tipo de promotor que se utilizará en el vector donde se desea clonar los adaptadores (promotor T7 para la síntesis de ARN *in vitro* o U6 para la generación de construcciones para la expresión en plantas). Este recurso también permite analizar los posibles *off-target* y los *mismatches* que presentan respecto a nuestra secuencia de interés. A priori, los *off-target* intrónicos son de menos importancia que los exónicos, ya que al tratarse de regiones no codificantes no se espera que

la modificación de estos sitios alteren el producto proteico final. Por otra parte, si no es posible evitar *off-target* en regiones exónicas, es conveniente que existan *mismatches*, preferentemente en la región proximal al PAM (semilla), ya que esta región es más relevante para la especificidad del sgARN. En el caso particular de este trabajo, se seleccionó el PAM de *Streptococcus pyogenes* (NGG). Por otra parte, no se optó por la opción ofrecida por el software de utilizar los promotores T7 o U6 para clonar los adaptadores en el vector debido a que los *overhangs* sugeridos no coinciden con los *overhangs* que dispone nuestro vector.

El vector utilizado para el clonado de los adaptadores fue el pEN-Chimera T7 Expression, el cual permite el clonado de un adaptador en posición 5' respecto al *scaffold* para el sgARN. El vector posee un promotor y un terminador T7, los cuales permiten la expresión de ARN *in vitro* utilizando una ARN polimerasa del fago T7. El pEN-Chimera posee un gen de resistencia a ampicilina (AmpR, β -lactamasa), antibiótico utilizado para seleccionar los transformantes (Figura 22.A). Es interesante notar que el clonado de los adaptadores en el vector pEN-Chimera depende del corte con la enzima de restricción *BbsI*. Esta enzima pertenece a la categoría de enzimas de restricción de tipo IIS, las cuales reconocen secuencias de ADN asimétricas y cortan a una distancia por fuera de su sitio de reconocimiento (uno a veinte nucleótidos; *Pingoud y Jeltsch, 2001*; Figura 22.B). Esta característica hace que estas enzimas sean herramientas muy versátiles para el clonado, ya que los *overhangs* que dejan dependen de la secuencia circundante. Además, una vez que la enzima reconoce su secuencia y produce el corte en el ADN, el sitio de reconocimiento para la enzima es eliminado por lo que ésta no puede volver a actuar. Esta particularidad es explotada en la metodología de clonado por Golden Gate, en la cual varios fragmentos pueden ser ensamblados direccionalmente en una única reacción de digestión/ligación (*Marillonnet & Werner, 2015*).

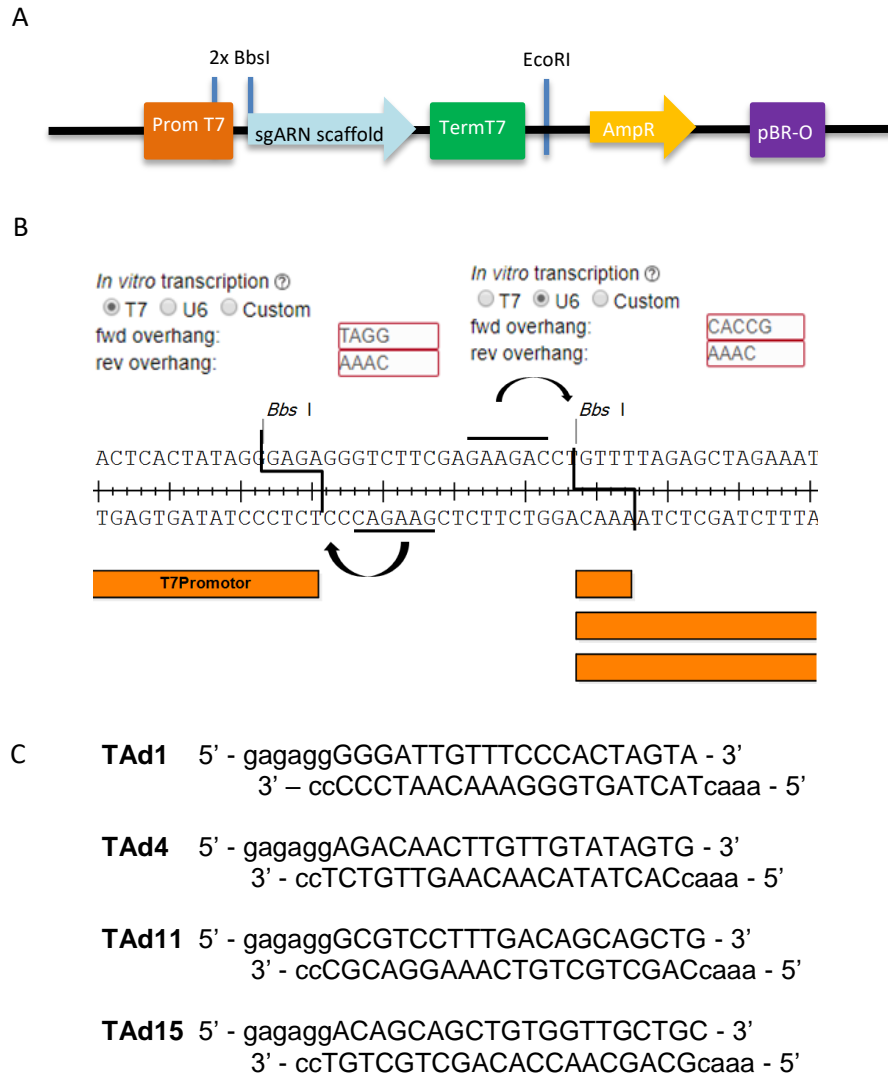


Figura 22. A) Esquema del vector pEn-ChimeraT7 Exp. Se muestra en rojo el *scaffold* del sgARN con los dos sitios *BbsI* que permiten el clonado del adaptador, en naranja el promotor y terminador T7, en amarillo el gen *AmpR* (β -lactamasa) de resistencia a ampicilina y en violeta el sitio de origen pBR. Se indica el sitio de restricción para la enzima *EcoRI*, utilizada para linealizar el vector previo a la síntesis del sgARN por IVT. **B)** Esquema de corte de la enzima de restricción de tipo IIS *BbsI* en el vector pEn-Chimera T7 Exp. Se observa que dicha endonucleasa genera un corte por fuera del sitio de reconocimiento. Los *overhangs* no coinciden con los de los promotores T7 o U6 sugeridos por el recurso CCTop. Es importante notar que una vez que el vector pEn-Chimera se digiere con la enzima *BbsI* los sitios de reconocimiento de la endonucleasa se eliminan, por lo que los clones que incorporan el adaptador no son digeridos nuevamente por *BbsI*. **C)** Adaptadores hibridados. Puede observarse que los *overhangs* generados al hibridar los oligonucleótidos coinciden con los dejados al digerir el vector pEn-Chimera con la enzima *BbsI*.

De los adaptadores sugeridos por el recurso CCTop se escogieron aquellos que presentaron un *score* de eficiencia medio-alto: TAd1, TAd4, TAd11, TAd15 (Figura 23). Para la selección también se tuvo en cuenta que aquellos que pudieran reconocer el parálogo de la *EPSPS1* en el cromosoma 3 (*EPSPS3*) tuvieran *mismatches* que impidieran o limitaran su reconocimiento. A las secuencias predichas por el programa CCTop se les agregaron *overhangs*

que coincidieran con los *overhangs* dejados en el vector tras el corte con la enzima *BbsI*. Esto es: Fw: 5'-GAGAGG + protoespaciador, y Rv: 5'- AAAC + protoespaciador + CC (Figura 22.C).

Se llevó a cabo la hibridación de los adaptadores y posterior clonado en el vector pEn-Chimera T7 Exp (ver punto 13). Se transformaron células quimiocompetentes de *E. coli* (ver punto 7) y se plaquearon en LB-agar con agregado de ampicilina.

De los cuatro adaptadores seleccionados, solo se obtuvo colonias para 3 de ellos. En el caso de TAd15 no se logró realizar la hibridación probablemente debido a la formación de horquillas internas en el primer T15EPSPSFw «TAggACAGGCAGCTGTGGTTGCTGC». A partir de las colonias obtenidas, se realizaron minipreparaciones de plásmido (ver punto 20) y se digirió el ADN plasmídico con la enzima *BbsI* para verificar el clonado (Figura 23). Como se mencionó anteriormente, una vez que el vector pEn-Chimera se digiere con la enzima *BbsI* los sitios de reconocimiento de la endonucleasa se eliminan, por lo que los clones que incorporan el adaptador no son digeridos nuevamente por *BbsI*. Como era esperable, en todos los casos se observa que los plásmidos no fueron digeridos (Figura 24). Se seleccionaron los clones TAd1.3, TAd4.1 y TAd11.1 para confirmar la presencia de los adaptadores por secuenciación. En la Figura 25 puede verse que en los tres casos el adaptador fue incorporado correctamente.

TAd1

Sequence: GGGATTGTTTCCCACTAGTAAGG

Efficacy score by CRISPRater: 0.61 MEDIUM

Coordinates	strand	MM	target_seq	PAM	distance	gene name	gene id
Chr01:46741625-46741647	-	0	GGGATTGT [TTCCCACTAGTA]	AGG	0	E Glyma.01G139600.1	Glyma.01G139600.Wm82.a2.v1
Chr04:6364682-6364704	-	3	GGGATAGT [TACCACTAGTA]	GGG	4114	- Glyma.04G076000.1	Glyma.04G076000.Wm82.a2.v1
Chr06:5969289-5969311	-	4	GGGATAT [TACCACTAGTA]	GGG	4728	- Glyma.06G077200.1	Glyma.06G077200.Wm82.a2.v1
Chr20:47122426-47122448	+	4	GGTCTTGT [TTCCCACTATTA]	GGG	0	E Glyma.20G239800.1	Glyma.20G239800.Wm82.a2.v1
Chr10:50874677-50874699	+	4	GGTTTGT [TTCCCACTATA]	GGG	0	E Glyma.10G290800.1	Glyma.10G290800.Wm82.a2.v1
Chr09:31714274-31714296	-	4	GGCATTGG [TTCCCACTAGAA]	AGG	0	E Glyma.09G127600.1	Glyma.09G127600.Wm82.a2.v1
Chr19:37392793-37392815	-	4	GGGATATT [TTCCCACTAGGA]	AGG	4587	- Glyma.19G117700.2	Glyma.19G117700.Wm82.a2.v1
Chr17:13606333-13606355	+	4	GGATTTT [TTCCCACTAGTT]	TGG	6321	- Glyma.17G158300.1	Glyma.17G158300.Wm82.a2.v1

TAd4

Sequence: AGACAACCTGTTGTATAGTGAGG

Efficacy score by CRISPRater: 0.62 MEDIUM

Coordinates	strand	MM	target_seq	PAM	distance	gene name	gene id
Chr01:46741739-46741761	-	0	AGACAACCT [TGTTGTATAGTG]	AGG	0	E Glyma.01G139600.1	Glyma.01G139600.Wm82.a2.v1
Chr14:45710540-45710562	+	4	ATAAACA [TTTGTATAGTG]	GGG	4573	- Glyma.14G192500.1	Glyma.14G192500.Wm82.a2.v1
Chr04:11846695-11846717	-	4	AGCCAAGT [TGCTTATAGTG]	AGG	15614	- Glyma.04G109700.2	Glyma.04G109700.Wm82.a2.v1
scaffold_276:13463-13485	+	4	AGCCAAGT [TGCTTATAGTG]	AGG	NA	- NA	NA
Chr14:24761436-24761458	+	4	AGAGACT [TGGTGTAAAGTG]	TGG	84256	- Glyma.14G137800.1	Glyma.14G137800.Wm82.a2.v1
Chr15:42697856-42697878	+	4	AGAGAGCT [TGTTGGAGAGTG]	AGG	0	E Glyma.15G228600.4	Glyma.15G228600.Wm82.a2.v1
Chr06:40965430-40965452	+	4	ATACCACT [TGTTGGACTGTG]	TGG	21515	- Glyma.06G244600.1	Glyma.06G244600.Wm82.a2.v1
Chr04:4346543-4346565	+	3	AGAGAACT [TGTTTATATG]	TGG	343	- Glyma.04G053300.3	Glyma.04G053300.Wm82.a2.v1
Chr16:4456071-4456093	-	4	TTACAACCT [TGTTTATATA]	GGG	2795	- Glyma.16G046900.1	Glyma.16G046900.Wm82.a2.v1
Chr20:39828160-39828182	+	4	CCACAACA [TGTTGTATAGTA]	GGG	1354	- Glyma.20G159800.1	Glyma.20G159800.Wm82.a2.v1
Chr03:39040762-39040784	+	4	AGAAAAT [TGTTGTACCCTG]	TGG	38	I Glyma.03G177600.2	Glyma.03G177600.Wm82.a2.v1
Chr11:15904283-15904305	+	4	AAACACT [TGTTGTACTGTG]	TGG	2468	I Glyma.11G165900.1	Glyma.11G165900.Wm82.a2.v1
Chr04:10157055-10157077	-	4	AGAGAGCT [TGTTGTAGATG]	GGG	27850	- Glyma.04G105400.1	Glyma.04G105400.Wm82.a2.v1
Chr05:29614138-29614160	+	4	AGAAAACA [TGTTGGATAGAG]	AGG	2556	- Glyma.05G111300.1	Glyma.05G111300.Wm82.a2.v1
Chr12:20746834-20746856	+	4	TGCCAACT [TGTTTATAGTT]	AGG	12408	- Glyma.12G148100.1	Glyma.12G148100.Wm82.a2.v1
Chr09:45116865-45116887	+	4	AGAGAAAT [TGTTGCATAGTT]	AGG	4239	- Glyma.09G226400.2	Glyma.09G226400.Wm82.a2.v1
Chr03:3942518-3942540	+	4	GGATAACT [TGTTGTATGGAG]	AGG	10200	- Glyma.03G033900.1	Glyma.03G033900.Wm82.a2.v1
Chr12:24610477-24610499	+	4	ATACCACT [TGTTGTATCGAG]	TGG	25282	- Glyma.12G155800.1	Glyma.12G155800.Wm82.a2.v1
Chr04:41000942-41000964	-	4	ATACAAT [TGTTGTAGTCT]	TGG	25167	- Glyma.04G164400.2	Glyma.04G164400.Wm82.a2.v1
Chr01:1048670-1048692	-	4	AGCCAAT [TGTTGTAAAGTT]	TGG	212	I Glyma.01G010900.1	Glyma.01G010900.Wm82.a2.v1

