

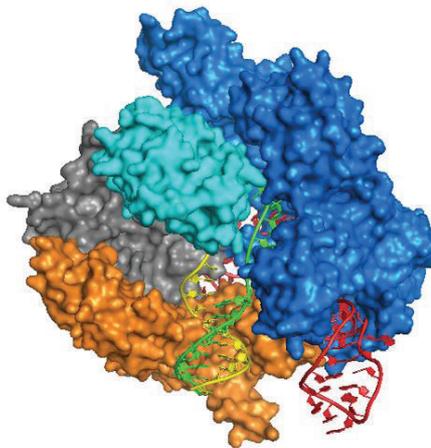
Grado en Bioloxía

Memoria do Traballo de Fin de Grao

**Descripción del sistema CRISPR-Cas y diseño de sgRNAs
(*single guide RNAs*) para edición de ADN genómico en
levaduras**

**Descripción do sistema CRISPR-Cas e deseño de sgRNAs
(*single guide RNAs*) para edición de ADN xenómico en lévedos**

**Overview of the CRISPR-Cas system and design of sgRNAs
(*single guide RNAs*) for genomic DNA editing in yeast**



Marta Pardo Piñón

Junio, 2017

Director: Esther Rodríguez Belmonte



UNIVERSIDADE DA CORUÑA

Dña. María Esther Rodríguez Belmonte, profesora titular del departamento de Biología Molecular y Celular de la Facultad de Ciencias de la Universidad de A Coruña, autoriza la presentación del Trabajo de Fin de Grado “Descripción del sistema CRISPR-Cas y diseño de sgRNAs (*single guide RNAs*) para edición de ADN genómico en levaduras” presentado por Marta Pardo Piñón para su defensa ante el tribunal calificador.

En A Coruña a 21 de Junio del 2017

RODRIGUEZ
BELMONTE
MARIA ESTHER
- DNI
32776974L

Firmado digitalmente por RODRIGUEZ
BELMONTE MARIA ESTHER - DNI 32776974L
Nombre de reconocimiento (DN): c=ES,
o=UNIVERSIDAD DE A CORUÑA,
ou=certificado electrónico de empleado
público, ou=BIOLOXIA CELULAR E
MOLECULAR, ou=6682,
serialNumber=32776974L, sn=RODRIGUEZ
BELMONTE, givenName=MARIA ESTHER,
cn=RODRIGUEZ BELMONTE MARIA ESTHER -
DNI 32776974L
Fecha: 2017.06.21 10:48:11 +02'00'

Asdo: Dra. María Esther Rodríguez Belmonte

RESUMEN

El descubrimiento de los sistemas CRISPR-Cas y concretamente del sistema CRISPR-Cas9 ha supuesto una revolución en el campo de la ingeniería genómica y la biotecnología, puesto que permite editar cualquier región del genoma de cualquier organismo de manera altamente eficaz, precisa y económica. Así, desde su descubrimiento como un sistema natural de inmunidad adaptativa en organismos procariotas, sus características estructurales y funcionales han sido estudiadas y redefinidas hasta obtener variantes de CRISPR-Cas9 que incrementan la versatilidad de esta maquinaria para su empleo en investigación e industria. En el presente trabajo se realiza una revisión teórica de las características generales, estructura y función de CRISPR-Cas9 en la naturaleza así como de las funciones y aplicaciones que lo perfilan como una de las más novedosas y eficaces herramientas biotecnológicas conocidas. Además, para ejemplificar su empleo como mecanismo de edición genómica, se llevará a cabo el diseño de un proyecto experimental mediante el diseño de un oligonucleótido que permitirá la introducción y expresión de la proteína GFP (*Green Fluorescent Protein*) en el organismo eucariota *Saccharomyces cerevisiae*.

Palabras clave: CRISPR-Cas, CRISPR-Cas9, edición genómica, Cas9, sgRNA

ABSTRACT

The discovery of CRISPR-Cas systems and specifically the CRISPR-Cas9 system has led to a revolution in the field of genomic engineering and biotechnology, since it allows to edit any region of the genome of any organism in a highly effective, precise and economical way. Thus, since its discovery as a natural system of adaptive immunity in prokaryotic organisms, its structural and functional characteristics have been studied and redefined until obtaining variants of CRISPR-Cas9 that increase the versatility of this machinery for its use in research and industry. In the present work a theoretical review of the general characteristics, structure and function of CRISPR-Cas9 in the nature as well as of the functions and applications that profile it as one of the most novel and effective known biotechnological tools. In addition, to exemplify its use as a genomic editing mechanism, the design of an experimental project will be carried out by designing an oligonucleotide that will allow the introduction and expression of the GFP (*Green Fluorescent Protein*) protein in the eukaryotic organism *Saccharomyces cerevisiae*.

Key words: CRISPR-Cas, CRISPR-Cas9, genomic edition, Cas9, sgRNA

RESUMO

O descubrimento dos sistemas CRISPR-Cas e concretamente do sistema CRISPR-Cas9 supuxo unha revolución no campo da inxeniería xenómica e a biotecnoloxía, posto que permite editar calquera rexión do xenoma de calquera organismo de maneira altamente eficaz, precisa e económica. Así, dende o seu descubrimento como un sistema natural de inmunidade adaptativa en organismos procariotas, a súas características estruturais e funcionais foron estudadas e redefinidas ata obter variantes de CRISPR-Cas9 que incrementan a versatilidades desta maquinaria para o seu emprego en investigación e industria. No presente traballo realízase unha revisión teórica das características xerais, estrutura e función de CRISPR-Cas9 na natureza así como das funcións e aplicacións que o perfilan como unha das máis novedosas e eficaces ferramentas biotecnolóxicas coñecidas. Ademáis, para exemplificar o seu emprego como mecanismo de edición xenómica, levaráse a cabo o deseño dun proxecto experimental mediante o deseño dun oligonucleótido que permitirá a introducción e expresión da proteína GFP (*Green Fluorescent Protein*) no organismo eucariota *Saccharomyces cerevisiae*.

Palabras clave: CRISPR-Cas, CRISPR-Cas9, edición xenómica, Cas9, sgRNA

GLOSARIO

- **bp:** par de bases (*base pair*)
- **Cas:** genes o proteínas asociados a CRISPR (*CRISPR-associated*)
- **CRISPR:** agrupación de repeticiones palindrómicas cortas regularmente interespaciadas (*clustered regularly interspaced short palindromic repeats*)
- **CRISPR-Cas:** agrupación de repeticiones cortas palindrómicas regularmente interespaciadas asociadas a nucleasas Cas.
- **CRISPR-Cpf1:** agrupación de repeticiones palindrómicas cortas regularmente interespaciadas de *Prevotella* y *Francisella 1*)
- **crRNA:** RNA transcrito de CRISPR (*CRISPR RNA*)
- **dCas:** Cas9 defectiva para la actividad nucleasa
- **Di-CRISPR:** CRISPR-Cas de integración en delta (*delta integration CRISPR-Cas*)
- **DSB:** doble corte en la cadena (*double-stranded break*)
- **GFP:** proteína verde fluorescente (*green fluorescent protein*)
- **HDR:** reparación directa por homología (*homology directed repeat*)
- **Indel:** inserción o delección
- **nCas:** Cas9 nicasa
- **NHEJ:** reparación de unión de extremos no homólogos (*non-homologous end joining*)
- **nt:** nucleótido
- **NUC:** lóbulo con actividad nucleasa de Cas9
- **HNH:** dominio nucleasa de Cas9
- **ORF:** pauta abierta de lectura (*open reading frame*)
- **PAM:** motivo adyacente al protoespaciador (*protospacer adjacent motif*)
- **PI:** dominio de reconocimiento de la secuencia PAM
- **REC:** lóbulo de reconocimiento de Cas9
- **SRSR:** repeticiones cortas regularmente interespaciadas (*short regularly spaced repeats*)
- **tracrRNA:** *noncoding trans-activating CRISPR-RNA*
- **TALENs:** nucleasas tipo activadores de transcripción (*transcription activator-like effector nucleases*)
- **Tm:** temperatura de fusión de oligonucleótidos (*melting temperature*)
- **ZFNs:** nucleasas de dedos de zinc (*zinc-finger nucleases*)

ÍNDICE

1. Introducción	1
2. Objetivos	1
3. Materiales y métodos	2
3.1. Búsquedas bibliográficas	2
3.2. Gestor bibliográfico	2
3.3. Diseño experimental de inserción de un marcador molecular en el genoma de un organismo eucariota como ejemplo de herramienta de edición genómica	2
<i>Análisis y manipulación experimental de las secuencias</i>	2
<i>Diseño de cebadores</i>	2
4. Resultados	2
4.1. Revisión bibliográfica: Descripción del Sistema CRISPR-Cas	2
4.1.1. Antecedentes	2
4.1.2. La Biología del Sistema CRISPR-Cas en la Naturaleza	4
<i>¿Qué es el sistema CRISPR-Cas?</i>	4
<i>Clasificación de los sistemas CRISPR</i>	5
<i>Etapas del funcionamiento del Sistema CRISPR-Cas</i>	6
<i>Estructura y modo de acción de la proteína Cas9</i>	8
4.1.3. El Sistema CRISPR-Cas9 como Herramienta de Edición Genómica	9
<i>Patrones de reparación tras la escisión de Cas9</i>	9
<i>nCas y dCas</i>	10
<i>Cpf1 como enzima nucleasa y sus ventajas en el Sistema CRISPR-Cpf1</i>	11
<i>¿Por qué CRISPR-Cas ha resultado ser una herramienta biotecnológica revolucionaria?</i>	12
4.2. Diseño teórico de un sgRNA para edición del genoma de un organismo eucariota	13
<i>Microorganismo</i>	13
<i>Localización de la inserción</i>	14
<i>Selección del marcador molecular a insertar</i>	14
<i>Diseño de cebadores y amplificación del casete gfp-delta para inserción en el genoma</i>	14
<i>Diseño del megaprimer codificante del sgRNA y construcción del plásmido Di-Delta-Marta</i>	16
<i>Transformación de S. cerevisiae</i>	17
<i>Comprobación de la inserción</i>	17
5. Discusión	18
5.1. Conclusiones generales sobre el sistema CRISPR-Cas9 y perspectivas futuras	18
5.2. La efectividad del diseño de sgRNAs	19
6. Conclusiones	20
7. Bibliografía	21

ÍNDICE DE FIGURAS

Figura 1.....	4
Figura 2.....	5
Figura 3.....	5
Figura 4.....	7
Figura 5.....	8
Figura 6.....	9
Figura 7.....	9
Figura 8.....	11
Figura 9.....	12
Figura 10.....	13
Figura 11.....	14
Figura 12.....	15
Figura 13.....	15
Figura 14.....	16
Figura 15.....	17
Figura 16.....	18

1. INTRODUCCIÓN

Tras la publicación de la estructura tridimensional del polímero de ADN por Watson y Crick en 1953 (Watson & Crick, 1953), el interés por desentrañar las claves que hacen de esta molécula el origen de la vida no dejan de crecer entre la comunidad científica. Sin embargo, no es hasta los años 70 del siglo pasado cuando tiene lugar el gran auge de la biotecnología y dentro de esta, de la ingeniería genómica. Con ella, no sólo se pretenden comprender los engranajes que posibilitan la vida tal y como la concebimos, sino que se busca dar un paso más allá, modificando la información genética a antojo del ser humano para obtener un objetivo concreto.

Paralelamente el desarrollo de técnicas que permiten la modificación del ADN, la investigación científica vuelca también un gran esfuerzo en el estudio del material genético de los organismos procariotas, que habitan la Tierra desde hace 4 millones de años y de los que el ser humano ha obtenido beneficio de múltiples formas (Mojica & Rodríguez-Valera, 2016).

La suma de las novedades obtenidas mediante el estudio general del material genético y el análisis de la información genética procariota da sus frutos en la década de los 90, cuando aparecen nuevas técnicas de secuenciación que abren el camino al conocimiento de los genomas completos de los organismos vivos. Gracias a estas herramientas, en 1995 se publica por vez primera el primer genoma completo de una bacteria (Fleischmann *et al.*, 1995). Además de esto, es en este momento cuando se descubren las primeras herramientas de edición genómica a través de mecanismos de recombinación homóloga empleando secuencias apareables con el sitio diana que se quiere modificar (HDR o *reparación directa por homología*) o incluyendo inserciones y deleciones a través del mecanismo de reparación de dobles roturas en el ADN (NHEJ o *reparación de unión de extremos no homólogos*). Así, se descubren sistemas de edición genómica tales como ZFNs o los TALENs (Gaj, 2014). Sin embargo, ninguna de estas tecnologías es tan eficaz, de tan fácil diseño y económicamente viable como la que trataremos en esta memoria: el sistema CRISPR-Cas empleado por los organismos procariotas como sistema de defensa inmune y reinventado para la edición genómica (Ran *et al.*, 2013). Este sistema aúna los conocimientos conseguidos a lo largo de los años mediante el estudio de la estructura de la molécula de ADN y las peculiaridades de la genómica procariota y se perfila con un gran protagonismo en el presente y futuro de la genética moderna.

