

**CENTRO DE INVESTIGACIÓN Y ESTUDIOS  
AVANZADOS DEL INSTITUTO POLITÉCNICO  
NACIONAL**

UNIDAD ZACATENCO

DEPARTAMENTO DE BIOTECNOLOGÍA Y BIOINGENIERÍA

Título: **“Análisis y estrategia de construcción del sistema CRISPR – Cas12k –  
Tn7 para su evaluación en *Pseudomonas chlororaphis*”**

Presenta

**Samuel Sotelo Gómez**

Para obtener el grado de

**MAESTRO EN CIENCIAS EN LA ESPECIALIDAD DE BIOTECNOLOGÍA**

Director de la Tesis: **Dr. Rodolfo Marsch Moreno**

Ciudad de México

Abril, 2021

Este proyecto se llevó a cabo en el Laboratorio de Ingeniería Genética y Biología Molecular de Suelos (Laboratorio 50) del Departamento de Biotecnología y Bioingeniería (DBB) en el Centro de Investigación y Estudios Avanzados (Cinvestav) del Instituto Politécnico Nacional (IPN), bajo la dirección del Dr. Rodolfo Marsch Moreno, Investigador Titular adscrito al DBB en Cinvestav – IPN. De forma personal, se agradece el apoyo de Conacyt a través de la beca 932096 otorgada al LBT. Samuel Sotelo Gómez para estudios de Maestría en Ciencias en la especialidad de Biotecnología y Bioingeniería.

## **Agradecimientos**

Agradezco a mi director de tesis, **Dr. Rodolfo Marsch Moreno**, por la oportunidad de recibir su tutoría, paciencia y apoyo para la realización de este proyecto. El conocimiento se adquiere y se pierde fácilmente, pero la capacidad para enfrentar retos de forma creativa es innata.

Valoro los consejos puntuales y agradezco todo el conocimiento brindado por la **Dra. María Eugenia Hidalgo Lara** y el **Dr. Jaime Ortega López**, a lo largo de los estudios de maestría.

Adicionalmente, agradezco a todo el personal del Laboratorio 50 por la guía y convivencia que fue parte del trabajo desarrollado. Particularmente, a la **Dra. Martha Barajas** y la **Dra. Yaneth Bartolo**.

## **Dedicatorias**

A mi madre, **María Victoria**, por su esfuerzo para enfrentarme a la vida, mostrándome que una lucha no se inicia tarde y que puedes encontrar fuerza en las cosas más simples. No abandones el valor nunca y no olvides lo que tus manos han moldeado.

A mi hermana, **Ana Victoria**, porque me llevó de la mano en la incertidumbre y supo brindarme tranquilidad. Apenas comienza tu camino, no tienes más que seguir y todo será mejor porque nos decidimos por la valentía.

A mi hermano, **Alejandro**, por la fortaleza que has mostrado. Espero nos sea permitido compartir más tiempo y cumplir todos los planes que hemos hablado.

A mi sobrino, **Samuel Rodrigo**, por brindarme momentos que no necesitan ser planeados, esperados o una realización perfecta. El mundo te dará lecciones, la vida lucirá difícil y las personas te dejarán huellas, pero tu fortaleza será mayor.

A mi cuñado, **Humberto**, porque un abrazo es efímero pero el significado cambia vidas.

# CONTENIDO

<b>AGRADECIMIENTOS .....</b>	<b>I</b>
<b>DEDICATORIAS.....</b>	<b>II</b>
<b>ÍNDICE DE TABLAS.....</b>	<b>1</b>
<b>ÍNDICE DE FIGURAS .....</b>	<b>2</b>
<b>LISTA DE ABREVIATURAS.....</b>	<b>3</b>
<b>RESUMEN .....</b>	<b>4</b>
<b>ABSTRACT .....</b>	<b>5</b>
<b>1. INTRODUCCIÓN.....</b>	<b>6</b>
1.1. <i>Sistemas CRISPR – Cas</i> .....	6
1.1.1. Organización modular.....	7
1.1.2. Sistemas CRISPR – Cas de Clase I .....	8
1.1.3. Sistemas CRISPR – Cas de clase II.....	10
1.1.3.1. CRISPR – Cas9 .....	10
1.1.4. Origen.....	11
1.1.4.1. Clase I.....	12
1.1.4.2. Clase II.....	14
1.1.4.3. Casposones.....	15
1.2. <i>Sistemas de transposición.....</i>	16
1.2.1. Transposasas de la familia de Tn7 .....	17
1.3. <i>Sistema CAST.....</i>	18
1.3.1. CRISPR – Cas12 – Tn7.....	18
1.4. <i>Pseudomonas chlororaphis.....</i>	20
<b>2. ANTECEDENTES.....</b>	<b>21</b>
<b>3. JUSTIFICACIÓN .....</b>	<b>23</b>
<b>4. HIPÓTESIS.....</b>	<b>24</b>
<b>5. OBJETIVOS.....</b>	<b>24</b>

5.1.	<i>General</i> .....	24
5.2.	<i>Específicos</i> .....	24
6.	ESTRATEGIA DE TRABAJO.....	25
7.	MATERIALES Y MÉTODO.....	28
7.1.	<i>Búsqueda y análisis bibliográfico del sistema CAST</i> .....	28
7.1.1.	Análisis de los componentes del locus CAST .....	28
7.1.2.	Papel del sistema CAST en la función celular .....	28
7.1.3.	Sistema INTEGRATE .....	29
7.1.4.	Comparación con el sistema INTEGRATE.....	29
7.1.5.	Otras asociaciones CRISPR – Cas – transposasa .....	29
7.2.	<i>Diseño in silico y análisis bioinformático</i> .....	29
7.2.1.	Búsqueda e identificación de las secuencias proteicas y elementos genéticos necesarios para las construcciones. ....	29
7.2.2.	Diseño de las construcciones.....	31
7.2.3.	Diseño de los oligonucleótidos para la amplificación de los fragmentos del plásmido pSS5001.....	31
7.2.4.	Diseño de los oligonucleótidos para la integración de los sgRNA.....	32
7.2.5.	Análisis de interacción entre los elementos del sistema CAST .....	32
7.2.6.	Descripción de las relaciones evolutivas y filogenéticas .....	33
7.3.	<i>Obtención del material genético</i> .....	33
7.3.1.	Plásmido con las proteínas del sistema CAST.....	33
7.3.2.	Plásmido con el transposón Tn7.....	34
7.3.3.	Plásmido Diana (Testigo +) .....	34
7.3.4.	Oligonucleótidos para amplificación de los elementos del plásmido pSS5001 .	34
7.4.	<i>Construcción del plásmido pSS5001</i> .....	35
7.4.1.	Amplificación de los elementos del plásmido pSS5001 .....	35
7.4.2.	Purificación y cuantificación de los amplicones .....	36
7.4.3.	Ensamble de los fragmentos.....	36
7.4.4.	Análisis del ensamble de los fragmentos.....	37

7.4.5.	Construcción de los plásmidos pSS5001 con los sgRNA .....	37
7.5.	<i>Ensayo funcional del sistema CAST</i> .....	37
7.5.1.	Transformación de <i>P. chlororaphis</i> .....	37
7.5.2.	Análisis de inserción sitio específico.....	38
8.	RESULTADOS .....	39
8.1.	<i>Búsqueda y análisis bibliográfico del sistema CAST</i> .....	39
8.1.1.	Análisis de los componentes del locus CAST .....	39
8.1.2.	Papel del sistema CAST en la función celular .....	41
8.1.3.	Sistema CRISPR – Cascade – Tn6677 (INTEGRATE).....	41
8.1.4.	Comparación con el sistema INTEGRATE.....	45
8.1.5.	Otras asociaciones CRISPR – Cas – transposasa .....	47
8.2.	<i>Diseño in silico y análisis bioinformático</i> .....	49
8.2.1.	Búsqueda e identificación de las secuencias proteicas y elementos genéticos necesarios para las construcciones .....	49
8.2.2.	Diseño de las construcciones.....	53
8.2.3.	Diseño de los oligonucleótidos para la amplificación de los fragmentos del plásmido pSS5001.....	56
8.2.4.	Diseño de los oligonucleótidos para la integración de los sgRNA.....	58
8.2.5.	Análisis de interacción entre los elementos del sistema CAST .....	59
8.2.6.	Descripción de las relaciones filogenéticas y evolutivas .....	62
8.3.	<i>Obtención del material genético</i> .....	66
8.3.1.	pHelper: Plásmido con los genes del sistema CAST.....	66
8.3.2.	pDonor: Plásmido con el transposón de Tn7.....	67
8.3.3.	pTarget: Plásmido Diana (Control +).....	67
8.3.4.	Oligonucleótidos para amplificación de los elementos del plásmido pSS5001 .	68
8.4.	<i>Construcción del plásmido pSS5001</i> .....	68
8.4.1.	Amplificación de los elementos del plásmido pSS5001 .....	68
8.4.2.	Purificación y cuantificación de los amplicones .....	69
8.5.	<i>Nota aclaratoria</i> .....	70

9.	DISCUSIÓN.....	71
10.	CONCLUSIÓN .....	75
11.	RECOMENDACIÓN.....	76
12.	BIBLIOGRAFÍA .....	77
13.	ANEXOS.....	86
13.1.	<i>Anexo 1</i> .....	86
13.2.	<i>Anexo 2</i> .....	87
13.3.	<i>Anexo 3</i> .....	88
13.4.	<i>Anexo 4</i> .....	91
13.5.	<i>Anexo 5</i> .....	93
13.6.	<i>Anexo 6</i> .....	94
13.7.	<i>Anexo 7</i> .....	95
13.8.	<i>Anexo 8</i> .....	96
13.9.	<i>Anexo 9</i> .....	97
13.10.	<i>Anexo 10</i> .....	98
13.11.	<i>Anexo 11</i> .....	100
13.12.	<i>Anexo 12</i> .....	102
13.13.	<i>Anexo 13</i> .....	103
13.14.	<i>Anexo 14</i> .....	104
13.15.	<i>Anexo 15</i> .....	105
13.16.	<i>Anexo 16</i> .....	106

## **Índice de tablas**

**Tabla 1.** Características de los sistemas CAST e INTEGRATE.

**Tabla 2.** Arreglo CRISPR del locus CAST.

**Tabla 3.** Posibles sitios blanco de los espaciadores contenidos en el arreglo CRISPR del locus CAST.

**Tabla 4.** Posibles sitios blanco en genomas víricos.

**Tabla 5.** Elementos para la integración del plásmido pSS5001.

**Tabla 6.** Estructura de los oligonucleótidos para la amplificación de los genes del sistema CAST.

**Tabla 7.** Oligonucleótidos para la amplificación de los 11 fragmentos del plásmido pSS5001.

**Tabla 8.** Secuencias guía para la integración de los sgRNA.

## Índice de figuras

**Figura 1.** Origen hipotético de los sistemas CRISPR – Cas de Clase I

**Figura 2.** Criterios de identificación de los extremos LE y RE del transposón

**Figura 3.** Estructura de los oligonucleótidos para ensamble por Golden Gate.

**Figura 4.** Programa utilizado para la amplificación por PCR.

**Figura 5.** Esquema propuesto para el sistema CAST.

**Figura 6.** Representación esquemática del sistema INTEGRATE

**Figura 7.** Modelo de la estructura de interacción de los elementos del sistema INTEGRATE.

**Figura 8.** Localización del locus CAST en el genoma completo de *S. hofmannii*.

**Figura 9.** Locus CAST.

**Figura 10.** Estructura de los extremos LE y RE del transposón de Tn7.

**Figura 11.** Diseño inicial del plásmido pSS5001.

**Figura 12.** Diseño final del plásmido pSS5001.

**Figura 13.** Secuencia del tracrRNA.

**Figura 14.** Interactoma de la proteína Cas12k.

**Figura 15.** Interactoma de TniQ.

**Figura 16.** Interactoma del sistema CAST.

**Figura 17.** Alineamiento de las secuencias aminoacídicas de Cas12k

**Figura 18.** Árbol filogenético de los locus identificados.

**Figura 19.** Gel de electroforesis de la digestión enzimática del plásmido pHelper.

**Figura 20.** Gel de electroforesis con el fragmento CMR.

## Lista de abreviaturas

**CRISPR – Cas:** Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Espaciadas Asociadas a la proteína Cas.

**MGE:** elemento genético móvil

**Acrs:** proteínas anti CRISPR

**Cascade:** complejo asociado al arreglo CRISPR para la defensa antivírica

**PAM:** sitio adyacente al protoespaciador

**ssDNA:** DNA de cadena sencilla

**TIR:** repetido invertido

**nt:** nucleótido

**kb:** kilobases

**pb:** pares de bases

**μl:** microlitros

**min:** minutos

**s:** segundos

**RNA:** ácidos ribonucleicos

**DNA:** ácidos desoxirribonucleicos

**mRNA:** RNA mensajero

**sgRNA:** RNA guía corto

**HPRT:** hipoxantina-guanina fosforibosiltransferasa

**ZFP:** proteína de dedos de zinc

**TALE:** efectores del tipo activador de la transcripción

**CAST:** transposasa asociada a sistemas CRISPR – Cas

**INTEGRATE:** Sistema CRISPR – Cascade – Tn6677

**nM:** nano molar

**PCR:** reacción en cadena de la polimerasa

**h:** hora

**MUCICAT:** Inserción cromosómica multicopia mediada por transposasas asociadas con CRISPR – Cas.

**LE:** extremo izquierdo del transposón

**RE:** extremo derecho del transposón

**NCBI:** National Center for Biotechnology Information

## Resumen

El sistema CAST promueve la inserción sitio específico de regiones genéticas, utilizando la maquinaria de transposición derivada de la familia de Tn7 y un complejo CRISPR – Cas12k. Este sistema proviene de *Scytonema hofmannii* y tiene una eficiencia de inserción del 60%, tanto en condiciones de ensamblaje in vitro, como a través de la transformación de una línea celular con el vector conteniendo los genes del complejo. En este trabajo se analizaron los componentes del sistema CAST de manera individual con el objetivo de proponer una hipótesis acerca de su papel en el entorno celular natural. Los resultados mostraron que el sistema podría estar implicado en la respuesta inmune adaptativa de la bacteria y que su principal objetivo serían las secuencias genéticas provenientes de virus. Por otra parte, es probable que el beneficio de la asociación entre la proteína Cas12k y la transposasa de Tn7, favorezca o le brinde una ventaja al transposón de propagarse o insertarse de forma más controlada. Además, se hizo la predicción de las interacciones entre los elementos del sistema CAST y la base de datos de las secuencias proteicas reportadas en la actualidad. El interactoma global nos permite deducir que este sistema requiere de distintas proteínas y factores para su correcto funcionamiento, siendo un factor para considerar en el perfeccionamiento de esta herramienta de edición genética. Toda esta información fue complementada con un análisis filogenético, que nos muestra que el sistema CAST de *S. hofmannii* es único en su tipo. Finalmente, se menciona que este proyecto tiene pendientes etapas por realizarse para culminar con el proceso experimental, las cuales son planeadas para realizarse en los estudios de doctorado.