Sequence: GCGTCCTTTGACAGCAGCTGTGG

Efficacy score by CRISPRater: 0.59 MEDIUM

TAd11

Coordinates	strand	MM	target_seq	PAM	distance	gene name	gene id
Chr01:46741553-46741575	-	0	GCGTCCTT [TGACAGCAGCTG]	TGG	0	E Glyma.01G139600.1	Glyma.01G139600.Wm82.a2.v1
Chr09:6551242-6551264	+	4	GAGAAAT [TGACAGCAGCTG]	TGG	0	E Glyma.09G066700.3	Glyma.09G066700.Wm82.a2.v1
Chr05:36141463-36141485	-	4	TCTTCCTG [TGACAGCAACTG]	TGG	0	E Glyma.05G171000.1	Glyma.05G171000.Wm82.a2.v1
Chr08:9966066-9966088	-	4	TCTTCCTG [TGACAGCAACTG]	TGG	0	E Glyma.08G129400.1	Glyma.08G129400.Wm82.a2.v1

Sequence: ACAGCAGCTGTGGTGTGCTGCAGG

Efficacy score by CRISPRater: 0.76 HIGH

TAd15

Coordinates	strand	MM	target_seq	PAM	distance	gene name	gene id
Chr01:46741543-46741565	-	0	ACAGCAGC [TGTGGTTGCTGC]	AGG	0	E Glyma.01G139600.1	Glyma.01G139600.Wm82.a2.v1
Chr01:45575554-45575576	-	4	AATGTTGC [TGTGGTTGCTGC]	TGG	0	E Glyma.01G134400.1	Glyma.01G134400.Wm82.a2.v1
Chr03:3016823-3016845	-	1	ACAGCAGC [TGTAGTTGCTGC]	AGG	0	E Glyma.03G027400.1	Glyma.03G027400.Wm82.a2.v1
Chr12:3688830-3688852	-	4	GTTGCAGC [TGTGTTGCTGC]	TGG	0	E Glyma.12G051200.1	Glyma.12G051200.Wm82.a2.v1
Chr01:3316612-3316634	-	4	AGAGAAGC [TATGTTGCTGC]	TGG	11487	- Glyma.01G031800.1	Glyma.01G031800.Wm82.a2.v1
Chr01:39690833-39690855	-	4	ACAGAAATA [TGTAGTTGCTGC]	CGG	358	- Glyma.01G114900.1	Glyma.01G114900.Wm82.a2.v1
Chr05:35533554-35533576	-	4	GCTGCAGC [TGTGCTGCTGC]	TGG	1180	- Glyma.05G164500.2	Glyma.05G164500.Wm82.a2.v1
Chr19:664206-664228	-	4	ACAGGTGC [TGTAGCTGCTGC]	TGG	0	E Glyma.19G006600.3	Glyma.19G006600.Wm82.a2.v1
Chr16:30640729-30640751	+	4	ACATGAGC [TGTGTGCTGC]	CGG	4476	- Glyma.16G145500.1	Glyma.16G145500.Wm82.a2.v1
Chr19:489166-489188	+	4	TCAGCAC [TGTGTGCTGC]	CGG	2226	- Glyma.19G005300.1	Glyma.19G005300.Wm82.a2.v1
Chr16:7107873-7107895	+	4	CCACCAGC [TGTGTTACTGC]	TGG	3834	I Glyma.16G071000.5	Glyma.16G071000.Wm82.a2.v1
Chr18:55087842-55087864	+	4	ACTGCATC [TGCGGTTGTTGC]	GGG	420	- Glyma.18G266500.1	Glyma.18G266500.Wm82.a2.v1
Chr19:31166559-31166581	+	4	AGAGTAGC [TGTGTTGTTGC]	TGG	0	E Glyma.19G087600.4	Glyma.19G087600.Wm82.a2.v1
Chr09:4315470-4315492	+	4	ATAGCAGT [TGTGTTGTTGC]	TGG	0	E Glyma.09G049700.2	Glyma.09G049700.Wm82.a2.v1
Chr15:13112443-13112465	+	4	ATAGCAGT [TGTGTTGTTGC]	TGG	0	E Glyma.15G156400.1	Glyma.15G156400.Wm82.a2.v1
Chr03:1337640-1337662	-	4	ACTGCGGC [TGTGTTGCTGC]	TGG	0	E Glyma.03G013200.1	Glyma.03G013200.Wm82.a2.v1

Figura 23. Adaptadores sugeridos por el recurso CCTop (Heidelberg, s.f.) con sus respectivas secuencias y valores de eficiencia. Se muestran los sitios con homología *off-target* en diferentes cromosomas (con sus ID) de soja y sus *mismatches* (MM) señalados en rojo.

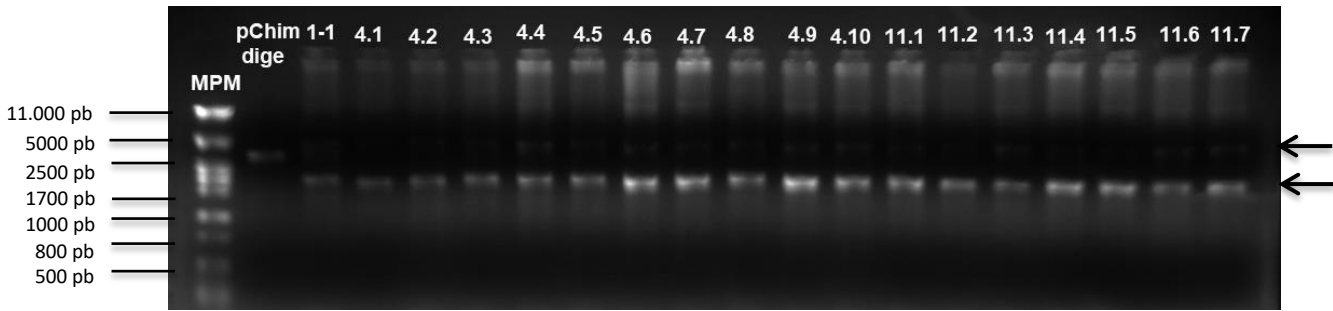


Figura 24. Electroforesis en gel de agarosa 1 % de digestión con *BbsI* de los clones pEn-ChimeraT7 Exp:TAd1/TAd4/TAd11. Se sembraron las digestiones del ADN plasmídico con *BbsI* de las colonias TA1, TA4 y TA11 junto con el vector pEn-Chimera T7 Exp digerido (pChim dige). Se observa por los patrones de banda que los clones conteniendo los adaptadores no fueron digeridos por la enzima ya que exhiben el típico patrón de migración de un plásmido cerrado. MPM: marcador de peso molecular (λ_{PstI}).

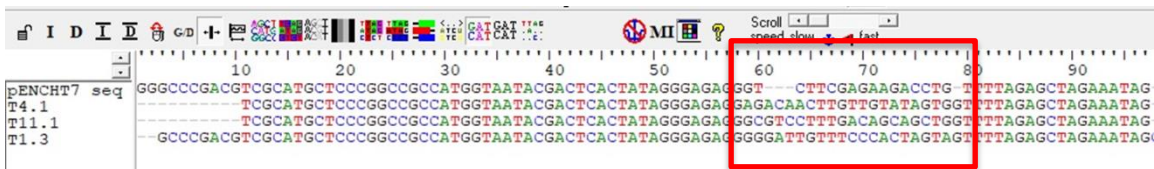


Figura 25. Resultados de secuenciación de los clones pEn-Chimera T7 Exp:TAD1/TAD4/TAD11. En el recuadro rojo se muestra las secuencias correspondientes a cada adaptador versus el vector sin modificación (pENCHT7 seq).

2. Expresión y purificación de Cas9

Para la obtención de la proteína Cas9 necesaria para los ensayos de validación *in vitro* de los sgARNs diseñados, se expresó y purificó la proteína recombinante a partir del vector pMJ922 (Figura 26). Dicho plásmido contiene el *cassette* His6-MBP-tagged Cas9-NLS-EGFP, permitiendo la expresión de la proteína SpCas9 presentando secuencias de localización nuclear (NLS) como fusión N-terminal a la proteína de unión a maltosa (MBP) junto a una cola de histidinas (HisTag) y una fusión C-terminal a GFP. La MBP está unida por un *linker* que contiene una secuencia de corte para la proteasa TEV. Por otro lado, la HisTag permite la purificación por cromatografía de afinidad por metales inmovilizados (IMAC), utilizando níquel en este caso. El vector presenta a su vez un gen de resistencia a ampicilina (AmpR, β -lactamasa) para seleccionar la presencia del plásmido en *E. coli*, y un promotor y terminador T7, que controlan la expresión de la proteína quimérica (Figura 26 y 27.A).

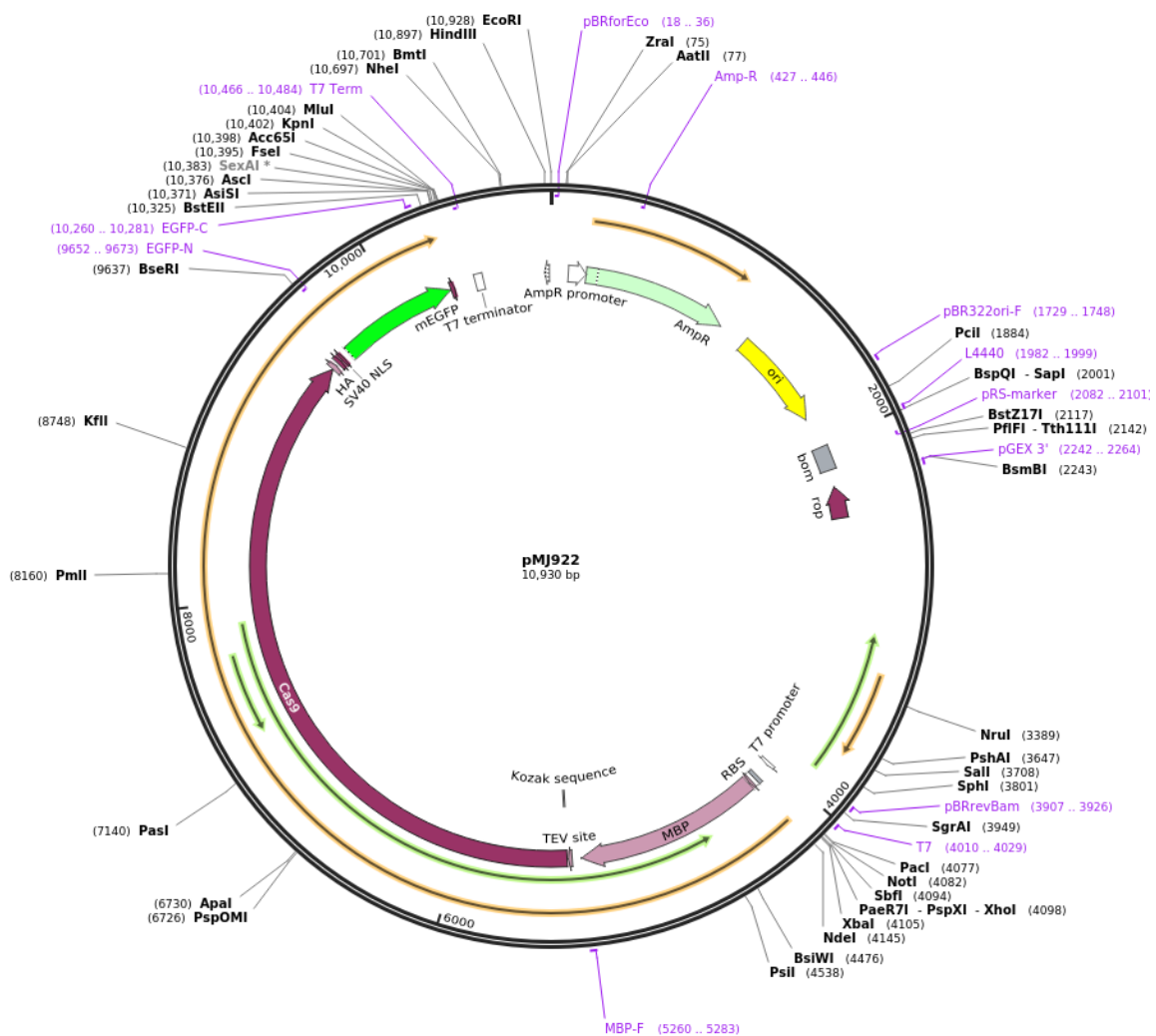


Figura 26. Esquema del vector PMJ922 obtenido del repositorio Addgene. Se muestra el cassette His6-MBP-tagged Cas9-NLS-EGFP, el gen de resistencia a ampicilina, promotor y terminador T7 y origen de replicación.

Se utilizó la cepa de *E. coli* Rosetta DE3, la cual es adecuada para la expresión de proteínas recombinantes ya que presenta una copia cromosómica de la T7 ARN polimerasa bajo el control de un promotor lacUV5. Es decir que la expresión de la T7 ARN polimerasa puede inducirse por el agregado de lactosa o de alguna molécula similar, como el isopropil- β -D-1-tiogalactopiranosido (IPTG). El IPTG tiene la ventaja de que no puede ser metabolizado por la bacteria, por lo que su concentración se mantiene constante durante el tiempo que dure la inducción (Wurm *et al.*, 2016). Por otra parte, la cepa Rosetta presenta un plásmido pRARE, que codifica para codones raros. Esto permite superar problemas de expresión que puedan

ocurrir debido al uso diferencial de codones entre el gen del organismo cuya proteína queremos expresar y *E. coli* (Carmignotto & Azzoni, 2019).

Se realizó una inducción con 200 μM IPTG durante 16 horas a 18 $^{\circ}\text{C}$ y 180 rpm. Para el proceso de lisis, las células se resuspendieron en buffer A IMAC y se agregaron MgCl_2 , DTT y benzonasa. Ya que la proteína Cas9 es una proteína de unión a ácidos nucleicos, el agregado de Mg^{2+} sirve para estabilizar estos dominios. Por otro lado, el DTT es utilizado como agente reductor para prevenir la oxidación de la proteína durante la purificación. Por último, la benzonasa es una nucleasa que degrada ADN y ARN y es fundamental para permitir una purificación correcta. Hemos observado que la Cas9 se une al ARN muy fuertemente y si éste no es eliminado durante el primer paso de purificación por cromatografía de afinidad por metales inmovilizados (IMAC), el paso siguiente de intercambio catiónico se ve alterado.

Se realizó un primer paso de purificación por IMAC (ver punto 15), realizando un lavado con buffer A IMAC muy extenso, de forma de eliminar lo máximo posible fragmentos de ácidos nucleicos y proteínas contaminantes. El segundo paso consistió en una cromatografía de intercambio iónico (IEX) utilizando una columna de intercambio catiónico. Finalmente, se utilizó la columna PD-10 para transferir la proteína hacia el buffer de almacenamiento (ver punto 15) (Figura 27.B).