2. OBJETIVOS

1. Ofrecer un acercamiento sobre el sistema CRISPR-Cas9 al lector, realizando un análisis detallado de su estructura y modo de acción en situaciones naturales así como de los avances que le han permitido convertirse en la herramienta de edición genómica más prometedora del presente y el futuro. Para ello, se realizará una extensa revisión bibliográfica que será gestionada con la herramienta de gestión bibliográfica Mendeley.
2. Diseñar un sgRNA que permita la edición del gen *gfp* en la levadura *Saccharomyces cerevisiae* con el fin de realizar una demostración práctica de la utilidad de este sistema. Para su ejecución se emplearán los diversos programas informáticos destinados a ello. Se seleccionará también el sistema más adecuado para expresar el gen *gfp* en dicha levadura y se diseñará un sistema que permita discriminar los organismos modificados del resto.
3. Redactar una memoria que incluya la información extraída de la revisión bibliográfica consultada y que conformará la parte teórica del trabajo, y la parte práctica con la explicación detallada del diseño experimental teórico de modificación genética y selección de candidatos.

3. MATERIALES Y MÉTODOS

3.1. Búsquedas bibliográficas

Las búsquedas bibliográficas, tanto de artículos de revisión como de investigación científica, se realizaron utilizando las bases de datos del PubMed (del NIH), como de Google Scholar. En algunos casos, para mejorar la comprensión de algunos de los mecanismos de CRISPR-Cas, se utilizaron figuras extraídas de casas comerciales, cuyas URLs están mencionadas a lo largo del texto.

3.2. Gestor bibliográfico

Las citas referenciadas a lo largo del texto y la bibliografía final se implementaron utilizando el gestor bibliográfico Mendeley, con el formato *Harvard Reference Format 1*.

3.3. Diseño experimental de inserción de un marcador molecular en el genoma de un organismo eucariota como ejemplo de herramienta de edición genómica

Análisis y manipulación experimental de las secuencias

Para el diseño experimental de inserción de un marcador molecular en el genoma de un organismo eucariota se utilizaron las secuencias de la cepa de *Saccharomyces cerevisiae* S288C, obtenida de la base de datos GenBank, y la secuencia codificadora de la proteína fluorescente verde, obtenida de la base de datos de vectores de la casa comercial Addgene. Para el reconocimiento de secuencias diana se utilizó el programa BLASTn del NCBI. Para análisis de cadenas complementarias se empleó el programa Reverse Complement del servidor informático Bioinformatics.org.

Diseño de cebadores

Para el diseño teórico del meganucleótido guía para la producción de sgRNA se utilizó el servidor CRISPy-web. Para el diseño de los cebadores de amplificación del casete de inserción en el genoma de *S. cerevisiae* se utilizó la secuencia del casete GFP del vector pCYC1m_yGFP obtenido de Addgene y las temperaturas de hibridación se calcularon con la herramienta *Tm Calculator* de New England Biolabs.

4. RESULTADOS

4.1. Revisión bibliográfica: Descripción del Sistema CRISPR-Cas

4.1.1. Antecedentes

En el año 1993 un incipiente investigador español, el alicantino Francis Mojica, elegía como protagonista de su tesis doctoral a una arquea extremófila que habita en las salinas de Santa Pola (Alicante) y que dota a las aguas de las salinas de una tonalidad rosácea cuando se incrementa la concentración de sal. Concretamente, Mojica se dedicó a estudiar la cepa R-4 de *Haloflex mediterranei*. El objetivo final de su trabajo era hallar los mecanismos regulatorios que hacían posible que este microorganismo fuese capaz de soportar las variaciones de salinidad en el ambiente (Mojica & Rodríguez-Valera, 2016). Lo que este joven investigador aún no sabía es que terminaría descubriendo un sistema que provee a organismos procariontes de inmunidad adaptativa y que la ciencia reinventaría para convertirla en una de las más revolucionarias herramientas de ingeniería genética conocidas (Doudna & Charpentier, 2014).

A lo largo de su estudio, Mojica se fijó en unas secuencias concretas del genoma de *H. mediterranei* que parecían sufrir modificaciones según la salinidad presente en el ambiente y que fueron seleccionadas para su secuenciación y análisis de expresión génica. Fue en estas secuencias donde se observaron los primeros patrones de repetición de secuencias nucleotídicas. En ellas, se observaron 14 repeticiones de pequeños segmentos de ADN de 30 bp de longitud y parcialmente palindrómicos que se sucedían a

distancias regulares y que a su vez se encontraban separados entre sí por otras regiones nucleotídicas. Estas regiones repetidas se encontraban localizadas en un área aparentemente no codificante próxima a las ORFs y no mostraban homología con ninguna proteína conocida (Mojica, Juez & Rodríguez-Valera, 1993). Algunos años antes del curioso hallazgo de Mojica, en 1987 un grupo de investigación liderado por Ishino se topaba también con 5 secuencias nucleotídicas repetidas en la región flanqueante 3' del gen *iap* de una cepa de *Escherichia coli* cuando realizaban un estudio nucleotídico de dicho gen. En este caso, las secuencias tenían una longitud de 29 nt y se disponían separadas entre ellas por secuencias de 32 nt de longitud (Ishino *et al.*, 1987). Este patrón de secuencias repetidas sería descrito también en otros organismos como *Shigella dysenteriae*, *Salmonella enterica* y otras cepas de *E. coli* (Nakata, Amemura & Makino, 1989), *Mycobacterium tuberculosis* (Van Soolingen *et al.*, 1991) y *Haloferax volcanii* (Mojica *et al.*, 1995).

La década de los 90 aporta al campo de la genética un panorama alentador debido al desarrollo de nuevas técnicas de secuenciación que facilitan el estudio del genoma de diversos organismos y aporta nuevos datos a cerca de estos (Mojica & Rodríguez-Valera, 2016). Estos nuevos avances permitieron describir las características generales de los particulares patrones de repeticiones de secuencias, conocidas como SRSR (*Short Regularly Spaced Repeats*) (Mojica *et al.*, 2000). Las SRSR se describen así como secuencias cortas de entre 24 y 40 bp que contienen en su estructura repeticiones palindrómicas terminales invertidas y altamente conservadas de hasta 11 bp de longitud. Este tipo de estructura se ha encontrado en muchas y muy diversas especies procariontas (concretamente en el 40% de las bacterias y aproximadamente el 90% de las arqueas secuenciadas hasta el año 2000), estando presentes en grupos filogenéticamente muy distantes. Además, se demuestra que estas secuencias están altamente conservadas en miembros del mismo grupo filogenético pero que además existe un alto porcentaje de similitud entre dominios, indicando un claro origen común (Mojica *et al.*, 2000).

Es en el año 2002, cuando observando las distintas nomenclaturas que se habían empleado para denominar estas secuencias en estudio, Jensen y Mojica llegan al acuerdo de referirse a estas como CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) (Mojica & Garrett, 2013).

Paralelamente al estudio de las características estructurales del loci CRISPR, se van llevando a cabo investigaciones que revelan la presencia de genes asociados al mismo. Jansen, durante su estudio del genoma de la bacteria Gram-negativa *Yersenia pestis*, revela la presencia de 4 genes estrictamente asociados a CRISPR y que pasarían a denominarse genes *cas* (Jansen *et al.*, 2002). En este estudio se vislumbran las posibles funciones de los productos codificados por estos genes, las denominadas proteínas *Cas*, y debido a la relación tan íntima de los genes *cas* con el loci CRISPR, se sugiere que existe una relación funcional entre ambas partes (Jansen *et al.*, 2002). Las proteínas codificadas por estos genes resultaron ser nucleasas o proteínas de interacción con el ADN, como por ejemplo, helicasas. En estudios posteriores se descubre que estos 4 genes no conforman la totalidad de los genes *cas*, sino que 4 años después se describen 41 familias más además de las 4 anteriormente descritas y también encontradas en la vecindad de CRISPR (Haft *et al.*, 2005).

Hasta aquel momento el estudio de CRISPR y de sus genes asociados no fue más que una mera descripción de sus componentes y su organización dentro del genoma. Sin embargo, es en el prolífico año 2005 cuando se publican estudios altamente relevantes acerca de la función biológica y origen del sistema CRISPR en procariontas. El análisis de las secuencias de las estructuras CRISPR en distintas especies confirma la homología de las secuencias espaciadoras con elementos extracromosomales (fagos y plásmidos) (Bolotin *et al.*, 2005; Mojica *et al.*, 2005; Pourcel, Salvignol & Vergnaud, 2005). Este hallazgo conduce a los investigadores a formular hipótesis sobre una función inmune

llevada a cabo por CRISPR en la que las secuencias espaciadoras serían las trazas de los elementos genéticos externos que infectaron la célula en tiempos pasados (Bolotin *et al.*, 2005; Pourcel *et al.*, 2005). En la línea de esta hipótesis, más tarde se demuestra que los organismos procariontes que contienen secuencias CRISPR, son menos susceptibles de ser infectados de nuevo por organismos patógenos que portan genomas con secuencias homólogas a las contenidas en dichas secuencias CRISPR (Mojica *et al.*, 2005). Sin embargo, no es hasta el año 2007 donde se demuestra la función inmune adaptativa del sistema CRISPR en *Streptococcus thermophilus*, al observar que esta deja de ser sensible a la infección por un fago al adquirir como espaciadores los protoespaciadores presentes en dichos fagos (Barrangou *et al.*, 2007).

Los estudios acerca de la estructura del sistema CRISPR-Cas y sus posibles funciones prosiguen, pero es en el año 2012, cuando la canadiense Jennifer Doudna y la francesa Emmanuelle Charpentier reinventan el sistema de defensa inmune procarionte CRISPR, modificándolo y convirtiéndolo en un mecanismo eficaz y viable en el ámbito de la ingeniería genética (Jinek *et al.*, 2012). A partir de este momento, son múltiples los estudios y artículos publicados que abalan este mecanismo como una promesa de la edición genética y muchas y muy variadas son ya las pruebas de su eficacia en la modificación de la información genética en organismos vivos (Figura 1).

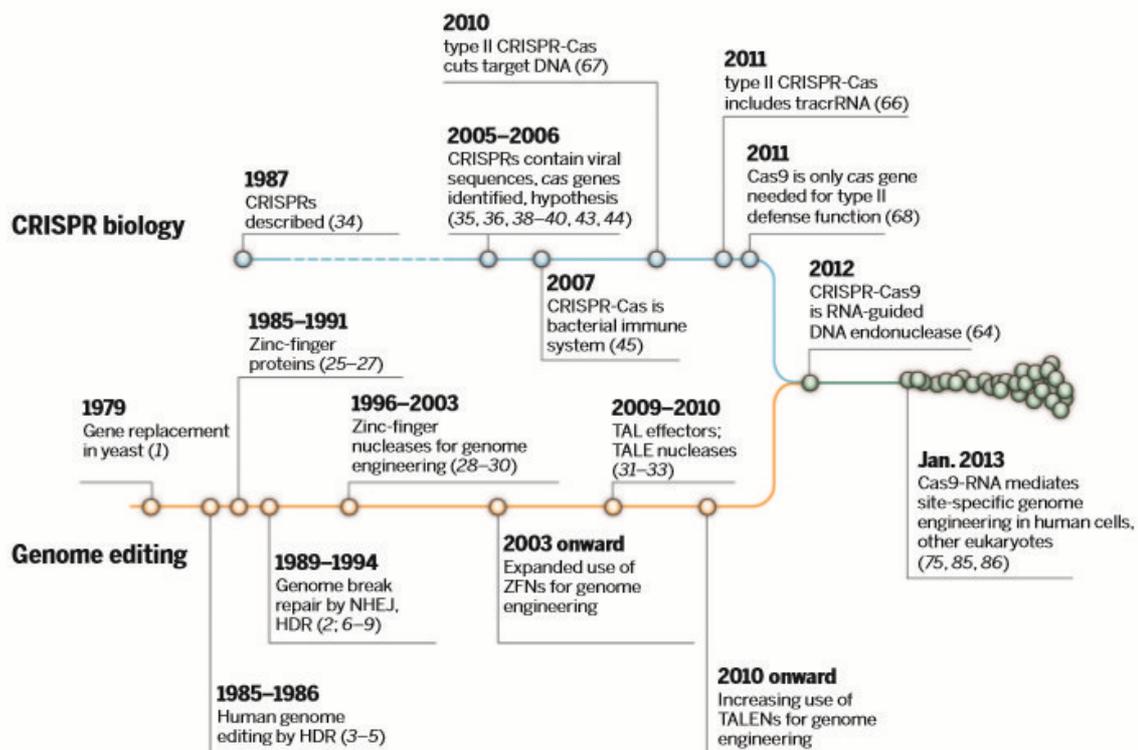


Figura 1. Cronograma con los hitos de mayor relevancia que han permitido la descripción biológica de CRISPR y su posterior empleo como herramienta de modificación genética hasta el año 2013 (Doudna & Charpentier, 2014).