## **Abstract**

The CAST system promotes site specific insertion of genetic regions, using the transposition machinery derived from the Tn7 family and a CRISPR – Cas12k complex. This system comes from *Scytonema hofmannii* and has an insertion efficiency of 60%, under *in vitro* assembly conditions and through the transformation of a cell line with the vector containing the genes of the complex. In this work, the components of the CAST system were analyzed individually with the aim to propose an hypothesis about their role in the natural cellular environment. The results showed that the system could be involved in the adaptive immune response of the bacteria and its main objective would be genetic sequences from viruses. On the other hand, the benefit of the association between the Cas12k protein and Tn7 transposase is likely to favor or give to the transposon an advantage of spreading or inserting in a more controlled fashion. In addition, the interactions between the elements of the CAST system and the database of the currently reported protein sequences were predicted. The global interactome allows us to deduce that this system requires different proteins and factors for its proper function, being it a factor to consider in the improvement of this genetic editing tool. All this information was complemented by a phylogenetic analysis, which shows that *S. hofmannii's* CAST system is one of a kind. Finally, it is mentioned that this project has pending steps to complete the experimental process, which are planned to be carried out in doctoral studies.

## 1. Introducción

### 1.1. Sistemas CRISPR – Cas

Uno de los descubrimientos más importantes de la última década reside en la revelación de que los organismos procariontes disponen de un sistema inmune adaptativo heredable. Nombrado CRISPR – Cas, tiene la capacidad de almacenar fragmentos de ácidos nucleicos de infecciones previas y, frente a una reincidencia del patógeno, desplegar nucleasas guiadas por RNA y llevar a cabo un silenciamiento sitio específico de fagos y otros elementos genéticos móviles (MGEs), como plásmidos y transposones (Hille *et al.*, 2018). Estos sistemas son encontrados del 50% y 90% en el genoma de bacterias y arqueas, respectivamente (Makarova *et al.*, 2015). El locus genómico del sistema CRISPR está compuesto por repeticiones idénticas alternadas que se encuentran entre espaciadores o secuencias genómicas diferenciales (Jansen *et al.*, 2002). Un estudio informático clave reveló que las secuencias espaciadoras coincidían con regiones de plásmidos y genomas víricos, siendo el primer indicio de que el arreglo CRISPR podía cumplir un rol dentro del sistema inmune adaptativo de los organismos procariontes (Pourcel *et al.*, 2005).

Adyacente al arreglo CRISPR hay una serie de genes que codifican para proteínas Cas implicadas en las tres fases de inmunidad adaptativa: adaptación, biogénesis del RNA basado en el arreglo CRISPR (crRNA) e interferencia. Durante la adaptación, los ácidos nucleicos foráneos son seleccionados, procesados e integrados dentro del arreglo CRISPR para proveer una memoria de la infección. Posteriormente, el arreglo CRISPR es transcrito para producir un precursor completo del crRNA (pre – crRNA) que es procesado en el sitio que contiene las

secuencias repetidas y se lleva a cabo la formación del crRNA maduro. Posteriormente a la infección, la maquinaria es guiada por los crRNA para reconocer una secuencia complementaria en los ácidos nucleicos foráneos, donde se lleva a cabo un corte enzimático (Hille *et al.*, 2018). De acuerdo con la variedad de genes *cas* y la naturaleza del complejo de interferencia, los sistemas CRISPR – Cas han sido divididos en dos clases, que a su vez están subdivididos en seis tipos. Los sistemas de Clase I (tipos I, III y IV) emplean complejos multiproteicos para la interferencia, mientras que en los sistemas de Clase 2 (tipos II, V y VI), la interferencia sólo involucra una proteína efectora (Macarova *et al.*, 2015).

#### **1.1.1. Organización modular**

La clasificación general de los sistemas CRISPR – Cas está determinada únicamente por la naturaleza efectora del complejo, es decir, si requieren una sola proteína o un conjunto de ellas. Sin embargo, los sistemas que emplean una única proteína efectora de forma natural dependen de la actividad de otras proteínas (Cas1, Cas2, Cas4, y RNasa III). Con base en su función, es posible agrupar las proteínas de los diferentes sistemas CRISPR – Cas en módulos: adaptación, expresión, interferencia, transducción de señales y auxiliares. Los módulos de adaptación e interferencia muestran una pronunciada autonomía no sólo con respecto a función y estructura, sino también a su evolución (Koonin *et al.*, 2019). En la topología del árbol filogenético de Cas1, la subunidad clave de todos los complejos de adaptación y estructuras repetidas es compatible pobremente con los módulos efectores y las filogenias individuales de cada una de sus proteínas, presumiblemente debido a que existe un intercambio frecuente entre los sistemas CRISPR – Cas de diferentes tipos y subtipos (Komor *et al.*, 2017). La presión evolutiva para el intercambio de módulos efectores parece ser la consecuencia de una carrera

armamentista contra virus. Estos codifican diversas proteínas antiCRISPR (Acrs) que tienen preferencia sólo por componentes de los complejos efectores y tienen una alta especificidad por variantes particulares de sistemas CRISPR – Cas (Bondy – Denomy *et al.*, 2013). La adaptación parcialmente independiente de los módulos efectores y de adaptación es corroborada por la presencia de estos módulos de forma solitaria en diversos genomas de bacterias y arqueas (Hudaiberdiev *et al.*, 2017).

### **1.1.2. Sistemas CRISPR – Cas de Clase I**

Entre los sistemas CRISPR – Cas, los sistemas de tipo I son los de mayor distribución y emplean un complejo multiproteico asociado al crRNA denominado complejo Cascade, que permite el reconocimiento del objetivo y mediante la actividad nucleasa de la proteína Cas3 lleva a cabo el proceso de interferencia (Brouns *et al.*, 2008). Dentro de los siete subtipos que han sido identificados (I – A hasta I – F y I – U), el sistema I – E de *E. coli* es el mejor descrito y caracterizado. El complejo Cascade del sistema CRISPR – Cas tipo I – E tiene un peso molecular de 405 kDa y está compuesto por las proteínas: Cas5, Cas6, Cas7, Cas8 y Cas11 (Jore *et al.*, 2011). En la mayoría de los sistemas de tipo I, el pre – crRNA es procesado por una RNasa de la familia de Cas6. En *E. coli*, el procesamiento del pre – crRNA permite la obtención de un crRNA maduro con una longitud de 61 nucleótidos que comprende la secuencia completa del espaciador, flanqueado en ambos sitios por fragmentos de la región repetida. El crRNA es una parte integral del complejo Cascade y está asociado a la proteína Cas5 en el extremo 5' (Zhao *et al.*, 2014). El esqueleto del complejo está compuesto por seis proteínas Cas7 conectadas, que adoptan una estructura con forma de mano y dominios similares a pulgares, que permiten

la conexión estrecha de las subunidades. Cas11 y Cas8 son definidas como las subunidades pequeña y grande del complejo Cascade, respectivamente (Mulepati *et al.*, 2014).

El reconocimiento del PAM (sitio adyacente al protoespaciador) en el DNA objetivo de doble cadena, es mediada por la subunidad grande, que también favorece el inicio del relajamiento en la doble cadena de DNA y la unión subsecuente del crRNA al DNA complementario del protoespaciador (Hayes *et al.*, 2016). Para la unión del complejo Cascade resultan cruciales los primeros ocho nucleótidos proximales al PAM del crRNA (nombrados secuencia semilla), con la excepción del sexto nucleótido que no se une al objetivo (Xiao *et al.*, 2017). La cadena que no se une al crRNA se mantiene asociada al complejo por dos subunidades de Cas11, conduciendo a la formación y estabilización de la estructura denominada “loop – R”, que es acompañada por cambios conformacionales de las subunidades grande y pequeña, permitiendo el reclutamiento de la nucleasa Cas3 (Hayes *et al.*, 2016). El dominio HD de Cas3 corta la cadena desplazada de DNA, induciendo cambios estructurales en la proteína que activa su actividad helicasa dependiente de ATP. Como resultado, Cas3 transloca y degrada sucesivamente la cadena de DNA en dirección 3' a 5', dejando una región de DNA de cadena sencilla de aproximadamente 200 – 300 nucleótidos en el genoma objetivo (Redding *et al.*, 2015). Sin embargo, la formación de un producto intermedio de degradación como la cadena parcial de ssDNA puede no conducir a la destrucción completa del invasor. Se piensa que la degradación completa del objetivo es mediada por otras nucleasas del huésped o por la actividad independiente de la nucleasa Cas3 sin encontrarse en asociación con el complejo Cascade, que ha sido observada *in vitro* (Wright *et al.*, 2016).

### **1.1.3. Sistemas CRISPR – Cas de clase II**

El sistema mejor caracterizado dentro de esta clasificación emplea la proteína Cas9 cuya naturaleza dual le permite tener una actividad endonucleasa guiada por RNA (Gasiunas *et al.*, 2012). Los tipos II – A, II – B y II – C son diferenciados por el tamaño del gen *cas9* y la presencia de un subtipo específico de genes (Fonfara *et al.*, 2014). Adicionalmente al crRNA, Cas9 requiere una región complementaria (tracrRNA), que es un RNA pequeño que permite la complementariedad a las regiones repetidas del crRNA (Deltcheva *et al.*, 2011). Una vez unidos (tracrRNA:crRNA) o sintéticamente producidos (sgRNA), la proteína Cas9 identifica el objetivo en el DNA a través del PAM y se produce el apareamiento entre bases complementarias del RNA guía y el DNA objetivo. Si el objetivo muestra suficiente complementariedad, la proteína Cas9 genera un corte a la cadena doble a una distancia de 3 nucleótidos del sitio que contiene el PAM, llevando a cabo el proceso de interferencia (Jinek *et al.*, 2012).

#### **1.1.3.1. CRISPR – Cas9**

El sistema CRISPR – Cas9 está clasificado dentro de los sistemas de clase 2 de tipo II. La proteína Cas9 tiene una estructura bilobulada con una hendidura central que permite el acomodo del dúplex de crRNA:DNA. El lóbulo de reconocimiento  $\alpha$ -hélice (REC) y el lóbulo nucleasa (NUC) están unidos por un enlace desordenado y por un puente altamente conservado rico en arginina que permite un contacto múltiple con el sgRNA. El lóbulo NUC contiene dominios con actividad nucleasa conservados de HNH (His-Asn-His) y RuvC, y un extremo C-terminal variable que interactúa con el PAM (Yamada *et al.*, 2017). El análisis estructural de la proteína Cas9 en su estado activo e inactivo, ha contribuido al entendimiento del mecanismo de interferencia

en sistemas CRISPR de tipo II, confirmando que la interacción de la proteína Cas9 y el sgRNA regula su actividad a través de cambios conformacionales en la proteína. Esto resulta en la disponibilidad de los residuos aminoacídicos para interactuar con los PAM y la secuencia semilla, permitiendo el reconocimiento del sitio objetivo de corte (Jiang *et al.*, 2015).

El apareamiento entre el RNA guía y el DNA objetivo acompañado de cambios conformacionales adicionales en Cas9 promueve la formación de la estructura “loop – R”. La complementariedad distal con el PAM y cationes divalentes son necesarios para la activación conformacional del dominio HNH en el estado competente de unión (Dagdas *et al.*, 2017). La activación conformacional del dominio HNH es acompañado del reacomodo de los péptidos de enlace entre los dominios HNH y RuvC. Esta comunicación alostérica entre los dominios con actividad nucleasa, resulta en la unión estabilizada de la proteína Cas9 sobre el DNA objetivo y permite llevar a cabo el proceso de interferencia (Sternberg *et al.*, 2015).

