



Facultad de Veterinaria
Universidad Zaragoza



Trabajo Fin de Máster en

Calidad, Seguridad y Tecnología de los Alimentos

Desarrollo de herramientas moleculares basadas en el sistema CRISPR-Cas9
para la modificación del aceite de *Camelina sativa*

Development of a CRISPR-Cas9 molecular tool for the modification of the seed
oil from *Camelina sativa*

Autor

Alejandro del Río García

Director/es

Miguel Alfonso Lozano

María Lourdes Sánchez Paniagua

Facultad de Veterinaria

2020

Tabla de contenido

1. RESUMEN/ <i>ABSTRACT</i>	4
2. INTRODUCCIÓN	6
2.1 <i>Camelina sativa</i> como fuente de origen vegetal de ácidos grasos omega-3.....	6
2.2 La ingestión de ácido erúcico puede acarrear problemas a la salud	8
2.3 Los genes y proteínas implicados en la biosíntesis de omega-3 y erúcico en el aceite de semilla de camelina son conocidos	11
2.4 La tecnología CRISPR/Cas9 como herramienta de edición genética.....	14
2.5 La tecnología CRISPR/Cas9 permitiría la inactivación simultánea de los tres genes <i>CsFAEI</i> de camelina.....	15
3. OBJETIVOS DEL ESTUDIO	17
4. METODOLOGÍA.....	18
4.1 Crecimiento de plantas de Camelina (<i>Camelina sativa L.</i>).....	18
4.2 ANÁLISIS LIPÍDICO DE LAS SEMILLAS DE CAMELINA	19
4.2.1 Extracción de lípidos totales	19
4.2.2 Obtención de ésteres metílicos de ácidos grasos totales.....	19
4.2.3 Análisis de la composición de ácidos grasos mediante cromatografía gas-líquido	20
4.3 Descripción del vector pHSE401	21
4.4 Extracción del ADN plasmídico y digestión con enzimas de restricción	22

4.5 Técnica de PCR.....	22
4.6 Cebadores utilizados en este trabajo	24
4.7 Electroforesis en gel de agarosa.....	25
4.8 Extracción y purificación de ADN en geles de electroforesis	26
4.9 Determinación de la concentración de los productos de PCR	26
5. RESULTADOS.....	27
5.1 Caracterización previa de la composición del aceite de Camelina en plantas crecidas en la EEAD-CSIC	27
5.2 Diseño y construcción del vector conteniendo las dianas, los ARNs de guiado y la nucleasa Cas9, dirigidos contra los 3 genes <i>CsFAE1</i> del genoma de <i>Camelina sativa</i>	28
5.3 Amplificación de los fragmentos por PCR splicing y armado del vector de edición	30
6. DISCUSIÓN	33
7. CONCLUSIONES	38
8. BIBLIOGRAFÍA	39

1. RESUMEN/ABSTRACT

La manipulación genética de los aceites vegetales es un objetivo prioritario para los tecnólogos de alimentos debido a sus múltiples aplicaciones. *Camelina sativa* posee un contenido inusual de ácidos grasos omega-3 (30-35%) en su aceite siendo una de las fuentes naturales de omega-3, diferentes de los peces, más abundantes del planeta. Sin embargo, el aceite de esta planta también posee otros ácidos grasos como el erúcico (22:1), en cantidades que pueden variar entre el 2 al 5% del total de ácidos grasos lo que supone un obstáculo para la generalización del aceite de camelina como aditivo alimentario. El gen *FAEI* es el responsable de la síntesis de ácido erúcico en la semilla y por tanto el gen diana a editar. Así pues, el objetivo de este proyecto a largo plazo es modificar genéticamente la composición química del aceite de camelina eliminando el contenido de ácido graso erúcico actuando sobre los tres genes *CsFAEI* presentes en Camelina. Para ello, hemos utilizado el sistema de edición CRISPR-Cas9. En este trabajo, se ha procedido al diseño y obtención del vector de edición, conteniendo las secuencias diana, el sistema de guiado y la proteína Cas9 como paso previo a su transformación en plantas de Camelina. Para ello se ha utilizado una estrategia de PCR splicing para la clonación de las secuencias diana en el vector pHSE401. Los resultados muestran un correcto armado del vector de edición previo a la transformación de plantas de camelina. Estudios anteriores demuestran que es posible la edición mediante CRISPR-Cas9 de genes involucrados en la biosíntesis de aceite y de ácidos grasos en el aceite de plantas. La eliminación del erúcico mediante una estrategia similar, abriría las puertas al uso definitivo del aceite de camelina como fuente de omega-3 para usos agroalimentarios.

Genetic manipulation of vegetable oils has become a priority objective for food technologists due to their multiple applications. Camelina sativa's oil has an unusual content of omega-3 fatty acids (30-35%) and is one of the major source of this fatty acid apart from fish. However, camelina's oil has other fatty acids erucic (22:1) in amounts that can vary between 2-5% of the total fatty acid content. The presence of erucic acid reveals as an obstacle to the generalization of camelina's oil as a food additive. It is known that FAE1 gene is responsible for the synthesis of erucic acid in seed plants so it is our target gene. The aim of this project is to genetically modify the chemical composition of camelinas's oil eliminating the acid erucic content by targeting the three CsFAE1 gene present in the camelina's genome. To accomplish this goal, a CRISPR-Cas9 editing system strategy has been chosen. In this work, we have proceed to the design and cloning of the target and guiding sequences into a pHSE401 vector containing the Cas9 protein. A PCR splicing strategy for cloning the target sequences into the pHSE401 vector has been used. The results show a correct assembly of the editing vector prior to the transformation of camelina plants. Moreover, previous studies show that the use of CRISPR-Cas9 to edit genes involved in seed oil fatty acid biosynthesis in plants is possible. The elimination of erucic acid would open the opportunity to the definitive use of camelina oil as a source of omega-3 for the agri-food sector.

2. INTRODUCCIÓN

La manipulación genética de los aceites vegetales es un objetivo prioritario para los biotecnólogos debido a sus múltiples aplicaciones. El uso principal de los aceites obtenidos de semillas o frutos de plantas es el uso alimentario, bien para el cocinado de los alimentos o en crudo como aditivo, como es el caso del aceite de oliva. Sin embargo, los aceites vegetales también tienen una gran cantidad de usos industriales, tales como componentes de pinturas, barnices, surfactantes, antidispersantes en tintas de impresora, líquidos hidráulicos o más recientemente para producir biocombustible (Jaworski & Cahoon, 2003).

2.1 Camelina sativa como fuente de origen vegetal de ácidos grasos omega-3

Camelina (*Camelina sativa*) es una planta perteneciente a la familia *Brassicaceae* cuyo cultivo despertó el interés del hombre ya en la época griega y romana, precisamente porque su aceite se utilizaba para usos medicinales o como combustible de lámparas. Se trata de una planta común en gran parte de Europa y es también conocida como falso lino. Mide entre 30 y 120 cm de altura, tiene las hojas lanceoladas y su fruto es una pequeña silicua que alberga entre 8 y 15 semillas (Miguel Gutiérrez López & Antonio Albalat Borrás, 2013). Camelina presenta excelentes propiedades agronómicas: es un cultivo oleaginoso natural, anual, de elevada rusticidad y adaptabilidad con capacidad de producción en zonas con baja precipitación. Además es un cultivo económico ya que tiene un ciclo de vida corto, reproducción autógama, requerimientos nutricionales bajos y una resistencia innata a patógenos microbianos y plagas de insectos (Zubr, 1997). Estas características hacen que su cultivo pueda constituir una alternativa en Aragón a nuestros secanos áridos, pudiendo integrarse en la rotación y sistema de cultivos de nuestros cereales de invierno con escasas necesidades de insumos. De hecho en la región el cultivo de camelina se ha extendido en los últimos años, sobre todo en comarcas del bajo Aragón (Miguel Gutiérrez López & Antonio Albalat Borrás, 2013).

Pero sin duda el gran interés del cultivo de Camelina viene dado por el alto contenido de aceite en su semilla, que puede llegar a un 43% del peso de la misma, así como a la inusual acumulación de ácido linolénico (18:3 u omega-3) en dicho aceite, que puede llegar a

representar un 37% del total de ácidos grasos (Zubr, 1997) (Figura 1). Esta elevada presencia de ácido linolénico omega-3 solo se encuentra en la naturaleza en dos especies, la propia camelina y el lino (*Linum ussitatissimum*), aunque éste no acumula tanto aceite como la camelina en su semilla. Esto ha provocado que la camelina haya emergido como una planta modelo para la investigación en el aceite de semilla y su producción a gran escala. Además, su proximidad genética con la planta modelo *Arabidopsis thaliana* (ambas son brassicáceas) y su facilidad para ser transformada por métodos similares han colocado a la camelina como una especie clave en los estudios sobre biosíntesis de aceite en la semilla de plantas. Su genoma está anotado y secuenciado (Kagale et al., 2014) y está disponible en (<http://www.camelinadb.ca/index.html>) lo que proporciona una clara ventaja desde el punto de vista de su manipulación genética.

En los últimos años el aceite de camelina se ha utilizado para la producción de biocombustibles, concretamente bioqueroseno presente en aviación civil (vuelo KLM-B474 en noviembre de 2019 o IBERIA Airbus 320 en octubre de 2011) o militar (US Navy Superhornet F18, abril 2010) (Miguel Gutiérrez López & Antonio Albalat Borrás, 2013).