4.1.2. La Biología del Sistema CRISPR-Cas en la Naturaleza

¿Qué es el sistema CRISPR-Cas?

CRISPR-Cas es un sistema de inmunidad adaptativa procarionte, presente en gran parte de las bacterias y la práctica totalidad de las arqueas que representa un ejemplo de memoria molecular transmisible de forma hereditaria. El sistema CRISPR actúa recopilando y almacenando fragmentos de material genético de bacteriófagos y plásmidos invasores, que transcriben en forma de crRNAs. Son estos crRNAs los que actuarán como guía de unas enzimas conocidas como endonucleasas Cas de la célula

hospedadora, que en procesos infectivos posteriores degradarán el ARN o ADN foráneo diana, siempre que éste presente una secuencia complementaria al crRNA (Mojica *et al.*, 2000); Mojica & Rodríguez-Valera, 2016; Wang, La Russa & Qi, 2016; de la Fuente-Núñez and Lu, 2017). Un ejemplo general de la estructura de un locus CRISPR se puede observar en la figura 2. Este locus presenta dos componentes diferenciados: el operón *cas*, que contiene las regiones codificantes de las endonucleasas y otras proteínas de unión a ADN; y la región CRISPR, que está formada por secuencias repetidas intercaladas con pequeñas secuencias espaciadoras, de origen exógeno (Figura 2).

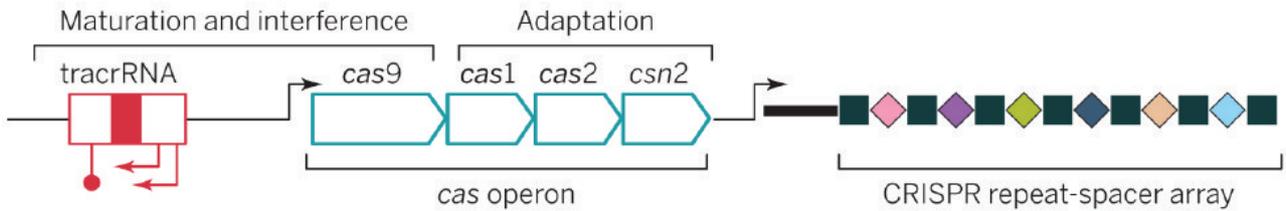


Figura 2. Estructura del locus CRISPR-Cas9 (Doudna & Charpentier, 2014).

Clasificación de los sistemas CRISPR

Existen diversos sistemas CRISPR que en la actualidad se encuentran clasificados en 2 clases, 5 tipos y 16 subtipos, siguiendo un criterio basado en la arquitectura de sus componentes y sus mecanismos de acción (Makarova *et al.*, 2011; Shmakov *et al.*, 2015). En primer lugar, según quien lleve a cabo la escisión de los ácidos nucleicos invasores, se distinguen sistemas CRISPR de clase 1 y clase 2. De esta forma, bajo la denominación de clase 1 se encuentran todos aquellos sistemas que requieren de un gran complejo de proteínas efectoras guiado por ARN para llevar a cabo la escisión del ADN o ARN

foráneo. Por el contrario, los sistemas CRISPR recogidos dentro de la clase 2, precisan de una única proteína endonucleasa efectora, que guiada por ARN lleve a cabo la neutralización del genoma invasivo, como es el caso de Cas9 y Cpf1 (Makarova *et al.*, 2011; Shmakov *et al.*, 2015). Dentro de estas dos clases, se agrupan los distintos tipos de sistemas CRISPR:
 Clase 1: tipos I, III y IV
 Clase 2: tipos II, V y VI

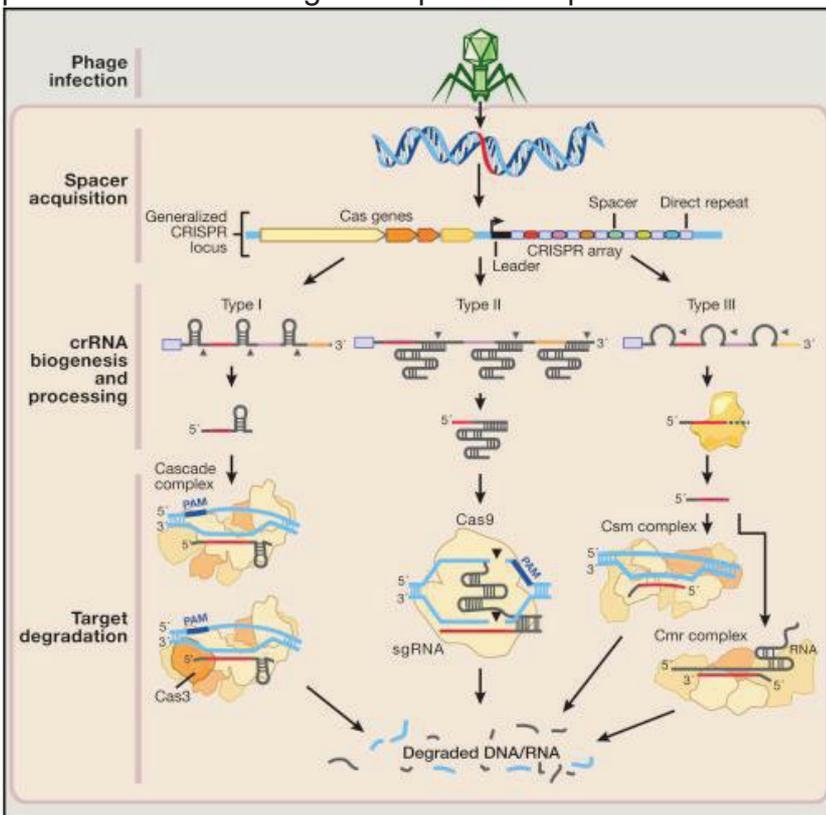


Figura 3. Esquematización comparativa de las distintas etapas de funcionamiento de los distintos tipos que conforman los sistemas CRISPR-Cas (tipos I, II y III). En esta figura podemos visualizar las diferencias en la biología de los tres tipos, observando que el tipo II difiere en gran medida de la maquinaria empleada por sus vecinos (Hsu, Lander & Zhang, 2014).

Etapas del funcionamiento del Sistema CRISPR-Cas

En general, el funcionamiento del sistema CRISPR para la consecución de la inmunidad adquirida se puede dividir en tres etapas bien diferenciadas: *etapa de adquisición*, *etapa de expresión* y *etapa de interferencia* (Figura 4). Cada uno de los diferentes tipos de sistema CRISPR presenta peculiaridades propias de su biología, sin embargo, este trabajo se centrará en analizar el *modus operandi* del sistema CRISPR-Cas9, de tipo II.

La **etapa de adquisición** tiene como objetivo la captación e incorporación de fragmentos de DNA foráneo (ya sea ADN o ARN, plasmídico o vírico, y a los que nos referiremos con el nombre de *protoespaciadores* cuando todavía se encuentran formando parte del genoma del ente invasor) al locus CRISPR de la célula hospedadora. Dentro de CRISPR, estos protoespaciadores serán insertados (pasando a denominarse espaciadores) entre las secuencias repetidas y en posición 5' respecto a la secuencia espaciadora y siempre tras la secuencia líder. En el caso del sistema CRISPR-Cas tipo II, esta fase corre a cargo de las proteínas Cas1 y Cas2 (Sapranauskas *et al.*, 2011), que son las encargadas de escindir el protoespaciador del resto del material genético foráneo e incluirlo dentro del locus CRISPR (Figura 4).

A continuación, en la **etapa de expresión**, CRISPR se transcribe dando lugar a un largo transcrito primario denominado pre-crRNA, que contiene al locus tracrRNA y a todos los espaciadores adquiridos en la etapa anterior separados por secuencias repetidoras. Este pre-crRNA es procesado mediante la endonucleasa RNAasa III, dando lugar a secuencias de ARN constituidas por una única región espaciadora flanqueada por una corta secuencia repetitiva, que servirá como zona de anclaje de Cas9 en la etapa posterior. Estas secuencias son los denominados crRNAs (Figura 4).

La **etapa de interferencia** tiene lugar cuando el genoma de un organismo foráneo reincidente intenta de nuevo infectar la célula. En este estadio, el crRNA maduro se une mediante complementariedad de bases Watson-Crick a la secuencia activadora tracrRNA, formando un híbrido de RNA (tracrRNA:crRNA), necesario para dirigir a la endonucleasa Cas9 hacia la secuencia diana del material genómico invasivo que debe degradar. Mediante el apareamiento de bases del crRNA con la secuencia diana del ente infectivo (formando el complejo cuaternario tracrRNA:crRNA:DNA), Cas9 logrará la neutralización del agente invasivo mediante la generación de una DSB en la secuencia diana del mismo (Tang *et al.*, 2002; Jinek *et al.*, 2012; Brouns *et al.*, 2008; Deltcheva *et al.*, 2011; Karvelis *et al.*, 2013; Kirchner & Schneider, 2015; Wang, La Russa & Qi, 2016; de la Fuente-Núñez & Lu, 2017) (Figura 4).

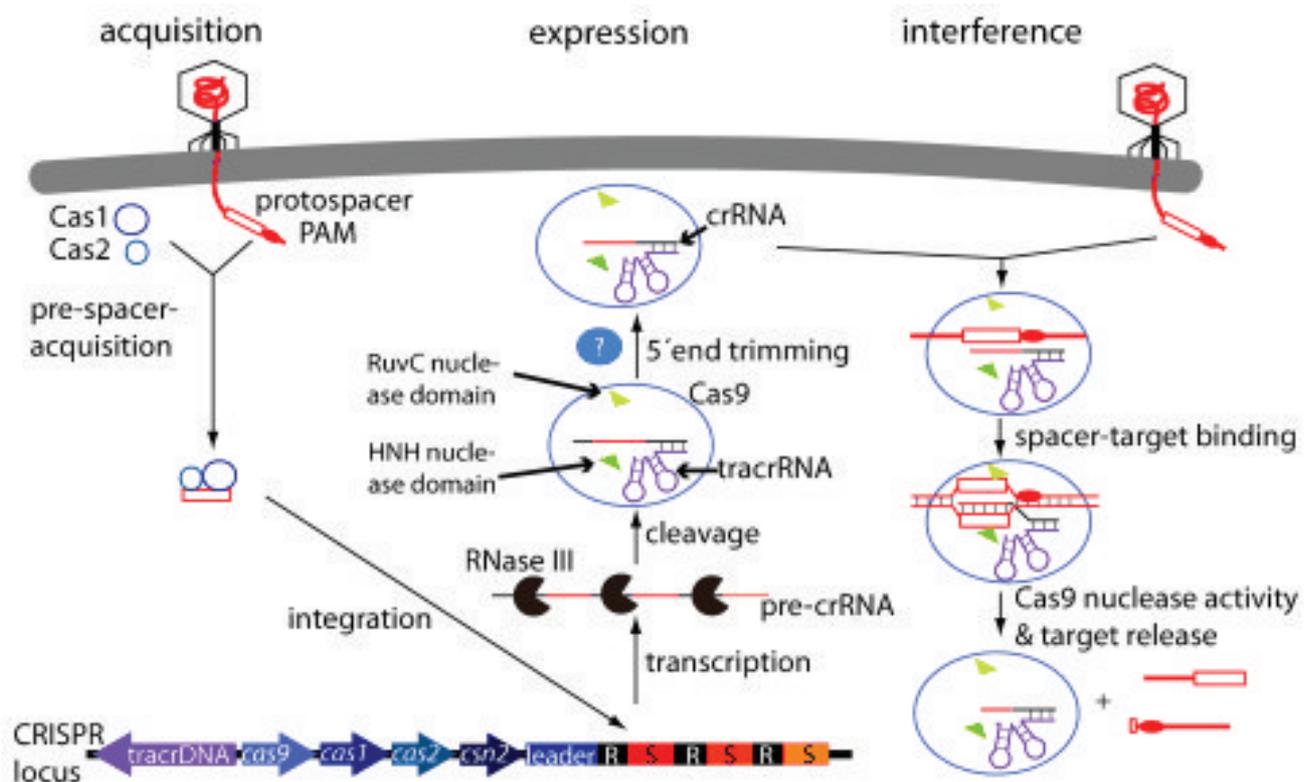


Figura 4. Visión general de los componentes y etapas de actuación del sistema CRISPR de tipo II en su defensa de la célula contra agentes invasivos. En la etapa de adquisición las proteínas Cas1 y Cas2 actúan escindiendo secuencias del material genético invasor, que será integrado en el locus CRISPR como secuencia espaciadora (S=espaciadora), entre secuencias repetidoras (R=repetidoras). Durante la etapa de expresión, el locus CRISPR dará lugar a un transcrito primario que será escindido en pequeños crRNAs por la RNAasa III. En la etapa de interferencia final, se formará un complejo efector entre los crRNAs, tracrRNA y Cas9, la cual escindirán aquellas regiones del ADN foráneo complementarias a la secuencia crRNA con la que esté formando complejo (Kirchner & Schneider, 2015).