#### **1.1.4. Origen**

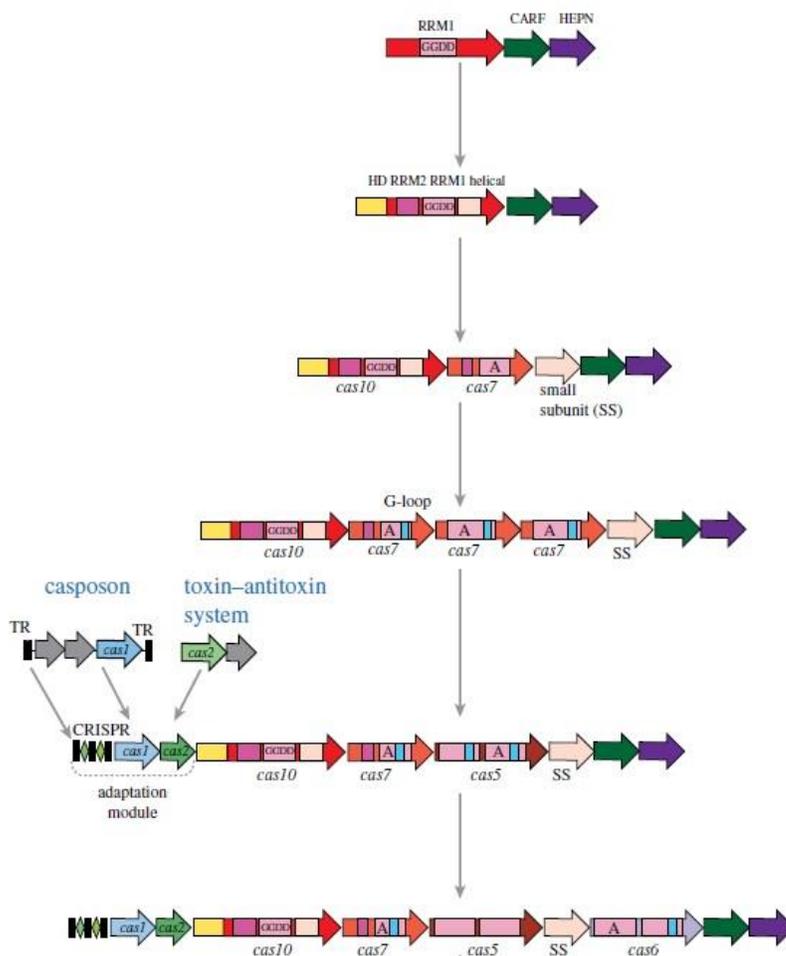
Se han encontrado en diversos estudios genes de proteínas Cas en regiones flanqueadas por TIRs (repetidos invertidos de transposones), lo cual sugiere que están asociados a elementos transponibles (Krupovic *et al.*, 2017). Los mecanismos bioquímicos de las reacciones catalizadas por Cas1 durante la integración de los espaciadores en los arreglos CRISPR y por las integrasas durante la conformación del transposón son muy similares, lo que supone que Cas1 es una integrasa de un nuevo grupo de transposones, que fueron nombrados casposones y las integrasas predichas se conocen como casposasas (Kapitonov & Jurka, 2006). El módulo completo de adaptación en los sistemas CRISPR parece tener su origen en un casposón que

podría también contribuir con genes adicionales de proteínas Cas (Koonin & Krupovic, 2015). La configuración del casposón ancestral incluye un gen homólogo de Cas2. Adicionalmente, las secuencias repetidas y la secuencia líder podrían haber evolucionado a partir de los TIRs o de un sitio duplicado dentro de la secuencia objetivo del casposón (Krupovic *et al.*, 2017). El evento que pudo haber generado el módulo de adaptación y, por lo tanto, la inmunidad adaptativa de procariontes podría involucrar la inserción de un casposón dentro de la vecindad de un locus de inmunidad innata ancestral, seguido por la inmovilización del casposón y la eliminación de algunos de sus genes, incluyendo la DNA polimerasa (Koonin & Krupovic, 2015). El sistema de inmunidad innato que dio lugar al módulo efector del sistema CRISPR podría haber funcionado directamente ensamblando RNA guías derivados de transcritos a partir de genomas foráneos, en analogía a los sistemas de defensa Argonauta en procariontes (Hur *et al.*, 2014).

#### **1.1.4.1. Clase I**

El origen hipotético (Figura 1) sugiere que los sistemas de Clase I se originaron a partir de tres eventos principales: transducción de señales, acoplamiento con un sistema de inmunidad innata ancestral y la formación del sistema con la asociación de proteínas con funciones diversas. Durante el primer evento se llevó a cabo la adquisición de un dominio HD (His-Asp) nucleasa y el origen del DNA objetivo, la duplicación de los dominios RRM (motivo de reconocimiento de RNA) y la formación del sitio de unión del RNA. El evento posterior generó la duplicación de genes de la familia de *cas10* seguida por su fisión, el origen de la familia de proteínas Cas7 y la subunidad pequeña, y la estabilización de los sitios de reconocimiento de RNA. Además, mediante diversas duplicaciones en genes y el origen de enlaces ricos en glicina,

se permitió el origen de la familia de proteínas Cas5. Por otra parte, se integró un casposón al lado del locus de inmunidad innata y a partir de los TIRs se generaron las secuencias repetidas del arreglo CRISPR se originó la proteína Cas1, implicada en el procesamiento de los crRNA y la adquisición de proteínas de la familia Cas2 provenientes del sistema toxina – antitoxina. Finalmente, mediante una serie de duplicaciones de genes *cas5* se generó la familia de proteínas Cas6 y se formó completo el sistema dejando las proteínas Cas5 sin actividad catalítica (Koonin *et al.*, 2019).



**Figura 1. Origen propuesto de los sistemas CRISPR – Cas de Clase I.** Evolución y adquisición de los elementos que participan activamente en el complejo Cascade. Tomado de Koonin *et al.* (2019).

#### 1.1.4.2. Clase II

El análisis de las secuencias proteicas de los efectores en sistemas de clase II muestra que todos ellos poseen homología con nucleasas codificadas por diferentes clases de MGEs (Shmakov *et al.*, 2017). Todos los efectores de tipo II y V (Cas9 y Cas12, respectivamente) comparten un dominio que pertenece a la familia de endonucleasas RuvC. Sin embargo, es baja la similitud en la secuencia entre los dominios RuvC de las proteínas Cas9 y Cas12, y dentro de los demás subtipos, por lo que estas proteínas pueden ser reconocidas como homólogas sólo por su perfil altamente sensible a secuencias génicas o a través de comparaciones estructurales (Shmakov *et al.*, 2015). Un hecho importante en el origen de Cas9 y Cas12 es la observación de que tienen la más alta similitud con proteínas TnpB que contienen dominios RuvC, que pertenecen a la familia de transposones IS605. El escenario evolutivo para los sistemas de clase II comienza con la inserción al azar de transposones codificando proteínas TnpB al lado de los arreglos CRISPR. La evolución subsecuente puede involucrar paralelamente, una maduración independiente de los efectores a través de la adquisición de dominios adicionales provenientes de fuentes externas que pueden implicar duplicaciones y recombinación (Shmakov *et al.*, 2017).

Adicionalmente a las dos distintas clases de MGEs (elementos genéticos móviles), casposones y transposones codificando proteínas tnpB, los módulos microbianos toxina – antitoxina (TA) parecen ser importantes contribuyentes a la evolución de sistemas CRISPR – Cas de clase II. El tipo más común es el módulo II TA que consiste en dos proteínas, una toxina y una antitoxina, de las cuales la antitoxina es menos estable y es eliminada por proteólisis bajo estrés, resultando en la activación de la toxina (Gerdes, 2012). Este tipo de toxinas contienen dominios

con actividad interferasa, es decir, RNAsas que indiscriminadamente rompen mRNAs dentro del ribosoma, resultando en dormancia o muerte celular (Yamaguchi & Inouye, 2009). Estas interferasas pertenecen a diversas familias de proteínas no relacionadas entre sí, que incluyen HEPN (en inglés: higher eukaryotes and prokaryotes nucleotide-binding domain), RelE y VapD, que es una variante distinta de la estructura RRM (Makarova *et al.*, 2009). Dado que algunas proteínas parecen derivar de casposones como las proteínas homólogas de Cas4, esta familia de proteínas que es parte de los módulos de adaptación en diversos subtipos de sistemas CRISPR – Cas, podrían tener su origen a partir de un casposón ancestral (Koonin *et al.*, 2019).

#### **1.1.4.3. Casposones**

Estos transposones comparten dos genes universales codificando para Cas1 y una DNA polimerasa de la familia B, además de diversos sets de genes de nucleasas y proteínas de unión a DNA que contienen dominios hélice – giro – hélice (HTH). La evidencia sugiere que los casposones son transposones autosintetizados, que dirigen su propia replicación durante la transposición a través de un mecanismo de copiar y pegar. Este tipo de transposón no ha sido encontrado directamente en procariontes, pero es común en eucariontes; como el caso de los polintones, que también codifican para DNA polimerasas de la familia B con una integrasa de tipo retrovirus que no está relacionada con Cas1 (Krupovic & Koonin, 2015).

Sin embargo, los casposones no son la clase de MGEs más abundantes en procariontes y el análisis comparativo de su organización génica revela la existencia de al menos cuatro familias distintas de casposones presentes en arqueas y algunas bacterias (Krupovic *et al.*, 2017). Notablemente, un grupo pequeño de casposones codifican para una proteína de cápside vírica,

indicando que algunos de estos elementos son realmente “caspovirus”, en una cercana analogía a los polintones que también poseen proteínas de cápside y la predicción dice que pueden participar en la formación de viriones (Yutin *et al.*, 2018). La transposición de los casposones no ha sido demostrada de forma directa, sin embargo, el análisis genómico comparativo de diversas cepas de la arquea *Methanosarcina mazei* ha conducido a la identificación de signos claros de movilidad reciente, indicando que al menos algunos de los casposones son transposones activos (Krupovic & Koonin, 2016).

## **1.2. Sistemas de transposición**

La transposición es un proceso de recombinación en la que secuencias de DNA llamadas transposones, se mueven de un sitio original en la molécula de DNA a un nuevo sitio. Adicionalmente, los transposones pueden causar, o estar asociados con otros tipos de reacomodos genéticos como deleciones, inversiones y fusiones de cromosomas. Estos elementos están presentes en organismos eucariontes y procariontes, estando implicados en la expresión de genes de resistencia y de factores de transcripción. Existen diversas vías de transposición, que favorecen la transmisión eficiente de los genes contenidos en el transposón: replicativa, conservativa y mixta. Los mecanismos de transposición conservativos y replicativos comparten muchas características básicas. Los transposones codifican dos funciones críticas requeridas en el proceso, los extremos y las proteínas de transposición llamadas transposasas. Estas proteínas son críticas para el sistema debido a que se unen en los extremos del transposón favoreciendo su oligomerización y permitiendo su conformación con potencial de llevar a cabo el proceso de transposición. Las transposasas de las diversas familias contienen dominios que llevan a cabo funciones específicas, como el ensamble del transposón,

estabilidad al complejo, reconocimiento del sitio blanco y el de transposición de acuerdo con la vía específica (replicativa, conservativa o mixta) (Reznikoff, 1993).

### **1.2.1. Transposasas de la familia de Tn7**

La familia de Tn7 se distingue por los niveles de control para decidir cuándo y dónde se ejerce la transposición y su capacidad para utilizar diversos sitios objetivo. Tn7 tiene la capacidad de insertarse utilizando dos vías de transposición que aseguran la supervivencia del huésped y maximizan su transferencia vertical y horizontal (Peters, 2014). En una vía de transposición se dirige a un sitio único neutral encontrado en bacterias nombrado “sitio adjunto” (attTn7), lo cual favorece la transmisión vertical del elemento a células hijas (Gringauz *et al.*, 1988). En la segunda vía la transposición tiene una preferencia por elementos móviles como los plásmidos conjugativos, lo que facilita su transferencia horizontal (Wolkow *et al.*, 1996). La transposición de Tn7 está mediada por cinco proteínas codificadas dentro del transposón; TnsA, TnsB, TnsC, TnsD y TnsE (TnsABCDE), así como, secuencias en los extremos terminales que actúan en *cis*. TnsA y TnsB constituyen el núcleo de la transposasa, TnsD y TnsE son factores que median la especificidad y TnsC es una proteína reguladora que comunica la transposasa y las proteínas que determinan el sitio de transposición (Peters, 2014). Tn7 se corta completamente de un DNA objetivo a través de un corte en la doble cadena de DNA sobre los extremos del elemento. En el mecanismo de cortar y pegar de Tn7, la proteína TnsA es responsable de reconocer los extremos 5' con una distancia de 3 pb fuera de elemento (Sarnovsky *et al.*, 1996). TnsB es responsable del corte en el extremo 3', se unirá directamente al DNA objetivo. Los extremos de Tn7 son similares, el extremo izquierdo tiene una longitud aproximada de 150 pares de bases y contiene tres sitios de unión no contiguos para TnsB de 22 pb, mientras que el extremo

derecho tiene una longitud de 90 pb en longitud con cuatro sitios yuxtapuestos (Holder & Craig, 2010).