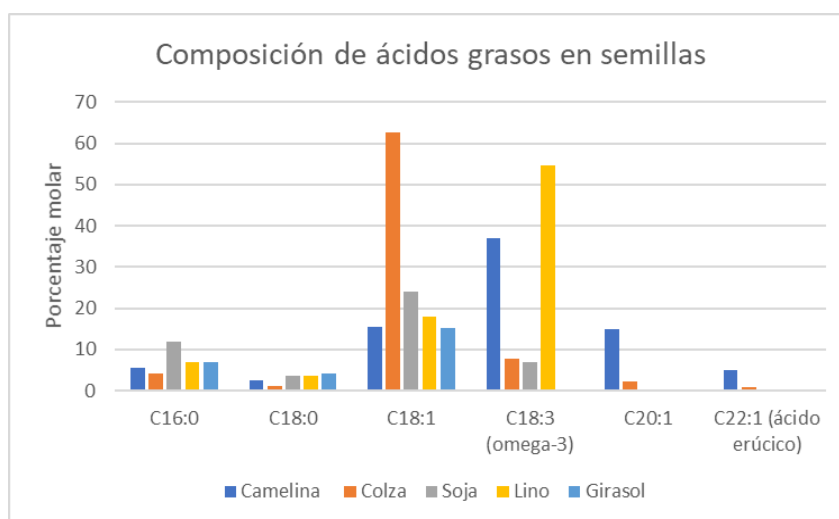


Figura 1 Composición de ácidos grasos del aceite de semilla de diferentes plantas. Datos extraídos de la Tabla 1 de (Zubr, 1997)

Los ácidos grasos omega-3 (18:3) presentes en el aceite de camelina, forman parte de los denominados ácidos grasos saludables por los beneficios de su ingesta en la dieta sobre el sistema cardiovascular y como precursores de dos derivados del omega-3 como son el EPA (ácido eicosapentanoico (20:5) y DHA (ácido docosahexaenoico (22:6). Estos derivados son necesarios para el correcto desarrollo del sistema neural y son potentes antiinflamatorios vía la generación ya sea de agentes antiinflamatorios o a través del bloqueo de agentes proinflamatorios. Además, la evidencia sugiere su tratamiento para enfermedades como la artritis reumatoide, enfermedades neurodegenerativas y enfermedades inflamatorias intestinales (**Cassileth, 2010**). Esto hace que en el mercado existan gran cantidad de productos en los que los ácidos grasos omega-3 y sus derivados se incluyen como aditivos. Las fuentes de estos ácidos grasos suelen ser peces o pequeños crustáceos, ricos en omega-3. Tanto la materia prima de origen como su procesamiento, encarecen el coste de producción y su uso. El aceite de camelina podría ser una clara alternativa de gran potencial por el origen vegetal del aceite. Sin embargo, su uso alimentario como fuente de ácidos grasos saludables omega-3 se ha visto dificultado por el hecho de que no solo contiene omega-3 sino otros ácidos grasos como el ácido erúxico (22:1), en cantidades que pueden llegar hasta el 3% (**Zubr, 1997**).

2.2 La ingestión de ácido erúxico puede acarrear problemas a la salud

La presencia de ácido erúxico supone un obstáculo para la generalización de la comercialización del aceite de camelina como aditivo alimentario y fuente de ácidos grasos saludables omega-3. El corazón es el principal órgano que padece los efectos tóxicos derivados de una exposición crónica. El efecto más crítico deriva en cuadros clínicos de lipodosis miocárdica. Es decir, los ácidos grasos se acumulan en el tejido cardíaco lo que provoca una disminución de la contractilidad del miocardio generando posteriores problemas cardiovasculares.

Así pues, el 21 de septiembre de 2016, la Comisión Técnica Científica de Contaminantes de la Cadena Alimentaria (CONTAM) de la Autoridad Europea de Seguridad Alimentaria (EFSA) adoptó un dictamen científico sobre la presencia de ácido erúxico y sus

consecuencias en los piensos y alimentos (**EFSA Panel on Contaminants in the Food Chain (CONTAM) et al., 2016**). En referencia a la toxicidad, la evaluación afirma que el corazón es el principal órgano que padece los efectos tóxicos derivados de una exposición crónica prolongada. El efecto más crítico deriva en cuadros clínicos de lipidosis miocárdica, que es reversible y transitoria. Además, la incidencia es diferente dependiendo de algunos parámetros como la especie, la edad y la capacidad de metabolizar el ácido erúxico. Además, se establece una ingesta diaria tolerable (IDT) de ácido erúxico de 7 mg/kg de peso corporal (pc) por día. Se detectó que los lactantes y otros niños con niveles de exposición alimentaria por encima de la IDT presentaban el nivel de exposición más alto y por tanto eran la población *target* con más riesgo. Si miramos la presencia de este compuesto en los alimentos, la EFSA especifica que la bollería es el grupo de alimentos que más contribuye a la exposición dietética al ácido erúxico (240-390 mg/kg) incluso en los grupos de niños, con medias de 21-28% del total de la exposición. Por último, con los datos disponibles no se pueden sacar conclusiones sobre la genotoxicidad y carcinogenicidad del ácido erúxico (**EFSA Panel on Contaminants in the Food Chain (CONTAM) et al., 2016**).

Con respecto a la ingesta y los efectos del ácido erúxico en los animales, La EFSA ha establecido un nivel de efectos adversos no observados (NOAEL) de 700 mg/kg de peso corporal en cerdos. Sin embargo, la exposición alimentaria en cerdos está por debajo de dicho nivel. En el caso de las aves de corral, se ha establecido un nivel bajo donde se observan efectos adversos de 20 mg/kg de peso corporal y día, valor que representa el doble del máximo de exposición (12 mg/kg de peso corporal y día). Esto da un pequeño margen de seguridad que podría indicar un riesgo para pollos en el caso de ingestas elevadas. No existen valores de referencia para las demás especies, pero lo que se ha evidenciado es que el ácido erúxico ingerido por los animales es transferido a los alimentos de origen animal y que esta transferencia es proporcional al nivel de ingesta (**EFSA Panel on Contaminants in the Food Chain (CONTAM) et al., 2016**).

Debido a las conclusiones de la evaluación de la EFSA explicada anteriormente, se adopta en noviembre de 2019 el Reglamento (UE) 2019/1870 que postula nuevos límites máximos de ácido erúxico presentes en algunos alimentos (Tabla 1).

Tabla 1. Contenido máximo de ácido erúxico en determinados alimentos (Reglamento (UE) 2019/1870)

Productos alimenticios ⁽¹⁾		Contenido máximo (g/kg)
«8.1.	Ácido erúxico, incluido el ácido erúxico presente en grasas	
8.1.1.	Aceites y grasas vegetales comercializados al consumidor final o para su uso como ingredientes alimentarios, a excepción del aceite de camelina, el aceite de mostaza y el aceite de borraja	20,0
8.1.2.	Aceite de camelina, aceite de mostaza(*) y aceite de borraja	50,0
8.1.3.	Mostaza (condimento)	35,0

⁽¹⁾ Con la aceptación de la autoridad competente, el contenido máximo no se aplicará al aceite de mostaza producido y consumido a escala local.»

Así pues, el contenido máximo en aceites y grasas vegetales es de 20g/kg exceptuando el aceite de camelina, mostaza y borraja que se mantiene en 50 g/kg. El mismo Reglamento (UE) 2019/1870 da una explicación de porque en el caso del aceite de camelina no se ha reducido su contenido máximo de erúxico como sí se ha hecho para la categoría de aceites vegetales:

“La presencia de ácido erúxico en vegetales indica que conviene reducir su contenido mediante diferentes técnicas enfocadas a la utilización de variedades con bajo contenido de este ácido graso. Sin embargo, en el caso del aceite de camelina, el aceite de mostaza y el aceite de borraja, se han aportado pruebas que demuestran que no es posible alcanzar un contenido más bajo mediante la aplicación de buenas prácticas, puesto que no existen variedades de estas especies cuyos aceites vegetales extraídos de estas plantas presenten un contenido de ácido erúxico inferior al contenido máximo que se ha propuesto para los demás aceites vegetales. Por lo tanto, y dado que estos aceites no tienen el mismo impacto en la exposición humana que otros aceites vegetales, el contenido máximo de ácido

erúcico en aceite de camelina, aceite de mostaza y aceite de borraja debe seguir siendo el mismo".

Sin embargo, sería aconsejable poder conseguir una reducción de los niveles de ácido erúcico en el aceite de camelina. Sin la existencia de variedades naturales de camelina bajas en ácido erúcico, la manipulación genética nace como una alternativa para conseguir este objetivo.

2.3 Los genes y proteínas implicados en la biosíntesis de omega-3 y erúcico en el aceite de semilla de camelina son conocidos

El mecanismo de biosíntesis de aceite en la semilla de camelina no es diferente al de otras especies con las que está filogenéticamente emparentada, tales como *Arabidopsis*. Como ya se ha comentado, ambas pertenecen a la familia de las brasicáceas y presentan un desarrollo similar. En ambos casos se dispone de información genética (ambos genomas están secuenciados y anotados) y los genes y proteínas que participan en esta ruta son conocidos y están caracterizados (**Kagale et al., 2014**).

El triacilglicerol o TAG es el componente predominante del aceite de las semillas. La molécula está formada por una molécula de glicerol esterificado con tres ácidos grasos. La función y el valor de los aceites vegetales dependerá de la composición de ácidos grasos del TAG. La vía central de biosíntesis de TAG en semillas es la ruta *de novo* o ruta de Kennedy (**Bates & Browse, 2012**) que utiliza directamente los ácidos grasos sintetizados y exportados desde el plastidio. De forma breve, la síntesis de aceite en la semilla se produce por la incorporación de tres grupos acilo-ácidos grasos esterificados a una molécula de CoA al esqueleto de glicerol mediante las reacciones ilustradas en la Figura 2.

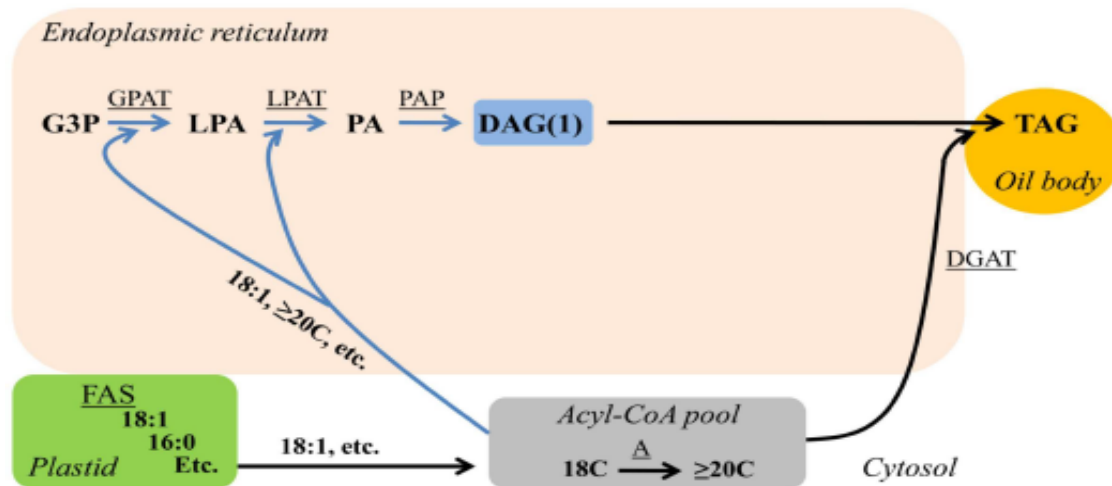


Figura 2. Biosíntesis de aceite en semilla por la ruta de Kennedy (Bates & Browse, 2012)

En primer lugar, se produce la acilación del gliceraldehído 3-fosfato (G3P) por la glicerol-3-fosfato aciltransferasa (GPAT) en la posición *sn*-1 del esqueleto de glicerol dando como producto ácido liso-fosfatídico (LPA). A continuación, la segunda acilación en la posición *sn*-2 del esqueleto de glicerol es realizada por la ácido lisofosfatídico aciltransferasa (LPAT) generando ácido fosfatídico (PA). Después de la segunda acilación, actúa la ácido fosfatídico-fosfatasa (PAP) que defosforila el PA para formar diacilglicerol (DAG). La tercera acilación es llevada a cabo por la diacilglicerol aciltransferasa (DGAT) para dar lugar al producto final, el triacilglicerol (TAG). A pesar de considerar la ruta de Kennedy como la vía central de biosíntesis de TAG en semillas, existen otras rutas interconectadas a ésta en la que están implicadas diversas enzimas (Claver Lacasa, 2018).