Como podemos observar, la secuencia tracrRNA es un pequeño RNA no codificante con dos funciones críticas, disparar el procesamiento del pre-crRNA por la enzima RNAasa III y subsecuentemente activar la guía del crRNA para la escisión llevada a cabo por Cas9. Aunque este sea el mecanismo natural de acción de CRISPR-Cas, en los procesos de edición genética (los cuales trataremos más adelante), se tiende a la simplificación de la maquinaria fusionando en una quimera a tracrRNA y crRNA y dando lugar así al denominado sgRNA. Este sgRNA retiene la capacidad de guiar a la nucleasa Cas9 hacia la secuencia diana que se pretenda variar. De esta forma, alterando simplemente la secuencia de 20 nt complementaria a la secuencia diana de ADN (que se corresponde con los primeros 20 nt de sgRNA en posición 5'), se podrá modificar cualquier región del genoma (Ran *et al.*, 2013; Hsu, Lander & Zhang, 2014; de la Fuente-Núñez & Lu, 2017).

Una vez analizada toda la biología del proceso de defensa inmune llevado a cabo por el sistema CRISPR-Cas, podemos afirmar que es la transcripción del locus CRISPR (que da lugar a los crRNAs) la base de este mecanismo de defensa procarionota contra el material genético exógeno (Brouns *et al.*, 2008).

En este proceso, una pequeña secuencia adyacente a la secuencia del material genético del organismo invasivo que es adquirida como espaciadora en el locus CRISPR juega una gran importancia. Esta es la denominada secuencia PAM, de la que hablaremos a continuación.

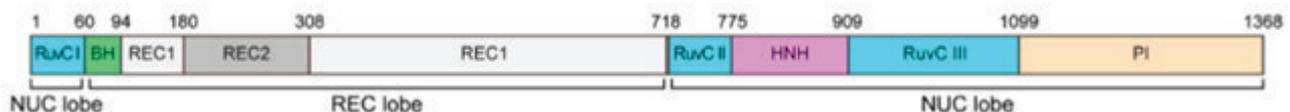
Estructura y modo de acción de la proteína Cas9

La endonucleasa Cas9 está presente en la última etapa del mecanismo. Cas9 presenta una arquitectura bilobular, constituida por el lóbulo de acción nucleasa NUC y el lóbulo de reconocimiento REC (Figura 5). Ambas estructuras se encuentran separadas por un canal central cargado positivamente donde se inserta el heterodúplex de carga negativa tracrRNA:crRNA:ADN en la etapa de interferencia. El lóbulo NUC, a su vez, presenta 4 dominios: los dominios nucleasa HNH y RuvC creadores del DSB en la secuencia diana del ADN invasor, el dominio PI que interacciona con la secuencia PAM y el dominio WED. La secuencia PAM consiste en una secuencia de tres a cinco nucleótidos, que varía siempre entre especies y que es fundamental para el reconocimiento de la diana por la endonucleasa (Leenay *et al.*, 2017). Por su parte, el lóbulo REC (que difiere según los distintos ortólogos de Cas9) agrupa las regiones que actúan reconociendo al complejo tracrRNA:crRNA:ADN, diferenciándose en él los dominios REC1, REC2 y BH (hélice puente) (Jinek *et al.*, 2012; Nishimasu *et al.*, 2014; (Wang, La Russa & Qi, 2016).

Tanto la delección del dominio PI en el lóbulo NUC como la del dominio REC1 en el lóbulo REC, abolen la actividad nucleasa de Cas9, indicando que ambos son críticos para el desarrollo del mecanismo de defensa CRISPR-Cas. Además, ambos dominios adoptan estructuras únicas no similares a ninguna proteína conocida, lo cual indica que son dominios de función específica de Cas9 (Nishimasu *et al.*, 2014).

Los fosfatos presentes en el esqueleto nucleotídico del ADN diana interactúan con los dominios REC1, RuvC y PI (Nishimasu *et al.*, 2014).

a)



b)

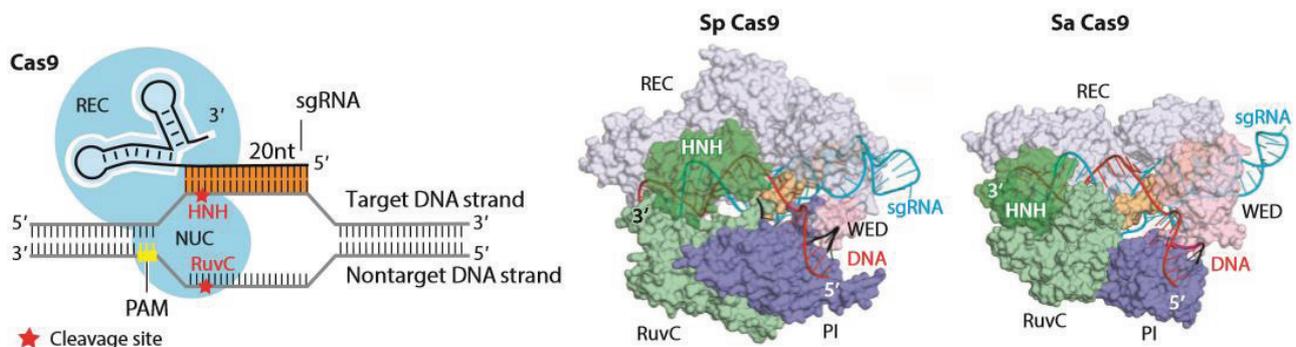


Figura 5. (a) Disposición de los lóbulos de SpCas9 (con sus correspondientes dominios) en el espacio, de forma lineal (Nishimasu *et al.*, 2014). (b) (imagen inferior izquierda) podemos visualizar de forma esquemática la estructura cuaternaria de Cas9 asociada al heterodúplex tracrRNA:crRNA:DNA en la etapa de interferencia del sistema CRISPR-Cas. A continuación (imagen inferior derecha) se muestra la estructura que Cas9 adoptaría en el espacio, señalizando los dominios de su arquitectura, comparando las nucleasas de dos organismos distintos y pudiendo observar que ambas estructuras son prácticamente idéntica (Wang, La Russa & Qi, 2016).

La neutralización del material genético plasmídico o viral comienza por el reconocimiento de la secuencia PAM por el dominio PI de Cas9 (en la cadena no complementaria de la secuencia que se pretende escindir) y del híbrido tracrRNA:crRNA complementario a dicha secuencia. De esta manera se formará un complejo ternario constituido por el complejo de ARN y la Cas9. Tras esto, ocurrirá el apareamiento de

bases entre el ADN infectivo y el crRNA, dando lugar al heterodúplex de aproximadamente 20 bp tracrRNA:crRNA:DNA, que se posicionará dentro de Cas9 en el surco central entre los lóbulos REC y NUC. Una vez formando este complejo cuaternario (tracrRNA:crRNA:ADN y Cas9) el dominio móvil HNH del lóbulo NUC se aproximará a la cadena complementaria de la secuencia diana provocando su escisión y lo mismo ocurrirá con el dominio RuvC con la cadena no complementaria. Como consecuencia, entre ambos dominios se da lugar a una DBS que destruirá el genoma del organismo infectante (Nishimasu *et al.*, 2014; Wang, La Russa & Qi, 2016) (Figura 6).

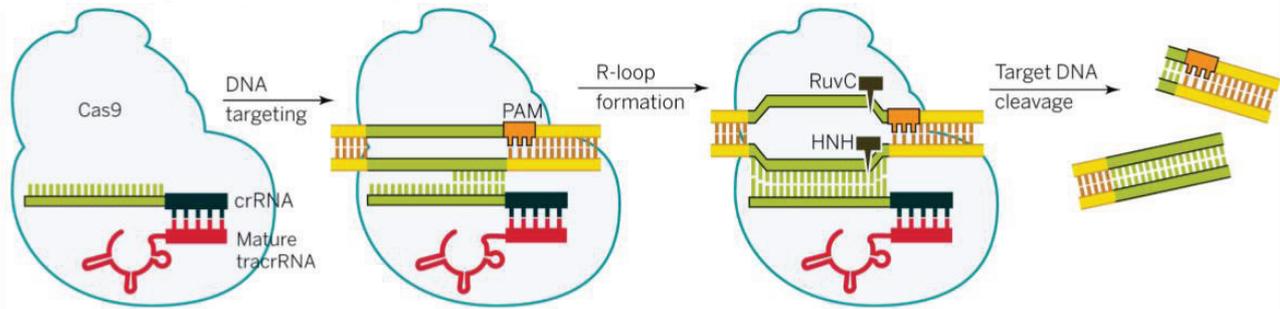
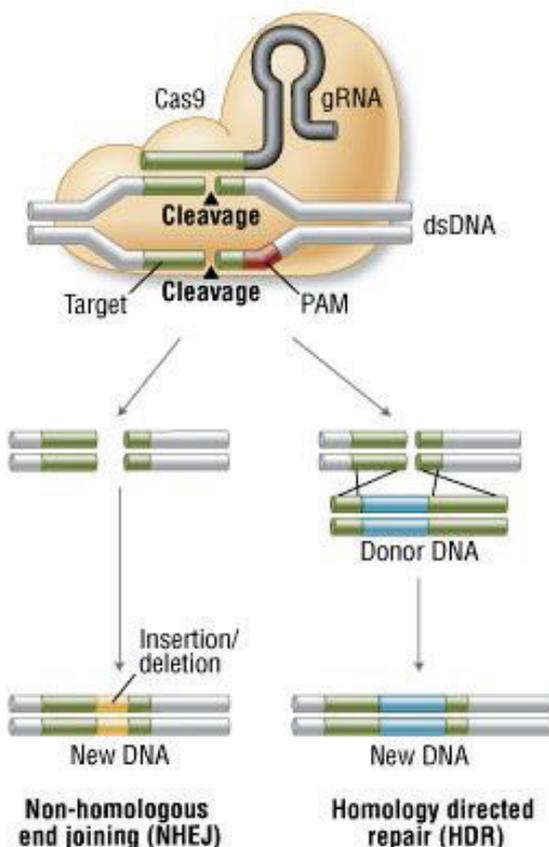


Figura 6. Secuencia de acción de Cas9 en la neutralización del agente infectivo mediante la generación de DSBs en la secuencia diana de su ADN. En este esquema se puede observar la localización que adopta el complejo ternario tracrRNA:crRNA:DNA en el surco medio entre los lóbulos REC y NUC de Cas9 y la creación de roturas en las cadenas del ADN infectante, originadas por los dominios HNH y RuvC en las cadenas complementaria y no complementaria a crRNA, respectivamente (Doudna & Charpentier, 2014).

4.1.3. El Sistema CRISPR-Cas9 como Herramienta de Edición Genómica

Patrones de reparación tras la escisión de Cas9



Antes de pasar a comentar las distintas modificaciones practicadas en el sistema CRISPR-Cas que impulsaron su avance como herramienta de edición genética a lo largo de estos últimos años, cabe realizar una breve mención a los mecanismos naturales de reparación de las DSBs, que además, nos servirán como nexo de unión entre la función natural de CRISPR-Cas ya comentada y la función que ejerce como herramienta de edición genómica, que veremos a continuación. Y es que los patrones naturales de reparación del ADN también han sido estudiados y empleados entre los grupos de investigación de manera controlada para inducir modificaciones genéticas.

Figura 7. Comparación gráfica de los mecanismos naturales de reparación de DSBs, NHEJ y HDR. El sistema de reparación NHEJ causa indels que llevan a la disrupción génica. El sistema HDR sólo tiene lugar cuando está presente una secuencia nucleotídica externa que recombina en el lugar de la escisión, por lo que permite introducir modificaciones concretas en el material genético (Doudna & Charpentier, 2014).