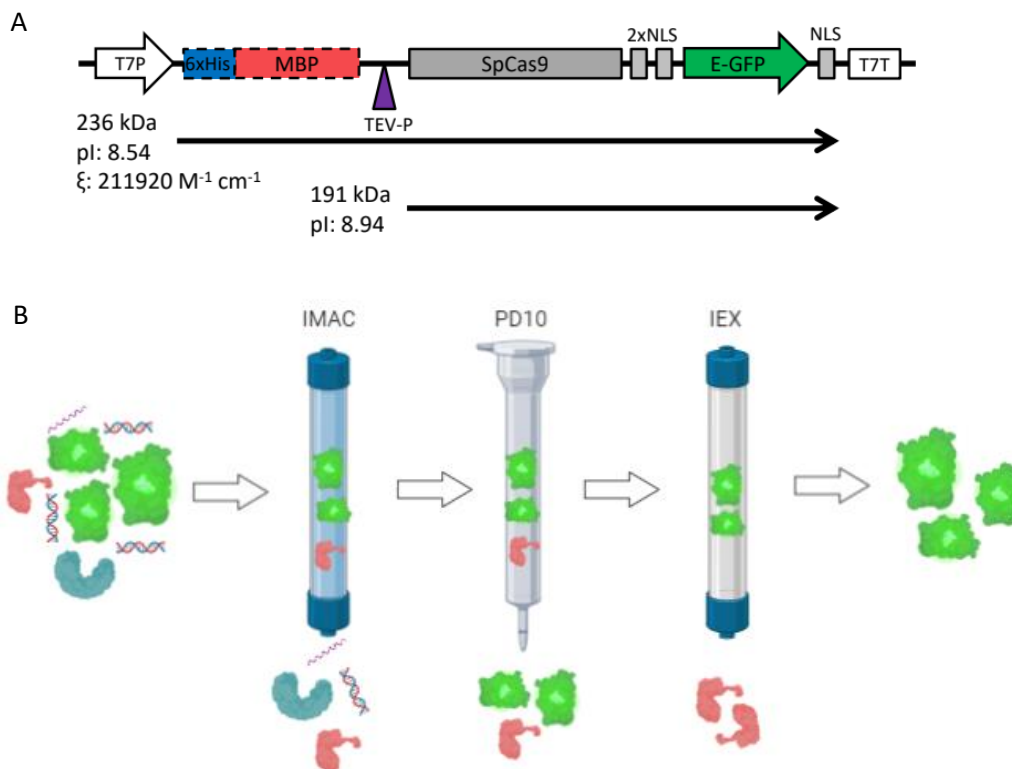


Figura 27. A) Esquema de la construcción His6-MBP-Cas9-NLS-EGFP en el vector PMJ922. La proteína SpCas9 conteniendo secuencias de localización nuclear (NLS) se expresa como fusión N-terminal a la proteína de unión a maltosa (MBP) junto a una cola de histidinas (6xHis) y una fusión C-terminal a EGFP. La MBP está unida por un *linker* que contiene una secuencia de corte para la proteasa TEV. La secuencia codificante de la proteína de fusión se encuentra flanqueada por el promotor y terminador para la T7 ARN polimerasa. Se muestra el peso de la proteína en KDa, el punto isoeléctrico para la proteína completa y carente de la porción MBP-6xHis y el coeficiente de absortividad molar predichos por el recurso ProtParam (<https://web.expasy.org/protparam/>). **B)** Esquema del proceso de purificación. Durante la IMAC, se une la proteína MBP-SpCas9-EGFP a la resina gracias a la afinidad de los residuos de histidina por el níquel inmovilizado. En los lavados, se eliminan ácidos nucleicos y proteínas contaminantes. En un segundo paso, se realiza un cambio de buffer utilizando una columna PD10. El cambio de buffer hacia buffer SP A se realiza porque la concentración de NaCl e imidazol en el buffer B IMAC no permitiría que la proteína se uniera a la resina de IEX. En un tercer paso, la proteína se carga en una columna de IEX. En particular, utilizamos una columna con grupos sulfopropil (SP), un intercambiador catiónico fuerte. Es importante notar que el pH del buffer SP A (7.5) es inferior al punto isoeléctrico de la proteína MBP-SpCas9-EGFP, por lo que la proteína estará cargada positivamente y podrá unirse a la resina en estas condiciones. La proteína se eluye de la resina de IEX utilizando una concentración de NaCl elevada (1 M). Finalmente, se realiza un cambio de buffer en columna PD10 para dejar la proteína en el buffer de almacenamiento.

La concentración de proteína se determinó por medida de absorbancia a 280 nm utilizando un coeficiente de absortividad molar predicho de $211920 \text{ M}^{-1} \text{ cm}^{-1}$, se congeló en N_2 líquido en alícuotas a una concentración de 1 mg/mL y se almacenó a -80°C .

Para verificar la pureza de la proteína y analizar el proceso de purificación se llevó a cabo una electroforesis desnaturizante en gel de poliacrilamida (SDS-PAGE). El dodecilsulfato de sodio (SDS) es un detergente aniónico que actúa desnaturizando las proteínas y cargándolas a todas con la misma densidad de carga negativa. Esto permite la migración en el gel de acuerdo al peso molecular. La incorporación de un agente reductor en el buffer de preparación de la muestra (β -mercapto etanol) también permite la reducción de enlaces disulfuro lo que garantiza la migración de las proteínas como monómeros (*Green & Sambrook, 2012*). En la figura 28 se puede observar que luego de estos dos pasos de purificación la proteína posee algunos contaminantes. No obstante, su grado de pureza es elevado (superior al 95 %). Quizás si esta proteína fuera a utilizarse para realizar ensayos de transformación de complejos ribonucleoproteicos *in vivo* sería conveniente agregar un paso adicional de *polishing* por filtración en gel para mejorar la pureza aún más. Otro punto a notar es que no se separó la porción 6xHis-MBP de la fusión por corte con proteasa TEV. Para otras aplicaciones, sería conveniente realizar un paso de corte con la proteasa de manera de trabajar con una proteína de tamaño sustancialmente menor (191 kDa en lugar de 236 kDa).

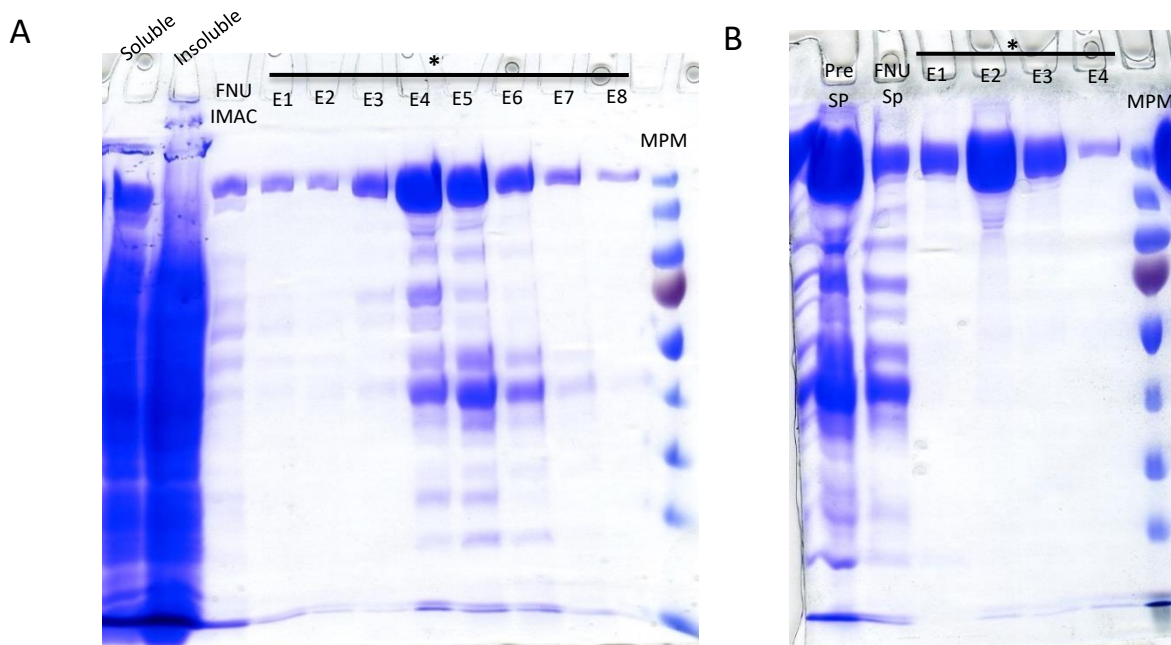


Figura 28. Electroforesis desnaturante en gel de poliacrilamida (SDS-PAGE) de la purificación de la proteína SpCas9. **A)** Se muestra la fracción soluble, fracción insoluble, fracción no unida de IMAC (FNU), diferentes fracciones del eluido del primer paso de purificación por IMAC (1, 2, 3, 4, 5, 6, 7, 8). **B)** Se muestra la fracción pre-Sp, FNU Sp, diferentes fracciones del eluido del segundo paso de purificación por IEX (1, 2, 3, 4). En la parte A y B se muestran las fracciones de interés (eluidos) con un asterisco. MPM: marcador de peso molecular (NEB, Cat. P77045).

3. Síntesis y validación in vitro de los sgARNs

Utilizando la proteína purificada, se validaron los sgARNs diseñados para editar el gen *EPSPS1* presente en el genoma de soja. Para ello, los sgARNs fueron sintetizados por IVT (ver punto 14) y se evaluó su eficacia de corte utilizando como blanco un fragmento de ADN genómico de la *EPSPS1* amplificado por PCR. También se analizó si los sgARNs son capaces de producir el corte en el gen parálogo del cromosoma 3, *EPSPS3*.

Para la síntesis de sgARN por transcripción *in vitro* se empleó el kit *TranscriptAid high yield* (Thermo K0441). Como molde para la síntesis se utilizaron las construcciones en el vector pEn-Chimera T7 Exp: TAd1, TAd4 o TAd11. Debido a la alta procesividad de la T7 ARN polimerasa empleada en el kit, los moldes circulares generan transcritos largos y heterogéneos (*Thermo Scientific, 2014*). Por este motivo, es importante que el molde de ADN para la reacción de IVT sea lineal. Asimismo, es recomendable que los extremos del molde sean romos o 5' protruyentes ya que se ha observado la generación de productos espurios cuando se utilizan moldes con extremos 3' protruyentes (*Thermo Scientific, 2014*). De manera de evitar estos problemas, se digirió el ADN plasmídico con la enzima de restricción *EcoRI* para

linealizar los vectores (Figura 29). Posteriormente, se llevó adelante la síntesis de los sgRNAs por IVT y los productos fueron purificados por columna (ver punto 14).

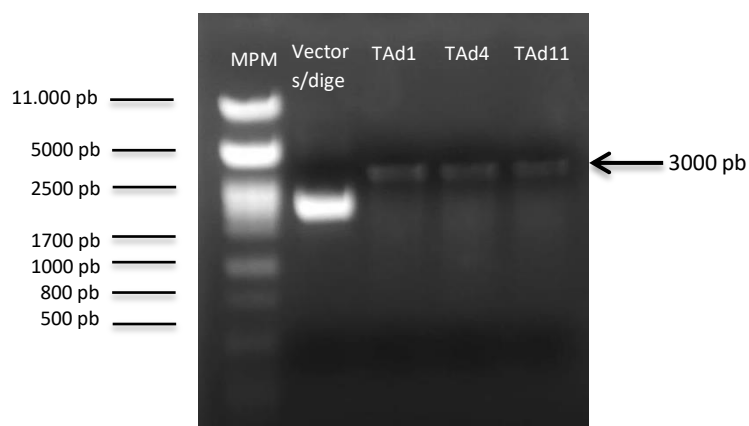
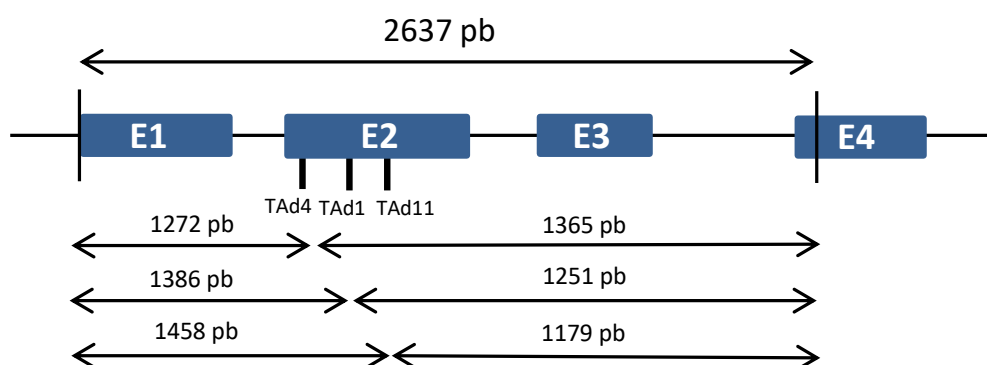


Figura 29. Electroforesis en gel de agarosa 1 % de digestión con *EcoRI* de los clones pEn-Chimera T7 Exp:TAd1/TAd4/TAd11. Se sembraron las digestiones del ADN plasmídico (TAd1, TAd4, TAd11) con *EcoRI* junto con el vector pEn-Chimera T7 Exp sin digerir (Vector s/dige). Se observa la migración del vector linealizado en las digestiones del pEn-Chimera Exp: TAd1/TAd4/TAd11 con un patrón de bandas de 3000 pb (aproximadamente) y 350 pb. MPM: λ_{PstI} .

Para el ensayo de validación *in vitro* de los sgARNs, los productos de PCR de una región genómica de los genes de la *EPSPS1* o la *EPSPS3* fueron incubados con complejos ribonucleoproteicos formados con la proteína Cas9 y los sgARNs sintetizados (sgARNT1, sgARNT4 y sgARNT11; ver punto 14). La reacción se incubó por 2 horas a 37 °C y se analizó por electroforesis (Figura 30). En la Figura R30.A se resumen los tamaños esperados para los cortes del producto de PCR de *EPSPS1*. Se observó que los tres sgARNs diseñados promueven el corte en el gen de la *EPSPS1* por la proteína Cas9 (carriles 4, 5 y 6, Figura 30.B), mientras que no se producen cortes en el producto de la *EPSPS3* (carriles 9, 10 y 11, Figura 29.B). Aparentemente, los sgARNs 1 y 4 tienen una mayor eficiencia que el sgARN11 (carriles 4, 5 y 6, Figura 30.B). Asimismo, puede observarse que el corte del producto de PCR es específicamente promovido por los sgARNs ya que la proteína en ausencia de sgARN no es capaz de producir el corte (carril 3, Figura 30.B). Para continuar con los trabajos posteriores se escogió el sgRNA1.