Generalmente los reservorios de grupos acilo contienen ácidos grasos de 18 átomos de carbono. Sin embargo, algunas plantas como la camelina, son capaces de introducir dobles enlaces en esos ácidos grasos a través de la actividad de las omega-6 (FAD2) y omega-3 (FAD3) desaturasas resultando en un incremento del contenido en ácidos grasos omega-3 en el aceite. Además, como se ha comentado anteriormente, camelina acumula cantidades de ácido erúico que oscilan entre un 2,5-5% del contenido total de ácidos grasos. Esto es posible gracias a la enzima *Fatty Acid Elongase* (FAE1) que es capaz de elongar los ácidos grasos de 18 átomos de carbono de los reservorios de grupos acilo para producir tanto 20:1-CoA como 22:1-CoA (ácido erúico) que posteriormente son incorporados a TAG.

Así pues, podemos definir tres grupos distintos de especies vegetales en función de la composición química de su aceite: por un lado aquellas especies que acumulan ácido oleico (18:1) como componente mayoritario, aunque pueden presentar cantidades relevantes de ácido linoleico (18:2). Es el caso de el olivo, la soja o el girasol, entre otras. En otro grupo de plantas, las omega-6 y omega-3 desaturasas presentan una mayor actividad, acumulando altas cantidades de 18:2 y sobre todo 18:3. Es el caso de la camelina o el lino. Por último, en un tercer grupo de especies vegetales, la elongasa FAE1 es capaz de elongar de forma eficiente el 18:1 para sintetizar 20:1 y 22:1. En este tercer grupo de especies podemos encontrar la colza (*Brassica napus*), la jatropha (*Jatropha curcas*), Crambe (*Crambe abissinica*), el Pennycress (*Thlaspi arvense*) o la planta de este estudio (*Camelina sativa*). En la figura 3 se representa, de forma esquemática las tres rutas de modificación de los ácidos grasos. Es importante señalar que, en todas ellas, el 18:1 es el sustrato común.

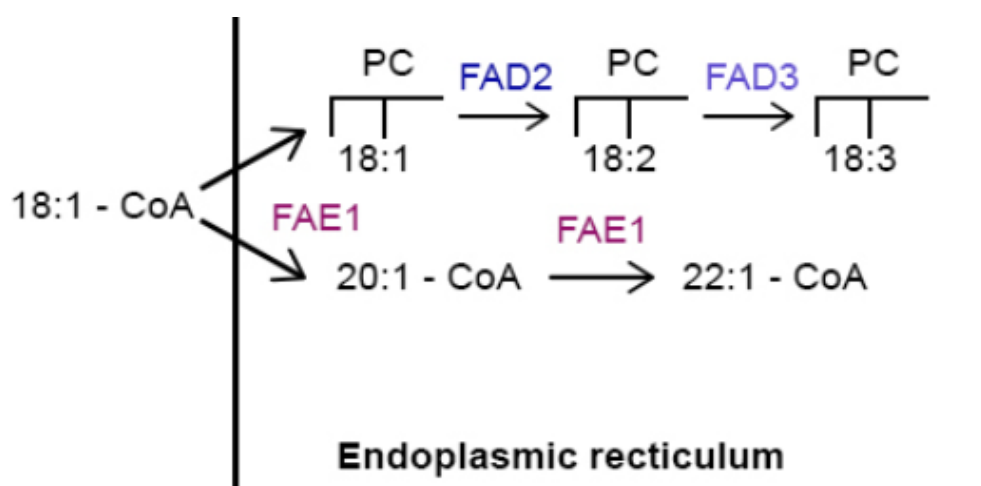


Figura 3. Rutas involucradas en la síntesis de omega-3 (18:3) y ácido erúcido (22:1). Imagen adaptada de (Claver Lacasa, 2018)

En algunas brasicáceas como *Arabidopsis* o colza (*Brassica napus*) se han obtenido mutantes deficientes en la actividad elongasa que es la responsable de la biosíntesis de erúcido. Estos mutantes en los que el gen *FAE1* se encuentra silenciado o es inactivo, muestran una ausencia total de erúcido en el aceite y un incremento concomitante de ácido graso oleico (Roscoe, Lessire, Puyaubert, Renard, & Delseny, 2001). En el caso de la colza, estas variedades sin erúcido como consecuencia de la mutación del gen *FAE1* se denomina canola y son utilizados

en la alimentación animal. Esto sugiere que la inactivación del gen *FAEI* es posible y permitiría eliminar o reducir drásticamente el contenido en erúxico del aceite de semilla.

2.4 La tecnología CRISPR/Cas9 como herramienta de edición genética

La herramienta CRISPR/Cas9 deriva de un mecanismo de inmunidad adaptativa que poseen determinadas bacterias para defenderse de los bacteriófagos. Algunas bacterias son capaces de eliminar y reparar un ADN de origen viral insertado en su genoma utilizando una endonucleasa, Cas9, que reconoce una serie de secuencias palindrómicas repetidas o CRISPs (del inglés *Clustered Regularly Interspaced Short Palindromic Sequences*). Este genoma viral se inserta en el genoma bacteriano como unidades repetitivas. La transcripción de estas unidades repetidas es reconocida por Cas9, que utiliza estos propios ARNs como guía para localizar ADN viral cuando se produce una nueva infección. El ARNs por complementariedad se unirá al ADN viral y quedará marcado para que la Cas9 corte el ADN viral y se elimine de tal forma que la infección no prospera. Así pues, a partir de este sistema de inmunidad adaptativa propio de algunas bacterias se desarrolló la herramienta de edición genética conocida como CRISPR/Cas9 (del inglés *Clustered Regularly Interspaced Short Palindromic Repeats*). El sistema CRISPR/Cas9 es simple y requiere solo dos componentes: la endonucleasa Cas9 y un ARN guía (ARNg o *sgRNA* en inglés). La secuencia de este ARNg determina la especificidad del evento de edición, es decir, qué gen concreto vamos a editar y cómo.

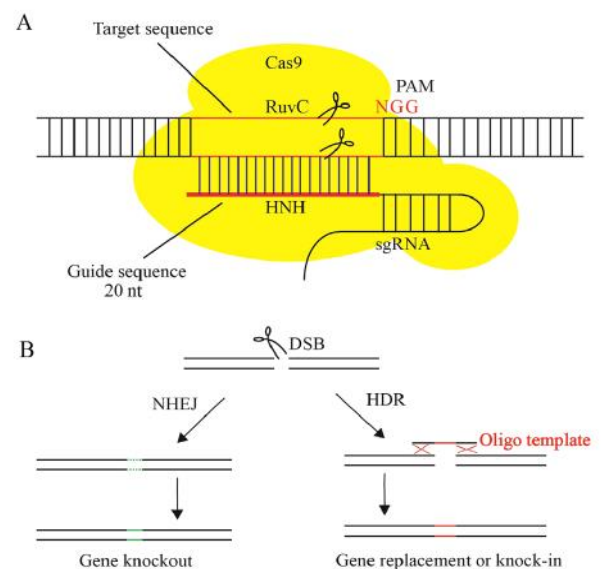


Figura 3. Diagrama esquemático del sistema de edición CRISPR/Cas9. (A)- boceto del sistema CRISPR/Cas9. (B)- Los dos mecanismos por los cuales se reparan los DSB (Liu, X. et al., 2017).

El ARNg es un ARN sintético de aproximadamente 100 nucleótidos. Su extremo 5' posee 20 nucleótidos que actúan como secuencia guía para identificar el gen diana (**Liu, X., Wu, Xu, Sui, & Wei, 2017**). Además, va seguido de la secuencia PAM (del inglés *Protospacer Adjacent Motif*) que está formada por el consenso NGG (N-cualquier nucleótido; G-Guanina). El brazo que se forma en el extremo 3' del ARNg puede “anclar” el gen diana con la secuencia guía y formar un complejo con la Cas9 (Figura 4A). La endonucleasa corta la doble cadena de ADN y se forma así la rotura de las dos hebras (*double-strand break* o DSB). Una vez se forma el DSB se pueden producir dos mecanismos posibles de reparación del ADN: la reparación por unión de extremos no homólogos (NHEJ) o la reparación por recombinación homóloga (HR) (Figura 4B). La NHEJ es una forma sencilla para producir *mismatches*, deleciones o inserciones lo que provoca una eliminación del gen. Cuando en cambio está presente un cebador como molde, HDR induce un reemplazo de los genes o un reemplazo con el gen foráneo (**Liu, X. et al., 2017**).

Esta novedosa herramienta de edición genética supone una clara ventaja frente a otras tecnologías actuales basadas en inserción de T-DNA o uso de mutágenos que no son específicos, pudiendo afectar a varios genes a la vez. Estos sistemas generan muchos defectos indeseados a nivel fenotípico. Además la tecnología CRISPR es de manipulación sencilla, no se necesitan muchos reactivos, es específica (reducción de mutaciones *off target*) y rápida.

2.5 La tecnología CRISPR/Cas9 permitiría la inactivación simultánea de los tres genes CsFAEI de camelina

El análisis del genoma de camelina muestra que, de manera similar a otros genes implicados en el metabolismo de lípidos, el gen *CsFAEI*, que codifica la elongasa responsable de la síntesis de ácidos grasos de 20 y 22 átomos de carbono, muestra tres copias del gen en el genoma de camelina. La existencia de tres copias de estos genes está asociada al hecho de que la camelina es un hexaploide (**Kagale et al., 2014**) y como consecuencia los tres genes se expresan en la semilla. La naturaleza hexaploide del genoma y la existencia de varias copias o isoformas funcionales de los genes, dificulta mucho la manipulación genética en camelina. Intentos de abordar la manipulación genética de camelina utilizando tecnologías como los

ARNi obtuvieron algunos resultados (genes *CsDGATI*, (Liu, J. et al., 2015)), aunque no pudieron alterar los tres genes ni eliminar la posibilidad de efectos “off-target” inherentes a esta técnica. Otras tecnologías como TALENs, usadas en biología animal, no han sido adoptadas por la comunidad vegetal por la necesidad de construir grandes proteínas modulares.

Así pues, cabe esperar que la tecnología CRISPR/Cas9 permita la edición de los tres genes *CsFAEI* a la vez dada la especificidad de la secuencia y así solventar los problemas que generan otras técnicas de edición genética mencionadas anteriormente. Mediante esta tecnología se podría introducir una delección en el gen *CsFAEI* produciendo un mutante por pérdida de función. Tal y como se muestra en la Figura3, la edición simultánea de los tres genes *CsFAEI* produciría un descenso o incluso la eliminación total del ácido graso erúxico. En paralelo, cabría esperar que como consecuencia de la mayor disponibilidad de 18:1 acil-CoA para la desaturación, se podría producir un incremento de 18:3 (omega-3).