Tras la escisión de la secuencia del ADN diana por Cas9 la cual ha generado un DSB, esta rotura en la doble cadena debe ser reparada por la maquinaria celular. Estos patrones de reparación son aprovechados por los investigadores para generar modificaciones en lugares concretos del genoma con el que se esté trabajando. Se dan mayoritariamente dos patrones de reparación, denominados NHEJ y HDR (Figura 7). El mecanismo NHEJ de reparación de DSBs es un patrón caracterizado por ser “propenso a errores” ya que se trata de un sistema de reparación que deja cicatrices en forma de indels en el lugar en el que actúa. Estas mutaciones en forma de inserciones o deleciones de nucleótidos, si tienen lugar en el interior de exones codificantes darán lugar a cambios en la pauta de lectura que pueden dar lugar a la aparición de codones de parada prematuros. Debido a esto, puede generarse el silenciamiento de genes o elementos genómicos. Por su parte, el patrón de reparación HDR sólo tendrá lugar ante la presencia de una secuencia nucleotídica añadida de forma exógena y que por tanto podrá ser usada como herramienta de modificación altamente precisa (ya que es el propio investigador el que decide cuál será la secuencia que será introducida en el lugar de la escisión). Dicha secuencia deberá ser de doble cadena y homóloga a las regiones flanqueantes de la región del ADN escindida donde vaya a incorporarse (Ran *et al.*, 2013; Wang & Qi, 2016; Xiong *et al.* 2016).

nCas y dCas

Una vez fueron conocidas la estructura de Cas9 y los dos típicos sistemas de reparación de las DSBs generadas en el ADN, con el objetivo de que no ocurriesen mutaciones no deseadas en las secuencias diana de ADN que se pretenden modificar a causa del mecanismo de reparación NHEJ, se invirtió un gran esfuerzo en investigar acerca de cómo se podría solventar este riesgo. Fruto de estas numerosas investigaciones nace nCas9, una nCas9 que tiene mutado uno de los dos dominios nucleasa (HNH o RuvC), de manera que sólo podrá escindir una de las cadenas de ADN (Figura 8). De esta forma, si el objetivo es realizar un DSB sin que ocurran mutaciones incontroladas, se deberán emplear dos nCas, cada una con una mutación en un dominio nucleasa distinto. Cada nCas escindirán una de las hebras, generando extremos protuberantes en los que no tendrá lugar el NHEJ pero sí se podrá llevar a cabo un HDR introduciendo una secuencia particular. Estas mutaciones se logran mutando el aspartato situado en la décima posición en el dominio RuvC a alanina (D10A) o la histidina en la posición 840 del dominio HNH a alanina (H840A) (de la Fuente-Núñez & Lu, 2017).

De la misma forma, también se pueden emplear nucleasas Cas9 con ambos dominios nucleasa mutados (Figura 8), de forma que esta enzima perderá su acción nucleasa pero no la de reconocimiento por complementariedad de bases Watson y Crick de la secuencia diana de ADN. Estas son las denominadas dCas (o dead Cas). dCas son empleadas como mecanismo de regulación de la transcripción, ya sea inhibiéndola o activándola (CRISPRi y CRISPRa, respectivamente) o el marcaje fluorescente mediante la fusión de dCas9 a proteínas como GFP para localizar mediante fluorescencia secuencias concretas en el genoma, entre otras (Hsu, Lander & Zhang, 2014; Kirchner & Schneider, 2015; Wang, La Russa & Qi, 2016; Xiong *et al.*, 2016; de la Fuente-Núñez & Lu, 2017).

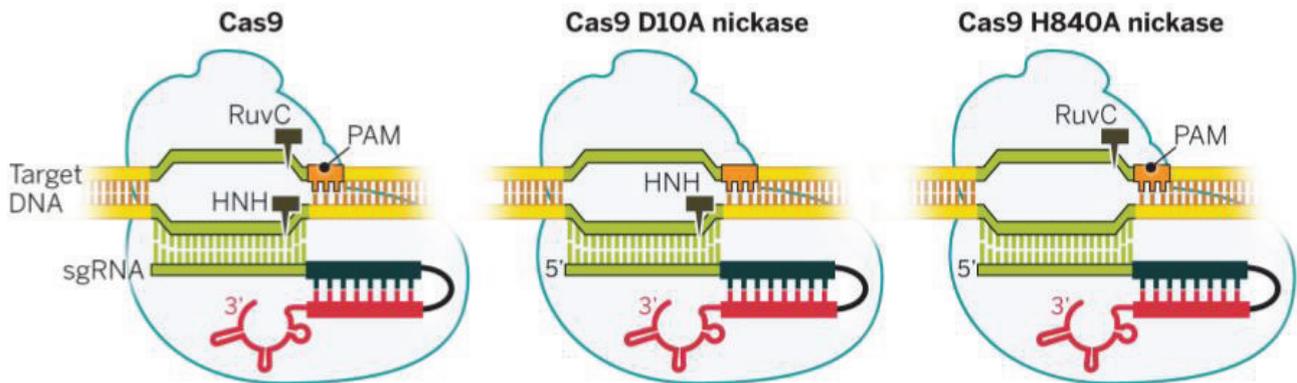


Figura 8. Representación gráfica y comparativa de las dos formas posibles de nCas9 y la forma no mutada de Cas9. Véase como la mutación en un dominio nucleasa, RuvC en el esquema central y HNH en el de la derecha, inhabilita dicho dominio, funcionando únicamente el dominio nucleasa contrario (Doudna & Charpentier, 2014).

Cpf1 como enzima nucleasa y sus ventajas en el Sistema CRISPR-Cpf1

La proteína Cpf1, descrita en organismos tales como *Francisella novicida* (Zetsche *et al.*, 2015) y *Lachnospiraceae bacterium* (Dong *et al.*, 2016) es una endonucleasa guiada por ARN de entre 1.200 y 1.500 aminoácidos de longitud. Junto a CRISPR forma el sistema CRISPR-Cpf1, dentro del tipo V de la clase 2 de los sistemas CRISPR-Cas. El sistema CRISPR-Cpf1 presenta diferencias estructurales frente al sistema CRISPR-Cas9 ya estudiado que, como veremos, facilitarán su uso como herramienta biotecnológica frente al segundo.

- ❖ La primera diferencia entre ambos sistemas es que en el sistema CRISPR-Cpf1, Cpf1 no requiere la presencia de tracrRNA ni para la maduración de los crRNAs ni en la etapa de interferencia, ya que el complejo formado por crRNA-Cpf1 es capaz de escindir el ADN por sí solo (Figura 9). Además, el crRNA empleado en este sistema es mucho más corto (aproximadamente 42 nt) que el utilizado en el sistema CRISPR-Cas9 (que ronda los 100 nt). Ambas características facilitan y abaratan el costo de producción de diseño de herramientas para la edición genómica (Zetsche *et al.*, 2015; Dong *et al.*, 2016).
- ❖ Por otro lado, al igual que Cas9 Cpf1 efectúa un DSB, pero a diferencia de la primera que genera extremos cohesivos, Cpf1 genera un corte escalonado en el ADN, generando extremos protuberantes en posición 5' de entre 4 y 5 nt de longitud (Figura 9). La existencia de extremos protuberantes, como hemos explicado, impide los NHEJ creadores de indels. De esta forma, pudiendo diseñar un inserto de ADN cuyos extremos sean complementarios a los extremos protuberantes dejados por Cpf1 en el ADN, este será capaz de integrarse en la orientación correcta (Zetsche *et al.*, 2015; Dong *et al.*, 2016).
- ❖ Relacionado con la característica anterior, CRISPR-Cpf1 se posiciona como una mejor herramienta biotecnológica frente a CRISPR-Cas9, ya que elimina el riesgo a la aparición de indels dentro de la secuencia diana de ADN que pudiesen eliminar toda opción a una nueva inserción de ADN en ese sitio en concreto. Esto se debe a que Cpf1 escinde el ADN en el extremo distal del protoespaciador y no dentro de la secuencia diana como Cas9. Gracias a esto, cualquier indel que se pudiese generar después de la reparación NHEJ no afectaría a la secuencia diana, que se mantendrá invariable después de varias rondas de escisión (Zetsche *et al.*, 2015; Dong *et al.*, 2016).
- ❖ Por último, otra ventaja de este sistema es que reconoce regiones PAM ricas en timina (en contraste con los PAM ricos en guanosina reconocidos por la familia Cas9). Esto permite la edición de genomas de organismos ricos en adenina y

timina y de regiones del genoma de organismos en general ricas en estos nucleótidos, que con el sistema CRISPR-Cas9 no podrían ser modificados (Zetsche *et al.*, 2015; Dong *et al.*, 2016).

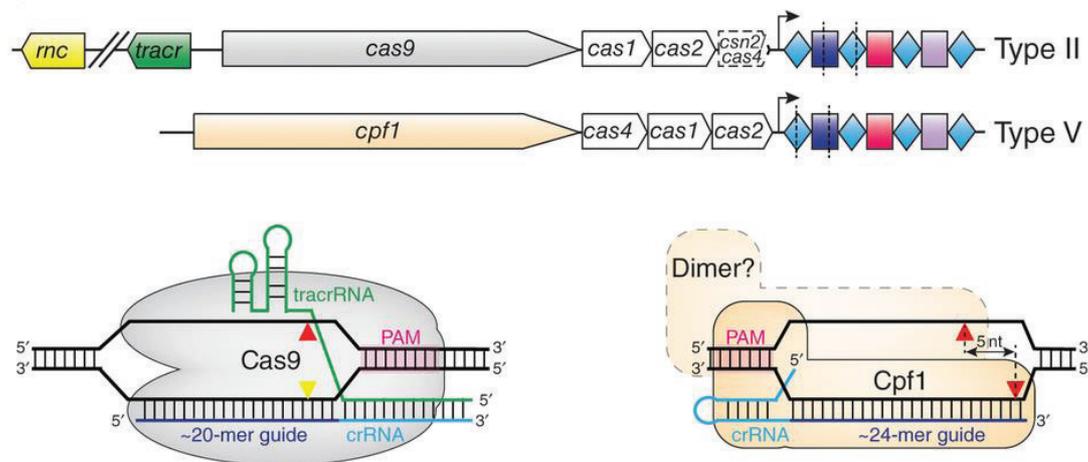


Figura 9. Esquema comparativo de la estructura de los sistemas CRISPR-Cas9 y CRISPR-Cpf1 (imagen superior) y de la etapa de interferencia llevada a cabo por ambos sistemas (imagen inferior). Es destacable la no presencia de *tracrRNA*, la escisión escalonada que da lugar a extremos protuberantes de 5 nt de longitud y la posición 5' de la secuencia PAM rica en T en el sistema CRISPR-Cpf1, características que difieren del sistema CRISPR-Cas9 anteriormente explicado (Sontheimer & Wolfe, 2015).

Debido a las características estructurales que incrementan las ventajas del sistema CRISPR-Cpf1 frente a otros sistemas de edición genética, este ha comenzado a ser utilizado en el campo de la ingeniería genética en diversos organismos y con distintas finalidades. Ejemplo de esto es su utilización en la corrección de la distrofia muscular en cardiomiocitos humanos y de ratones (Zhu *et al.*, 2017), la modificación de la bacteria *Corynebacterium glutamicum*, muy empleada a nivel industrial para la síntesis de aminoácidos y ahora, mediante el sistema CRISPR-Cpf1, también para la obtención de otros productos como biofuels (Jiang *et al.*, 2017) o en diversas modificaciones genéticas en ratones (Watkins-Chow *et al.*, 2017).

¿Por qué CRISPR-Cas ha resultado ser una herramienta biotecnológica revolucionaria?

Ahora que conocemos la estructura y el modo de actuación del sistema CRISPR-Cas, cabe preguntarnos por qué este sistema inmune adaptativo procarionta, reinventado como herramienta biotecnológica, ha resultado ser una de los mayores y más revolucionarios descubrimientos en el campo de la ingeniería genética. Y es que CRISPR-Cas no ha sido la primera tecnología de modificación del material genético, sino que antes de ella fueron empleados otros mecanismos, tales como los ya mencionados ZFNs y los TALENs. En este apartado abordaremos las ventajas que han convertido a CRISPR-Cas en una alternativa sencilla, eficaz y barata de edición genética, por delante de sus predecesores.

Por un lado, el empleo de CRISPR-Cas para editar múltiples fragmentos de ADN es fácilmente viable ya que simplemente basta con personalizar la secuencia nucleotídica del sgRNAs de tan solo 20 nt de longitud que se apareará con el ADN diana que se quiera modificar. Por el contrario, el empleo de ZFNs y TALENs requiere del diseño de nuevos genes cada vez que se quiera modificar una región distinta del genoma (Figura 10). Esto supone un incremento en la dificultad, en los costes y en el tiempo de desarrollo del proceso. Además, debido a esto, CRISPR-Cas es un sistema altamente eficaz ya que

puede ser empleado para la modificación simultánea de múltiples loci, únicamente incluyendo en la célula a modificar los distintos sgRNAs correspondientes a las secuencias objetivo de la edición genética, lo que incrementa la eficacia de este método de edición (Ran *et al.*, 2013; Hsu, Lander & Zhang, 2014). Por ello, CRISPR-Cas se ha alzado como una de las grandes promesas de la Biotecnología contemporánea, prometiendo grandes mejoras en campos tan dispares como la Medicina o la alimentación.