### **1.3. Sistema CAST**

Los análisis genómicos comparativos han mostrado que diversas clases de elementos genéticos móviles están relacionados con sistemas CRISPR – Cas, ya sea de forma evolutiva o funcional. Estos complejos tienen como características los componentes proteicos requeridos para el procesamiento del pre – crRNA y el reconocimiento de una secuencia blanco, sin embargo, a diferencia de los sistemas de clase II, estos carecen de actividad nucleasa. Se ha propuesto que estos complejos conforman una transposición más específica vía los sistemas CRISPR – Cas. El estudio de los componentes de los sistemas CRISPR – Cas codificados por MGEs, comprueba la plasticidad funcional de estos sistemas y de las rutas de coevolución de los MGEs con los mecanismos de defensa del huésped (Faure *et al.*, 2019).

#### **1.3.1. CRISPR – Cas12 – Tn7**

Los transposones de la familia de Tn7 asociados a sistemas CRISPR – Cas, contienen genes *tnsA*, *tnsB*, *tnsC* y *tniQ* de forma similar al complejo TnsABC. Además, emplean una proteína Cas que carece de actividad nucleasa debido a que tiene inactivado de forma constitutiva el dominio RuvC. La asociación entre los transposones de la familia de Tn7 y los sistemas CRISPR – Cas, sugiere que los transposones han aprovechado este hecho para generar “loops – R” en los sitios específicos y facilitar la propagación del transposón a través de plásmidos y fagos (Strecker *et al.*, 2019). A la fecha sólo han sido probados dos sistemas CRISPR – Cas12 – Tn7 que han resultados exitosos en la capacidad de insertar secuencias génicas sitio específico, los

cuales provienen de las cianobacterias *Scytonema hofmanni* (UTEX B 2349) y *Anabaena cylindrica* (PCC 7122). El locus CAST en ambas líneas celulares tiene una longitud de aproximadamente 20 kb, conteniendo los genes de la transposasa Tn7 en uno de los extremos del transposón y el arreglo CRISPR con el gen de la proteína Cas12k en el otro extremo. Dentro del loci CAST se han identificado diversos tracrRNA y particularmente, en *S. hofmannii*, se han detectado crRNAs de entre 28 y 34 nucleótidos, consistiendo en 11 a 14 nucleótidos de los repetidos directos y entre 17 a 20 nucleótidos del espaciador. Se ha analizado el sistema por si puede operar solo o requiere la participación de otros factores en el huésped, y se encontró que promueve una inserción eficiente cuando es expresado de forma intracelular y también cuando se ensambla *in vitro* y posteriormente es introducido en la célula (Strecker *et al.*, 2019).

Este sistema tiene la capacidad de insertar secuencias de forma unidireccional, teniendo una preferencia por PAMs con estructura NGTN y realizando la inserción a una distancia de entre 60 y 66 nucleótidos del sitio que contiene el PAM. Se ha propuesto el papel hipotético de cada subunidad de la transposasa: TnsB participa en la conformación del transposón a través de los extremos LE y RE, TniQ tiene la capacidad de reconocer secuencias de DNA relativamente específicas participando en la asociación con el DNA blanco y TnsC brinda estabilidad en conformación de la transposasa y la proteína Cas12. Sin embargo, la eficiencia del sistema ronda dentro del 60 % y 80 %, hay diferencias debido a que por sí sola la maquinaria de la transposasa puede llevar a cabo una inserción independiente de Cas12 (Strecker *et al.*, 2019).

#### 1.4. *Pseudomonas chlororaphis*

Son bacterias Gram negativas aeróbicas con forma redonda que poseen propiedades entomopatógenicas y que han sido propuestas como fuente de genes con propiedades insecticidas para cultivos genéticamente modificados (Kupferschmied *et al.*, 2013). *P. chlororaphis* es ubicua en el ambiente, con una distribución amplia en agua y suelo, participando en diversas funciones ecológicas y de control biológico. Se han encontrado en la rizósfera, asociadas en consorcios microbianos en raíces y colaborando directamente en procesos benéficos para la planta, compitiendo con diversos patógenos y protegiendo de plagas fúngicas (Anderson y Kim, 2018). *P. chlororaphis* ha sido reportada promoviendo el crecimiento de plantas, estimulando las comunidades microbianas y protegiendo a la planta a través de la producción de compuestos (antibióticos de tipo fenazina, quitinasas y proteasas) que inhiben el crecimiento fúngico, de insectos y nemátodos (Mendes *et al.*, 2011). Además, ha sido utilizada en una gran variedad de aplicaciones agrícolas, incluyendo el control biológico de fitopatógenos, promoción del crecimiento vegetal, la solubilización de fosfatos y biorremediación de compuestos orgánicos (Moore *et al.*, 2006). La aplicación amplia de este microorganismo se basa también en su denominación como organismo GRAS, debido a que su crecimiento se inhibe a 37°C (Velivelli *et al.*, 2014).

## 2. Antecedentes

Se ha reportado ampliamente la asociación de sistemas CRISPR-Cas y otras proteínas efectoras como factores de transcripción, desaminasas, proteínas fluorescentes, etc., (Richter *et al.*, 2017). Es de particular interés el trabajo realizado por Luo *et al.* (2017), que hicieron la fusión de la proteína dCas9 y un sistema de transposición. En ese trabajo se construyó una proteína quimérica dCas9-PB (piggyBac) que potencialmente emplea la transposición a un locus específico mediado por un RNA guía. Además, se comparó con otros sistemas quiméricos contruidos con diversas nucleasas; ZFP-PB y TALE-PB, cuyo objetivo fue el gen que codifica la enzima HPRT (hipoxantina-guanina fosforibosiltransferasa). Encontrando que el sistema dCas9-PB transponía aún a pesar de que no fuera agregado el RNA guía de la proteína Cas9. Sin embargo, al comparar las frecuencias de transposición con los otros sistemas se encontró que la eficiencia de dCas9-PB fue muy baja con respecto a TALE-PB y ZFP-PB. Posteriormente, se analizó una proteína Cas9 catalíticamente activa y se encontró que la eficiencia de transposición mejoraba ligeramente, pero no de forma significativa. Mientras que la transposición sitio específico (HPRT) se encontró sólo en una eficiencia muy baja con ambos sistemas (dCas9-PB y Cas9-PB).

Recientemente, Peters *et al.* (2017), reportaron mediante un análisis *in silico* la existencia de diversos locus CAST que codificaban las proteínas requeridas para la transposición (transposasas) y sistemas CRISPR-Cas. Este análisis permitió el planteamiento de diversas estrategias para el uso del sistema CAST como una herramienta para la edición genética. Por su parte, Strecker *et al.* (2019) evaluaron dos sistemas CAST (*S. hoffmani* y *A. cilíndrica*) para

comprobar si participaban activamente en la inserción de genes mediada por sgRNA, encontrando que la eficiencia de este sistema era de aproximadamente el 60%. En otro análisis, Klompe *et al.* (2019), reportaron una aplicación similar de los sistemas CAST, pero basándose en un complejo de proteínas Cas (Cascade), encontrando una eficiencia de inserción similar. La diferencia principal entre ambos sistemas es que en el primero la inserción está mediada por una proteína Cas12k, mientras que en el segundo se utiliza un complejo multiproteico. Es importante señalar que la inserción del transposón se lleva a cabo a una distancia aproximada de 60 pares de bases corriente abajo del PAM, por lo que es fundamental considerar este factor dentro del diseño de los sgRNA. Por otra parte, Halpin – Healy *et al.* (2019), determinaron la naturaleza de interacción del sistema CAST que utiliza un complejo multiproteico (CASCADE), encontrando que la interacción de la transposasa y la proteína Cas se efectúa a través de la subunidad TniQ de la transposasa. Además, la formación del “loop – R” causado por el proceso de reconocimiento de una secuencia diana por la proteína Cas, permite que el transposón se inserte de forma más eficiente.

### **3. Justificación**

El sistema CRISPR-Cas ha sido utilizado como una alternativa moderna para la manipulación y edición genética de organismos. Las aplicaciones de esta herramienta tienen una naturaleza diversa incluyendo la generación de proteínas quiméricas que permitan modificaciones puntuales del genoma, favorecer o inhibir la expresión de proteínas y el establecimiento de plataformas de terapia génica. Sin embargo, aún se han tenido limitaciones directas sobre la especificidad de este sistema y restricciones de la eficiencia que brinda la reparación dependiente y no dependiente de homología. Con el descubrimiento de la asociación natural de sistemas CRISPR-Cas y transposones, se ha abierto la posibilidad de insertar genes o regiones génicas en un locus específico con una mayor eficiencia y con un mayor tamaño en la región a insertar. Por lo tanto, se plantea generar una estrategia para el establecimiento y aplicación de la herramienta de edición genética basada en el sistema *CAST* en *P. chlororaphis* y evaluar la capacidad de inserción génica en un locus específico.

#### 4. Hipótesis

La aplicación del sistema CAST con capacidad de asociarse *in vitro* e *in vivo*, permitirá la inserción génica en un locus específico del genoma de *P. chlororaphis*.

#### 5. Objetivos

##### 5.1. General

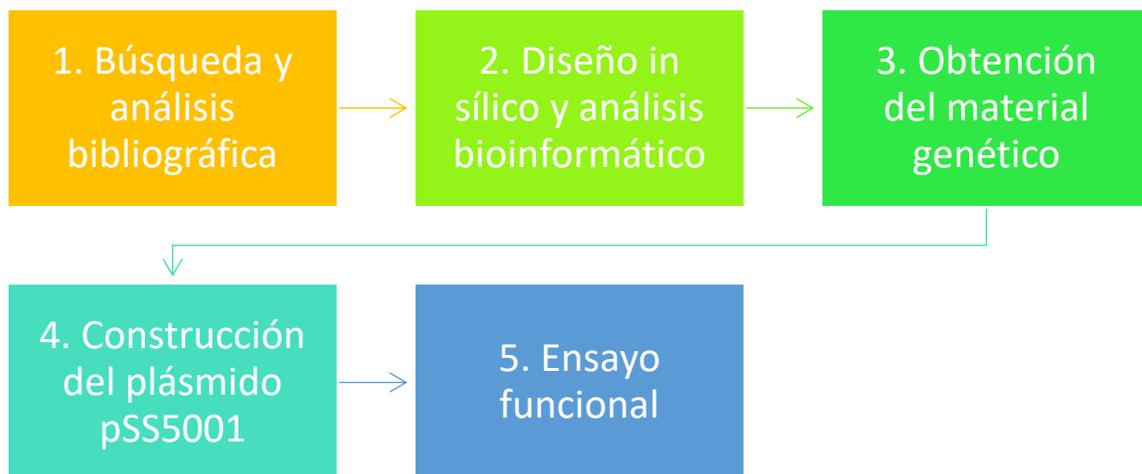
Establecer una estrategia de construcción y análisis del sistema quimérico CAST para la inserción de genes sitio específico en *Pseudomonas chlororaphis*.

##### 5.2. Específicos

- Análisis individual de los elementos del sistema CAST.
- Obtención del material genético:
  - Plásmido con los genes *tnsB*, *tnsC*, *tniQ* y *cas12k*.
  - Plásmido con los extremos de Tn7 y el marcador de selección.
  - Plásmido Diana (Testigo +).
- Diseño y síntesis de los sgRNA.
- Diseño y construcción del plásmido pSS5001.
- Evaluación de la eficiencia del sistema CAST modificado.
- Transformación de *P. chlororaphis* por dos distintos métodos y selección de mutantes
- Clonación y secuenciación para determinar la eficiencia del sistema.
- Descripción de las relaciones evolutivas del sistema CAST.

## 6. Estrategia de trabajo

Considerando las diversas situaciones que tuvieron un efecto sobre este trabajo, se planteó complementarlo con una parte teórica y la descripción a detalle de las etapas pendientes de realizar, como una forma de plan metodológico de trabajo futuro. La primera etapa comprende el análisis teórico correspondiente al sistema empleado en este trabajo, las etapas 2 y 3 representan un análisis *in silico* y la primera etapa experimental, respectivamente, y las etapas 4 y 5 contienen los protocolos y la propuesta metodológica pendiente por realizar.



1. Búsqueda y análisis bibliográfico de las características del sistema CAST.

- 1.1 Análisis de los componentes del locus CAST.
- 1.2 Papel del sistema CAST en la función celular.
- 1.3 Sistema INTEGRATE.
- 1.4 Comparación con el sistema INTEGRATE.
- 1.5 Otras asociaciones CRISPR – Cas – transposasa.

## 2. Diseño *in silico* y análisis bioinformático.

2.1 Búsqueda e identificación de las secuencias proteicas y elementos genéticos necesarios para las construcciones.

2.2 Diseño de las construcciones en el software Gene Construction Kit™ versión 4.5.

2.3 Diseño de los oligonucleótidos iniciadores para la amplificación de cada fragmento que conformará el plásmido pSS5001.

2.4 Diseño de los sgRNA de acuerdo con los criterios requeridos para el ensamble (Golden Gate).

2.5 Análisis de interacción de los elementos del sistema CAST.

2.6 Descripción de las relaciones filogenéticas y evolutivas.

## 3. Obtención del material genético.

3.1 pHelper: Plásmido con los genes *tnsB*, *tnsC*, *tniQ* y *cas12k*.

3.2 pDonor: Plásmido con los extremos de *Tn7* y el marcador de selección.

3.3 pTarget: Plásmido Diana (Testigo +).