3. OBJETIVOS DEL ESTUDIO

Camelina (*Camelina sativa*) es una planta de elevado potencial tanto biotecnológico como agronómico. Como se ha comentado, el aceite de su semilla es muy codiciado por su alto contenido en ácidos grasos omega-3. Sin embargo, la presencia de ácido erúcico obstaculiza su uso a gran escala como aditivo alimentario. Por lo tanto, buscar alternativas como la manipulación genética mediante CRISPR del gen que codifica el ácido erúcico se presenta como una opción prometedora.

El objetivo fundamental de este proyecto es modificar genéticamente mediante la tecnología CRISPR/Cas9 la composición química del aceite de *Camelina sativa*, eliminando o reduciendo drásticamente el contenido en ácido erúcico 22:1.

El estudio de este objetivo principal se desarrolla a partir de los siguientes objetivos específicos:

1. Diseño y construcción del vector conteniendo las dianas, los ARNs de guiado y la nucleasa Cas9, dirigidos contra los tres genes *CsFAEI* del genoma de *Camelina sativa*.
2. Transformación de líneas comerciales silvestres de camelina con el vector.
3. Segregación y caracterización de las líneas editadas. Análisis fenológico de su crecimiento, producción de semilla y aceite. Análisis químico de la composición de ácidos grasos del aceite de semilla en las plantas editadas.

En este Trabajo de Fin de Máster nos hemos centrado en el primer objetivo.

4. METODOLOGÍA

4.1 Crecimiento de plantas de *Camelina* (*Camelina sativa* L.)

Las plantas de *Camelina* utilizadas en este trabajo fueron obtenidas de la empresa *Camelina Company-España*, líder en la distribución y obtención de *Camelina* en España.

De forma habitual, las semillas de *Camelina* se germinaron en una placa de Petri sobre sustrato de papel Whatman humedecido con agua destilada y mantenido en la cámara climática. Tras una semana, fueron visibles las primeras plántulas que se trasvasaron a pequeñas macetas con una mezcla de sustrato (Turba Kekkila) y vermiculita en proporción 3:1. Las plantas se cultivaron en cámara climática (Fitoclima 10000 EHHF Aralab) bajo condiciones controladas de temperatura ($22^{\circ}\text{C}/18^{\circ}\text{C}$ día/noche); humedad relativa (45%); intensidad de luz ($150\text{-}200 \mu\text{moles m}^{-2} \text{s}^{-1}$) y con un fotoperíodo de 16/8 h de luz/oscuridad. Las plantas de *Camelina* mantuvieron el crecimiento vegetativo durante unas 6-8 semanas transcurridas las cuales aparecieron las inflorescencias y posteriormente (2 semanas adicionales) aparecieron las vainas que contenían las semillas.



Figura 5. Fotografías de las plantas de *Camelina* (*Camelina sativa*), crecidas en la cámara de cultivo de la EEAD-CSIC. En la foto se muestran diferentes estadios desde las etapas iniciales de crecimiento, la posterior formación de inflorescencias y la aparición de las vainas en las que se acumulan las semillas.

4.2 ANÁLISIS LIPÍDICO DE LAS SEMILLAS DE CAMELINA

4.2.1 Extracción de lípidos totales

Los lípidos totales fueron extraídos de semillas de *Camelina sativa* tal y como está descrito por **(Bligh & Dyer, 1959)**. En primer lugar, se homogeneizaron 0,5 g de material vegetal fresco en mortero de cerámica con N₂ líquido y seguidamente se realizó la extracción o se congeló el material a una temperatura de -80 °C hasta su uso. A cada muestra se le añadieron 2 ml de cloroformo:metanol 1:2 (v/v), y se agitó. Este proceso se repitió tres veces. Seguidamente, se añadieron 3 ml de KCl 0,7% (p/v), se agitó y centrifugó a 150 g, durante 5 min. La fase inferior lipídica se recogió con una pipeta Pasteur. A partir de la fase superior acuosa se realizó una segunda extracción del mismo modo con 3 ml de cloroformo y se volvió a recoger la fase inferior lipídica con una pipeta Pasteur mezclándola con la obtenida anteriormente. Los lípidos se evaporaron a sequedad, bajo una corriente de N₂ (gas) y se disolvieron en 1 ml de cloroformo. Se tomaron dos alícuotas de 150 µl para el análisis de ácidos grasos totales.

4.2.2 Obtención de ésteres metílicos de ácidos grasos totales

Para analizar la composición de los ácidos grasos totales, estos fueron transformados en sus correspondientes ésteres metílicos mediante una reacción de transesterificación. Para ello, se utilizó el método de transmetilación ácido catalizada descrito por **(Garcés, Sarmiento, & Mancha, 1992)**. Mediante esta reacción se consigue la metilación tanto de los ácidos grasos libres (AGL) como de los ácidos grasos que forman parte de los glicerolípidos. Las alícuotas de lípidos totales se evaporaron a sequedad bajo una corriente de N₂ (g) y se añadieron 3 ml de mezcla (metanol:tolueno: H₂SO₄, 88:10:2 v/v/v). En algunos casos se añadieron 40 µl de ácido heptadecanoico (17:0), 0,1 µg/µl en metanol:tolueno (1:1 v/v) como patrón interno. Las muestras fueron incubadas en tubos de vidrio herméticamente cerrados, a 80 °C durante 1 h y se dejaron enfriar. Para la extracción de los ésteres metílicos formados se añadieron 5 ml de Na₂SO₄ 6,7 % (p/v) y 1 ml de heptano. Tras agitar, las muestras fueron centrifugadas a 1500 g durante 2 min y se recogió la fase superior. Este paso se repitió 2/3 veces y se mezclaron las fases superiores en otro tubo para evaporar a sequedad bajo corriente de N₂ (g). Esta fracción, conteniendo los ésteres metílicos de los ácidos grasos, se disolvió en 50-100 µl de

heptano y se introdujo en viales con inserto de vidrio para ser analizado por cromatografía de gases.

4.2.3 Análisis de la composición de ácidos grasos mediante cromatografía gas-líquido

Como ya se ha explicado en el apartado anterior, los ésteres metílicos de ácidos grasos de los lípidos totales producidos por transesterificación catalizada, fueron analizados por cromatografía de gases (GC). En esta técnica los compuestos que se quieren separar deben ser volatilizados y transportados por una corriente de gas inerte (fase móvil) a través de una columna que contiene la fase estacionaria líquida unida a un soporte sólido. El residuo seco de los ésteres metílicos de ácidos grasos totales, o de cada clase de lípido, se disolvió en 50-100 µl de heptano y se introdujo en viales con inserto de vidrio. Las muestras se analizaron en un cromatógrafo de gases modelo 7890A (Agilent Technologies, Santa Clara, CA USA) equipado con una columna capilar (60m de longitud, 0,25 mm de diámetro interno, 0,5 µm de espesor de película) de sílice fundida (Supelco, Bellefonte, PA USA) y con un detector de ionización de llama (FID). El gas portador utilizado fue el helio, con un flujo lineal de 1,2 ml/min y una relación de Split de 1/100. La temperatura del inyector fue 250 °C y la temperatura del detector fue 260 °C. La temperatura del horno fue modificada de tal manera: 170 °C durante 30 min, la temperatura subió a 5 °C/min hasta 200 °C. Para la identificación de los picos del cromatograma, se utilizaron patrones que contenían diferentes ácidos grasos conocidos, previamente transesterificados, como el aceite de girasol, el aceite de lino alto oleico o un extracto de espinaca y se compararon los tiempos de retención de los picos de las muestras con los identificados anteriormente con los patrones.

4.3 Descripción del vector pHSE401

El plásmido utilizado para insertar las secuencias diana junto con el ARNg y transformar las plantas de *Camelina* fue el vector binario de transformación pHSE401 (Xing et al., 2014), obtenido de forma comercial en la empresa Addgene. Este plásmido contiene los elementos necesarios para la transformación de plantas de *Camelina* por la tecnología CRISPR/Cas9 (Figura 6).

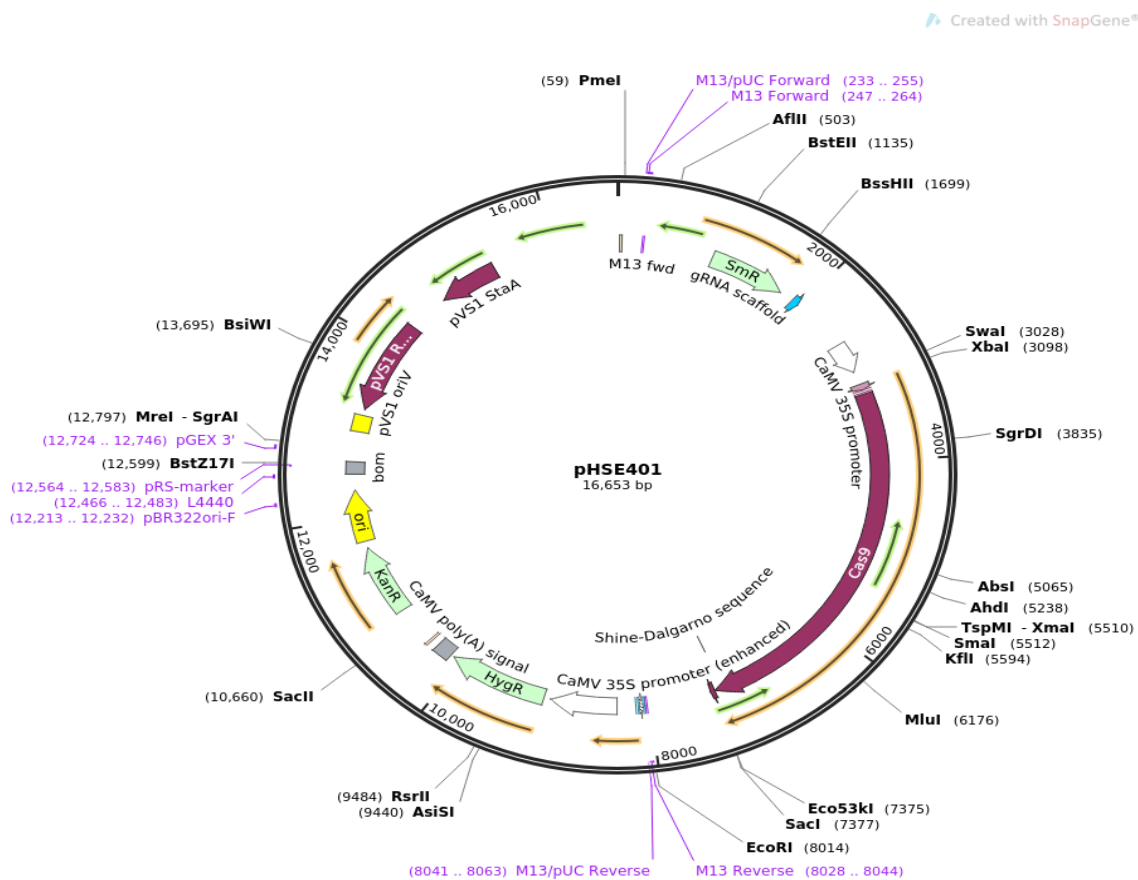


Figura 6. Imagen del vector pHSE401 utilizado en este trabajo. En él se pueden ver todos los elementos junto con sus dianas de restricción. pHSE401 was a gift from Qi-Jun Chen (Addgene plasmid # 62201 ; <http://n2t.net/addgene:62201> ; RRID:Addgene_62201)

Por un lado la secuencia diana junto con el gen que codifica el ARNg se encuentran bajo el control del promotor de *Arabidopsis* U6-26. Nuestra secuencia diana se insertará en sustitución del gen de resistencia a la espectinomicina. El gen que codifica la Cas9 se encuentra bajo el control del promotor 35S. El marcador de selección que utiliza es la higromicina fosfotransferasa (HygR) que confiere resistencia a la higromicina y está bajo el control del promotor CaMVd35S.