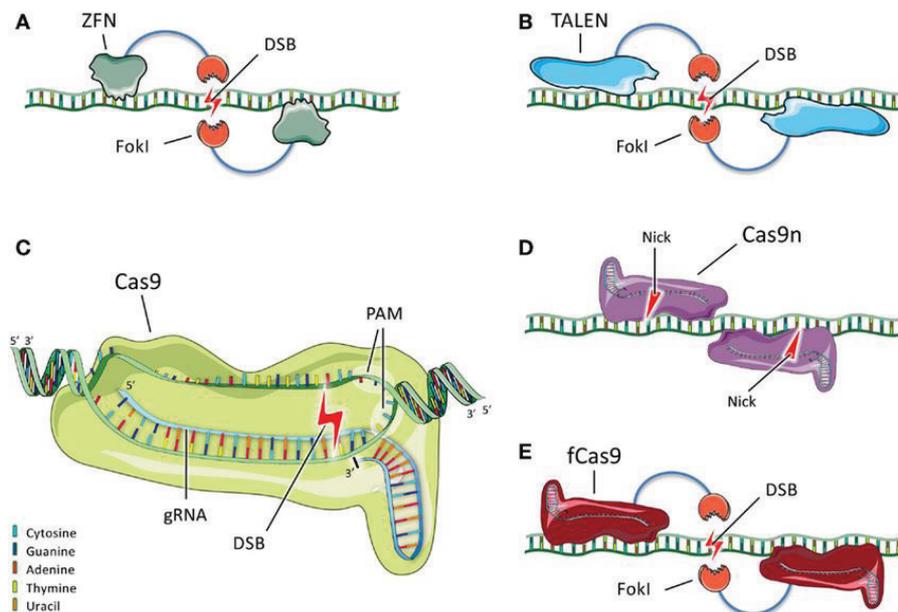


Figura 10. Diferencias fundamentales entre el sistema CRISPR y los sistemas ZFN y TALEN. En el primer sistema basta con diseñar un sgRNA para localizar y degradar la secuencia diana. En los dos últimos, es necesario modificar las ORFs que codifican para las endonucleasas con el fin de que éstas reconozcan las secuencias diana y puedan degradarlas (Ott De Bruin, Volpi & Musunuru, 2015).

4.2. Diseño teórico de un sgRNA para edición del genoma de un organismo eucariota

Tras haber profundizado de manera teórica en la estructura y mecanismos del sistema CRISPR-Cas9, a continuación se expone un ejemplo práctico de su uso como herramienta de edición genética en el organismo eucariota *Saccharomyces cerevisiae*. El objetivo es mostrar, paso a paso, de qué forma se puede introducir en el genoma de esta levadura el gen que codifica para la proteína GFP (*Green Fluorescent Protein*) en las denominadas secuencias *delta*, dispersas por todo el genoma de este hongo. Para ello es necesario el diseño de un sgRNA que conduzca a la nucleasa Cas9 hasta dichas secuencias, donde generará un DSB gracias al cual se puede llevar a cabo la introducción del gen *gfp* por recombinación homóloga.

a. **Microorganismo:** *Saccharomyces cerevisiae*. *S. cerevisiae* es un hongo unicelular del que se conoce su genoma completo (es el organismo eucariota mejor conocido por lo que es tomado como modelo biológico del dominio). Su genoma consta de 6.000 genes repartidos en 16 cromosomas (Reynaud, 2010). La elección de *S. cerevisiae* como protagonista de nuestro diseño se debe a que es una especie robusta y tolerante a condiciones adversas, que ha sido empleada a lo largo de la historia en múltiples campos como la alimentación (pan, vino o cerveza) o la síntesis de compuestos farmacéuticos. Debido a ello, muchos son los avances en ingeniería genética que son testados en este organismo (Shi *et al.*, 2016). Particularmente, en este caso se emplea

la cepa *S. cerevisiae* S288C, cuya secuencia completa se encuentra disponible en el GenBank.

- b. **Localización de la inserción.** Para realizar la inserción del marcador molecular se utilizan los sitios *delta*. Se trata de secuencias de aproximadamente 330 bp ampliamente distribuidas por todo el genoma de los organismos del género *Saccharomyces* (Gafner, Robertis & Philippsen, 1983). Para localizar las secuencias *delta* en la cepa S288C de *S. cerevisiae*, se realiza una búsqueda de homología en la base de datos SGD (*Saccharomyces Data Base*) con secuencias *delta* obtenidas de la bibliografía (Shi *et al.*, 2016). Se selecciona una de las secuencias *delta* del cromosoma XIV (YNLCdelta2), localizada entre las posiciones 102.192 y 102.525.
- c. **Selección del marcador molecular a insertar.** En este ejemplo se selecciona el gen que codifica para la proteína GFP, la cual es una proteína monomérica de unos 230 aminoácidos cuya peculiar estructura es la responsable de la emisión de la bioluminiscencia verde que la caracteriza (Tsien, 1998).
- d. **Diseño de cebadores y amplificación del casete *gfp-delta* para inserción en el genoma.** El plásmido pCYC1m_yeGFP (Addgene) se utiliza como ADN molde para la amplificación del casete *gfp-delta*. Este vector contiene la ORF codificante de GFP, flanqueada por un promotor (CYC1) y terminador (ADH1) que son funcionales en *S. cerevisiae* (Figura 11).

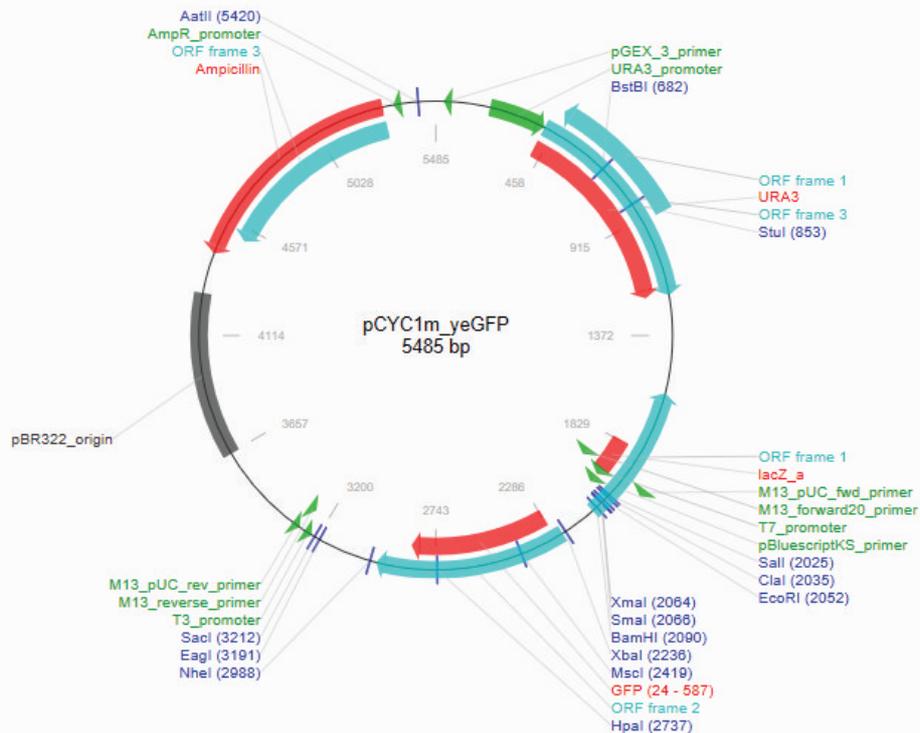


Figura 11. Esquema del vector pCYC1m-yeGFP de Addgene utilizado como molde para la amplificación (addgene.org).

La amplificación se hace utilizando unos cebadores quiméricos que contienen una región que hibrida con la secuencia que pertenece al promotor (en 5'), en el caso del cebador *forward*, o con la secuencia del terminador (en 3'), en el caso del cebador *reverse*, y la secuencia *delta* para la recombinación homóloga (Figura 12). El objetivo es amplificar el gen *gfp* mediante PCR para luego introducirlo en las secuencias *delta* de *S. cerevisiae* mediante recombinación homóloga. Para ello, debemos diseñar los cebadores

que nos permiten amplificar la secuencia nucleotídica completa del gen *gfp* junto con el promotor *CYC1p* y el terminador *ADH1t*. Para localizar la secuencia codificante de *gfp* en la secuencia del plásmido *pCYC1m_yeGFP* se realiza un BlastN (blast.ncbi.nlm.nih.gov/Blast.cgi), entre la secuencia completa del vector y la secuencia codificante de *gfp* obtenida de la base de datos GenBank. Esto permite conocer la posición exacta de la *ORF* del gen en el plásmido. Tras localizar la región codificante, se toman aproximadamente 1.000 bp 5' y 3' flanqueantes para tener la certeza de que las regiones promotora y terminadora de *gfp* son amplificadas también.

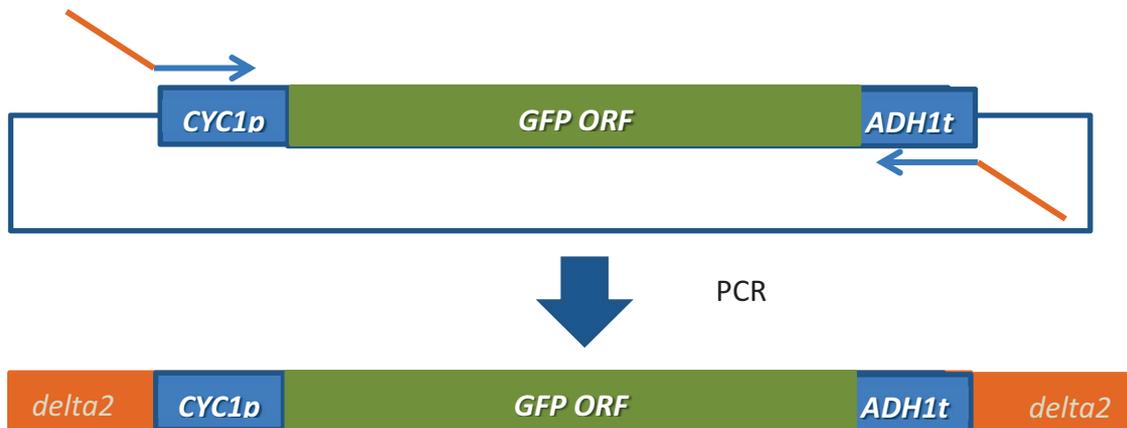


Figura 12. Esquema del casete *gfp-delta* tras amplificación por PCR con los cebadores MegaCebadorDelta1 y MegaCebadorDelta2.

En estas secuencias flanqueantes, se seleccionan 19 nucleótidos para el diseño del cebador *forward* y 21 nucleótidos para el cebador *reverse*:

Cebador *gfp-deltaForward*: 5'-ATGCGGTGTGAAATACCGC-3'

Cebador *gfp-deltaReverse*: 5'-GGGCCATTCTGTGCTGAATAG-3'

La diferencia de longitud en la secuencia de ambos oligonucleótidos está determinada por la T_m , que no debe diferir en más de 1°C (Figura 13).

Note: Recent changes to T_m calculations

Product Group
Q5

Polymerase/Kit
Q5 High-Fidelity DNA Polymerase

Primer Concentration (nM)
500 [Reset concentration](#)

Primer 1
ATGCGGTGTGAAATACCGC

Primer 2
GGGCCATTCTGTGCTGAATAG

Anneal at
67 °C

Primer 1
19 nt
53% GC
T_m: 66 °C

Primer 2
21 nt
52% GC
T_m: 66 °C

Figura 13. Captura de pantalla en la que se observa la secuencia nucleotídica de los dos cebadores (*Primer 1* = *gfp-deltaForward* y *Primer 2* = *gfp-deltaReverse*) y la T_m final de cada uno de los cebadores, que es exactamente la misma. Además, el programa informático tmcalculator.neb nos da de forma automática la temperatura de anillamiento que tendrá que tener la PCR (tmcalculator.neb).