3.4 Diseño y síntesis de los sgRNA.

## 4. Construcción del plásmido pSS5001.

4.1 Amplificación de los elementos del plásmido pSS5001.

4.2 Purificación y cuantificación de los amplicones.

4.3 Ensamble de los fragmentos.

4.4 Análisis del ensamble de los fragmentos.

4.5 Construcción de los plásmidos pSS5001 con los sgRNA.

5. Ensayo funcional del sistema CAST modificado.

5.1 Transformación de *P. chlororaphis*.

5.2 Análisis de inserción sitio específico.

## **7. Materiales y método**

### **7.1. Búsqueda y análisis bibliográfico del sistema CAST**

Se hizo una revisión bibliográfica para describir las características fundamentales del sistema CAST considerando su posible relación con diversas vías y procesos celulares. Además, se consideró toda la información disponible y actualizada sobre el mecanismo de inserción del sistema CAST y todas las posibles implicaciones dentro de su aplicación directa como una herramienta de edición genética.

#### **7.1.1. Análisis de los componentes del locus CAST**

Se llevó a cabo una descripción individual de cada uno de los elementos del sistema CAST, que incluyen las proteínas TnsB, TnsC, TniQ y Cas12k, los extremos LE y RE del transposón, y las proteínas codificadas por los genes cargo o secuencias genéticas contenidas dentro del locus. Además, se identificaron los elementos que se encuentran cercanos al locus CAST para describir posibles interacciones.

#### **7.1.2. Papel del sistema CAST en la función celular**

Tomando en consideración la información obtenida mediante la descripción de los componentes del sistema CAST, se propuso el papel natural de este sistema dentro de los procesos celulares. También se planteó una posible vía de interacción de este sistema con diversos mecanismos celulares, para observar si participa en otros procesos o depende de la presencia de elementos adicionales.

### **7.1.3. Sistema INTEGRATE**

Para establecer la proyección y elección del sistema CAST, se describió de forma breve los avances y mecanismos de interacción del sistema INTEGRATE más sobresalientes. Considerando los elementos que lo componen y están implicados en su funcionamiento, las limitaciones que posee y los requerimientos principales para su uso como sistema de edición genética.

### **7.1.4. Comparación con el sistema INTEGRATE**

Se estableció una comparación entre ambos sistemas para identificar las ventajas de cada uno de ellos y comparar la eficiencia relativa de inserción, con base en las observaciones más recientemente descritas. También se señalan las posibles diferencias en la aplicación de estos sistemas y cómo pueden ser mejorados.

### **7.1.5. Otras asociaciones CRISPR – Cas – transposasa**

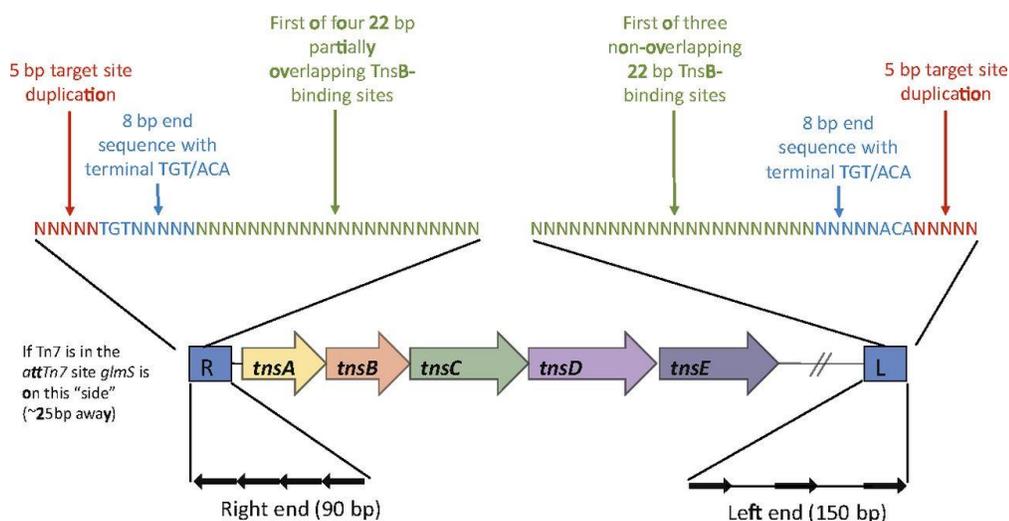
Los sistemas de asociación entre un locus CRISPR y proteínas con actividad de transposición, se distribuyen en diversas especies de bacterias y arqueas, por esto, se llevó a cabo la descripción de sistemas con una posible aplicación como herramientas de edición genética.

## **7.2. Diseño *in silico* y análisis bioinformático**

### **7.2.1. Búsqueda e identificación de las secuencias proteicas y elementos genéticos necesarios para las construcciones.**

Se obtuvieron las secuencias del genoma completo de *Scytonema hofmannii* (Nucleotide: NZ\_KK073768, Cepa: UTEX B 2349) y *Pseudomonas chlororaphis* (CDBB-92) en la base de datos

del NCBI (National Center for Biotechnology Information). Las secuencias de las proteínas TnsB, TnsC, TniQ y Cas12k (WP\_084763316.1, WP\_029636336.1, WP\_029636334.1 y WP\_029636312.1, respectivamente) provienen del reporte de Strecker *et al.* (2019). Se llevó a cabo la identificación del locus CAST y cada uno de sus elementos dentro del genoma de *S. hoffmannii*, mientras que para el reconocimiento de los CRISPR se utilizó la herramienta CRISPRFinder (Grissa *et al.*, 2007), realizando un alineamiento simple para buscar posibles sitios blanco de los espaciadores. Se identificaron los extremos RE y LE del transposón Tn7 en el plásmido pDonor y el genoma de *S. hoffmannii* de acuerdo con los criterios propuestos (Figura 2) por Peters *et al.* (2017). El origen de replicación derivado de pBR322 fue obtenido del plásmido pUIRM2304 (Anexo 13; construido por el Dr. Rodolfo Marsch) y el gen de resistencia a cloranfenicol proviene del plásmido pSUP2021 (Anexo 12; Simon *et al.*, 1983), ambos disponibles dentro de la colección interna del laboratorio.



**Figura 2. Criterios de identificación de los extremos LE y RE del transposón.** Requiere la búsqueda de la secuencia repetida invertida, la secuencia repetida directa y los sitios de reconocimiento por TnsB. Tomado de Peters *et al.* (2017).

### **7.2.2. Diseño de las construcciones**

Cada una de las secuencias requeridas para la construcción, fue analizada y manipulada mediante los programas Gene Construction Kit™ versión 4.5 y SnapGene versión 5.1.5. Los genes y elementos fueron ensamblados de forma hipotética revisando los marcos de lectura, la disposición correcta y la no redundancia en los sitios de restricción. Se optó por colocar el origen de replicación dentro del transposón para facilitar la evaluación de la eficiencia del sistema CAST.

### **7.2.3. Diseño de los oligonucleótidos para la amplificación de los fragmentos del plásmido pSS5001**

Cada oligonucleótido fue diseñado cumpliendo con los criterios requeridos (Figura 3) para el ensamble tipo Golden Gate (siguiendo las recomendaciones del fabricante NEB®), incluyendo sitios de restricción para la enzima BsaI, cuatro nucleótidos que permitan la unión complementaria de cada fragmento, un sitio de unión a ribosoma (RBS) al inicio de cada gen del sistema CAST y dos sitios de restricción flanqueando cada gen para facilitar una futura modificación. Las regiones de alineamiento en cada iniciador se plantearon para tener una temperatura ( $T_m$ ) entre 58°C y 62°C, se analizó la formación de dímeros o estructura secundaria y la presencia de una timina o adenina en el extremo 3'. Se revisaron de forma manual y utilizando la función Gateway Cloning en Snapgene, para comprobar que permitirían la amplificación e integración exitosa de cada uno de los fragmentos.



**Figura 3. Estructura de los oligonucleótidos para ensamblaje por Golden Gate.** Contiene cinco regiones: 1. Nucleótidos aleatorios para facilitar el reconocimiento del sitio Bsal, 2. Sitio Bsal, 3. Nucleótidos complementarios para el ensamblaje, 4. Sitio de restricción único en el vector y 5. Región de alineamiento.

#### 7.2.4. Diseño de los oligonucleótidos para la integración de los sgRNA

Se contempló la misma estrategia que en el caso de los oligonucleótidos para la amplificación de los fragmentos de plásmido pSS5001, sin embargo, el sitio de restricción utilizado fue para la enzima Lgul. La estructura de los oligonucleótidos sigue los requerimientos para la técnica de Golden Gate, no añadiendo características adicionales y únicamente considerando una longitud de 20 nucleótidos para la región de alineamiento en el sgRNA. La selección de regiones codificantes y no codificantes se hizo comparando los marcos de lectura en el genoma completo y los genes reportados en la base de datos del NCBI.

#### 7.2.5. Análisis de interacción entre los elementos del sistema CAST

Para identificar las posibles interacciones de los elementos del sistema CAST entre sí y con otras proteínas, se hizo el mapeo del interactoma de forma individual y en conjunto mediante las herramientas disponibles en el software STRING. De este modo, una vez identificadas las posibles interacciones, se describió individualmente cada uno de los elementos implicados y su papel dentro del proceso celular.

### **7.2.6. Descripción de las relaciones evolutivas y filogenéticas**

La estrategia para describir las posibles relaciones evolutivas del sistema CAST, se basó en la identificación de las características filogenéticas de la proteína Cas12k. Esta consideración se debió a que los elementos del mecanismo de transposición de la familia de Tn7, están ampliamente distribuidos y es posible encontrarlos en una proporción muy diversa dentro de los distintos organismos, aun cuando no estén relacionados con proteínas Cas12k. Se comenzó buscando la presencia de locus CAST en el genoma de diversos microorganismos mediante una búsqueda BLAST. Posteriormente, se aislaron las secuencias de los locus para cada organismo identificado y se realizó un alineamiento

### **7.3. Obtención del material genético**

#### **7.3.1. pHelper: Plásmido con las proteínas del sistema CAST**

El plásmido pHelper fue adquirido del repositorio Addgene bajo el número de catálogo #127921. Para recuperar el material genético se utilizó el protocolo modificado (Anexo 1) de solubilización de plásmido a partir de un liofilizado en papel filtro. Una vez recuperado se propagó transformando células de *E. coli* DH5 $\alpha$  mediante un protocolo mixto (Anexo 2) de choque térmico y permeabilidad química. Transcurridas 12 horas, se realizó la selección de mutantes utilizando como marcador de células transformantes la resistencia a ampicilina y las colonias seleccionadas fueron sujetas a resiembra por cuadrantes. La extracción del DNA plasmídico fue mediante el protocolo modificado de Birnboim y Doly (Anexo 3) y el análisis del

material obtenido se llevó a cabo a través de electroforesis en gel de agarosa y diversas digestiones enzimáticas para corroborar la presencia del plásmido.

### **7.3.2. pDonor: Plásmido con el transposón Tn7**

Se adquirió el plásmido pDonor (#127924) bajo el mismo procedimiento descrito para el plásmido pHelper. La recuperación del material genético fue mediante el mismo protocolo (Anexo 1) descrito para el plásmido pHelper. La amplificación de los fragmentos se hizo directamente sobre el plásmido recuperado del papel filtro en el cual fue enviado.

### **7.3.3. pTarget: Plásmido Diana (Testigo +)**

El plásmido pTarget (#127926) fue adquirido bajo el mismo procedimiento del plásmido pHelper. La recuperación del plásmido, su propagación y análisis, se hicieron mediante los mismos protocolos descritos para el plásmido pHelper. Sin embargo, como marcador de selección se utilizó la resistencia a cloranfenicol que le provee a la cepa el gen *cat* (cloranfenicol acetil transferasa), que contiene el plásmido pTarget.

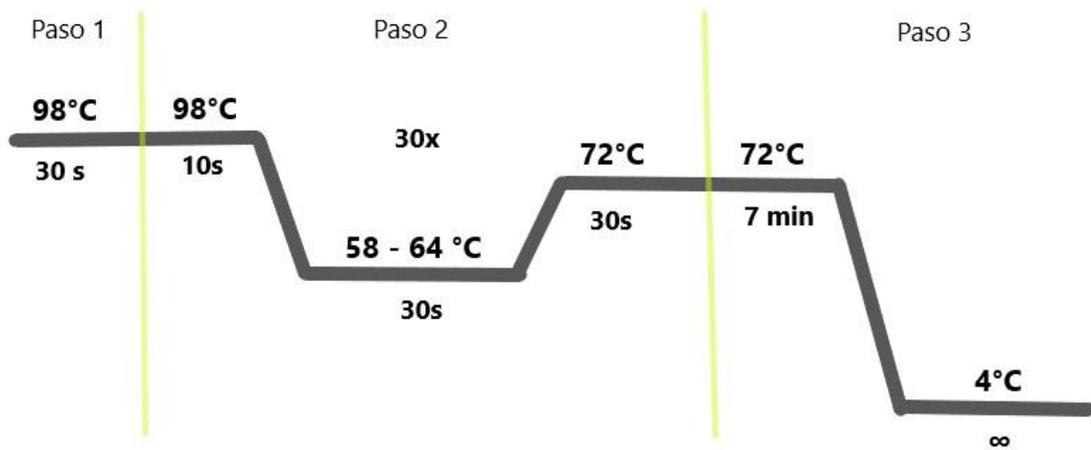
### **7.3.4. Oligonucleótidos para amplificación de los elementos del plásmido pSS5001**

Una vez diseñados y analizados a detalle, cada uno de los oligonucleótidos para amplificar los fragmentos que mediante Golden Gate permitirían integrar el plásmido pSS5001, fueron solicitados para su síntesis a la compañía Oligo T4. En función al tamaño del oligonucleótido, fue seleccionado un método de purificación; para los de un tamaño mayor a 35 pb se eligió una purificación por casete, mientras que para los más pequeños un desalado simple. El DNA liofilizado fue disuelto en agua inyectable estéril con un volumen variable para ajustar una concentración final de 100 nM en cada tubo.