A



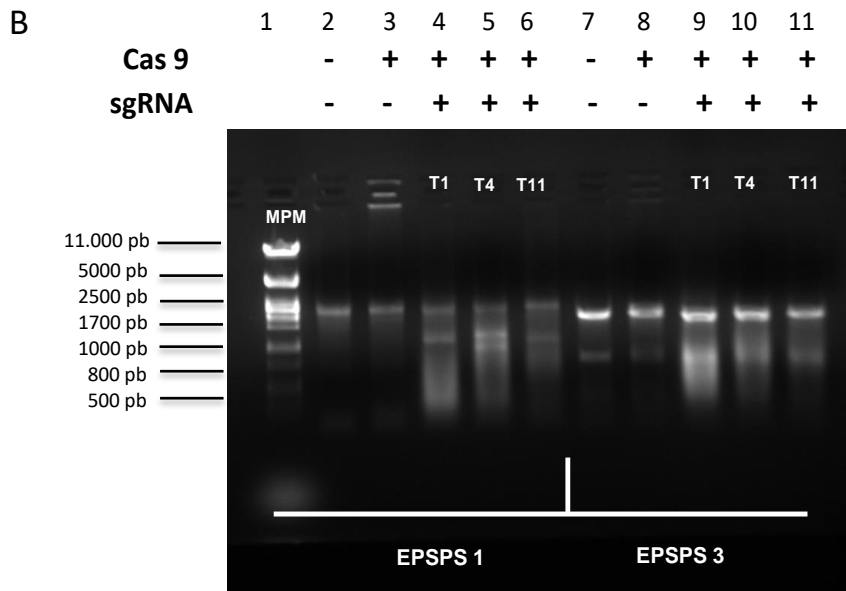


Figura 30. A) Esquema de producto PCR de ADN genómico de *EPSPS1*. Se muestran los sitios de corte para los sgARN1 (1386 pb), 4 (1272 pb) y 11 (1458 pb) en el exón 2 y los tamaños de las bandas esperados tras el corte. **B)** Electroforesis en gel de agarosa 1 % del ensayo de validación *in vitro* de los sgARNs. Se sembraron controles sin sgARNs ni Cas9 (carriles 2 y 7), sin sgARNs (carriles 3 y 8) y las reacciones correspondientes a las digestiones con los sgARNs T1 (carril 4 y 9), T4 (carril 5 y 10) y T11 (carril 6 y 11) para cada fragmento genómico (*EPSPS1* o *EPSPS3*). Se observa que en el gen *EPSPS1* se produce el corte, siendo más eficientes los sgARNs 1 y 4. En el caso del gen *EPSPS3* no se observó digestión, al igual que en los carriles control. MPM: λ_{PstI} .

4. Generación del vector de remplazo

Como se mencionó anteriormente, para promover el remplazo alélico es necesario proveer un fragmento de ADN con homología al sitio blanco para que, una vez que se genere el daño, el sistema de reparación celular pueda tomarlo como molde para la reparación, introduciendo la modificación deseada. Con este objetivo, se generó un vector de remplazo para ser utilizado durante los ensayos de remplazo alélico. El diseño de este vector se realizó teniendo en cuenta determinadas características:

1. Que contuviera un fragmento largo con homología de secuencia al gen de la *EPSPS1*.
2. Que contuviera las mutaciones que se desean introducir (TIPS o TIPA).
3. Que no contuviera la secuencia de reconocimiento compatible con los sgARNs diseñados, de forma que el complejo Cas9-sgARN no pueda actuar sobre el molde para la reparación, ni en el tubo de ensayo ni una vez que se produzca el remplazo alélico en la planta.

Para la producción de este vector, inicialmente se clonó un fragmento de ADN genómico de la *EPSPS1* amplificado por PCR en el vector pCRII-Blunt TOPO (ver punto 17). Los sitios de corte para los sgARNs diseñados se encuentran aproximadamente en medio de este fragmento genómico que tiene 2423 pb (1200 aproximadamente hacia cada lado). El vector pCRII-Blunt TOPO permite el clonado de productos de PCR con extremos romos dentro de la

secuencia de un gen LacZ- α lo que permite la selección blanco/azul. El clonado en este vector tiene la particularidad de que no es necesario utilizar T4 ligasa debido a la presencia de topoisomerasas en los extremos del vector. Asimismo, posee un gen letal *ccdB*, lo que impide la recuperación de clones que no hayan incorporado un inserto. El gen *nptII* confiere resistencia a kanamicina, antibiótico utilizado para seleccionar los transformantes (Figura 31).

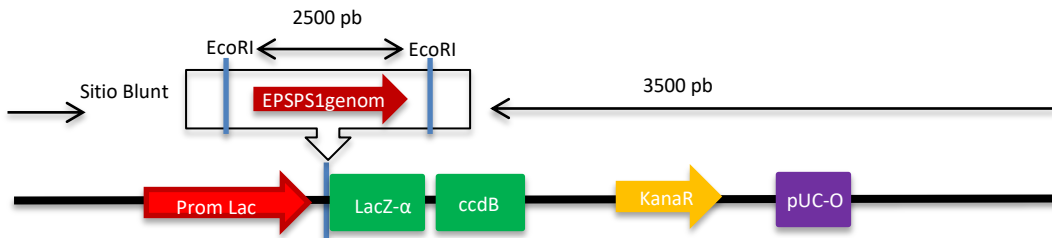


Figura 31. Esquema del vector pCRII:EPSPS1_{genom}. El vector permite el clonado de fragmentos romos dentro del gen LacZ- α . El gen letal *ccdB* garantiza la obtención de clones que hayan incorporado inserto. KanaR (NptII) permite la selección de los clones resistentes a kanamicina. pUC-O: sitio de origen de replicación.

Se realizaron minipreparaciones de plásmido a partir de las colonias blancas obtenidas y el ADN plasmídico se digirió con *EcoRI* para verificar el clonado. *EcoRI* corta en ambos extremos del sitio de clonado, por lo que se esperaban fragmentos de 3500 pb (correspondiente al vector lineal) y 2500 pb (correspondiente al inserto). Los clones que presentaron un patrón de digestión correcto (1, 2, 4 y 5 Figura 32) fueron confirmados por secuenciación (Figura 33).

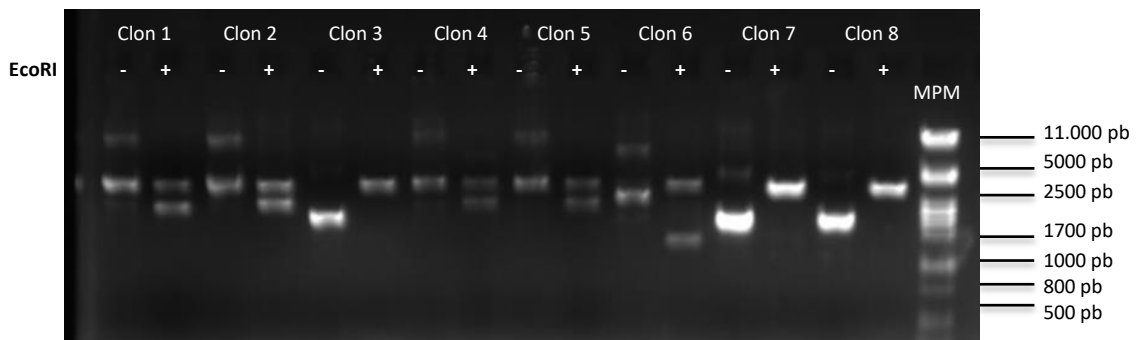


Figura 32. Electroforesis en gel de agarosa 1 % de la digestión con *EcoRI* del ADN plasmídico de los clones pCRII:EPSPS1_{genom}. Se sembraron muestras sin digerir (- *EcoRI*) y digeridas (+*EcoRI*). En el caso del clon 1, 2, 4 y 5 se observa el patrón de digestión esperado de 3500 pb y 2500 pb. MPM: marcador de peso molecular (λ_{PstI}).

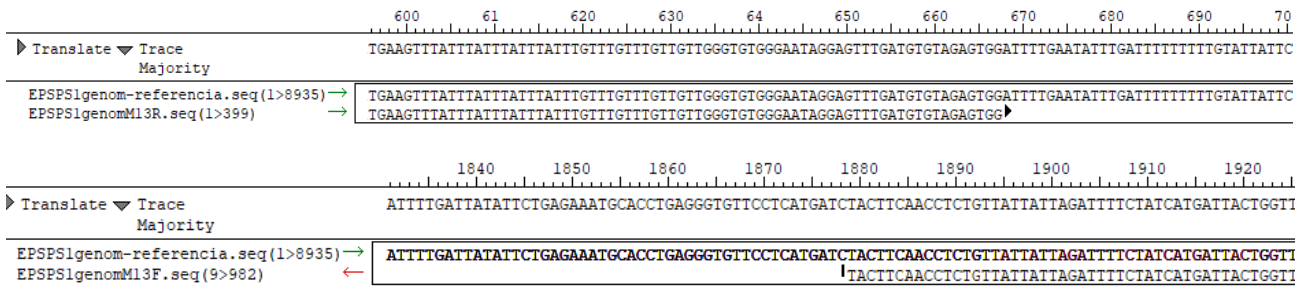


Figura 33. Resultados de secuenciación del clon pCRII:EPSPS1_{genom}. Mediante el recurso SeqMan de Lasergene se realizó un contig con las secuencias. En la imagen superior e inferior se muestra la coincidencia del fragmento genómico del gen *EPSPS1* respecto al fragmento secuenciado. El resultado de secuenciación no presentó diferencias con la secuencia de referencia.

Como se mencionó anteriormente, el objetivo de construir un vector de remplazo es que sirva como molde para reparar la región genómica específica mediante recombinación homóloga. Por este motivo, una vez obtenida la construcción pCRII:EPSPS1_{genom} se introdujeron por PCR las mutaciones de interés TIPS y TIPA mediante tecnología QuickChange, utilizando *primers* parcialmente complementarios (ver punto 18), de igual manera que se realizó para la generación de construcciones para sobreexpresión en *A. thaliana*. En la Figura 34 se muestra el resultado de la amplificación por PCR de los clones pCRII:EPSPS1_{genom}-TIPS y TIPA, observándose amplificación en el entorno de 4000 pb en los carriles positivos (TIPS+ y TIPA+). Se transformaron células quimiocompetentes de *E. coli* (ver punto 7) y se plaquearon en LB-agar con agregado de kanamicina como medio selectivo. Se realizaron minipreparaciones de plásmidos (ver punto 20) de varios clones y los mismos fueron secuenciados para confirmar las mutaciones (Figura 35). En ambos casos la mutación fue exitosa.

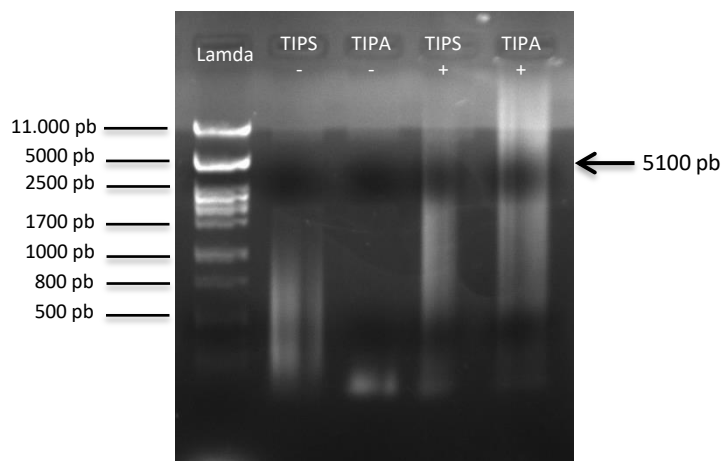


Figura 34. Electroforesis en gel de agarosa 1 % de amplificación por PCR de clones pCRII:EPSPS1_{genom}-TIPS (TIPS+) o TIPA (TIPA+) con *primers* parcialmente complementarios. Se sembraron controles negativos de TIPS y TIPA sin molde (-). Se observa amplificación en los carriles positivos a los 5100 pb. MPM: marcador de peso molecular (λ_{PstI})

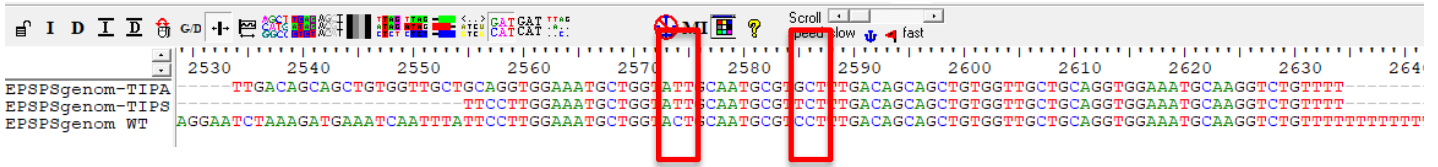


Figura 35. Resultados de secuenciación de los clones pCRII:EPSPS1_{genom}-TIPS (inferior)/TIPA (superior). En el recuadro rojo se muestra las secuencias correspondientes a la mutaciones TIPS y TIPA: ACT (Thr-102)→ ATT (Ile) y CCT (Pro-106) → TCT (Ser) / GCT (Ala).

Además de incorporar las mutaciones TIPS y TIPA, es importante que una vez que se produzca el remplazo, el sistema CRISPR/Cas9 no permanezca actuando sobre el genoma editado. Por este motivo, se introdujo una mutación silenciosa lisina-lisina (K-K) en la secuencia PAM asociada al sgARN1. Otro motivo para mutar la secuencia PAM es que si uno desea utilizar el molde de remplazo en combinación con complejos ribonucleoproteicos, los complejos cortarían el molde, por lo que el remplazo no ocurrirá. Tomando en cuenta esto, se realizó una mutagénesis utilizando como molde las construcciones pCRII:EPSPS1_{genom}-TIPS o pCRII:EPSPS1_{genom}-TIPA, junto con *primers* parcialmente complementarios que permiten la introducción de la mutación AAG (K87)→AAA (K) (ver punto 18). Se transformaron células quimiocompetentes de *E. coli* (ver punto 7) y se plaqueó en LB-Agar con agregado de kanamicina. Las construcciones pCRII:EPSPS1_{genom}-TIPS-KK y pCRII:EPSPS1_{genom}-TIPA-KK se evaluaron por electroforesis para verificar su tamaño (Figura 36). Los clones de tamaño correcto se enviaron a secuenciar para confirmar la mutación (Figura 37). En ambos casos se logró introducir la mutación deseada en el PAM.