Siguiendo con el diseño de los cebadores, a continuación se diseña la fracción nucleotídica complementaria a las secuencias *delta* y que posibilita la recombinación homóloga en el genoma de *S. cerevisiae* (Figura 12). En este caso estas secuencias tendrán 30 nt de longitud y serán las siguientes:

- ❖ Secuencia *delta* 5': 5'-TGTTGGAATAAAAATCCACTATCGTCTATC-3'
- ❖ Secuencia *delta* 3': 5'-TGTTAATAGAGTTGTAAGTGGGTAAAGAGT-3'

Estas secuencias se añaden a las secuencias de los cebadores *gfp-delta* en posición 5' para la obtención de los megacebadores que son utilizados para la amplificación del casete mediante PCR (Figura 12) y posterior transformación de la cepa S288C de *S. cerevisiae* y que se inserta en el sitio *delta* por recombinación homóloga:

- ❖ MegaCebadorDelta1:
5'- TGTTGGAATAAAAATCCACTATCGTCATGCGGTGTGAAATACCGC-3'
- ❖ MegaCebadorDelta2:
5'-TGTTAATAGAGTTGTAAGTGGGTAAAGAGTGGGCCATTCTGTGCTGAATAG-3'

e. Diseño del megaprimer codificante del sgRNA y construcción del plásmido Di-Delta-Marta. Para el diseño del sgRNA se utiliza la herramienta Web CRISPy-Web (<http://crispy.secondarymetabolites.org>). Para ello se descarga la secuencia completa del del cromosoma XIV de la cepa S288C, en formato GenBank y se carga en la aplicación CRISPy-Web. Esta aplicación realiza una búsqueda automática de secuencias PAM dentro del rango de búsqueda que el investigador acota. En este caso se utiliza el rango 102.262-102.301, localizado dentro de la secuencia *delta2* en este cromosoma.

El *output* son varias secuencias nucleotídicas, de 21 nucleótidos de longitud (los necesarios para la generación de un sgRNA) y 3 nucleótidos correspondientes a la secuencia PAM (en este caso, la secuencia NGG, para el sistema Cas9) (Figura 14).

Start	End	Strand	ORF	Sequence	PAM	0bp mismatches	1bp mismatches	2bp mismatches	Download
11	34	1	-	GAAGTCTCCTCGAGGATAT	AGG	1	6	2	
4	27	1	-	TATACTAGAAGTCTCCTCG	AGG	8	0	0	

Figura 14. *Output* del análisis de la región 102.262 -102.301 del cromosoma XIV de la cepa S288C de *S. cerevisiae* realizado por la herramienta de búsqueda CRISPy-Web (crispy.secondarymetabolites.org).

Aunque el programa genera automáticamente varias opciones, se seleccionan aquellas que presentan menos posibilidades de *off-targets*. En el ejemplo propuesto, la secuencia inferior presenta 0 posibilidades con 1 o 2 *mismatches*.

Para la síntesis del megaoligonucleótido se añaden al diseño del mismo unas colas conteniendo la secuencia de reconocimiento de la enzima de restricción que permiten la inserción del megadúplex en el vector de expresión del sgRNA. En este caso se utiliza el vector pDi-CRISPR (Shi *et al.*, 2016), siendo el sitio de restricción el sitio de reconocimiento *Bsal* (Figura 15).

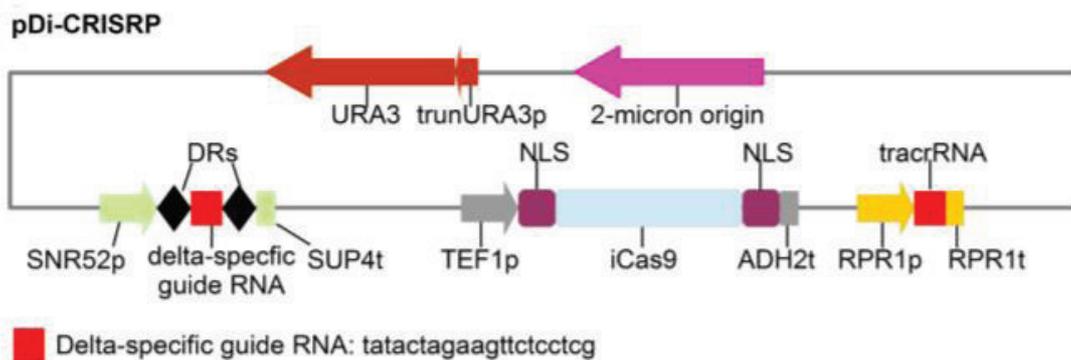


Figura 15. Esquema del vector pDi-CRISPR. Este vector contiene la secuencia codificadora de iCas9, bajo el control de un promotor y un terminador funcionales en levaduras. Contiene también la secuencia codificadora del tracrRNA, también bajo el control de promotor y terminador funcionales en levaduras. Además, presenta un sitio *Bsal* de inserción del fragmento de ADN codificador del sgRNA entre un promotor (SNR52p) y un terminador (SUP4t) funcionales en levaduras (Shi *et al.*, 2016).

Secuencia de los megaoligonucleótidos para inserción en el sitio *Bsal* de pDi-CRISPR:

delta.gRNA1_MartaF

5'- GGTCTCCCAAATATACTAGAAGTTCTCCTCG**AGGTTTAGAGCGAGACC** - 3'

delta.gRNA1_MartaR

5'- GGTCTCGCTCTAACCTCGAGGAGAACTTCTAGTATATTTGGGG**GAGACC** - 3'

Tras la síntesis de estos dos meganucleótidos, se incuban para que autohíbriden, se digieren con *Bsal* junto con pDi-CRISPR, que también se digiere con *Bsal*, y se ligan entre sí para obtener el plásmido Di-Marta-CRISPR.

f. Transformación de *S. cerevisiae*.

Para conseguir la modificación se realiza la co-transformación de la cepa S288C con el plásmido conteniendo el fragmento de ADN que codifica el sgRNA, junto con el fragmento amplificado por PCR del casete de *gfp-delta*, mediante el método con acetato de litio descrito por Shi y colaboradores (Shi *et al.*, 2016) o mediante electroporación (Dicarlo *et al.*, 2013).

El plásmido Di-Marta-CRISPR expresará en el interior de las células transformadas el gen que codifica para la endonucleasa iCas9, el tracrRNA y el sgRNA diseñado para la inserción del casete *gfp-delta* en las secuencias *delta2* de S288C. El complejo dirige a Cas hacia el sitio *delta2* produciendo los nicks en esta región reconocida por el sgRNA. Mediante recombinación homóloga, la levadura repara estos nicks introduciendo el casete *gfp-delta* que se ha introducido junto con Di-Marta-CRISPR (Figura 16).

g. Comprobación de la inserción:

Con el fin de comprobar si la inserción se ha realizado con éxito, se observan las células transformadas bajo un microscopio de fluorescencia para analizar la emisión de fluorescencia verde frente a un control negativo de células sin transformar.

Para comprobar si la inserción se ha realizado en los sitios esperados se realizan amplificaciones mediante PCR utilizando cebadores internos (dentro de la región codificante de *gfp*) y externos (en la región *delta*).

Para comprobar la múltiple inserción de *gfp* se utiliza un espectrofotómetro de fluorescencia midiendo la intensidad de fluorescencia frente a un control negativo sin

transformar y/o un control de copia única (*S. cerevisiae* transformada con un vector centromérico que contenga el gen que codifica GFP).

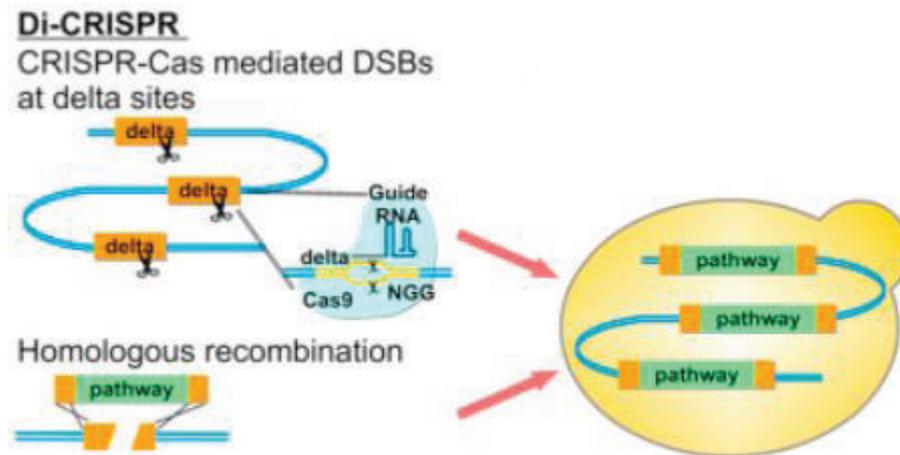


Figura 16. Representación gráfica de los pasos seguidos en la introducción del gen *gfp* en *Saccharomyces cerevisiae* mediante recombinación homóloga con las secuencias *delta*.

5. DISCUSIÓN

5.1. Conclusiones generales sobre el Sistema CRISPR-Cas9 y perspectivas futuras

Tras una extensa labor de investigación bibliográfica, hemos podido conocer y comprender en profundidad la estructura y actuación de los distintos tipos de sistemas CRISPR y más concretamente dentro de estos, del sistema CRISPR-Cas9. Y es que lejos de su función natural como sistema inmune en organismos procariontes, desde el año 2012 CRISPR-Cas9 se ha redefinido como una herramienta de edición genética sin parangón en el campo de la ingeniería genética y la biotecnología. La elevada eficacia y precisión de este sistema se debe a sus características estructurales. CRISPR-Cas9 está constituido por dos componentes que llevan a cabo distintas funciones: el casete CRISPR, de función de reconocimiento, que almacena secuencias nucleotídicas de elementos genómicos externos que invadieron a la célula en el pasado y la proteína Cas9, de función endonucleasa, que degrada cualquier secuencia reconocida por CRISPR mediante complementariedad de bases. Gracias a esto, CRISPR-Cas9 dota a la célula de inmunidad adaptativa ya que disminuye las probabilidades de que un elemento genético extraño que previamente haya infectado la célula lo vuelva a hacer. Las características estructurales de este sistema se comprobaron que eran idóneas para su empleo en experimentos de modificación genética, ya que bastaría con introducir en la célula la secuencia nucleotídica que querríamos modificar y que actuaría de guía de Cas9, para provocar un cambio en dicha secuencia en el ADN o ARN celular. Pero esto no se quedó aquí, sino que distintos ensayos de modificación de Cas9, permitieron dar lugar a variantes de esta endonucleasa que incrementarían la eficacia de este sistema. Surgen así la nCas9 y dCas9, que provocan el corte en una sola cadena o no efectúan corte alguno, respectivamente, en la secuencia diana. Esto dio lugar al aumento del espectro de funciones que CRISPR-Cas9 podría desenvolver, ya que ahora no sólo podría ser empleado como mecanismo de degradación, sino que fusionando a Cas9 distintos factores podríamos actuar sobre otras funciones como la transcripción o la localización de cualquier secuencia nucleotídica dentro del genoma de cualquier organismo. Por otra parte, se han descrito sistemas CRISPR más eficaces y de más fácil manejo en ingeniería como el sistema CRISPR-Cpf1. En resumen, los sistemas CRISPR-Cas se alzan como la

herramienta genómica más eficaz de entre los complejos moleculares editores conocidos, abriendo un amplio campo de investigación.

Por todas las razones y evidencias anteriormente comentadas, son muchas las esperanzas puestas en la efectividad de CRISPR-Cas9 como herramienta de edición genética. Sin embargo, son muchas las luces y sombras que se arrojan a este sistema desde el punto de vista ético, tanto en su empleo en células somáticas pero sobre todo en su aplicación en células germinales, debido a que se desconocen los efectos que podrían acarrear las modificaciones genéticas a largo plazo en la población. Y es que la vertiginosa velocidad a la que crece la investigación de los sistemas CRISPR no ha permitido una valoración clara acerca de las consecuencias de su uso (Benston, 2017). Un caso que ejemplifica el actual desconocimiento de las consecuencias de su empleo *in vivo*, es un proyecto de investigación publicado este mismo año que empleaba la tecnología CRISPR en ratones, revela la aparición de hasta 1.700 mutaciones nucleotídicas por ratón en zonas del genoma fuera de la región de homología del sgRNA y de las que se desconoce cómo han tenido lugar (Schaefer *et al.*, 2017). Debido a casos como este, existe una gran controversia respecto al empleo de esta herramienta editora de material genético.

5.2. La efectividad del diseño de sgRNAs

En la actualidad existen diversos programas, que permiten la búsqueda de secuencias PAM en regiones determinadas de un genoma, como el caso que fue utilizado en este ejemplo de diseño teórico: CRISPy-Web. Sin embargo, serían necesarias herramientas más potentes para poder realizar análisis de homología entre secuencias de pequeño tamaño (menores de 30 nucleótidos) frente a todo un genoma completo de un organismo para poder localizar regiones de alta homología que pudieran favorecer la formación de mutaciones *off-target*.