## 7.4. Construcción del plásmido pSS5001

### 7.4.1. Amplificación de los elementos del plásmido pSS5001

Los fragmentos fueron obtenidos mediante la técnica de PCR (Anexo 4) y una polimerasa de alta fidelidad contenida en el kit Phusion™ de Thermo Fisher Scientific. Se estandarizaron las condiciones de reacción de cada uno de los fragmentos realizando gradientes de temperatura y la adición de DMSO, cuando fue necesario. Las reacciones de amplificación se llevaron a cabo utilizando controles (positivo y negativo) y fueron visualizadas mediante electroforesis en gel de agarosa al 1% y comparadas con un marcador de tamaños de fragmento Gene Ruler 1 kb plus de Thermo Fisher Scientific. La Figura 4 muestra el programa utilizado en el termociclador con las temperaturas requeridas para la reacción.



**Figura 4. Programa utilizado para la amplificación por PCR.** Las temperaturas de alineamiento varían con respecto a cada par de iniciadores.

Al optimizar las reacciones para cada uno de los fragmentos, fue posible establecer la temperatura de alineamiento que mejor rendimiento de DNA brinda, el número de ciclos, si se requiere la adición de DMSO o una variante específica de PCR (Hot start o anidada).

#### **7.4.2. Purificación y cuantificación de los amplicones**

Los productos de PCR se purificarán por dos métodos diferentes: diálisis y en columna (Anexos 5 y 6, respectivamente). El método de columna se probó con las reacciones de estandarización y mostró que permite recuperar la mayor parte del material genético con una buena eficiencia en la remoción de contaminantes no deseados. La cuantificación de la concentración y calidad del DNA se hará mediante dos métodos: intensidad de banda en gel de agarosa y por espectrofotometría. La finalidad de la purificación es la eliminación de sales y otras moléculas que puedan interferir en la integración de los fragmentos.

#### **7.4.3. Ensamble de los fragmentos**

El plásmido pSS5001 se construirá siguiendo el protocolo de ensamble Golden Gate (Anexo 7), mezclando cantidades equimoleculares de cada uno de los fragmentos o amplicones, con los componentes de reacción del ensamble y un programa con un número variable de ciclos que favorezca la integración eficiente del plásmido. La digestión enzimática de cada uno de los amplicones se hará con la enzima Bsal mientras que la ligación con la enzima DNA ligasa de T4, ambas de Thermo Fisher Scientific. Una vez finalizado el programa, se llevará a cabo la transformación de células *E. coli* DH5 $\alpha$  (Anexo 2), la selección de colonias resistentes a cloranfenicol y la resiembra por cuadrantes. Finalmente, se hará la extracción de DNA plasmídico por el protocolo modificado de Birnboim y Doly (Anexo 3) y su visualización en un

gel de agarosa al 1%, seleccionando los plásmidos linearizados que tengan un tamaño aproximado de 8600 pb.

#### **7.4.4. Análisis de integración de los fragmentos**

Los plásmidos candidatos serán sujetos a un análisis de restricción para validar de forma cualitativa la integridad de cada una de las regiones y la construcción correcta del plásmido pSS5001. Posteriormente, se secuenciará el plásmido usando los iniciadores de los genes que codifican para las proteínas del sistema CAST y los extremos del transposón Tn7, para verificar su correcta disposición.

#### **7.4.5. Construcción de los plásmidos pSS5001 con los sgRNA**

Los sgRNA se insertarán en el plásmido pSS5001 mediante un ensamble Golden Gate con Lgul. Para insertar los sgRNA, se diseñaron oligonucleótidos que serán sintetizados por la compañía Oligo T4. El ensamble Golden Gate se hará de acuerdo con el protocolo adaptado (Anexo 7), los plásmidos obtenidos serán propagados en células de *E. coli* DH5 $\alpha$  (Anexo 2) y su ensamble correcto se comprobará por un análisis de restricción y mediante electroforesis en gel de agarosa al 1% y secuenciación de los sitios de ensamble.

### **7.5. Ensayo funcional del sistema CAST**

#### **7.5.1. Transformación de *P. chlororaphis***

Se plantean tres estrategias de transformación bacteriana; protocolo mixto de permeabilidad química y choque térmico (Anexo 2), electroporación (Anexo 8) y conjugación bacteriana (Anexo 9). Se buscará el método que mejor eficiencia brinde y que no se vea afectado por los

sistemas de restricción bacterianos de *P. chlororaphis*. La selección de transformantes implica la resistencia a cloranfenicol y aquellas que muestren un crecimiento apropiado después de incubarlas 12 h a 27°C, representarán la primera evidencia de funcionamiento del sistema debido a que el origen de replicación en el plásmido pSS5501 es derivado de pBR322 y no funciona en *P. chlororaphis*.

### **7.5.2. Análisis de inserción sitio específico**

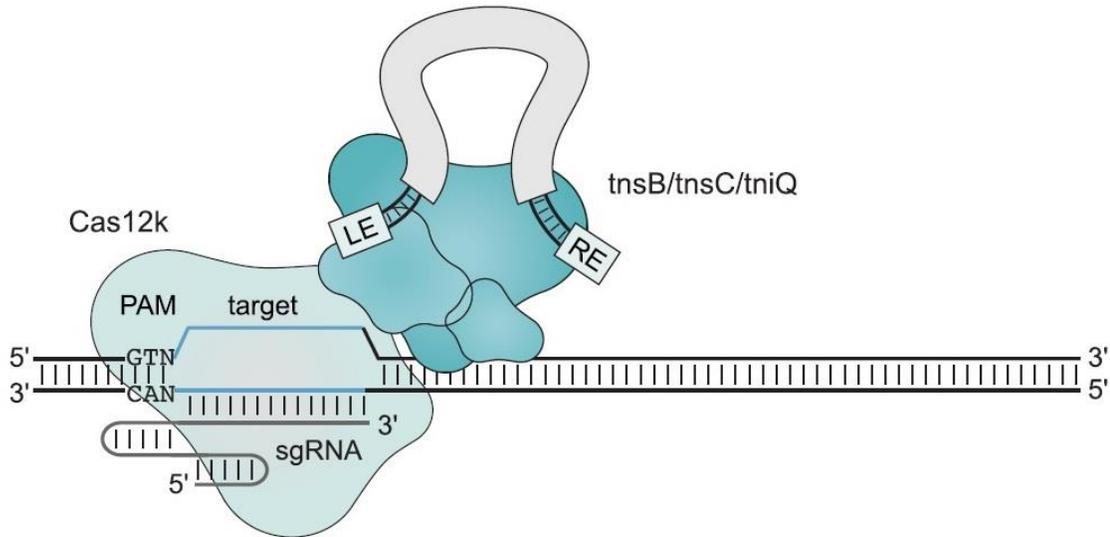
Para corroborar que la inserción del transposón se haya realizado en el sitio específico donde el sgRNA fue diseñado, se llevará a cabo la extracción del DNA genómico mediante el protocolo descrito por Wang *et al.* (2011), incluido en el Anexo 10 y la secuenciación de las regiones que flanqueen el fragmento insertado. Inicialmente se plantearon 10 sitios de prueba para el sistema, sin embargo, considerando la eficiencia de inserción reportada por Strecker *et al.* (2019), y debido a que se eligieron sitios que apuntan a regiones codificantes para enzimas clave en la ruta de biosíntesis de aminoácidos, es posible que resulte mortal para la célula y que no se refleje de forma correcta la eficiencia obtenida bajo las condiciones planteadas. Sin embargo, una opción sería utilizar un medio rico para que las células auxótrofas tengan un crecimiento observable. El siguiente paso sería implementar otros sitios de inserción sobre regiones no codificantes y enriquecer el medio de cultivo con los aminoácidos que potencialmente no sean sintetizados por la bacteria.

## 8. Resultados

### 8.1. Búsqueda y análisis bibliográfico del sistema CAST

#### 8.1.1. Análisis de los componentes del locus CAST

El sistema CAST (Figura 5) emplea una proteína efectora Cas12k, la cuál es una enzima que carece de actividad nucleasa en contraste con otras proteínas Cas, pero que tiene como función focalizar la maquinaria de transposición a un sitio específico a través de un RNA guía. La maquinaria de transposición está compuesta por tres subunidades: TnsB, TnsC y TniQ. La subunidad TnsB permite el reconocimiento de secuencias repetidas en los extremos del transposón y su ensamble, cuando se lleva a cabo el evento de inserción también participa directamente en la colocación del transposón en un sitio específico. TnsC favorece a la estabilidad del complejo multiproteico y permite la asociación de TnsB. Mientras que la interacción del complejo con el DNA se da a través de TniQ, la subunidad pequeña. En el sistema CAST, los genes de TniQ y TnsC se traducen con un solo marco de lectura, sin embargo, las adecuaciones realizadas en este trabajo consideraron la síntesis individual de cada elemento. La interacción entre el transposoma y la proteína Cas12k, es presumiblemente a través de TnsC (Strecker *et al.*, 2019). Sin embargo, la evidencia encontrada en la estructura cristalográfica para el sistema INTEGRATE (Halpin – Healy *et al.*, 2019), sugiere una asociación TniQ – Cas.



**Figura 5. Esquema propuesto para el sistema CAST.** La proteína Cas12k reconoce un sitio de complementario al sgRNA, genera la formación de un “loop – R” y la maquinaria de transposición compuesta por TnB, TnsC y TniQ, llevan a cabo la inserción del transposón. Tomado de Strecker *et al.* (2019).

Adicionalmente a los genes que codifican para las proteínas del sistema CAST, el locus contiene genes codificantes para las subunidades del sistema de restricción/modificación (R – M), caspasas CHAT (Hetf asociadas a Tprs) y una proteína homóloga a CorA. El sistema R – M está compuesto por las subunidades HsdS, HsdM y HsdR, tiene la capacidad de realizar la modificación (actividad metiltransferasa de HsdM) y restricción (actividad endonucleasa de HsdR) sobre DNA foráneo, representando una forma de inmunidad adaptativa frente a MGEs (Murray, 2000). Las caspasas CHAT presentes en una gran parte de las cianobacterias, están involucradas en la formación de heterocistos, un estadio morfológico que poseen estos organismos a lo largo de su desarrollo. Una delección en el gen *hetf* resulta en la incapacidad para formar heterocistos y morfologías variables irregulares (Risser & Callahan, 2008). La proteína homóloga a CorA podría estar implicada en el transporte de  $Mg^{+2}$  como los miembros

de esta familia (Murray, 2000). En una función hipotética podría proponerse que el sistema R – M funciona en conjunto con la proteína homóloga a CorA, debido a que la subunidad HsdR requiere de entre otros factores  $Mg^{+2}$  para llevar a cabo su actividad de endonucleasa.

### **8.1.2. Papel del sistema CAST en la función celular**

Actualmente, no se cuenta con un reporte que indique cuál es el papel del sistema CAST en el entorno celular natural. Tomando en consideración los elementos que contiene el locus, podría hacerse una hipótesis doble; el sistema CAST se encuentra implicado en la formación y maduración de los heterocistos y que este sistema participa activamente en la respuesta inmune adaptativa de *S. hofmannii*. La participación de los sistemas R – M podría verse beneficiada por los sitios blanco que poseen los espaciadores, debido a que los transcritos muestran múltiples objetivos en secuencias genómicas víricas. Esta verificación se hizo mediante el alineamiento de cada uno de los transcritos con potencial para conformar un tracrRNA maduro. Es importante mencionar que los transcritos tuvieron tres posibles sitios blanco dentro del genoma de *S. hofmannii*, pero ninguno para los MGEs reportados en las bases de datos. De los siete espaciadores que contiene el arreglo CRISPR, todos tuvieron al menos dos posibles sitios en genomas víricos, además de que ninguno de ellos apareció en otras cepas de *S. hofmannii*.