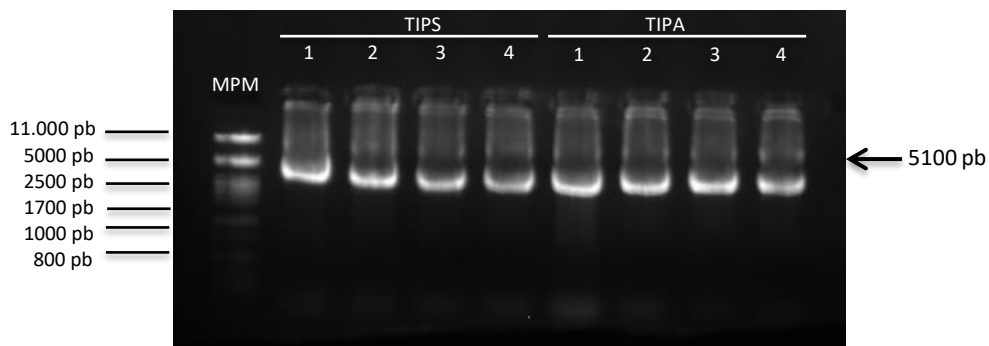


Figura 36. Electroforesis en gel de agarosa 1 % del ADN plasmídico de los clones pCRII:EPSPS1_{genom}-TIPS-KK y pCRII:EPSPS1_{genom}-TIPA-KK. Se sembró ADN plasmídico de TIPS 1, 2, 3, 4 y TIPA 1, 2, 3, 4. Se observa que los clones poseen el tamaño correcto (aproximadamente 5100 pb). MPM: marcador de peso molecular (λ_{PstI}).

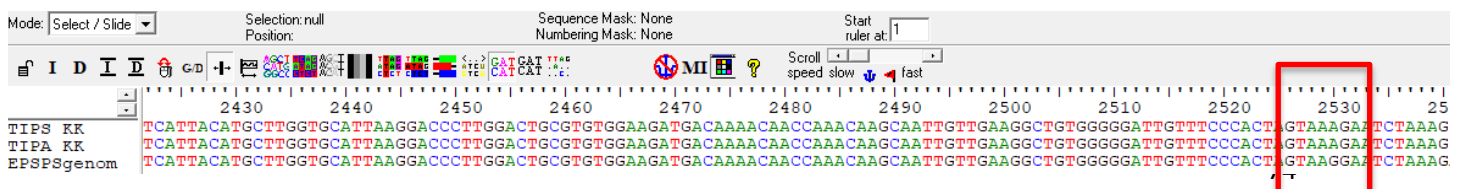


Figura 37. Resultados de secuenciación de los clones pCRII:EPSPS1_{genom}-TIPS-KK y TIPA-KK. En el recuadro rojo a los 2530 pb se muestran las secuencias correspondientes a la mutación K-K (AAG->AAA)

Existe evidencia que muestra que uno de los factores que aumenta significativamente la eficiencia de recombinación homóloga en plantas es la presencia de un fragmento de ADN lineal (Hanin & Paszkowski, 2003). Por este motivo, se introdujo por PCR la secuencia correspondiente al TA1+PAM en los bordes de la secuencia de homología del vector de remplazo (ver punto 18), de modo que al momento del remplazo alélico, los complejos ribonucleoproteicos formados por la proteína Cas9 y el sgARN1 reconozcan dicha secuencia y produzcan un corte, liberando un fragmento de ADN lineal que pueda ser utilizado para el remplazo alélico. Los *primers* utilizados para introducir esta secuencia hibridan con regiones intrónicas hacia su 3' y presentan en su región 5' la secuencia TA1+PAM (5'GGGATTGTTTCCCACTAGTAAGG-región intrónica-3'). Se decidió realizar la amplificación desde regiones intrónicas ante la posibilidad de que al ocurrir la reparación del ADN ocurra algún error en el proceso de recombinación que pueda afectar las regiones exónicas.

Se utilizó como moldes para esta amplificación las construcciones pCRII:EPSPS1_{genom}-TIPS-KK y pCRII:EPSPS1_{genom}-TIPA-KK. Los productos obtenidos fueron ligados en el vector pCRII-Blunt TOPO. Se transformaron células quimiocompetentes de *E. coli* (ver punto 7) y se plaquearon en LB-Agar en presencia de kanamicina y X-gal. En la Figura 38 se esquematiza la versión final del vector de remplazo.

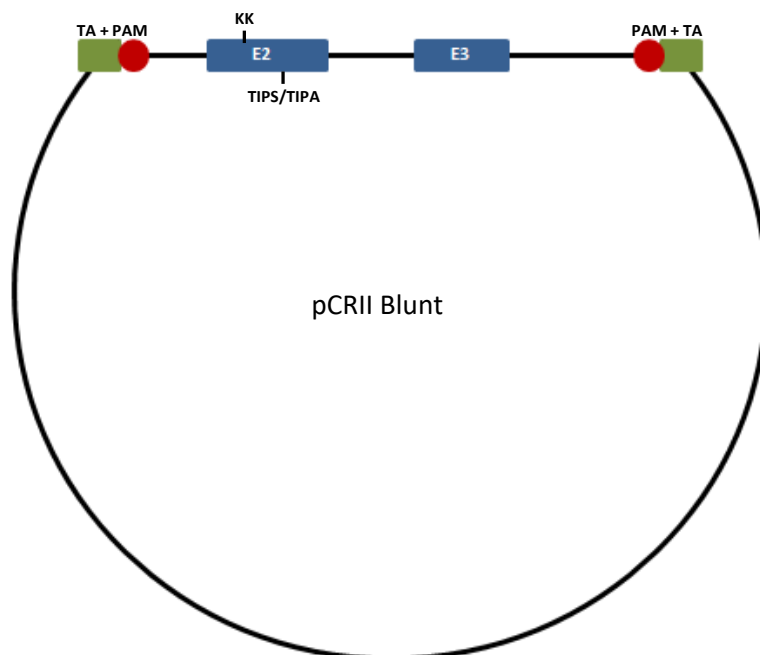


Figura 38. Esquema de construcción final del vector de remplazo (pCRII:EPSPS1_{genom}-TIPS-KK-TA1+PAM o pCRII:EPSPS1_{genom}-TIPA-KK-TA1+PAM). Se muestran los sitios con las mutaciones TIPS o TIPA y K-K y las secuencias TA1+PAM flanqueando el fragmento genómico de EPSPS1. En el momento del remplazo alélico la proteína Cas9 y sgARN reconocerán la secuencia TA1+PAM, produciendo un corte y liberando un fragmento de ADN lineal que luego podrá recombinarse en el genoma de soja, introduciendo las mutaciones de interés TIPS y TIPA.

Conclusiones y perspectivas

En este trabajo se desarrollaron una serie de herramientas moleculares con el objetivo de optimizar metodologías de edición génica en soja basadas en HR. Las metodologías de edición génica pueden utilizarse para modificar el genoma de diversas formas. Los procedimientos más sencillos implican la generación de mutantes *knock-out* (KO) ya que en estos casos solamente es necesario generar una lesión en el ADN y luego permitir que la maquinaria de reparación repare el daño; al utilizar mecanismos propensos a errores para hacerlo, es relativamente fácil generar estos mutantes. Otros procedimientos son más complejos debido a que implican direccionar de alguna manera la reparación. Entre éstos encontramos el remplazo alélico o el *gene-targeting*, esto es, la inserción de un fragmento de ADN en un sitio específico, haciendo uso de los mecanismos de reparación del ADN por HR. Estos procedimientos tienen un grado de complejidad mayor a los que buscan efectuar mutaciones al azar en sitios específicos, y por lo tanto, al momento, el principal uso que se le ha dado a las metodologías de edición génica en soja es la generación de plantas KO para distintos genes (*Jacobs et al., 2015*).

En este trabajo se han generado herramientas, que permitirán optimizar una metodología de remplazo alélico en soja. Se han generado versiones mutadas de la EPSPS1 de soja y se han transformado en *A. thaliana*. A la brevedad, se analizará si estas mutaciones logran producir variantes de la EPSPS1 insensibles al glifosato. Teniendo en cuenta el grado de conservación de estas proteínas, especialmente a nivel del sitio de unión del PEP, es esperable que así sea. Otro factor prometedor es que todas las versiones mutantes insensibles al glifosato originadas naturalmente cuentan con una mutación de la Pro106 (*Sammons & Gaines, 2014*).

A su vez, en este trabajo se generaron un conjunto de construcciones génicas que servirán como herramientas para modificar el genoma de soja. Específicamente, se diseñaron y evaluaron sgARNs para el reconocimiento de la secuencia de EPSPS1. De los sgARNs analizados, se seleccionó el sgARN1 para continuar con la estrategia ya que corta el gen de la EPSPS1 de manera eficiente y cerca del sitio donde se quieren introducir las mutaciones TIPS o TIPA, en tanto que no corta el gen parálogo EPSPS3. También se generaron vectores de remplazo que podrán utilizarse para producir el remplazo alélico del gen EPSPS1 endógeno, por la versión mutada de esta enzima. Este vector podría utilizarse como ADN circular o bien podría amplificarse por PCR exclusivamente la región de homología (Figura 39).

Las estrategias para realizar edición génica son varias. Para soja, pueden pensarse una serie de opciones (Figura 39). La Cas9 y los sgARNs pueden entregarse como construcciones,

como RNPs o como ARNs. Estos componentes pueden a su vez entregarse sobre embriones somáticos o sobre otros tejidos, por ejemplo, protoplastos. En la actualidad, el laboratorio de Biología Molecular Vegetal tiene la capacidad de transformar embriones somáticos con construcciones mediante biolística. También se ha logrado bombardear proteína sobre embriones somáticos de manera exitosa. Por otra parte, se ha logrado generar protoplastos de soja, aunque sólo se ha avanzado en etapas tempranas de la regeneración. Asimismo, se ha optimizado el protocolo para la obtención de proteína Cas9 activa, en abundancia y con alto grado de pureza. Sumado a esto, se ha generado una construcción que puede ser utilizada como molde para la síntesis *in vitro* del Cas9-mARN optimizado para su expresión en leguminosas (Fleitas, Gallino, sin publicar). Recientemente también se ha optimizado un protocolo de transformación del ovario de la flor de soja (Gallino, sin publicar) y se han obtenido plantas editadas generadas por esta metodología utilizando construcciones de ADN (Fleitas, sin publicar). Esta es una metodología promisoriosa para ser utilizada para la edición génica utilizando RNPs o ARNs.

Utilizando todas las herramientas disponibles, el trabajo futuro apuntará a dilucidar cuál es la forma más eficiente de entregar la proteína Cas9 y los sgARNs (construcciones, RNPs, ARNs) y qué tejido es más adecuado para realizar el remplazo alélico en soja (embriones somáticos, ovario de la flor, protoplastos). También existen alternativas en cuanto a qué molde utilizar para promover la reparación por recombinación homóloga. En este caso, hemos generado un vector de remplazo que podría utilizarse también como fragmento de doble hebra lineal. Alternativamente, en otros organismos se han utilizado moldes de reparación de simple hebra largos (lsADN) o cortos (ssADN) (Quadros *et al.*, 2017). Otra estrategia utilizada comúnmente en plantas para realizar remplazo alélico o *gene-targeting* es el uso de vectores virales basados en geminivirus (Figura 39, Čermák *et al.*, 2015); esta es una estrategia de entrega que no podemos descartar. Finalmente, en otros organismos se han analizado algunas estrategias que permiten potenciar la recombinación homóloga, como el uso de moléculas que inhiben la ligasa IV (SCR7) o el silenciamiento de las moléculas clave en la vía NHEJ (KU70, KU80 ligasa IV; Chu *et al.*, 2015). Estas posibilidades serán analizadas durante mi trabajo de maestría.

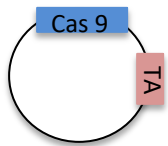

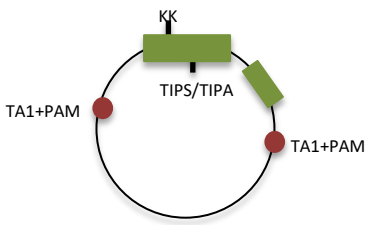
HERRAMIENTAS EDICIÓN GÉNICA		
	Pueden utilizarse en diversas modalidades	
Adaptadores	Como construcción 	<ul style="list-style-type: none"> • Generados por IVT • Comerciales: sgARN o gARN+CrARN
Cas 9		<ul style="list-style-type: none"> • mRNA generado por IVT • Proteína
PUEDE COMBINARSE CON		
Molde de remplazo	dsADN	
	Vector	
	Otras opciones: ssADN, lsADN, etc.	
PUEDE TRANSFORMARSE EN		
Embriones somáticos	Mediante biolística	
Flores	Mediante ovary drip (ovario en flor)	
Protoplastos	Mediante PEG/Heatshock	

Figura 39. El cuadro muestra las herramientas disponibles para realizar edición génica generadas por nuestro laboratorio y otras soluciones comerciales, las posibilidades manejadas para realizar la entrega del ADN con homología para la reparación y las metodologías disponibles en nuestro laboratorio para realizar la entrega de estos componentes en soja.