El sistema CRISPR-Cas9, es un sistema relativamente simple en cuanto al diseño, tal y como se ha demostrado en este proyecto teórico de edición genómica que ejemplifica de forma práctica la introducción de un sgRNA en una célula, que actúa guiando una nucleasa Cas9 hacia una secuencia homóloga en el ADN del genoma del organismo que queremos modificar. Para ello, se ha diseñado un ARN guía que contiene 21 nucleótidos de la secuencia *delta2*, allí donde queremos introducir el gen *gfp* en *S. cerevisiae*. Este diseño teórico ha sido planteado para que la recombinación se produzca en múltiples secuencias *delta* extendidas por todo el genoma de la levadura, y de alta homología entre ellas.

Sin embargo, para determinados casos experimentales, una edición muy precisa y puntual es absolutamente crucial, como puede ser en posibles casos de terapia génica con células humanas. En este sentido, actualmente están siendo desarrolladas nuevas tecnologías de alta sensibilidad para hacer un chequeo de genomas completos en búsqueda de *off-targets* (Park *et al.*, 2017; Tsai *et al.*, 2017) que posiblemente ayudarán a perfeccionar este sistema en el futuro en cuanto a su fiabilidad y no sólo en cuanto a facilidad de diseño, eficacia o bajo coste.

6. CONCLUSIONES

- ❖ Tras una exhaustiva búsqueda bibliográfica en la que se han abordado todos los aspectos relacionados con la biología de los sistemas CRISPR, se ha logrado dar una profunda visión de la estructura y funcionamiento de estos sistemas, no sólo en la naturaleza sino también abordando sus variantes y usos en los campos de la ingeniería genómica y la biotecnología.
- ❖ Para ejemplificar de forma práctica y consolidar los conocimientos obtenidos de forma teórica, se ha querido abordar una aplicación práctica del sistema CRISPR-Cas9 mediante el diseño de una secuencia oligonucleotídica que permitiera la expresión del gen *gfp* en *Saccharomyces cerevisiae*. Esto ha requerido del aprendizaje de la utilización de diversos programas informáticos, bases de datos y casas comerciales a partir de los cuales hemos obtenido las herramientas necesarias para llevar a cabo el proyecto teórico.
- ❖ Por tanto, el desarrollo de esta memoria de fin de grado además de permitirme ahondar en el conocimiento del mundo CRISPR me ha dotado de múltiples competencias transversales. Gracias a esto he tenido la oportunidad de conocer de primera mano la metodología a seguir en el diseño de un proyecto dentro del campo de la biología molecular y sacar adelante un proyecto de edición genómica empleando la herramienta biotecnológica más revolucionaria conocida.

6. CONCLUSIONS

- ❖ After an exhaustive bibliographical search that has addressed all the issues related to the biology of CRISPR systems, it has been possible to give a deep vision of the structure and operation of these systems, not only in nature but also addressing of genomic engineering and biotechnology.
- ❖ In order to practice and consolidate the theoretical knowledge, a practical application of the CRISPR-Cas9 system has been sought through the design of an oligonucleotide sequence that allows the expression of the *gfp* gene in *Saccharomyces cerevisiae*. This has required the learning of the use of various computer programs, databases and commercial houses from which we have obtained the necessary tools to carry out the theoretical project.
- ❖ Therefore, the development of this end-of-grade memory in addition to allowing me to deepen the knowledge of the CRISPR world has given me multiple transversal competences. Thanks to this I have had the methodology to follow in the design of a project within the field of molecular biology and to carry out a project of genomic editing using the most revolutionary biotechnological tool known.

6. CONCLUSIÓNS

- ❖ Tras unha exhaustiva búsqueda bibliográfica na que se abordaron todos os aspectos relacionados coa bioloxía dos sistemas CRISPR, lógrase dar una profunda visión da estrutura e funcionamento destes sistemas, non só na natureza senon tamén abordando a súas variantes e usos nos campos da enxeñaría xenómica e a biotecnoloxía.
- ❖ Para exemplificar de forma práctica e consolidar os coñecementos obtidos de forma teórica, quíxose abordar unha aplicación práctica do sistema CRISPR-Cas9 mediante o deseño dunha secuencia oligonucleotídica que permitira a expresión do xen *gfp* en *Saccharomyces cerevisiae*. Isto requeriu da aprendizaxe da utilización de diversos programas informáticos, bases de datos y casas comerciais a partir das cales obtivemos as ferramentas necesarias para levar a cabo o proxecto teórico.

- ❖ Por tanto, o desenvolvemento desta memoria de fin de grao ademais de permitirme afondar no coñecemento do mundo CRISPR dotoume de múltiples competencias transversais. Grazas a isto tiven a oportunidade de coñecer de primeira man a metodoloxía a seguir no deseño dun proxecto dentro do campo da bioloxía molecular e sacar adiante un proxecto de edición xenómica empregando a ferramenta biotecnolóxica máis revolucionaria coñecida.

7. BIBLIOGRAFÍA

Barrangou, R. *et al.* (2007) 'CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes', *Science*, 315(5819), pp. 1709–1712.

Benston, S. (2017) 'Everything in moderation, even hype: learning from vaccine controversies to strike a balance with CRISPR.', *Journal of medical ethics*, p. medethics-2016-103666.

Bolotin, A. *et al.* (2005) 'Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin', *Microbiology*, 151(8), pp. 2551–2561.

Brouns, S. J. J. *et al.* (2008) 'Small CRISPR RNAs guide antiviral defense in prokaryotes.', *Science (New York, N.Y.)*, 321(5891), pp. 960–4.

de la Fuente-Núñez, C. and Lu, T. K. (2017) 'CRISPR-Cas9 technology: applications in genome engineering, development of sequence-specific antimicrobials, and future prospects', *Integrative Biology*, 9(2), pp. 109–122.

Deltcheva, E. *et al.* (2011) 'CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.', *Nature*, 471(7340), pp. 602–7.

Dicarlo, J. E. *et al.* (2013) 'Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems', *Nucleic Acids Research*, 41(7), pp. 4336–4343.

Dong, D. *et al.* (2016) 'The crystal structure of Cpf1 in complex with CRISPR RNA', *Nature*, 532(7600), pp. 522–526.

Doudna, J. A. and Charpentier, E. (2014) 'The new frontier of genome engineering with CRISPR-Cas9', *Science*, 346(6213), pp. 1258096–1258096.

Fleischmann, R. *et al.* (1995) 'Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd', *Science*, 269(5223), pp. 496–512.

Gafner, J., Robertis, E. M. De and Philippsen, P. (1983) 'in the 5' Non-Coding Region of', 2(4), pp. 583–591.

Gaj, T. (2014) 'ZFN, TALEN and CRISPR/Cas based methods for genome engineering', 31(7), pp. 397–405.

Haft, D. H. *et al.* (2005) 'A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/cas subtypes exist in prokaryotic genomes', *PLoS Computational Biology*, 1(6), pp. 0474–0483.

Hsu, P. D., Lander, E. S. and Zhang, F. (2014) 'Development and applications of CRISPR-Cas9 for genome engineering', *Cell*, 157(6), pp. 1262–1278.

Ishino, Y. *et al.* (1987) 'Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene

- product.', *Journal of Bacteriology*, 169(12), pp. 5429–5433.
- Jansen, R. *et al.* (2002) 'Identification of genes that are associated with DNA repeats in prokaryotes', *Molecular Microbiology*, 43(6), pp. 1565–1575.
- Jiang, Y. *et al.* (2017) 'CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*.', *Nature communications*, 8, p. 15179.
- Jinek, M. *et al.* (2012a) 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.', *Science*, 337(6096), pp. 816–21.
- Jinek, M. *et al.* (2012b) 'A Programmable Dual-RNA – Guided', 337(August), pp. 816–822.
- Karvelis, T. *et al.* (2013) 'crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*', *RNA Biology*, 10(5), pp. 841–851.
- Kirchner, M. and Schneider, S. (2015) 'CRISPR-Cas: From the Bacterial Adaptive Immune System to a Versatile Tool for Genome Engineering', *Angewandte Chemie - International Edition*, 54(46), pp. 13508–13514.
- Leenay, R. T. *et al.* (2017) 'CRISPR-Cas systems', 62(1), pp. 137–147.
- Makarova, K. S. *et al.* (2011) 'Evolution and classification of the CRISPR-Cas systems', *Nature Reviews Microbiology*, 9(6), pp. 467–477.
- Mojica, F. J. *et al.* (1995) 'Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning.', *Molecular microbiology*, 17(1), pp. 85–93.
- Mojica, F. J., Juez, G. and Rodríguez-Valera, F. (1993) 'Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites.', *Molecular microbiology*, 9(3), pp. 613–21.
- Mojica, F. J. M. *et al.* (2000) 'Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria', *Molecular Microbiology*, 36(1), pp. 244–246.
- Mojica, F. J. M. *et al.* (2005) 'Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements', *Journal of Molecular Evolution*, 60(2), pp. 174–182.
- Mojica, F. J. M. and Garrett, R. A. (2013) 'Discovery and Seminal Developments in the CRISPR Field', in Barrangou, R. and van der Oost, J. (eds) *CRISPR-Cas Systems: RNA-mediated Adaptive Immunity in Bacteria and Archaea*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–31.
- Mojica, F. J. M. and Rodríguez-Valera, F. (2016) 'The discovery of CRISPR in archaea and bacteria', *FEBS Journal*, 283, pp. 3162–3169.
- Nakata, A., Amemura, M. and Makino, K. (1989) 'Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome', *Journal of Bacteriology*, 171(6), pp. 3553–3556.
- Nishimasu, H. *et al.* (2014) 'Crystal structure of Cas9 in complex with guide RNA and target DNA', *Cell*, 156(5), pp. 935–949.
- Ott De Bruin, L. M., Volpi, S. and Musunuru, K. (2015) 'Novel genome-editing tools to model and correct primary immunodeficiencies', *Frontiers in Immunology*, 6(MAY), pp. 1–11.
- Park, J. *et al.* (2017) 'Digenome-seq web tool for profiling CRISPR specificity.', *Nature methods*, 14(6), pp. 548–549.

- Pourcel, C., Salvignol, G. and Vergnaud, G. (2005) 'CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies', *Microbiology*, 151(3), pp. 653–663.
- Ran, F. A. *et al.* (2013) 'Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity', *Cell*, 154(6), pp. 1380–1389.
- Reynaud, E. (2010) 'Life with 6000 genes.', *Nature Education*, 3, p. 546,563-567.
- Sapranauskas, R. *et al.* (2011) 'The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*', *Nucleic Acids Research*, 39(21), pp. 9275–9282.
- Schaefer, K. A. *et al.* (2017) 'Unexpected mutations after CRISPR-Cas9 editing in vivo.', *Nature methods*, 14(6), pp. 547–548.
- Shi, S. *et al.* (2016) 'A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in *Saccharomyces cerevisiae*', *Metabolic Engineering*, 33, pp. 19–27.
- Shmakov, S. *et al.* (2015) 'Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems', *Molecular Cell*, 60(3), pp. 385–397.
- Sontheimer, E. J. and Wolfe, S. A. (2015) 'Cas9 gets a classmate.', *Nature biotechnology*, 33(12), pp. 1240–1241. doi: 10.1038/nbt.3426.
- Van Soolingen, D. *et al.* (1991) 'Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: Evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis', *Journal of Clinical Microbiology*, 29(11), pp. 2578–2586.
- Tang, T.-H. *et al.* (2002) 'Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*', *Proceedings of the National Academy of Sciences*, 99(11), pp. 7536–7541.
- Tsai, S. Q. *et al.* (2017) 'CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets.', *Nature methods*, 14(6), pp. 607–614.
- Tsien, R. Y. (1998) 'the Green Fluorescent', *Proteins*, 67(11), pp. 509–44.
- Wang, H., La Russa, M. and Qi, L. S. (2016) 'CRISPR/Cas9 in Genome Editing and Beyond.', *Annual review of biochemistry*, 85(1), pp. 227–64.
- Watkins-Chow, D. E. *et al.* (2017) 'Highly Efficient Cpf1-Mediated Gene Targeting in Mice Following High Concentration Pronuclear Injection.', *G3*, 7(2), pp. 719–722.
- Watson, J. D. and Crick, F. H. C. (1953) 'Molecular structure of nucleic acids', *Nature*, pp. 737–738.
- Xiong, X. *et al.* (2016) 'CRISPR/Cas9 for Human Genome Engineering and Disease Research.', *Annual review of genomics and human genetics*, 17(1), pp. 131–54.
- Zetsche, B. *et al.* (2015) 'Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System', *Cell*, 163(3), pp. 759–771.
- Zhu, P. *et al.* (2017) 'CRISPR/Cas9-Mediated Genome Editing Corrects Dystrophin Mutation in Skeletal Muscle Stem Cells in a Mouse Model of Muscle Dystrophy.', *Molecular therapy. Nucleic acids*, 7, pp. 31–41.