4.4 Extracción del ADN plasmídico y digestión con enzimas de restricción

Para extraer el ADN plasmídico (pHSE401) de colonias individuales crecidas en medio LB sólido se cultivaron en 3 ml de medio LB líquido con marcador selectivo (kanamicina 100 ug/ml) a 37 °C, durante toda la noche y con agitación (200 rpm). Al día siguiente, entre 3-5 ml de cultivo fueron centrifugados a 14000 g para precipitar las células y extraer el ADN plasmídico. La extracción fue mediante el método de lisis alcalina y la purificación en columna, utilizando el kit comercial GenJET Plasmid Miniprep Kit (Fermentas, Thermo ScientificTM), siguiendo las instrucciones del fabricante. El ADN plasmídico obtenido se cuantificó espectrofotométricamente (apartado 4.9). Las muestras obtenidas fueron seguidamente utilizadas o almacenadas a -20 °C hasta su utilización.

Una vez extraído el ADN plasmídico, se realizó su digestión mediante el uso de enzimas de restricción (en nuestro caso *XbaI* y *AflIII*; Fermentas, Thermo Scientific). Todas las digestiones se realizaron partiendo de 1 µg de ADN y utilizando un volumen de enzima equivalente a 5-20 U/µg de ADN, según las condiciones determinadas por el programa DoubleDigest Calculator, en la página web de la casa Thermo ScientificTM. Generalmente las digestiones se realizaron a 37 °C durante 1 h 30 min.

4.5 Técnica de PCR

La reacción en cadena de la polimerasa (PCR, *polymerase chain reaction*) consiste en un primer paso de activación de la enzima polimerasa a 95-98 °C durante 5 min, seguida de un número de ciclos, entre 25 y 40, dependiendo del tamaño del fragmento a amplificar. Los ciclos se componen de tres etapas: una primera etapa de desnaturalización a 95-98 °C durante 10-30 s dependiendo de la enzima polimerasa, otra de hibridación de los cebadores al DNA molde durante 30 s y otra de extensión a 72 °C durante 30 s/kb. Todos los programas finalizaron con una etapa de extensión a 72 °C durante 10 min. La temperatura de hibridación (Th) es específica de cada pareja de cebadores y depende de su temperatura de fusión (Tm). Por otro lado, el tiempo de extensión depende de la longitud del fragmento que se quiere amplificar.

En este trabajo se utiliza la *PCR-driven overlap extension* o *PCR splicing* para la inserción

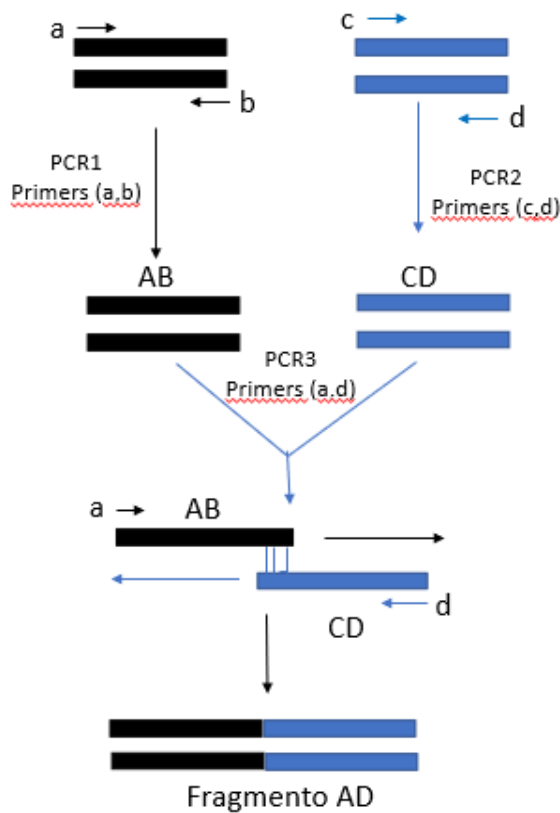


Figura 7. Diagrama esquemático de la *PCR-driven overlap extension* o *PCR splicing* (elaboración propia)

de la secuencia diana junto con el ARNg, usando el vector pHSE401 como molde, que indicará a la endonucleasa Cas9 el sitio por donde debe cortar el ADN (gen *FAEI*). Esta variante de la PCR tradicional es idónea para clonar fragmentos de genes grandes amplificando primero los dos extremos por separado para después empalmar los dos fragmentos en una tercera PCR y así generar el gen completo. Es decir, en un primer paso se amplifican los extremos del gen, usando los cuatro primers (a,b,c y d) (Figura7). Por un lado tenemos la PCR1 (a-b) y por otro la PCR2 (c-d). Posteriormente, cuando las hebras de los amplicones AB y CD se desnaturalizan en la PCR3, aquellas que tengan un extremo 3' libre se unen entre ellas por complementariedad de bases. En esta PCR3 los cebadores a y d permitirán la amplificación y se generará el producto completo AD (gen completo) (Figura 7).

En este trabajo se utilizó la Phusion® High Fidelity DNA Polymerase (Finnzymes, Thermo Scientific). Dicha polimerasa presenta actividad 5'→3' exonucleasa por lo que permite una amplificación más fiable y precisa debido a su actividad correctora de errores. Esta característica es muy importante puesto que necesitamos una enzima que genere el mínimo número de mutaciones espontáneas ya que es primordial copiar sin errores las secuencias diana. En la tabla 2 se detallan los componentes y concentraciones de la mezcla de reacción para cada una de las PCRs secuenciales de la *PCR-driven overlap extension*.

Tabla 2. Componentes de la mezcla de reacción de cada una de las PCRs realizadas necesarias para la inserción de la secuencia diana junto con el ARNg

Reactivos	PCR1B	PCR2B	PCR3B
pHSE401	0,5µl	0,8µl	-
PCR1B	-	-	2µl
PCR2B	-	-	2µl
Tampón 5x Phusion GC	4µl	4µl	4µl
Primer F_PCR1	2µl	-	2µl
Primer R_PCR1A	-	-	-
Primer R_PCR1B	2µl	-	-
Primer F_PCR2A	-	-	-
Primer F_PCR2B	-	2µl	-
Primer R_PCR2	-	2µl	2µl
dNTPs	1µl	1,5µl	1,5µl
Phusion® High Fidelity DNA Polymerase	0,8µl	0,8µl	1µl
H ₂ O (hasta volumen total de 20 µl)	9,7µl	9,8µl	6,5µl

En el caso de la polimerasa Phusion® High Fidelity DNA Polymerase (Finnzymes, Thermo Scientific) la temperatura de desnaturalización es de 94 °C, el tiempo de extensión de 15 s/kb y para calcular la T_h se utilizó el programa ofrecido por la casa comercial <http://www.thermoscientificbio.com/webtools/tmc/>, ya que esta polimerasa necesita unas consideraciones especiales de T_h . Todas las amplificaciones se realizaron en un termociclador GeneAmp® PCR System 9700 (Applied Biosystems, Thermo Scientific).

4.6 Cebadores utilizados en este trabajo

En este apartado se muestran los primers utilizados en las reacciones de amplificación de la *PCR splicing* (Tabla 3). La síntesis de los cebadores se encargó a la casa comercial Isogen Life Science.

Tabla 3. Secuencia de los primers utilizados en las reacciones de amplificación para la PCR splicing. También se muestra la temperatura de fusión de cada primer (T_m).

Gen	PCR splicing	Cebador	Secuencia 5'-3'	T_m (°C)
	PCR1	F_PCR1	CCAGTGCCAAGCTTCGACTTGCC	56
		R_PCR1A	GCAAGTAACGCCGTTAACGGAATCACTACTTCGACTCTAG	63
<i>CsFAE1</i>		R_PCR1B	TCAGGATCCAACCCAGTTTCAATCACTACTTCGACTCATG	61
	PCR2	F_PCR2A	CCGTTAACGGCGTACTTGCCTTTAGAGCTAGAAATAG	60
		F_PCR2B	CAACATCAAAAAGCTTTAGTCTGTTTTAGAGCTAGAAATAG	56
		R_PCR2	TCCATCTAGATTACTCGAGTATCGTTCG	53

Para la PCR3 final se utilizaron el primer Forward de la PCR1 (F_PCR1) y el Reverse de la PCR2 (R_PCR2).

4.7 Electroforesis en gel de agarosa

La electroforesis en gel de agarosa se realizó para analizar y verificar los productos amplificados por PCR. Para preparar el gel, la concentración de agarosa utilizada fue del 0.5% en tampón 1x de TBE (Tris-HCl 90 mM, pH 8,0, H₃BO₃ 90 mM y EDTA 2 mM). Para solubilizar la agarosa en el tampón se calentó hasta fundirse en un microondas. Después se agita suavemente y una vez que está enfriada la solución se añaden 5 µl del agente intercalante SYBR® Safe DNA Gel Stain (Life Technologies). Este reactivo es esencial para permitir la posterior visualización del ADN bajo luz UV. La mezcla final se añade a un molde en el cual se ha puesto un peine para formar los pocillos y se deja solidificar la solución. Una vez solidificado bien el gel se retira el peine y se coloca el molde en la cubeta de electroforesis (BioRad) con tampón TBE 1x. Antes de cargar las muestras se añade el tampón de carga (Orange G 0,25% p/v, Sigma-Aldrich y glicerol 30% p/v) con una proporción aproximada de 1 µl por cada 5 µl de muestra.