Bibliografía

- Aida, T., Chiyo, K., Usami, T., Ishikubo, H., Imahashi, R., Wada, Y., y otros. (2015). Cloning-free CRISPR/Cas system facilitates functional cassette knock-in in mice. *Genome Biology*, 16(1), doi: 10.1186/s13059-015-0653-x.
- Amin, N., Ahmad, N., Wu, N., Pu, X., Ma, T., Du, Y., y otros. (2019). CRISPR-Cas9 mediated targeted disruption of FAD2-2 microsomal omega-6 desaturase in soybean (*Glycine max.*L). *BMC Biotechnology*, 19(9), doi: 10.1186/s12896-019-0501-2.
- Baltes, N., Gil-Humanes, J., Cermak, T., Atkins, P., & Voytas, D. (2014). DNA replicons for plant genome engineering. *The Plant Cell*, 26(1), 151-163.
- Bao, A., Burritt, D., Chen, H., Zhou, X., Cao, D., & Tran, L. (2019). The CRISPR/Cas9 system and its applications in crop genome editing. *Critical Reviews in Biotechnology*, 39(3), 321-336.
- Bao, A., Chen, H., Chen, L., Chen, S., Hao, Q., Guo, W., y otros. (2019). CRISPR/Cas9-mediated targeted mutagenesis of GmSPL9 genes alters plant architecture in soybean. *BMC Biotechnology*, 19(131), doi: 10.1186/s12870-019-1746-6.
- Bao, A., Tran, L., & Cao, D. (2020). CRISPR/Cas9-Based Gene Editing in Soybean. *Methods in Molecular Biology*, 2107, 349-364.
- Bennardo, N., Cheng, A., Huang, N., & Stark, J. (2008). Alternative-NHEJ Is a Mechanistically Distinct Pathway of Mammalian Chromosome Break Repair. *PLOS Genetics*, 4(6), doi: 10.1371/journal.pgen.1000110.
- Bitinaite, J., Wah, D., Aggarwal, A., & Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences*, 95(18), 10570–10575.
- Bonawitz, N., Ainley, W., Itaya, A., Chennareddy, S., Cicak, T., Effinger, K., y otros. (2019). Zinc finger nuclease-mediated targeting of multiple transgenes to an endogenous soybean genomic locus via non-homologous end joining. *Plant Biotechnology Journal*, 17(4), 750-761.
- Carmignotto, G., & Azzoni, A. (2019). On the expression of recombinant Cas9 protein in *E. coli* BL21(DE3) and BL21(DE3) Rosetta strains. *Journal of Biotechnology*, 306, 62-70.
- Čermák, T., Baltes, N., Čegan, R., Zhang, Y., & Voytas, D. (2015). High-frequency, precise modification of the tomato genome. *Genome Biology*, 16(232), doi: 10.1186/s13059-015-0796-9.
- Chen, B., Gilbert, L., Cimini, B., Schnitzbauer, J., Zhang, W., Li, G., y otros. (2013). Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell*, 155(7), 1479-1491.
- Chen, M., & Qi, L. (2017). Repurposing CRISPR System for Transcriptional Activation. *Advances in Experimental Medicine and Biology*, 983, 147-157.

- Cheng, Q., Dong, L., Su, T., Li, T., Gan, Z., Nan, H., y otros. (2019). CRISPR/Cas9-mediated targeted mutagenesis of GmLHY genes alters plant height and internode length in soybean. *BMC Biotechnology*, *19*(1), doi: 10.1186/s12870-019-2145-8.
- Chu, V., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., y otros. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology*, *33*(5), 543-548.
- Clough, S., & Bent, A. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal*, *16*(6), 735-743.
- Curtin, J., Xiong, Y., Michno, J., Campbell, B., Stec, A., Čermák, T., y otros. (2018). CRISPR/Cas9 and TALENs generate heritable mutations for genes involved in small RNA processing of *Glycine max* and *Medicago truncatula*. *Plant Biotechnology Journal*, *16*(6), 1125-1137.
- Curtin, S., Xiong, Y., Michno, J., Campbell, B., Stec, A., Čermák, T., y otros. (2018). CRISPR/Cas9 and TALENs generate heritable mutations for genes involved in small RNA processing of *Glycine max* and *Medicago truncatula*. *Plant Biotechnology Journal*, *16*(6), 1125-1137.
- Curtin, Z. F., Sander, J., Haun, W., Starker, C., Baltes, N., Reyon, D., y otros. (2011). Targeted Mutagenesis of Duplicated Genes in Soybean with Zinc-Finger Nucleases. *Plant Physiology*, *156*(2), 466-473.
- Du, H., Zeng, X., Zhao, M., Cui, X., Wang, Q., Yang, H., y otros. (2016). Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. *Journal Biotechnology*, *217*, 90-97.
- Edelheit, O., Hanukoglu, A., & Hanukoglu, I. (2009). Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnology*, *9*(61), doi: 10.1186/1472-6750-9-61.
- Elison, G., Song, R., & Acar, M. (2017). A precise genome editing method reveals insights into the activity of eukaryotic promoters. *Cell Reports*, *18*(1), 275-286.
- Endo, M., Mikami, M., & Toki, S. (2016). Biallelic Gene Targeting in Rice. *Plant Physiology*, *170*(2), 667-677.
- Fausser, F. S. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *The Plant Journal*, *79*(2), 348-359.
- Finer, J. (2016). Generation of Transgenic Soybean (*Glycine max*) via Particle Bombardment of Embryogenic Cultures. *Current Protocols*, *1*(4), 592-603.
- Firdous, S., Iqbal, S., Anwar, S., & Jabeen, H. (2018). Identification and analysis of 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene from glyphosate-resistant *Ochrobactrum intermedium* Sq20. *Pest Management Science*, *74*(5), 1184-1196.

- Fisher, T. (s.f.). *Thermo Fisher Scientific*. Recuperado el Enero de 2020, de <https://www.thermofisher.com/uy/en/home/life-science/cloning/gateway-cloning.html>
- Gaj, T., Sirk, S., Shui, S., & Liu, J. (2016). Genome-Editing Technologies: Principles and Applications. *Cold Spring Harbor perspectives in biology*, 8(12), doi: 10.1101/cshperspect.a023754.
- Garg, B., Vaid, N., & Tuteja, N. (2014). In-silico analysis and expression profiling implicate diverse role of EPSPS family genes in regulating developmental and metabolic processes. *BMC*, 7(58), doi: 10.1186/1756-0500-7-58.
- Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C., Sánchez-León, S., y otros. (2017). High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *The Plant Journal*, 89(6), 1251-1262.
- Green, M., & Sambrook, J. (2012). En M. Green, & J. Sambrook, *Molecular Cloning: a laboratory manual I*. New York: Cold Spring Harbor Laboratory Press .
- Grefen, C., Donald, N., Hashimoto, K., Kudla, J., Schumacher, K., & Blatt, M. (2010). A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. *The Plant Journal*, 64(2), 355-365.
- Hanin, M., & Paszkowski, J. (2003). Plant genome modification by homologous recombination. *Current Opinion in Plant Biology*, 6(2), 157-162.
- Haun, W., Coffman, A., Clasen, B., Demorest, Z., Lowy, A., Ray, E., y otros. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnology Journal*, 12(7), 934-940.
- Healy-Fried, M., Funke, T., Priestman, M., Han, H., & Schönbrunn, E. (2007). Structural Basis of Glyphosate Tolerance Resulting from Mutations of Pro101 in Escherichia coli 5-Enolpyruvylshikimate-3-phosphate Synthase. *The Journal of Biological Chemistry*, 282(45), 32949-32955.
- Heidelberg, C. f. (s.f.). *CCTop*. Recuperado el 2019, de <https://crispr.cos.uni-heidelberg.de/>
- Hsu, P., Lander, E., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6), 1262-1278.
- Igoucheva, O., Alexeev, V., & Yoon, K. (2001). Targeted gene correction by small single-stranded oligonucleotides in mammalian cells. *Gene Therapy*, 8(5), 391-399.
- Jacobs, T., LaFayette, P., Schmitz, R., & Parrott, W. (2015). Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnology*, 15(16), doi: 10.1186/s12896-015-0131-2.
- Jurica, M., & Stoddard, B. (1999). Homing endonucleases: structure, function and evolution. *Cellular and Molecular Life Sciences*, 55(10), 1304-1326.

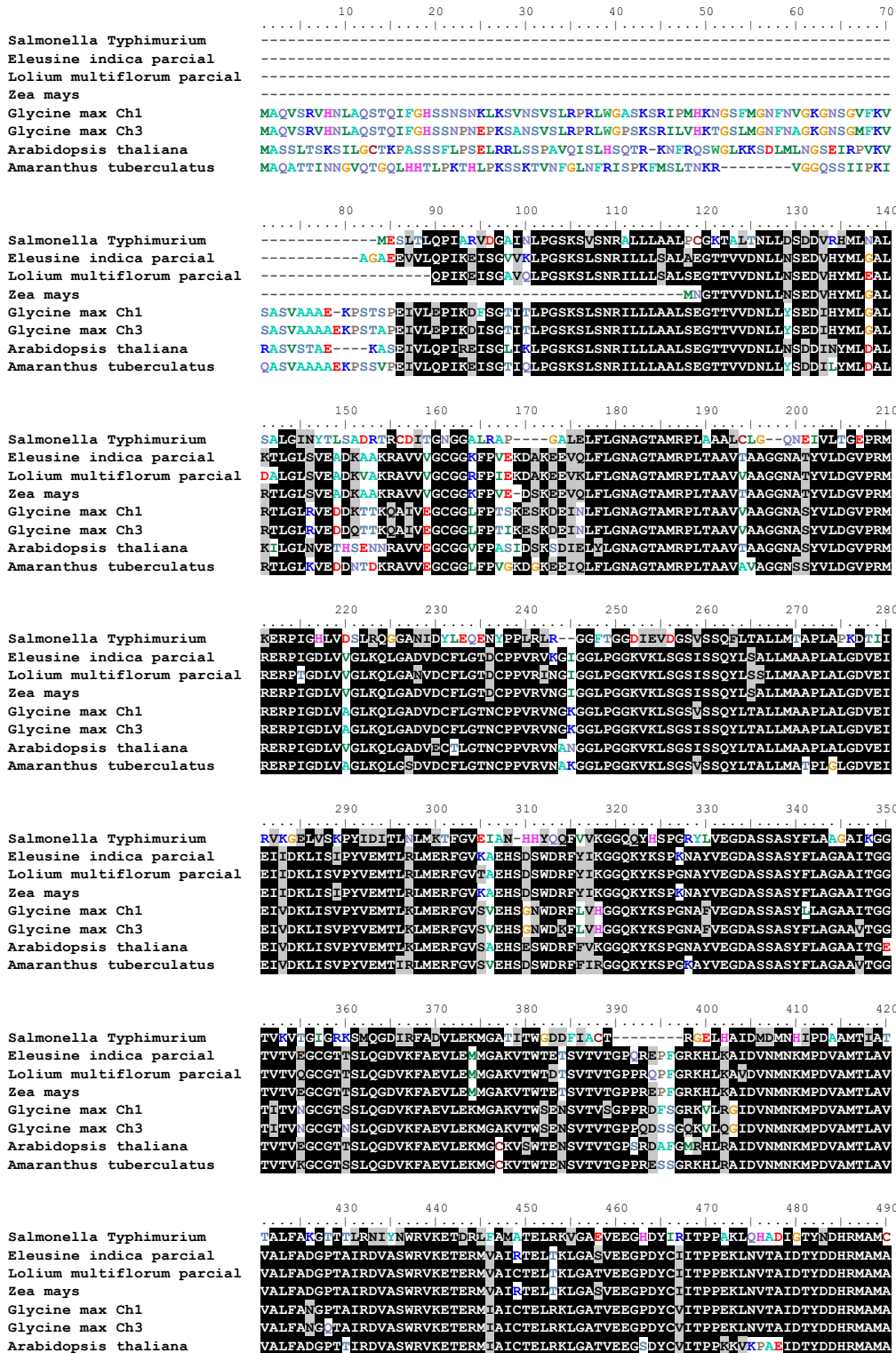
- Kanchiswamy, C. (2016). DNA-free genome editing methods for targeted crop improvement. *Plant Cell Reports*, *35*(7), 1469-1474.
- Khan, S. (2019). Genome-Editing Technologies: Concept, Pros, and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application. *Molecular Therapy — Nucleic Acids*, *16*, 326-334.
- Kim, H., & Kim, J. (2014). A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics*, *15*(5), 321-334.
- Kim, H., Kim, S., Ryu, J., Kang, B., Kim, J., & Kim, S. (2017). CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nature Communications*, *8*(14406), doi: 10.1038/ncomms14406.
- Li, J., Meng, X., Zong, Y., Chen, K., Zhang, H., Liu, J., y otros. (2016). Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. *Nature Plants*, *2*(16139), doi: 10.1038/nplants.2016.139.
- Li, X., Ma, S., Shan, N., Zhang, X., Sui, X., & Zhang, Z. (2017). A Protocol for Agrobacterium-mediated Transformation of Cucumber (*Cucumis sativus* L.) from cotyledon explants. *Protocol Exchange*, doi: 10.1038/protex.2017.107.
- Lin, C., Hsu, C., Yang, L., Lee, L., Fu, J., Cheng, Q., y otros. (2018). Application of protoplast technology to CRISPR/Cas9 mutagenesis: from single-cell mutation detection to mutant plant regeneration. *Plant Biotechnology Journal*, *16*(7), 1295-1310.
- Liu, H., & Naismith, J. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology*, *8*(91), doi: 10.1186/1472-6750-8-91.
- Liu, J., Gunapati, S., Mihelich, N., Stec, A., Michno, J., & Stupar, R. (2019). Genome Editing in Soybean with CRISPR/Cas9. *Methods in molecular biology*, *1917*, 217-234.
- Liu, L., & Fan, X. (2014). CRISPR-Cas system: a powerful tool for genome engineering. *Plant Molecular Biology*, *85*(3), 209-218.
- Ma, X., Zhu, Q., Chen, Y., & Liu, Y. (2016). CRISPR/Cas9 Platforms for Genome Editing in Plants: Developments and Applications. *Molecular Plant*, *9*(7), 961-974.
- Makarova, K., Wolf, Y., & Koonin, E. (2018). Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? *The CRISPR Journal*, *1*(5), 325-336.
- Manova, V., & Gruszka, D. (2015). DNA damage and repair in plants - from models to crops. *Frontiers in Plant Biology*, *6*(885), doi: 10.3389/fpls.2015.00885.
- Mao, Z., Bozzella, M., Seluanov, A., & Gorbunova, V. (2008). DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle*, *7*(18), 2902-2906.
- Marillonnet, S., & Werner, S. (2015). Assembly of Multigene Constructs Using Golden Gate Cloning. *Methods in Molecular Biology*, *1321*, 269-284.