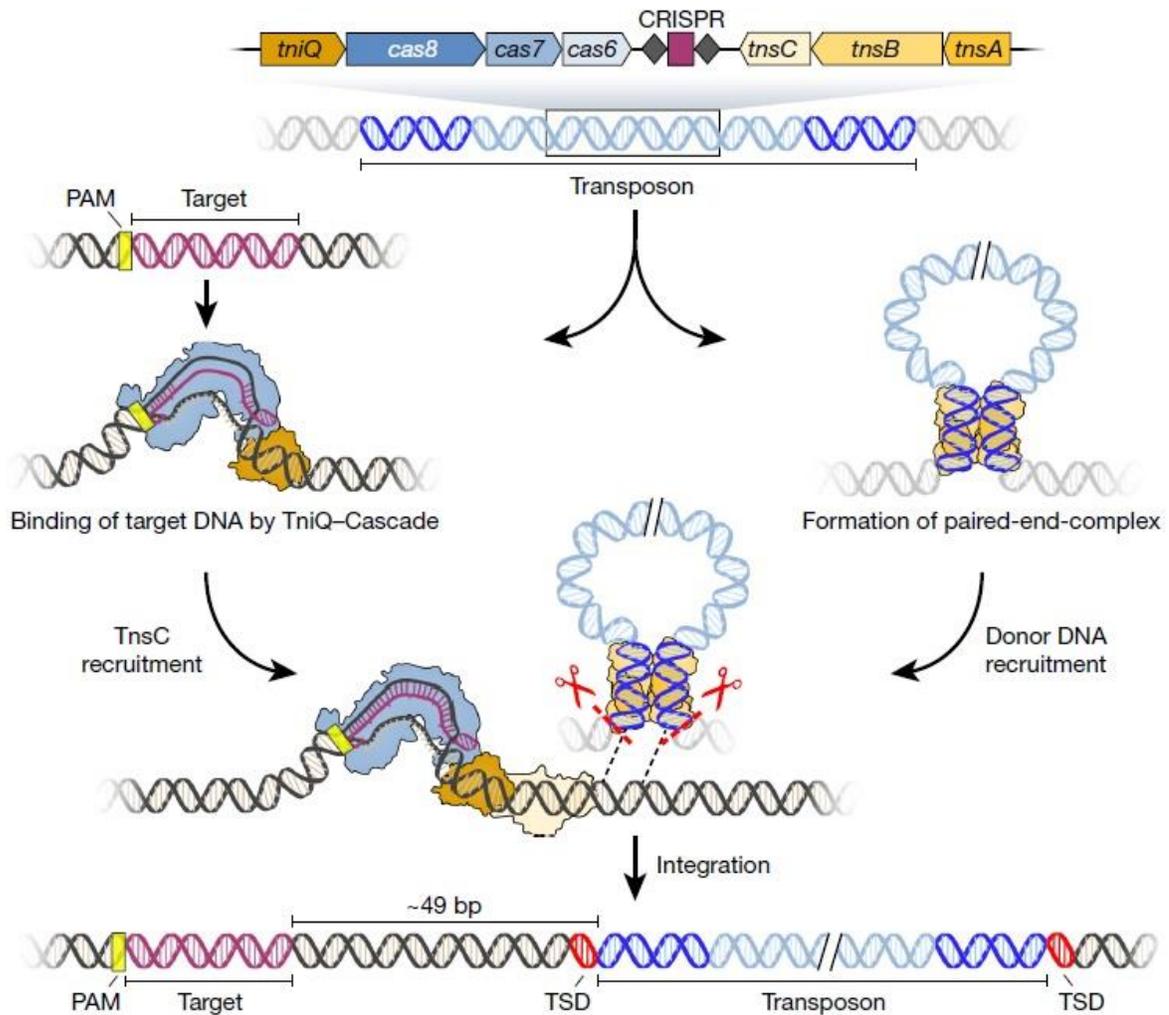
### **8.1.3. Sistema CRISPR – Cascade – Tn6677 (INTEGRATE)**

Simultáneamente al reporte del sistema que es empleado en este trabajo, fue publicado un trabajo en el que se empleó un transposón de la familia de Tn7 asociado a un complejo de proteínas Cas denominado Cascade. De acuerdo con Klompe *et al.* (2019), el sistema puede

efectuar la inserción dirigida de fragmentos en un locus específico. El sistema se encuentra de forma natural en *Vibrio cholerae*, está compuesto por las subunidades TnsA, TnsB, TnsC y TniQ, que es una proteína homóloga a TnsD. En el arreglo de genes del operón se encuentran los tres primeros en la cercanía del sitio CRISPR, mientras que el gen *tniQ* se encuentra en la vecindad de los genes *cas*. El complejo Cascade está formado por tres genes: *cas6*, *cas7* y la fusión de *cas8* – *cas5*, los cuales, al ser traducidos a proteínas, son responsables del procesamiento y acoplamiento del crRNA o RNA guía. Al igual que el resto de la familia de Tn7, el transposón Tn6677 está formado por extremos con secuencias TIR dentro de las que se encuentran un número variable de sitios de reconocimiento para TnsB.

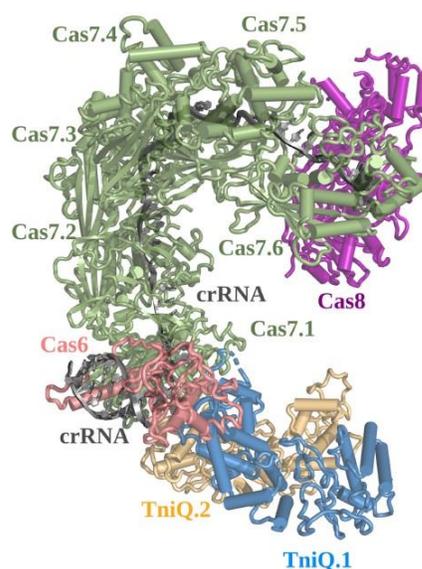
Para establecer el mecanismo de aplicación del sistema, Klompe *et al.* (2019), transformaron células de *E. coli* con tres plásmidos que codifican para los componentes del transposón de *V. cholerae*: el donador del transposón (pDonor), el operón *tns* (pTnsABC) y el operón *cas* (pQCascade) con el arreglo CRISPR para la síntesis del crRNA. La selección de los sitios blanco consideró un PAM con estructura 5' – CC – 3' y los eventos de inserción fueron en la región que comprende entre 58 y 60 bp del sitio con el PAM. Este sistema fue eficaz en la inserción tanto en un plásmido vacío, como a nivel de genoma. Las pruebas de la dependencia individual de proteínas indicaron que los eventos de inserción se interrumpían cuando se deletaba cualquier gen. Particularmente, considerando que el gen *tniQ* se encuentra en la región del operón *cas* se planteó desde el comienzo una interacción entre el complejo Cascade y la subunidad TniQ de la transposasa, por lo que al analizar directamente las interacciones proteicas se encontró una interacción entre la proteína Cas6 y TniQ, siendo ésta exclusiva y proponiéndose como el mecanismo de estabilización entre el complejo Cascade y la

transposasa de Tn6677. La mejor eficiencia en la inserción del plásmido donador se encontró con un tamaño aproximado de 775 bp, donde la eficiencia se reduce cuando se incrementa o reduce el tamaño del plásmido. Klompe *et al.* (2019) proponen un modelo (Figura 6) donde se esquematiza el mecanismo de integración genómica del sistema CRISPR – Cascade – Tn6677.



**Figura 6. Representación esquemática del sistema INTEGRATE.** Las proteínas codificadas en el operón *cas* se asocian con TniQ, se lleva a cabo el reconocimiento de DNA blanco y el reclutamiento de la proteína TnsC. Simultáneamente, se lleva a cabo el ensamblaje del transposoma, que mediante un mecanismo de corte y pegue, permite la inserción del transposón aproximadamente 49 bp corriente abajo del sitio con el PAM. Tomado de Klompe *et al.* (2019).

La estructura de interacción entre cada uno de los elementos del sistema INTEGRATE (Figura 7) descrita por Halpin – Healy *et al.* (2019), se analizó por criomicroscopía electrónica y se encontró que la subunidad grande, compuesta por Cas8 – Cas5, forma una protuberancia que permite el reconocimiento del extremo 5' del crRNA. Esta subunidad se conecta con TniQ a través de dominios flexibles en el extremo opuesto al sitio de ensamble del crRNA. El crRNA es protegido en su mayoría por seis subunidades Cas7 que forman una estructura tipo filamento a lo largo de la cadena, a través de motivos en forma de “dedo” que se unen al crRNA en intervalos regulares de 6 bp. La proteína Cas6 interactúa con una de las subunidades Cas7 mediante una hoja  $\beta$  formada por la contribución de una lámina  $\beta$  y de Cas6 y dos láminas  $\beta$  de la subunidad Cas6. La asociación Cas7 – Cas6 permite la interacción del complejo Cascade con TniQ, que genera el reclutamiento del transposoma.



**Figura 7. Modelo de la estructura de interacción de los elementos del sistema INTEGRATE.** La subunidad TniQ interactúa con las proteínas Cas7 y Cas6, que se encuentran formando el complejo Cascade. TniQ permite el reclutamiento de la maquinaria de transposición. Tomado de Halpin – Healy *et al.* (2020).

#### **8.1.4. Comparación con el sistema INTEGRATE**

En consideración de que ambos sistemas cumplen con una función similar, las diferencias en los métodos de inserción son notables. Ambos sistemas emplean transposones de la familia de Tn7, que utiliza un mecanismo de inserción de tipo corte y pegue, sin embargo, la diferencia estructural indica que mientras el sistema CAST emplea una única proteína Cas, el otro sistema emplea un complejo multiproteico Cascade. Al validar ambos sistemas los autores no encontraron la necesidad de proteínas adicionales o factores para el procesamiento, acoplamiento y conducción del complejo a un sitio blanco. La interacción de las subunidades de la transposasa en el caso de Tn6677 y el complejo Cascade, se encuentra atribuida a TniQ, mientras que en el sistema CAST la hipótesis de interacción sugiere un papel similar de TniQ que no fue validado experimentalmente y no coincide con las predicciones realizadas en el presente trabajo. Las características particulares de cada sistema (Tabla 1), permiten observar que el sistema CAST tiene una organización estructural menos compleja, mientras que cuentan con un tamaño de inserto evaluado muy similar. El tamaño de los genes que codifican para las subunidades del transposón y las proteínas Cas, es mayor en el sistema INTEGRATE en comparación con CAST.

**Tabla 1. Características de los sistemas CAST e INTEGRATE.** Comparación entre los elementos y características funcionales de cada sistema.

<i>Característica</i>	<i>CAST</i>	<i>INTEGRATE</i>
<i>Proteína Cas</i>	Cas12k	Cas6, Cas7 y Cas8 – Cas5
<i>Elementos transposasa</i>	TnsB, TnsC y TniQ	TnsA, TnsB, TnsC y TniQ
<i>PAM 5' – 3'</i>	NGTN	CC
<i>Sitio de inserción*</i>	60 – 66 bp	58 – 60 bp
<i>sgRNA</i>	tracrRNA	crRNA
<i>Plásmidos</i>	pDonor y pHelper	pDonor, pTnsABC y pQCascade
<i>Evaluado</i>	pInsert, genoma y ensamble proteico in vitro	pInsert y genoma
<i>Tamaño de inserto</i>	795 bp	775 bp

\*Corriente abajo del PAM

La eficiencia de ambos sistemas sólo ha sido evaluada por Zhang *et al.* (2020), en un estudio comparativo de CAST e INTEGRATE para observar si estos sistemas podían efectuar la inserción genética en múltiples sitios dentro del cromosoma bacteriano, en una herramienta que ellos denominaron MUCICAT. Para evaluar la aplicabilidad del sistema CAST, se seleccionaron secuencias de tipo multicopia (ISs), sin embargo, con este sistema no se encontró una buena eficiencia para la inserción múltiple permanente, perdiéndose copias del gen con los pases celulares. Cuando se utilizó el sistema INTEGRATE la inserción en múltiples sitios se vio incrementada, los mismos resultados fueron obtenidos utilizando un arreglo CRISPR con múltiples espaciadores. Además, se cuestionó si esta eficiencia era mantenida entre diferentes líneas celulares, para lo cual se transformaron *E. coli* MG1655  $\Delta endA$   $\Delta recA$  y BL21, encontrando que las eficiencias decaían cuando existía la delección de *recA* y *endA*. Además, se hizo un ensayo donde se evaluó directamente la capacidad del sistema INTEGRATE para insertar el casete de expresión para la enzima glucosa deshidrogenasa controlado por el promotor T7lac (T7-*gdh*). Los resultados fueron que existía una notable eficiencia de inserción

con al menos ocho copias del inserto en cada colonia seleccionada, generando también un incremento en la actividad de la enzima GDH al ser evaluada en un fermentador. En la última etapa, llevaron a cabo la transformación de *Tatumella citrea*, encontrando una eficiencia cercana al 100% atribuida a las características propias de la cepa o incluso la temperatura de crecimiento de 30°C.

#### **8.1.5. Otras asociaciones CRISPR – Cas – transposasa**

Actualmente, sólo se cuenta con reportes de la existencia de sistemas CAST en bacterias y en algunas arqueas. Podemos distinguir dos clases de sistemas que efectúan la inserción en sitios específicos mediadas por la actividad de una o varias proteínas Cas y una enzima integradora de DNA: los sistemas CAST (transposasa) y las casposasa. A pesar de que existen diferencias estructurales que diferencian a las integrasas denominadas casposasas de las proteínas de transposición, y el mecanismo mediante el que realizan la integración de fragmentos es diferente, ambas proteínas cumplen con una función y objetivo en común. No se cuenta con una clasificación formal y aceptada en torno a estos sistemas, debido a que han sido predichos una diversidad importante de elementos que los componen. Desde múltiples repeticiones de proteínas Cas o la participación de diferentes proteínas efectoras requeridas para el procesamiento, conformación o adquisición de la estructura funcional de inserción génica. Además, estas asociaciones ocurren con múltiples familias de transposones o de casposasas de estructura variable. Sin embargo, su origen podría referir a un modelo propuesto (Kapitonov and Jurka, 2005) donde se hipotetiza que estos sistemas provienen de un antiguo transposón de DNA, en analogía con los sistemas inmunes adaptativos de los organismos eucariontes superiores.