Por último, se carga la muestra del producto amplificado. Además, en uno de los pocillos se cargará 3.5 µl del marcador molecular 1 kb GeneRuler DNA Ladder Mix (Thermo Scientific™). La electroforesis se lleva a cabo a una intensidad y voltaje de 100v constante durante 20-50min

dependiendo del tamaño de los fragmentos que se quieran separar. Las imágenes se obtienen bajo luz UV usando el equipo Gel Doc™ (BioRad) que va asociado al servidor Quantity One (BioRad).

4.8 Extracción y purificación de ADN en geles de electroforesis

Una vez realizada la electroforesis y localizado nuestro fragmento de ADN de interés resultado de las PCRs se procede a su extracción y purificación. Para ello se corta con un cúter en el gel el fragmento que nos interesa y que contiene la banda a purificar. Seguidamente se procede a eluir y purificar usando el kit GeneJET™ Gel Extraction (Fermentas, Thermo Scientific™) siguiendo el manual de la casa comercial.

4.9 Determinación de la concentración de los productos de PCR

La determinación de la concentración de los productos de PCR una vez purificados se realiza a partir de 2 µl de muestra usando el microespectrofotómetro NanoDrop 2000 (Thermo Scientific™). Previamente se realiza un blanco usando 2 µl de agua destilada. Los resultados de concentración se expresan en ng/µl. De forma escueta, la concentración de ADN se obtiene a partir de la absorbancia a 260 nm de longitud de onda, que es la longitud de onda a la cual se detectan los ácidos nucleicos. La pureza de la muestra se mide por un lado mediante el coeficiente Abs260/Abs280 de forma que valores de ratio cercanos o iguales a 1.8 significa que el ADN es bastante puro (ausencia de moléculas proteicas). Por otro lado el ratio Abs260/Abs230 en un valor entre 1.8-2.2 también nos indica buena pureza (ausencia de sales por ejemplo).

5. RESULTADOS

5.1 Caracterización previa de la composición del aceite de *Camelina* en plantas crecidas en la EEAD-CSIC

Antes de los eventos de edición, se procedió a determinar la composición del aceite de *Camelina* a partir de las semillas obtenidas en la cámara de cultivo del EEAD y crecidas en las condiciones descritas anteriormente. Se comparó dicha composición con semillas de *Camelina* crecidas en campo. Los resultados se muestran en la Figura 8.

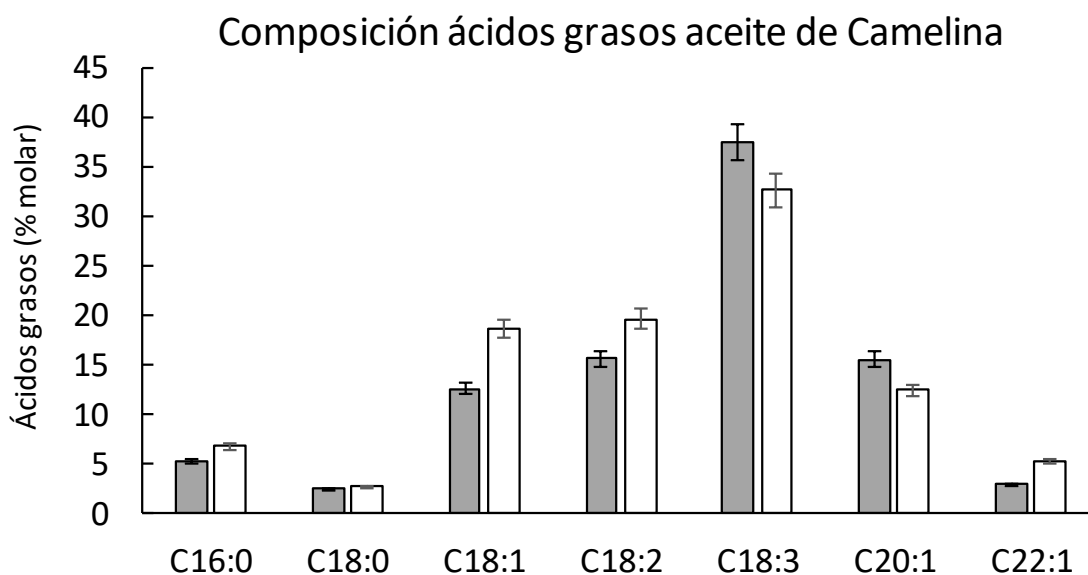


Figura 8. Composición de ácidos grasos del aceite de semillas de *Camelina* silvestres crecidas en nuestra cámara de cultivo (barras blancas) y cultivadas en campo (barras grises). Se realizaron tres determinaciones independientes por muestra.

Los resultados indican que el ácido graso mayoritario en el aceite tanto de plantas crecidas en campo como de plantas crecidas en la cámara bioclimática es similar, siendo el C18:3 u omega-3 el ácido graso mayoritario, con un porcentaje molar de 32,6 % en el caso de las plantas de cámara y un 37,5 % en el caso de las plantas de campo. También se detectó la presencia de ácido graso erúxico, con un porcentaje del 5,3% en el caso de las plantas de cámara, siendo este porcentaje algo inferior (2,9%) en el caso de las plantas de *Camelina* cultivadas en campo.

5.2 Diseño y construcción del vector conteniendo las dianas, los ARNs de guiado y la nucleasa Cas9, dirigidos contra los 3 genes CsFAEI del genoma de Camelina sativa.

Tras esta caracterización previa del fondo genético a utilizar, se procede a la construcción del vector de transformación, tomando como base el vector comercial de Addgene pHSE401.

Como ya se ha comentado, existen tres genes *CsFAEI* en el genoma de camelina denominados *CsFAEIA* (*Csa11g007400*), *CsFAEIB* (*Csa10g007610*), y *CsFAEIC* (*Csa12g009060*) (Hutcheon et al., 2010). Los tres genes se expresan en la semilla, siendo mayoritaria la expresión de *CsFAEIA* (Hutcheon et al., 2010). Para conseguir eliminar o reducir drásticamente el contenido en ácido erúxico 22:1 hay que editar los tres genes *CsFAEI*. Así pues, el primer paso de este trabajo consiste en analizar las tres secuencias de los genes *CsFAEI* utilizando herramientas bioinformáticas (CRISP-R y CRISPR-Plant) para el diseño de las secuencias diana apropiadas para un evento de edición exitosa. Los parámetros que se tomaron en cuenta para elegir las secuencias diana adecuadas fueron: i) como se pretende editar los tres genes de forma simultánea, hay que elegir una región de la secuencia idéntica en los tres. Esa secuencia no debe presentar homología con otras regiones del genoma para evitar mutaciones “run-off“. Estas secuencias diana deben ser adyacentes a los denominados sitios PAM de reconocimiento de Cas9; ii) deben de contener, preferentemente, un sitio de restricción dentro de la secuencia diana, que facilitará el genotipado de las líneas editadas por ausencia de respuesta al enzima de restricción; y iv) debe presentar en su proximidad algún polimorfismo de secuencia de cada uno de los tres genes *CsFAEI* que permita la identificación y confirmación de la correcta edición de cada uno de los tres genes mediante secuenciación.

Se identificaron 17 sitios PAM comunes entre los tres genes. De todos ellos y teniendo en cuenta los criterios mencionados se identificaron dos regiones (evento de edición A y B) en las secuencias de los tres genes *CsFAEI* que podrían actuar como secuencias diana para la edición. Estas secuencias se muestran en la Figura 9.

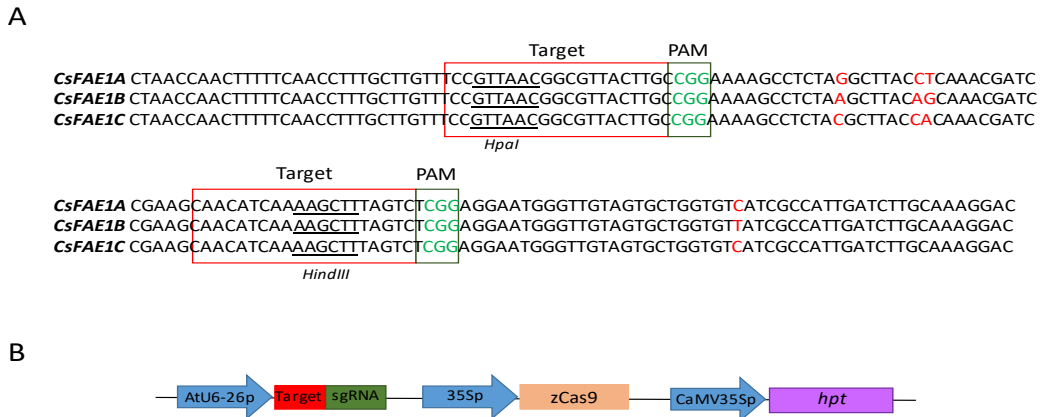


Figura 9. Diseño del vector para la edición mediante CRISPR de los tres genes *CsFAEI* de *Camelina sativa*. (A) Alineamiento de las secuencias homólogas de los 3 genes *CsFAEI* de camelina en las zonas tentativas escogidas como secuencias diana para la edición mediante Cas9. (B) Esquema tentativo del constructo a utilizar. La secuencia diana y el subsiguiente RNA de guiado se expresarán bajo el control del promotor U6 de *Arabidopsis*. La expresión de Cas9 y del gen marcador de selección (higromicina) estará controlado por un promotor 35S independiente.

Como se observa en la Figura 9, las secuencias diana de 20 nucleótidos elegidas (en el recuadro rojo) contienen un sitio de restricción (*HpaI* o *HindIII*) y son adyacentes al sitio PAM (caja verde). Se indican además polimorfismos cercanos en la secuencia (SNPs) en rojo que permitirían distinguir en el análisis por secuenciación qué copia del gen se ha editado.

La estrategia que utilizaremos para la edición mediante CRISPR/Cas9 será la de recombinación no homóloga o “*non-homologous end joining*” (NHEJ) que permite la delección o inserción de una secuencia determinada sin necesidad del uso de un “donador”, que complicaría la posterior segregación de las líneas editadas.

Posteriormente, las secuencias diana encontradas en la Figura 9 se clonarán en el vector pHSE401 mediante la estrategia de “PCR splicing” explicada anteriormente. Básicamente, se utilizará el vector pHSE401 vacío como molde y se llevarán a cabo las PCRs secuenciales (PCR1 y PCR2). Así, estas dos PCRs amplificarán la secuencia antes y después de la secuencia diana. Esta secuencia es introducida a través de las secuencias de los cebadores reverse y forward, respectivamente (Figura 10).