- Martinez, M., Limones, V., Cabrera, J., & Herrera, L. (2012). Improving transformation efficiency of *Arabidopsis thaliana* by modifying the floral dip method. *Plant Molecular Biology Reporter*, *22*, 63–70.
- Metje-Sprink, J., Menz, J., Modrzejewski, D., & Sprink, T. (2019). DNA-Free Genome Editing: Past, Present and Future. *Frontiers in Plant Science*, *9*(1957), doi: 10.3389/fpls.2018.01957.
- Ministerio de Ganadería, A. y. (2019). Anuario Estadístico Agropecuario. Montevideo: Oficina de Estadísticas Agropecuarias (DIEA).
- Nishimasu, H., Ran, F., Hsu, P., Konermann, S., Shehata, S., Dohmae, N., y otros. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, *156*(5), 935-949.
- Orosz, A., Boros, I., & Venetianer, P. (1991). Analysis of the complex transcription termination region of the *Escherichia coli* rrn B gene. *European Journal of Biochemistry*, *201*(3), 653-659.
- Pingoud, A., & Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Oxford University Press*, *29*(18), 3705-3727.
- Pollegioni, L., Schonbrunn, E., & Siehl, D. (2011). Molecular basis of glyphosate resistance: Different approaches through protein engineering. *FEBS Journal*, *278*(16), 2753-2766.
- Puchta, H., & Fauser, F. (2013). Gene targeting in plants: 25 years later. *The International Journal of Developmental Biology*, *57*(6-8), 629-637.
- Qi, L., Larson, M., Gilbert, L., Doudna, J., Weissman, J., Arkin, A., y otros. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*, *152*(5), 1173-1183.
- Quadros, R., Miura, H., Harms, D., Akatsuka, H., Sato, T., Aida, T., y otros. (2017). Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. *Genome Biology*, *18*(1)(92), doi: 10.1186/s13059-017-1220-4.
- Que, Q., Chen, Z., Kelliher, T., Skibbe, D., Dong, S., & Chilton, M. (2019). Plant DNA Repair Pathways and Their Applications in Genome Engineering. *Methods in molecular biology*, *1917*, 3-24.
- Raji, J., Frame, B., Little, D., Santoso, T., & Wang, K. (2018). Agrobacterium- and Biolistic-Mediated Transformation of Maize B104 Inbred. *Methods in Molecular Biology*, *1676*, 15-40.
- Reece-Hoyes, J., & Walhout, A. (2018). Gateway Recombinational Cloning. *Cold Spring Harbor Protocols*, *2018*(1), doi: 10.1101/pdb.top094912.
- Roth, N., Klimesch, J., Dukowic-Schulze, S., Pacher, M., Mannuss, A., & Puchta, H. (2012). The requirement for recombination factors differs considerably between different

- pathways of homologous double-strand break repair in somatic plant cells. *The Plant Journal*, 72(5), 781-790.
- Sammons, R., & Gaines, T. (2014). Glyphosate resistance: state of knowledge. *Pest Management Science*, 70(9), 1367-1377.
- Sanagala, R., Moola, A., & Kumari, B. (2017). A review on advanced methods in plant gene targeting. *Journal of Genetic Engineering and Biotechnology*, 15(2), 317-321.
- Schaefer, D., & Zrýd, J. (1997). Efficient gene targeting in the moss *Physcomitrella patens*. *The Plant Journal*, 11(6), 1195-1206.
- Schmidt, C. (2005). Genetically Modified Foods: Breeding Uncertainty. *Environmental Health Perspectives*, 113(8), doi: 10.1289/ehp.113-a526.
- Schonbrunn, E., Eschenburg, S., Shuttleworth, W., Schloss, J., Amrhein, N., Evans, J., y otros. (2000). Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. *Proceedings of the National Academy of Sciences*, 98(4), 1376-1380.
- Scientific, T. (2014). *Thermo Scientific TranscriptAid T7 High Yield Transcription Kit*.
- Shan, Q., Balthes, N., Atkins, P., Kirkland, E., Zhang, Y., Baller, J., y otros. (2018). ZFN, TALEN and CRISPR-Cas9 mediated homology directed gene insertion in Arabidopsis: A disconnect between somatic and germinal cells. *Journal of genetics and genomics*, 45(12), 681-684.
- Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., y otros. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology*, 31(8), 686-688.
- Shy, B., MacDougall, M., Clarke, R., & Merrill, B. (2016). Co-incident insertion enables high efficiency genome engineering in mouse embryonic stem cells. *Nucleic Acids Research*, 44(16), 7997-8010.
- Sledz, C., & Williams, B. (2005). RNA interference in biology and disease. *Blood*, 106(3), 787-794.
- Spampinato, C. (2017). Protecting DNA from errors and damage: an overview of DNA repair mechanisms in plants compared to mammals. *Cellular and Molecular Life Sciences*, 74(9), 1693-1709.
- Stemmer, M., Thumberger, T., Sol, K. M., & Wittbrodt, J. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLOS One*, 10(4), doi: 10.1371/journal.pone.0124633.
- Sun, X., Hu, Z., Chen, R., Jiang, Q., Song, G., Zhang, H., y otros. (2015). Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Scientific Reports*, 5(10342), doi: 10.1038/srep10342.

- Svitashev, S., Young, J., Schwartz, C., Gao, H., Falco, S., & Cigan, A. (2015). Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA. *Plant Physiology*, *169*(2), 931-945.
- UniProt. (s.f.). *UniProt*. Recuperado el Enero de 2020, de <https://www.uniprot.org/uniprot/P11043>
- Voytas, D. (2013). Plant genome engineering with sequence-specific nucleases. *Annual Review of Plant Biology*, *64*(1), 327-350.
- Wang, M., Lu, Y., Botella, J., Mao, Y., Hua, K., & Zhu, J. (2017). Gene Targeting by Homology-Directed Repair in Rice Using a Geminivirus-Based CRISPR/Cas9 System. *Cell Press*, *10*(7), 1007-1010.
- Waqas, M., Kaya, C., Riaz, A., Farooq, M., Nawaz, I., Wilkes, A., y otros. (2019). Potential Mechanisms of Abiotic Stress Tolerance in Crop Plants Induced by Thiourea. *Frontiers in Plant Science*, *10*(1336), doi: 10.3389/fpls.2019.01336.
- Woo, J., Kim, J., Kwon, S., Corvalán, C., Cho, S., Kim, H., y otros. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature Biotechnology*, *33*(11), 1162-1164.
- Wurm, D., Veiter, L., Ulonska, S., Eggenreich, B., Herwig, C., & Spadiut, O. (2016). The E. coli pET expression system revisited—mechanistic correlation between glucose and lactose uptake. *Applied Microbiology and Biotechnology*, *100*(20), 8721-8729.
- Xie, K., Minkenberg, B., & Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *PNAS*, *112*(11), 3570-3575.
- Yang, Z., Edwards, H., & Xu, P. (2019). CRISPR-Cas12a/Cpf1-assisted precise, efficient and multiplexed genome-editing in *Yarrowia lipolytica*. *Metabolic Engineering Communications*, *10*(e00112), doi: 10.1016/j.mec.2019.e00112.
- Yu, C., Zhang, Y., Yao, S., & Wei, Y. (2014). A PCR Based Protocol for Detecting Indel Mutations Induced by TALENs and CRISPR/Cas9 in Zebrafish. *Plos One*, *9*(6)(e98282), doi: 10.1371/journal.pone.0098282.
- Zhang, Y., Zhang, F., Li, X., Baller, J., Qi, Y., Starker, C., y otros. (2013). Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiology*, *161*(1), 20-27.
- Zhou, J., Deng, K., Cheng, Y., Zhong, Z., Tian, L., Tang, X., y otros. (2017). CRISPR-Cas9 Based Genome Editing Reveals New Insights into MicroRNA Function and Regulation in Rice. *Frontiers in Plant Science*, *8*(1598), doi: 10.3389/fpls.2017.01598.

Anexo



Alineamiento total de las EPSPs de soja y sus homólogos en diferentes especies (*G. max* Ch1 (Glyma.01G139600.1), *G. max* Ch3 (Glyma.03G027400.1), *A. thaliana* (AT1G48860.1), *Z. mays* (Zm00008a033811_T01), *E. indica* parcial (AAN63155.1), *L. multiflorum* parcial (AAZ79230.2), *A. tuberculatus* (ACV67278.1), *S. typhimurium* (AAA27028.1) y *E. coli* (NEY26641.1).

Secuencia genómica del gen EPSPS1 de *Glycine max* y con respectivos primers utilizados

>Glyma.01G13960 | Chr01:46734395..46743155 reverse
 CTGTTTACACTATGTATAACGAAATTAATATATTCATTTAATATGATTTTATTTTCTTTAATTTA
 AAAACATGAATTATTTGTTTCTTATTTTATTACCATTTTGTTCACCTTCTCAATTAATAAAATAAAAA
 TATATTCTACAATTATAGTAAAAAGGTTTAATTTTATATAATGTTAGCATAACAGATTTTCTCTATT
 AACTAATTAATTAATAATATCCTTATTTTAAAGGAAATTAATAATATCATATATATAAGTTTAA
 AATTAATTATCTTATATATGTACCAAAAAGTTTAAAGCAATTATTATAAAAAATTAATAAATTTATC
 ATATAAAATAATTTATAATTAATTTTAAATTATCAATTCATTAATTAATTTTAAATTTTGT
 AATGATAATATAATAATTTTATCCTCTACTAAGTCCCAACGTTTCTTATTTTATCCACTTTTAGCA
 ATAAATTTTGTCTATAAACACTTATAACAAAAAAGTAAGTAAAAAATAAAAAAAGTTTTTCAATAA
 AGTATAAACTAATTTGTATAAACTTTTAGAAAAAATAAAGTTATACATTGATAATATAAATTTTTTA
 CATAATTATCCGATCAACTCATTATATATGATAAATTTATTGATTTTTTAAATAATTATCTTAAAA
 TAATTTAAACAATGATTTGCAATTAGATGATAATATAAAATTTTACACACTACATGTATTAAC
 TCAAACTTTTATATATAGTTTTTCTAAAACTAATTTTAACTCAAAAAAATGTTACTTATAATT
 TTCTTATCTTCTTTTTTTTATAAGTATTTTTTAAAGAAATTTATTGAAACATGACCATGCTTGGGTCAA
 TAATACTACTCTCTTAGACACCAACAACCCTTCCCAAACATAATCTAATCCAAAAGCCATCATTC
 ATTTTCTTGGTAGGTAAAGTTCCAAGACCTTACCAACTTTTTCACTCAATTGTTTGGTGTAAGC
 AATTCGACATGTGTTAGTGTAGTTGGCAACCAAAAAATCCCTTTATGTGACTCAATCCAACAACCAC
 TCACACCACCAACCCCATACCATTTCTCACAATACCCTTCATTTACACATTATCATCACCAAAAA
 TAAATAAAAAAACCTCTCATTTTCAGAGAGAGAGAGAGACTTCACAGACCAAAGTGCAGAGAACA
 AC¹AAAGTTCACAACCTTTAAGGAAAAATGAAATGGCCCAAGTGAGCAGAG²TGCACAATCTTGCTCAA
 AGCACTCAAATTTTTGGCCATTCTTCCAACCTCCAACAACTCAAATCGGT (GAATTCGGTTTCATTG
 AGGCCACGCCTTTGGGGGGCCTCAAATCTCGCATCCGATGCATAAAAAATGGAAGCTTTATGGGAA
 ATTTTAATGTGGGGAAGGAAATTCGGCGTGTTTAAGGTTTCTGCATCGGTGCGCGCCGAGAGAA
 GCCGTCAACGTCGCCGAGATCGTGTGGAACCCATCAAAGACTTCTCGGGTACCATCACATTGCCA
 GGGTCCAAGTCTCTGTCCAATCGAATTTTGCTTCTTGCTGCTCTCTGAGGTGAAGTTTATTTATT
 TATTTATTTGTTTGTGTTGTTGGGTGTGGGAATAGGAGTTTGATGTGTAGAGTGGATTTGAATA
 TTTGATTTTTTTTTGTATTATCTGTGAAAATGAAGCATCATGTCCCATGAAAGAAAATGGACACGAA
 ATTAAGTGGCTTATGATGTGAAATGAGGATAGAAATGTGTGTAGGGTTTTTTAATGGGTAGCAATAA
 GCATATTCATATCTGGATTGATTTGGACGTTTCTGTATAAAGGAGTATGCTAGCAATGTGTTAATG
 TATGGCTTGCTAAAATACTCCTAAAATCAAGTGGGAGTAGTATACATATCTACAGCAAATGTATTA
 GGTGAGGCATTTGGCTTCTCTATTGTAAGGAACAAAATAATATCAGTTAATGTGAAAATCAATGGTTG
 ATATTCCAATACATTCATGATGTGTTATTTATATGTACCTAATATTGACTGTTGTTTTTCTCCGCAA
 TGACCAAGATTATTTATTTATCCTCTAAAGTGACTAATTGAGTTGCTTACTTTAGAGAAGTTGGAC
 CCATTAGGTGAGAGCGTGGGGGAACTAATCCTGAAATACAACTGAGTCTTGATTATCCAAGTAT
 GGTGTATGAACAATGTTAGCTCTAGAAGATAAACCCCTCCCCAAAACACAAAATTAGAATGACATTT
 CAAGTTCCATGTATGTCACCTTCATTCTATTATTTTTACAACTTTTAGTTACTTAACAGATGTCTTG
 TTCAGCATAAATTATAAATTTATCTGTTTTTTTTTAGGGAACAACCTGTTGTAGACAACCTGTTGTAT
 AGTGAGGATATTCATTACATGCTTGGTGCATTAAGGACCCTGGACTGCGTGTGGAAGATGACAAAA
 CAACCAACAAGCAATTGTTGAAGGCTGTGGGGGATGTTTCCCACTAGTA³AGGAATCTAAAGATG
 AAATCAATTTATTCCTTGAAATGCTGGTACTGCAATGCGTCCTTTGAC⁴AGCAGCTG⁵TGGTTGCTG
 CAGGTGGAAATGCAAGGTCTGTTTTTTTTTTTTTTTGTTCAGCATAATCTTTGAATGTTTCCTCGTAT

¹ FwGmEPSPS1 CAGACCAAAGTGCAGAGAACAAC
² FwGmEPSPS1-SalI TTTGTCGACATGGCCCAAGTGAGCAGAG
³ FwEPSPS1TA1+PAM GGGATTGTTCCCACTAGTAAGG
⁴ ReTIPsb GTCAAAGaACGCATTGCAaTACCAGCATTTC
⁵ FwTIPsb CTGGTAtTGCAATGCGTtCTTTGACAGCAGCTG
 FwTIPAb CTGGTAtTGCAATGCGTgCTTTGACAGCAGCTG
 ReTIPAb GTCAAAGcACGCATTGCAaTACCAGCATTTC