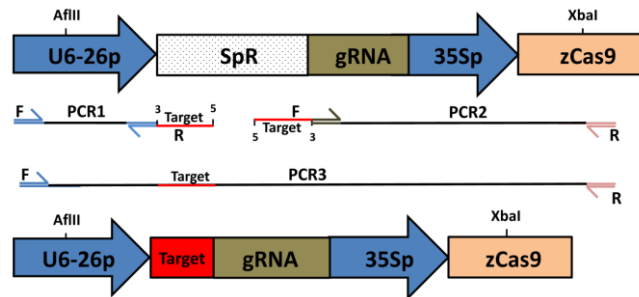


Figura 10. Estrategia de amplificación por PCR splicing para la generación de los fragmentos conteniendo las secuencias diana y los ARNg usados para la edición de los genes *CsFAE1*

Una vez obtenidos, los productos de PCR se mezclan y se lleva a cabo la tercera PCR con cebadores de los extremos que permiten obtener en un solo fragmento todos los diferentes componentes necesarios para el armado del vector (promotor U6 + diana y RNA de guiado + promotor 35S + gen Cas9).

5.3 Amplificación de los fragmentos por PCR splicing y armado del vector de edición

Siguiendo el protocolo diseñado por el grupo en el apartado anterior, hemos procedido a la amplificación de los diferentes fragmentos: (PCR1, PCR2 y PCR3) como primera etapa del procedimiento de edición génica. Los resultados del gel de electroforesis posterior al proceso de amplificación se muestran en la Figura 11.

Como se puede observar, en el resultado de la PCR2 del evento B se ven ciertas bandas adicionales justo por debajo del producto de interés. Esto nos generaba un problema en el posterior proceso de purificación de la banda. Para que la PCR3 tuviera éxito una de las condiciones es que la mezcla de las bandas de PCR1 y PCR2 tiene que ser lo más pura posible sin la existencia de productos que pudieran entorpecer la amplificación final. Se sospechó que esas bandas adicionales podrían ser debido a estructuras *supercoiling* que se pueden generar de forma natural en el plásmido circular y entorpecer el proceso de amplificación. Por ello, paralelamente a este experimento se realizó una linealización del

plásmido pHSE401 con las enzimas de restricción adecuadas previo al proceso de amplificación por PCR splicing. Los resultados no mejoraron el problema de las bandas inespecíficas en la PCR2B. Dado que el resultado fue el mismo se decidió seguir con la estrategia inicial para ahorrar material y reactivos. Además, otro paso crítico en la PCR3 recae en las cantidades de PCR1B y PCR2B en la mezcla de reacción de amplificación (Heckman & Pease, 2007). Los dos productos formados en estas PCRs sirven como *template* para la PCR3 por lo que es necesario que se encuentren en concentraciones equimolares. El volumen exacto a añadir en la PCR3 dependerá de la cantidad de ADN que tengamos después de la purificación de los productos de la PCR1 y PCR2. Este problema se ha visto plasmado en este trabajo, como se ve en la Figura 11 la PCR1 ha dado una banda mucho más intensa que la PCR2 y por lo tanto una mayor concentración de ADN. A esto se le suma, como ya se ha comentado, el hecho que las bandas inespecíficas que aparecen disminuyen la purificación del producto de nuestra PCR2 y a que el producto de la PCR2 tiene mayor longitud de secuencia (1188bp) frente las 483bp de la PCR1.

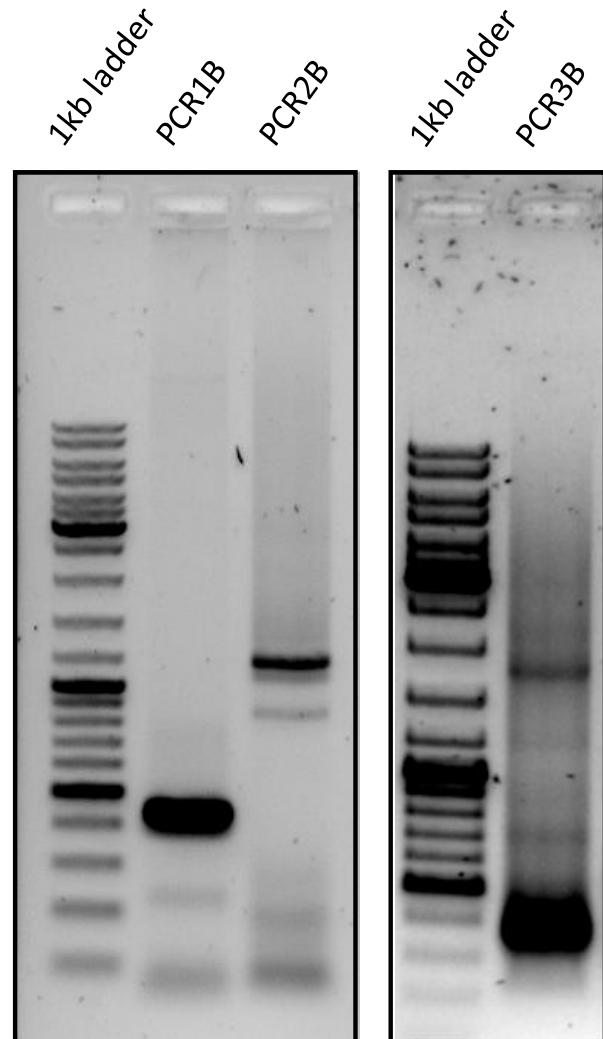


Figura 11. Construcción y armado del vector de edición: PCR1B (483 bp), PCR2B (1.188 bp) y PCR3B (1643 bp).

Una vez en disposición de los productos de PCR (PCR1, PCR2 y PCR3), que contienen las secuencias diana de Cas9 del gen *CsFAE1* insertadas junto a los RNAs de guiado, hemos procedido a su clonaje en los sitios *AflII* y *XbaI* del vector pHSE401. La eliminación del cassette de resistencia a espectinomicina permitirá la selección de los clones positivos en la bacteria. Tras finalizar la construcción del vector, todos los elementos de secuencia serán chequeados mediante secuenciación. Una vez conseguido este hito, procederíamos a la transformación de plantas de Camelina y posterior selección de las líneas higromicina-resistentes conteniendo los genes *CsFAE1* editados. Estos últimos objetivos no se han podido realizar en el laboratorio para este trabajo por falta de tiempo, derivado del confinamiento provocado por la pandemia de Covid-19 y las medidas de cuarentena.

6. DISCUSIÓN

En este trabajo, se está siguiendo una estrategia de CRISPR/Cas9 para editar los tres genes *CsFAEI* en el genoma de *Camelina sativa*. Frente a otras estrategias de modificación o inactivación de genes, basadas bien en la inserción de un T-DNA o bien la utilización de RNAs de interferencia (RNAi), CRISPR-Cas9 ofrece una especificidad muchísimo mayor y una mayor simplicidad en la técnica: en un solo evento de edición se pueden inactivar los tres genes. Ninguna de las técnicas mencionadas anteriormente puede tener éxito, más aún en el caso de *Camelina sativa*, que posee un genoma triple-diploide (en la práctica similar a un hexaploide), en el que la mayoría de los genes presentan como mínimo tres copias funcionales.

El contenido en ácido erúico presente en nuestras semillas de camelina se encuentra entre el 2-5% del total de ácidos grasos. La eliminación o reducción sustancial de ese porcentaje es el objetivo biotecnológico final de este proyecto.

Una de las primeras etapas necesarias para la consecución de este objetivo es el armado del vector para su posterior transformación en camelina. La estrategia de PCR splicing ha tenido éxito en la formación del fragmento con los componentes necesarios para el posterior armado del vector (promotor U6 + diana y RNA de guiado + promotor 35S + gen Cas9). De hecho no es la primera vez que se usa esta estrategia con este objetivo. (Aznar-Moreno & Durrett, 2017) modificaron los genes *DGATI* y *PDATI* de Camelina usando la misma estrategia de PCR splicing para generar el fragmento que contenía la secuencia diana junto con el ARNg y Cas9. Estos genes son esenciales en la síntesis de triacilglicerol (TAG) en el aceite de las semillas. Los ácidos grasos presentes en el aceite de la semilla se unen al TAG. En el estudio, los autores obtuvieron líneas transgénicas en la T2 con un menor contenido en aceite en su semilla. Más concretamente, la línea D mostró un 31% menos de contenido de aceite respecto el *wild-type*. Además, las mutaciones en *DGATI* y *PDATI* tuvieron efectos fenotípicos negativos en la planta puesto que los mutantes T2 mostraron semillas más oscuras y más pequeñas que las plantas silvestres.

En otras investigaciones en vez de utilizar la técnica de PCR splicing se opta por introducir el RNA de guiado y el *PAM site* en sendos oligos que luego se hibridan y posteriormente se clonan en los vectores de destino utilizando sitios de restricción específicos introducidos en la secuencia de dichos oligos. A priori no hay una razón científica para optar por un modelo u otro. En nuestro caso optamos por la estrategia de PCR splicing porque ya había sido usada con éxito en Camelina por Aznar Moreno y colaboradores para plantas editadas en los genes *DGATI* y *PDATI* y usando el mismo vector de transformación.

El diseño de dos secuencias diana de este trabajo (evento A y B) nos asegura una probabilidad de éxito más elevada en el proceso de edición por CRISPR. Es decir, más opciones de que la nucleasa Cas9 nos genere cortes y por lo tanto mutaciones en el gen diana *FAEI*. Sin embargo, por falta de tiempo, en este trabajo solo tenemos resultados del evento de edición B. Hay que mencionar que los resultados del evento A tienen que ser muy similares a los realizados en este trabajo con el evento B. La transformación y segregación de los transformantes es un proceso que dura unos 10-14 meses y que se continuará tras este trabajo.

Para analizar en profundidad si la estrategia de CRISPR/Cas9 de este proyecto es una buena herramienta para la edición de los ácidos grasos de la semilla en plantas se realizó una búsqueda bibliográfica de artículos científicos donde se usase esta tecnología con éxito en la manipulación de aceites vegetales. Los resultados se muestran en la Tabla 4.