AACTAATCACAACAGAGTACGTGTTCTTCTTCCTGTTATAATCTAAAAATCTCATCCAGATTAGTCA
TCCTTTCTTCTTAAAAGGAACCTTTAATTATCAATGTATTTATTTAATATTTAAATTAGCTTGTCAA
AGTCTAGCATATACATATTTTGATTATATTCTGAGAAAATGCACCTGAGGGTGTTCCTCATGATCTAC
TTCAACCTCTGTTATTATTAGATTTTCTATCATGATTACTGGTTTGAGTCTCTAAGTAGACCATCTT
GATGTTCAAATATTTTCAGCTACGTACTTGATGGGGTGCCCCGAATGAGAGAGAGGCCAATTGGGGA
TTTGGTTGCTGGTCTTAAGCAACTGGTGCAGATGTGATTGCTTTCTTGGCACAACTGTCCACCT
GTTCGTGTAATGGGAAGGGAGGACTTCCTGGCGGAAAGGTATGGTTTGGATTTCAATTTAGAATAAG
GTGGAGTAACTTTCTGGATCAAATTTCTAATTTAAGAAGCCTCCCTGTTTTCTCTCTTTAGAATA
AGACTAAGGGTAGGTTTAGGAGTTGGGTTTTGGAGAGAAAATGGAAGGGAGAGCAATTTTTTCTTCT
TCTAATAAATATTTCTTAATTTGATACATTTTTTAAGTAAAAGAATATAAAGATAGATTAGCATAAC
TTAATGTTTTAATCTTTTATTTATTTTTATAAATATTATATACCTGTCTATTTAAAAATCAAATATT
TGTCTCCATTCCCTTTCCCTTCAAACCTCAGTTCCAAATATACCGTAGTTGAATTATATTTTGGGA
AGGCCTATTGGTTGGAGACTTTTCCTTTTTCAGAGATTATCCCTCACCTTTATTATAGCCTTTCTATT
TTTAAACTTCATATAGACGCCATTCTGTTTTTAAAAACACTAAGTTTTCTTTTAGTTATACCTTCC
CCTCCTTATTCTCTCCAAAGTAAATATTGTAGCAAGTGTGGTTGATCCATTGTGAATCTTATTATC
CTATTTAACATGCAGGTGAACTGTCTGGATCAGTTAGCAGTCAATACTTGACTGCTTTGCTTATGG
CAGCTCCTTTAGCTCTTGGTGATGTGGAAAATTGAGATTGTTGATAAACTGATTTCTGTTCCATATGT
TGAAATGACTCTGAAGTTGATGGAGCGTTTTGGAGTTTCTGTGGAACACAGTGGTAA) TTGGGATAG
GTTCTTGGTCCATGGAGGTCAAAGTACAAGTAGGTTTCTATGTTTTAGTGTTACATCACTTTTAGT
ATCCAAAATGCAATGAAATTAACCTCATGTTTCTTATAGGTCTCCTGGCAATGCTTTTGTGGAAG
GTGATGCTTCAAGTGCCAGTTATTTACTAGCTGGTGCAGCAATTACTGGTGGGACTATCACTGTTAA
TGGCTGTGGCACAAGCAGTTTACAGGTATTTCATGAATGACACTATCTTCTATTATCTTTTATTTTAT
AGCTTACACTATCATTCCCTAATTCAACCCCTTCTTTTCTCTTTCCCTCACTCCCAAATAGAGCT
TATCATTGTATCCTTAATAGCTCTTTTATGTGCTGTTTCATCTCTTGAGCAGATTCTACATGATTAT
CCTTTTTAGTGAACCTTTTTGTCTTTTAAATAATTGATATCATCATTTGAACTTGCACATCATCTCT
ATCCTTCCCTCATTTTATTATATTTTATTAAAGAGTGATAAAATTACATTGTCAAGTAAAGCAAGTTAG
CGTATCAGATAATGATGCTTCGATTGCAGAATTGGTGAAGCTTGCACAATTCATTAGCTTAGTGAC
AATTGTGATGTAGATAGCCATAATCTCCATGCAACTTATTATGCTTTTTGTCTAAAACCTCTTTTG
CACTTAAACATGAATTAGAGAGTTAGGATGAACAATAGTAACAGACACAAAGCTTATTTCTAGACCA
CCCAGTGGTGAAGACGAAGCAGCATGAGTGAGATTAATGGCAGCATCTGATTGAAGCTTCTTTTGGT
ATTATTGGAGATGAAATCTTGAGCCAAATTAGGGCTTTTGCCCTCCTCAATAGCCATTGGTTCAAAT
TGCTTTCAATAAAATGATTACAGAATCAAATTTCTTAAGAAAGACCTTCAAGAACAACATCAAGATGT
TTCTGTGCAAAACAGGCTCACCTACAAAAGAGTGTCTACAAGCACCTTTACACTTGCAAAAAATTC
CTTAACGAACAATTGCCAAGAGTCATGTTGCAAACCTGAGATTGGAGTTGATTAGATTTGGTGTGT
GTGAGATTCTGGAAATGAGCATGAAGTTTTTTGCAAAGCTCATAAAATGAATGCAGCCAAGAATCCT
AGATAGGATTGAATTTGAGAGAGAGGCTTGAAGTCAAGATAACCGCAACTGATCTTGCTGTTCCCAT
GCTGCATAGGACTCGTTGATTCTACCAAGATAACAATCTTCTGTGGTGAGAATATGAGTGGGAATTG
TCAGGCAACAAGAAAATGAAGCTATGAGCCTTTGTTAAGGTGCAGAGAGAAAACAGAAAATGAAG
AACTTGTGTATTGAAGGAATTGAAAAGCCAATTACAATTTGTGTTGCAGAGAGATATTTATAGTTCTG
AAGTGATAACTAACTAAACAGAATGCTACCAACCTAGAGGCAGTTGAGAATTGCTCCCTTATTACA
ATACTCTCACAGCCTTGATGATAGGCTTGATTTGCTGCCACCGAAGAAGAACATTTATTGTTGGAGAG
TTTTTCAGTGATTGCATCAGAGAAAGTTACAAGTGTGATTGGAGATGAAGGACTTGCGGTCGTCGAA
GGTTAAGTGTAAAGTGGTGGAAATGTGTATCGGAGGTGGTTGCCTAGCCCAAGAAGCTGGTTGCAGTG
GAAGCGGATGCGGCAGAAGCCATGGGTTGAGATCAGAACTATGTTCTAGATACCATATTAGGGTTAG
CTTTCAAGAGAAAACCTTATGAACTGAACTGTAACTGTTGTATTATTGATTAGCTGATCTAGTAGT
GATGGAAGCTTGCTTGCGAAGCTTCTATGGAGGCTGGATCATTGAGCTTCAATGAGGTCCTTCAATG
GTGATTTTTTACCATGGAGATGCAGCGGAAGATAAATGAAGAGGTGAGAGGAGGCGTCATCCACTAG
GAAATAAATATGGAAGAAGGAGCTTACCACCAAGAGAGTGTCTTGGATAAGAAGCTTAGAGAGGA
AGCTTCAATGGAGGAATAGAAAAGAGAGAGGGGGAACAAAAATTTGAAGGAGGAAAAGAGGGAGAGAA
GTTGAACTTTGAAGTGTGTCTCACAAAGACTCATTCAATTAAGTTACAACAAGTGTACACATGCTTC
TATTTATAGTCTAGGTAACCTCCTTGAGAAGCTAGAACTTAACTACACACACCTCTCTAATAACTAA

GATAACCTCCTTGAGAAATTCCTTGAGAAGCTTCCTTAAGAAGATTCCTAGAGAAGCTAAAGCTTA
GTTACACACACCTCTCTAATAACTAAGTTCACCTCCTTGAGATGAGAAGCTAGAGCCTTAGCTACAC
ACACCCTTATAATAGCTAAACTCACTCTATTCCAAAATACATGAAAATACAAAAAAGTTCCTACTAC
AAAGATTACTCAAATGTCTGAAATACAAGACTAAAACCTATACTACTAGAATGGTCAAATACA
AGGTCCAAAAGAAGGAAAAACCTATTCTAATATTTACAAAGAGAGTGGACCCAACCTTAGTCCATGG
GCTCAGAAATCTACCCTGAGGTTCATGGGAATCCTAGGGTCTTCTTTAGTAGCTCTAGTCCAATTTT
CTTGGAGTCTTCTATCCAATATCCTTGGGGGTAGGATTGCATCATGTAGTACAAGAGAGGGGAATAT
ATATACAGTTGTAAGTTGTAACCTACCTAATCAAACCCAACCTGACTCACCATAACAGAAAAACAATG
TACAAAGGATAATAGAAAACTGAAACTGAGATATACTCATTACATAGAAGAATTACATGATGAATC
TGAAAATTTAATTGGGAACTGAGGACAATAACAAGGCTTTCAGAGTTCATTTTGACAATAGGTAA
ATTATGGACTTCAACTTAGACCTCAAGGGTATGCTTGGGAGTGGGGTTTTAGAGGAAAAGGGGAAGG
AGGGAAATTTTTAATTCTAGTAAATATTCTCTATTTTGATTAATTTTTAAAATAAAACAAATAAAC
TTATAGATTAGCATTCACTAATGTCCTGATCTTTTATTTGTTTCATGAAGTTTTTATAGCCTTTC
TGTTTCTGTTAAAAGAAATGGAGGGTGTTCATAATCTATATGCCCATATATATGTATTGTCCTGTT
GTCTGTGTTTCTGTTTTCTTTCTTTAGTGTCTCGCCTTTCCTTAATTCCTTATCAACACCTTGTTTA
TTTTTGAAAACCTAAATGTCCTCATTTTTTTTACCTGTACTATAACTTTCCAAACAAATAGTAGAAA
ATGTTTCAAATATTCCCCAGAAGTGTTCAGTGAACAAAGTTCAAGCCAGAAAAAAGCAAGCAGAG
GCTAAAGAAGTATATCTTATTGTGAAGAGTAATCTCAATTGGTTAATTATCGAATTGTATCATTTTA
CAATAGAAAATTAGAAGTGGTTGATTAGTTTATACGAACAATTATTTAAATAGGGTTGATGGTGTG
GTTAATTACTGATCATATGTTAAGACTATGAGTCGATGACCAGCCTCAGTGCAGTCATTTAAGTTC
CATTATTTTTAATTGTGCTTATGTTAACTTTGTCTTCTTTTAAAGGGAGATGTAAAATTTGCTG
AAGTTCCTGAAAAGATGGGAGCTAAGGTTACATGGTCAGAGAACAGTGTCACTGTTTCTGGACCACC
ACGAGATTTTTCTGGTCGAAAAGTCTTGCAGGCATTGATGTCAATATGAACAAGATGCCAGATGTT
GCCATGACACTTGCTGTTGTTGCACATTTTGCTAATGGTCCCACTGCTATAAGAGATGGTATGGTTT
AGCTGTTTGTGTTTTGCTCAAGAGCTTGTGTTTACATATTTGTGGTCAAGAGCGGTTTAGCTGTTTTT
TTTTTTAAATTGTGGTTAATGAAATTATTATAATTATACTAATATTGAAATTGGATCACTAACCATG
TGAAATGCTTGAATTTTTAGCTTGAAAACCTCTTTTGTGGATGATTTTTCTTTTATCTATGTAGTGT
GGGGCATCTATGTTGTTCTTTAAGCTTTGTGTCCATAGTTGTACAACCTGGTCTGTGGGCTGGTCCA
GTTCTTACTCATGACTGTCTTTTGCAGCAAACCTGGTTAACTTGGGTGACCTGTTGACCCACCAGA
TTGGCCTGTTCTTTTTTAAATAAACTCCAAAAGGCTACTTATAAGATTTCAACATGATACTTGAAT
GACCACCCTTACTTTAAAGGTCATTATGCCATCAGACCCCTTGATTCTCTGTTCACTTATGTTCTTA
TTCTTATTATTTTTAATTATATTATTAATCCAACGATCTGATCTTTGTACATGATGCTGGACAATC
ATCAAAGCTAGTTTGGTTACATCATATAATACCTTTTTTTTTTATAAATATTTTTTCCATAATATA
TTTATATGTATTTTTTTAATACAATATTCTGGGAAATTTTTTTTTTCTGTTTTCTTTTTCTGGTT
TGAATACTATTTCAATGACTGGGCCTTATTATTTCTGGTTTGGGTTATACAACTGTCCTTTGATGT
ACTTTTGAATGCTTCAATAGTATTTTTTATAATCATGGAAGGCACTAGAGTGCAGTGCATATTGCAC
AAGTTCCTTTAATGGCATCTTGAATTTGTTAGAACATGGAGTTATGATATAACAACCATTCCTATGTC
TTCATCTAATAGCTAAAGCTTTTGGCGTAGTTGGTTAACATGGTATTGGAATCTATTTTTAAAGGTC
CATCTTATTCGCAAGTACAAGGCAGGTGGACCTGCGCATTATTCACGTGCCAGGCTCAAAGGGATTT
TGATGAGGGGGAATTTTTGGAATATAATATATAGAATCATTATCATGTGGTTACCTATCATGTGTCA
ACTCAAGTTTTTGTCTAGTTGTATTATTGCAGTGTCTAACTAGGCATGATCTTCTGTGGAAGGAA
ATTTAACCTAAGGAACCTTGAATGCTTCTTGTCTGAGATCTTTATTTTTTTTTCTCCAATCTTTTTAA
TACTTGAAAGGTTTTAGTTGTATCTATAAAGGGTGCAGAAATTAGGGGGTCACTTTTCTGTTGTAAG
CGTGAATTAAGTGCATTTTATGTTGTTCCCTATTTTTATACATTTGAACTATCAACCATATCATTTAG
CAAACCTCTTTGTTGCTCAGATTCCATTTATTCTTGTTCCTCTATTTTTTAGTGGCAAGTTGGAG
AGTTAAAGAGACTGAGAGGATGATAGCAATCTGCACAGAACTCAGAAAGGCTTGCTAATTCCTTTA
TGGTTTCTATATTACTGGACTTTTTTACATCTCACTGACCTACTACTGTCTTGGTATTTTCACTAGGA
GCAACAGTTGAAGAAGGTCCTGATTACTGTGTGATTACTCCACCTGAGAAATTGAATGTCACAGCTA
TAGACACATATGATGACCACAGAATGGCCATGGCATTCTCTCTTGTGCTGCTTGTGGGGATGTTCCAGT

AACCATCAAGGATCCTGGTTGCACCAGGAAGACATTCCTGACTACTTTGAAGTCCTTGAGAGGTTA
ACAAAGCACTAA⁶GCATTTTGTACATTGAGTAGGGAAGAGAGAGATATATATAAATACTCACAAA
GCAGTGAGTATATGGCTTGCTTTTGTCTAAACATTCCTTTATAAGGACTTGAGATAATTTGTATT
GTTTAAAAGAGCCATTGGTCTGTTGTATCAAAGTAATTTAATTTTCAGTACCGACTTGGCTCTACT
CATATGTGG⁷TGGTGTACATTGCATCTTGCATTTGATTCCATATCTATTGGGATGGGACCATACAT
TTTTTGTATTATTAAAGTCAAACATATCAGATAATATGTGGTGACCAGAAA

⁶ ReGmEPSPS1-NotI TGCGGCCGCTTAGTGCTTTGTAAACCTC

⁷ ReGmEPSPS1 CCACATATGAGTAGAGCCAAGTCG