Tabla 4. Ejemplos en los que la tecnología CRISPR-Cas9 ha tenido éxito en la manipulación de aceites vegetales

Referencias	Planta objetivo	Gen diana	Vector de transformación	Resultados
(Xie, Meesapyodsuk, & Qiu, 2019)	<i>Arabidopsis thaliana</i>	<i>KASI</i>	Los ARNg se insertan en un vector intermediario Este plásmido después se insertó en el vector pYLCRISPR/ Cas9Pubi-B vector	Mutaciones en el gen dieron lugar a plantas con un crecimiento anormal tanto en las plántulas jóvenes como en los estadios de crecimientos posteriores
(Okuzaki et al., 2018)	<i>Brassica napus</i> (colza)	<i>FAD2</i>	<i>Target sequence cloning vector pUC19_AtU6oligo.</i> <i>Binary vector pZD_AtU6gRNA_FFCas9_NPTII</i>	Delección de 4bp en el gen diana. Se aumentó el contenido de ácido oleico.
(Wang et al., 2019)	<i>Glycine max</i> (soja)	<i>LOX</i>	pGES201	En dos generaciones han obtenido la mutación en los tres loci del gen lox
(Do et al., 2019)	<i>Glycine max</i> (soja)	<i>FAD2</i>	pFGC594	Mutantes GmFAD2-40% de las plantas T0. Las plantas mostraron un incremento del ácido oleico por encima del 80%
(Jiang et al., 2017)	<i>Camelina sativa</i>	<i>FAD2</i>	pCAMBIA1302	Incremento ácido oleico de un 16% a un 50% del total de ácidos grasos
(Aznar-Moreno & Durrett, 2017)	<i>Camelina sativa</i>	<i>DGATI</i> y <i>PDATI</i>	pHSE401	Reducción del contenido de aceite de hasta un 31% en la línea transgénica D

Como se puede observar, con la tecnología CRISPR-Cas9 se han modificado diferentes genes relacionados con la acumulación de ácidos grasos en la semilla de las plantas (*FAD2*, *DGATI*, *LOX*). Uno de los genes más estudiados ha sido el gen *FAD2*. Este gen juega un papel muy importante en la formación de ácidos grasos poliinsaturados. El gen codifica una desaturasa, es decir, una enzima que genera una insaturación en el ácido oleico (18:1) dando lugar a un ácido graso linoleico (18:2). En el caso de la Camelina esta enzima es muy importante porque a partir del 18:2 que genera *FAD2* actuará *FAD3* que formará una insaturación dando lugar al ácido graso omega-3 tan característico de este cultivo. Mutaciones en el gen *FAD2* de camelina (**Jiang et al., 2017**) ha generado un incremento significativo del contenido en ácido graso oleico. Como la enzima *FAD2* no puede desaturar se produce una mayor acumulación de ácido graso oleico 18:1. Estos resultados concuerdan con mutaciones en el gen *FAD2* de otras especies vegetales como en la colza o en la soja (Tabla4). Tal como habíamos propuesto en la introducción de este trabajo a partir de estos resultados de *FAD2* cabe suponer que si generamos una mutación en el gen *FAEI*, a parte de una disminución de la cantidad de ácido erúxico se pueda producir también un aumento de la concentración de ácido graso omega-3 como consecuencia de una mayor disponibilidad de 18:1 (Figura3). De esta forma, obtendríamos un beneficio adicional en nuestras plantas transformadas.

(**Wang et al., 2019**) mutaron mediante CRISPR los genes *LOX* que codifican una enzima, la lipooxigenasa. Esta enzima se expresa en las semillas maduras de soja y cataliza la oxidación de ácidos grasos insaturados como el ácido linoleico y el linolénico produciendo compuestos conjugados que se convierten en compuestos volátiles y generan un aroma desagradable. El gen tiene tres copias funcionales en el genoma de la planta de la soja. Los resultados muestran como de las 60 plantas transgénicas de la T0, dos de ellas son triples mutantes, es decir, que los tres genes funcionales han sido modificados. Como consecuencia la actividad lipooxigenasa se perdió. Este estudio demuestra que la tecnología CRISPR-Cas9 es una herramienta efectiva en la edición de genes que tienen múltiples copias funcionales como el caso de nuestra Camelina.

Por último, **(Xie et al., 2019)** usaron CRISPR-Cas9 para obtener mutantes del gen *KASI* en *Arabidopsis thaliana*. Este gen es esencial en la biosíntesis de ácidos grasos. Mutaciones en ese gen dieron lugar a plantas con un crecimiento anormal tanto en las plántulas jóvenes como en los estadios de crecimientos posteriores.

En algunos estudios como **(Aznar-Moreno & Durrett, 2017)** y **(Xie et al., 2019)** hemos visto que las mutaciones generadas han provocado alteraciones fenotípicas en las semillas. Desconocemos si la manipulación de la ruta de biosíntesis de ácido erúico presentada en este trabajo podría acarrear efectos secundarios no deseados como alteraciones en el crecimiento de la planta o en la producción del aceite haciendo el producto final inviable. Todos estos aspectos se seguirán estudiando tras este trabajo.

7. CONCLUSIONES

En este trabajo se ha estudiado el diseño de una estrategia basada en el sistema CRISPR-Cas9 para la edición del gen *FAEI* de *Camelina sativa*.

- Se han diseñado dos secuencias diana diferentes de forma exitosa
- Se ha conseguido con éxito el armado del vector de edición previo al proceso de transformación
- Es posible usar la tecnología CRISPR-Cas9 para manipular el aceite de especies vegetales con fines alimentarios.

8. BIBLIOGRAFÍA

- Aznar-Moreno, J. A., & Durrett, T. P. (2017). Simultaneous targeting of multiple gene homeologs to alter seed oil production in camelina sativa. *Plant & Cell Physiology*, 58(7), 1260-1267. doi:10.1093/pcp/pcx058
- Bates, P. D., & Browse, J. (2012). The significance of different diacylglycerol synthesis pathways on plant oil composition and bioengineering. *Frontiers in Plant Science*, 3, 147. doi:10.3389/fpls.2012.00147
- Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, 37(8), 911-917. doi:10.1139/o59-099
- Cassileth, B. (2010). Omega-3. *Oncology (Williston Park, N.Y.)*, 24(1), 106.
- Claver Lacasa, A. (2018). *Caracterización bioquímica y molecular de la biosíntesis de aceite y de ácido graso erúcido en germoplasma de origen europeo de la brasicácea thlaspi arvensis (pennycress)* Retrieved from <https://digital.csic.es/handle/10261/174803>
- Do, P. T., Nguyen, C. X., Bui, H. T., Tran, L. T. N., Stacey, G., Gillman, J. D., . . . Stacey, M. G. (2019). Demonstration of highly efficient dual gRNA CRISPR/Cas9 editing of the homeologous GmFAD2-1A and GmFAD2-1B genes to yield a high oleic, low linoleic and α -linolenic acid phenotype in soybean. *BMC Plant Biology*, 19(1), 311. doi:10.1186/s12870-019-1906-8

- EFSA Panel on Contaminants in the Food Chain (CONTAM), Knutsen, Helle Katrine Alexander, Jan, Barregård, L., Bignami, M., Brüschweiler, B., Ceccatelli, S., . . . Vleminckx, C. (2016). Scientific opinion on erucic acid in feed and food. *EFSA Journal*, *14*(11), e04593. doi:10.2903/j.efsa.2016.4593
- Garcés, R., Sarmiento, C., & Mancha, M. (1992). Temperature regulation of oleate desaturase in sunflower (*helianthus annuus* L.) seeds. *Planta*, *186*(3), 461-465. doi:10.1007/BF00195328
- Heckman, K. L., & Pease, L. R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols*, *2*(4), 924-932. doi:10.1038/nprot.2007.132
- Hutcheon, C., Ditt, R. F., Beilstein, M., Comai, L., Schroeder, J., Goldstein, E., Kiser, J. (2010). Polyploid genome of camelina sativa revealed by isolation of fatty acid synthesis genes. *BMC Plant Biology*, *10*, 233. doi:10.1186/1471-2229-10-233
- Jaworski, J., & Cahoon, E. B. (2003). Industrial oils from transgenic plants. *Current Opinion in Plant Biology*, *6*(2), 178-184. doi:10.1016/s1369-5266(03)00013-x
- Jiang, W. Z., Henry, I. M., Lynagh, P. G., Comai, L., Cahoon, E. B., & Weeks, D. P. (2017). Significant enhancement of fatty acid composition in seeds of the allohexaploid, camelina sativa, using CRISPR/Cas9 gene editing. *Plant Biotechnology Journal*, *15*(5), 648-657. doi:10.1111/pbi.12663
- Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W. E., Tuteja, R., . . . Parkin, I. A. P. (2014). The emerging biofuel crop camelina sativa retains a highly undifferentiated

hexaploid genome structure. *Nature Communications*, 5(1), 1-11.

doi:10.1038/ncomms4706

Liu, J., Tjellström, H., McGlew, K., Shaw, V., Rice, A., Simpson, J., . . . Ohlrogge, J. (2015).

Field production, purification and analysis of high-oleic acetyl-triacylglycerols from transgenic camelina sativa. *Industrial Crops and Products*, 65, 259-268.

doi:10.1016/j.indcrop.2014.11.019

Liu, X., Wu, S., Xu, J., Sui, C., & Wei, J. (2017). Application of CRISPR/Cas9 in plant

biology. *Acta Pharmaceutica Sinica. B*, 7(3), 292-302. doi:10.1016/j.apsb.2017.01.002

Miguel Gutiérrez López, & Antonio Albalat Borrás. (2013). EL CULTIVO DE LA

CAMELINA EN ARAGÓN. primeras experiencias de su cultivo en el bajo aragón.

Dirección General De Alimentación Y Fomento Agroalimentario- Servicios Recursos Agrícolas De, (243), 1-8.

Okuzaki, A., Ogawa, T., Koizuka, C., Kaneko, K., Inaba, M., Imamura, J., & Koizuka, N.

(2018). CRISPR/Cas9-mediated genome editing of the fatty acid desaturase 2 gene in brassica napus. *Plant Physiology and Biochemistry: PPB*, 131, 63-69.

doi:10.1016/j.plaphy.2018.04.025

Reglamento (UE) 2019/1870 de la comisión de 7 de noviembre de 2019 que modifica y

corrige el reglamento (CE) nº 1881/2006 en lo que respecta al contenido máximo de

ácido erúxico y ácido cianhídrico en determinados productos alimenticios, Diario Oficial de la Unión Europea, L289/37, (2019). Retrieved from

<http://data.europa.eu/eli/reg/2019/1870/oj/spa>

Roscoe, T. J., Lessire, R., Puyaubert, J., Renard, M., & Delseny, M. (2001). Mutations in the fatty acid elongation 1 gene are associated with a loss of beta-ketoacyl-CoA synthase activity in low erucic acid rapeseed. *FEBS Letters*, 492(1-2), 107-111.

doi:10.1016/s0014-5793(01)02243-8

Wang, J., Kuang, H., Zhang, Z., Yang, Y., Yan, L., Zhang, M., . . . Guan, Y. (2019).

Generation of seed lipoxygenase-free soybean using CRISPR-Cas9. *The Crop Journal*,

doi:10.1016/j.cj.2019.08.008

Xie, X., Meesapyodsuk, D., & Qiu, X. (2019). Enhancing oil production in arabidopsis

through expression of a ketoacyl-ACP synthase domain of the PUFA synthase from

thraustochytrium. *Biotechnology for Biofuels*, 12, 172. doi:10.1186/s13068-019-1514-8

Xing, H., Dong, L., Wang, Z., Zhang, H., Han, C., Liu, B., . . . Chen, Q. (2014). A

CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biology*, 14,

327. doi:10.1186/s12870-014-0327-y

Zubr, J. (1997). Oil-seed crop: Camelina sativa. *Industrial Crops and Products*, 6(2), 113-

119. doi:10.1016/S0926-6690(96)00203-8