Existen múltiples propuestas de sistemas CAST o casposas con predicción *in silico* de actividad de inserción, sin embargo, sólo algunos han sido probados con éxito. Por ejemplo, Hickman *et al.* (2020), reportaron la actividad de una casposasa (repeticiones de Cas1 flanqueadas por los TIRs de un casposón y en la cercanía de una integrasa de DNA no caracterizada) en *Methanosarcina mazei*, con capacidad de realizar la integración en sitios específicos de DNA de cadena simple. Este trabajo fue evaluado *in vitro*, llevándose a cabo la inserción de oligonucleótidos de alrededor de 30 pb dentro de un plásmido (pUC19) y también dentro de oligonucleótidos de entre 500 y 600 pb. Las diferencias de este sistema con respecto al utilizado dentro de este trabajo residen en el tamaño de fragmento a insertar, la evaluación únicamente *in vitro*, la especificidad relativa y la actividad de la integrasa con capacidad de insertar de una forma no regulada.

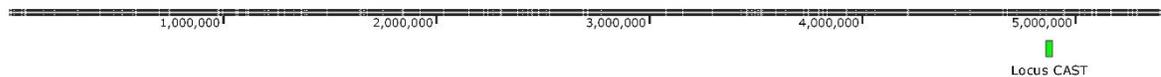
Se han reportado asociaciones entre proteínas Cas y enzimas con capacidad de efectuar transposición no sólo de forma natural, sino mediante ingeniería de proteínas, al hacer la fusión en un sistema quimérico que promueve la integración de fragmentos. El primer sistema probado con éxito fue el que asociaba una proteína Cas9 con el dominio nucleasa inhibido (dCas9) con un transposón PiggyBac. Recientemente Chen & Wang (2019), reportaron la asociación de la proteína dCas9 con un transposón de la mosca cornuda *Haematobia irritans* denominado Himar1. A diferencia de los sistemas CRISPR – Cas9 que requieren de la maquinaria de reparación dependiente de homología (HDR), el sistema quimérico desarrollado CasTn permitió una mejora en la frecuencia de inserción empleando únicamente las proteínas fusionadas. Este sistema permite la inserción de fragmentos del orden de las kilobases y fue

evaluado a nivel de genoma y en plásmidos receptores, en ambos casos en *E. coli*, mostrando una frecuencia de inserción 300 veces mayor al uso de una transposasa aleatoria.

## 8.2. Diseño *in silico* y análisis bioinformático

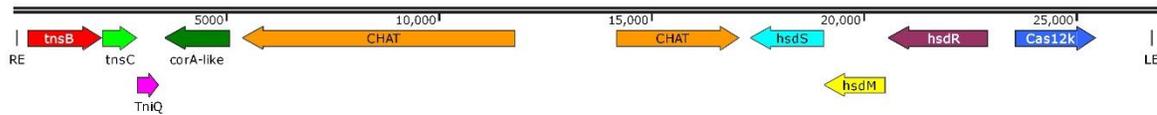
### 8.2.1. Búsqueda e identificación de las secuencias proteicas y elementos genéticos necesarios para las construcciones

Se identificó la posición del locus CAST dentro del genoma completo de *S. hofmannii* para el análisis de cada uno de sus componentes, éste se localiza en la posición 4,892,266 – 4,866,122 (Figura 8). Los genes del locus CAST fueron alineados y se hizo una comparación mediante una búsqueda efectuada con la herramienta BLAST (Basic Local Alignment Search Tool) dentro del sitio del NCBI, con las secuencias reportadas para corroborar su identidad.



**Figura 8. Localización del locus CAST en el genoma completo de *S. hofmannii*.** La longitud total del genoma es de aproximadamente 5400 kb y el locus se encuentra en la posición aproximada de 4800 kb.

El locus CAST (Figura 9) tiene una longitud de 26,895 pb, conteniendo los genes de las proteínas de transposición, de la proteína CAS y diversos genes cargo dentro de los que destacan los que codifican para proteínas del sistema de restricción/modificación (*hdsS*, *hdsM* y *hdsR*), con función dentro de la respuesta inmune adaptativa de la bacteria, y de caspasas de la familia de HetF asociadas con Tprs (CHAT). El locus está flanqueado por los extremos LE y RE del transposón, con el arreglo CRISPR cercano al extremo LE.



**Figura 9. Locus CAST.** Contiene los genes para el sistema CAST (*tnsB*, *tnsC*, *tniQ* y *cas12k*), genes cargo (*CHAT*, *hsdS*, *hsdM*, *hsdR* y *corA*), el arreglo CRISPR y los extremos RE y LE del transposón.

La identificación de los extremos LE y RE del transposón obedece a la búsqueda de la secuencia de repetidos directos (5 nt), repetidos invertidos (8 nt) y los 3 – 4 sitios de reconocimiento para la conformación del transposoma (22 nt). La Figura 10 muestra la estructura de los extremos LE y RE, identificados para el caso específico del sistema CAST la cual corresponde con los criterios propuestos por Peters *et al.* (2017). Estas secuencias tienen una especificidad relativa para una transposasa en particular, permitiendo que se regule y controle la inserción. Considerando los 3 – 4 sitios de reconocimiento para la conformación del transposoma, los extremos LE y RE muestran una longitud aproximada de 250 nt, los cuales coinciden al alinear las secuencias con el plásmido pDonor y el genoma completo de *S. hofmannii*.

RE: 5' – CTACGTGTACAGTGACTAATTATATGTCGTTGTGA – 3'

LE: 3' – CACCGACGACAGATAATTTGTCACTGTACACTACG – 5'

**Figura 10. Estructura de los extremos LE y RE del transposón de Tn7.** En rojo se muestran las secuencias de repetidos directos, en azul los repetidos invertidos y en verde el primer sitio de reconocimiento para la transposasa.

Al hacer la identificación del arreglo CRISPR (Tabla 2) se observó que contiene un total de ocho protoespaciadores o secuencias repetidas, y siete espaciadores, que son las secuencias que

permiten la síntesis del RNA guía. La longitud total del arreglo es de aproximadamente 545 n, y se localiza en la región comprendida entre el extremo LE y el gen *cas12*.

**Tabla 2. Arreglo CRISPR del locus CAST.** Los protoespaciadores son secuencias repetidas alternadas entre los espaciadores, que son secuencias génicas generalmente adquiridas en eventos de inmunidad adaptativa.

Protoespaciador	Espaciadores
GTGGCAACAACCTTCCAGGTACTAGGTGGGTTGAAA	GGCTATTGCCTTATCTAATTGCAGTCGAGTCAATGG
G	GCTTGGGATATAAAAACAAGGGATATTCCAAACCC
	CTATTCTCGCTTCATCCCAACAGCGGTTTAAGGAC
	TGCAAAGGGATCATCTTTATTTTCATCCTTTGCAG
	GTTGATTTTTCGATATCGATGGCTCGTTAATCGACAATCG
A	ATCCCGCAACCCGCACTCCCTGCTTTTCGGCAG
	AAGCTATCTTTGTCTTTTTTCGACAAAGCGATCGC

La identificación de los espaciadores permitió su búsqueda y alineamiento dentro del genoma de *S. hoffmannii* para reconocer posibles sitios blanco de inserción en un escenario natural. Se seleccionaron sólo los espaciadores que cumplen con un tamaño específico de 20 n, la presencia de un PAM con estructura NGTN y que alinearan en la totalidad de nucleótidos. Al alinear los espaciadores se identificaron tres posibles sitios blanco (Tabla 3) dentro del genoma de *S. hoffmannii*, los cuales se encuentran en regiones codificantes. Además, fueron rastreados

**Tabla 3. Posibles sitios blanco de los espaciadores contenidos en el arreglo CRISPR del locus CAST.** Un total de tres sitios fueron localizados en el genoma de *S. hoffmannii*, sobre regiones codificantes.

Espaciador	Sitio	Blanco
3	4,856,136 – 4,856,167	NADH - oxidoreductasa
5	432,360 – 432,391	Proteína hipotética
6	4,606,236 – 4,606,265	Proteína hipotética

De la misma forma, cada uno de los espaciadores fue validado para los genomas víricos disponibles en la base de datos del NCBI, buscando posibles sitios blanco. El uso de la herramienta online ViPR predijo la existencia de posibles sitios blanco para los espaciadores, representando un indicio que aporta a la hipótesis del papel del sistema CAST dentro de la respuesta inmune adaptativa natural en *S. hofmannii*. La selección de estos sitios blanco se llevó a cabo bajo los mismos criterios empleados en el análisis del genoma de *S. hofmannii*, se encontró que estos sitios cumplen con los requerimientos mínimos para su reconocimiento por la proteína Cas12k asociada al sgRNA. En la Tabla 4 se observan los organismos hipotéticos donde se encuentran los sitios blanco hallados, para evitar posibles redundancias, la búsqueda y alineamiento de las secuencias sólo considera aquellas con un alineamiento del 100%, sin embargo, es probable que se pueda formar un híbrido sgRNA – DNA aún con porcentajes menores (Strecker *et al.*, 2019). Los espaciadores 4 y 6 no coincidieron con ningún posible sitio blanco en bacterias, pero con el espaciador 6 si se encontró un posible blanco dentro del genoma de *S. hofmannii*.

**Tabla 4. Posibles sitios blanco en genomas víricos.** Se muestra un ejemplo de organismo, cepa y acceso a GenBank, debido a que cada espaciador coincide con sitios en varias cepas del mismo organismo

<i>Espaciador</i>	<i>Coincidencias</i>	<i>Organismo</i>	<i>Cepa</i>	<i>Hospedero</i>	<i>GenBank</i>
1	2	Sapporo virus	Ah-1	Cerdo	JX678943
2	47	Myxoma virus	A-sa-pa-10-2014	Conejo	MK388144
3	49	Sapporo virus	Parkville	Humano	U73124
5	4	Entomopoxvirus	HF679133	Polilla	HF679133
7	18	Virus de Norwalk	GII.P15	Humano	MN473468

Por otra parte, se seleccionó cada uno de los elementos del plásmido pSS5001 tomando como base diversos vectores para facilitar la amplificación de cada uno de los fragmentos. La estructura de cada uno de los plásmidos se incluye en los Anexos 14, 15 y 16, mostrando en la Tabla 5 cada uno de los fragmentos y el plásmido del que fue amplificado.

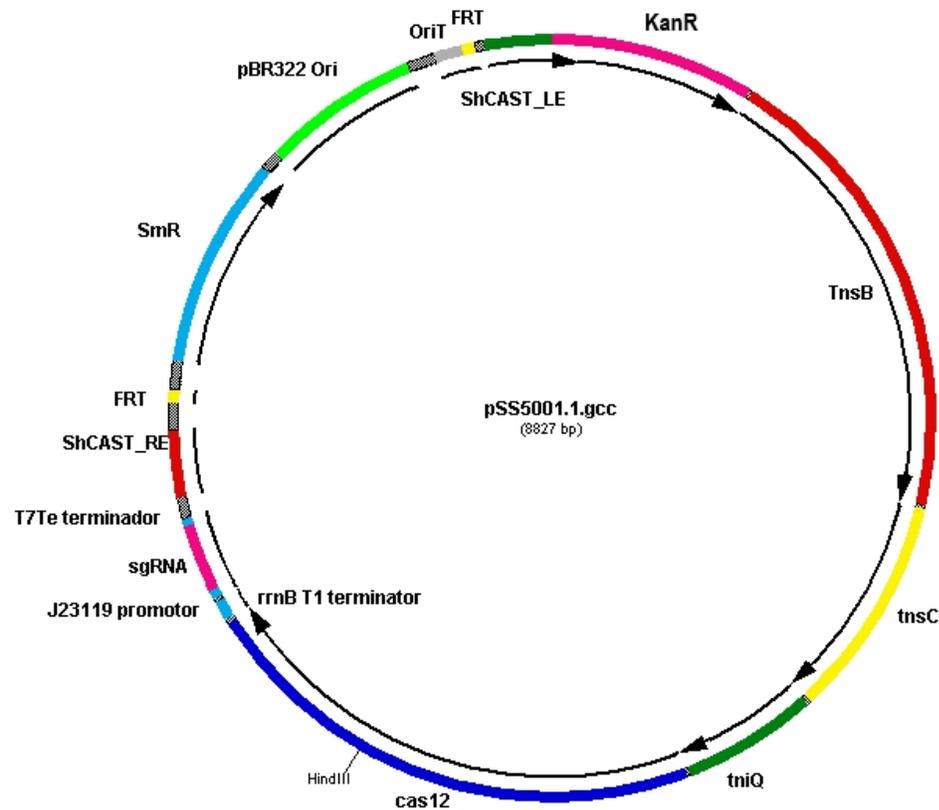
**Tabla 5. Elementos para la integración del plásmido pSS5001.** Cada fragmento contiene el origen del que fue obtenido.

<i>Elemento</i>	<i>Fuente</i>
<i>tnsB</i>	<b>pHelper</b>
<i>tnsC</i>	
<i>tniQ</i>	
<i>cas12</i>	
<i>Sitio sgRNA</i>	
<i>RE y LE</i>	<b>pDonor</b>
<i>Transposón</i>	
<i>FRT</i>	
<i>oriT</i>	<b>pUIRM2304</b>
<i>Origen de replicación</i>	
<i>cmR</i>	<b>pSUP2021</b>

### 8.2.2. Diseño de las construcciones

El diseño incluyó dos variantes iniciales: una con doble gen de resistencia a antibiótico y una con un solo gen. La variante de doble gen (Figura 11) estaba planteada para observar las posibles coinserciones que son características de los transposones de la familia de Tn7, observando una coinserción completa del plásmido cuando las colonias resultantes

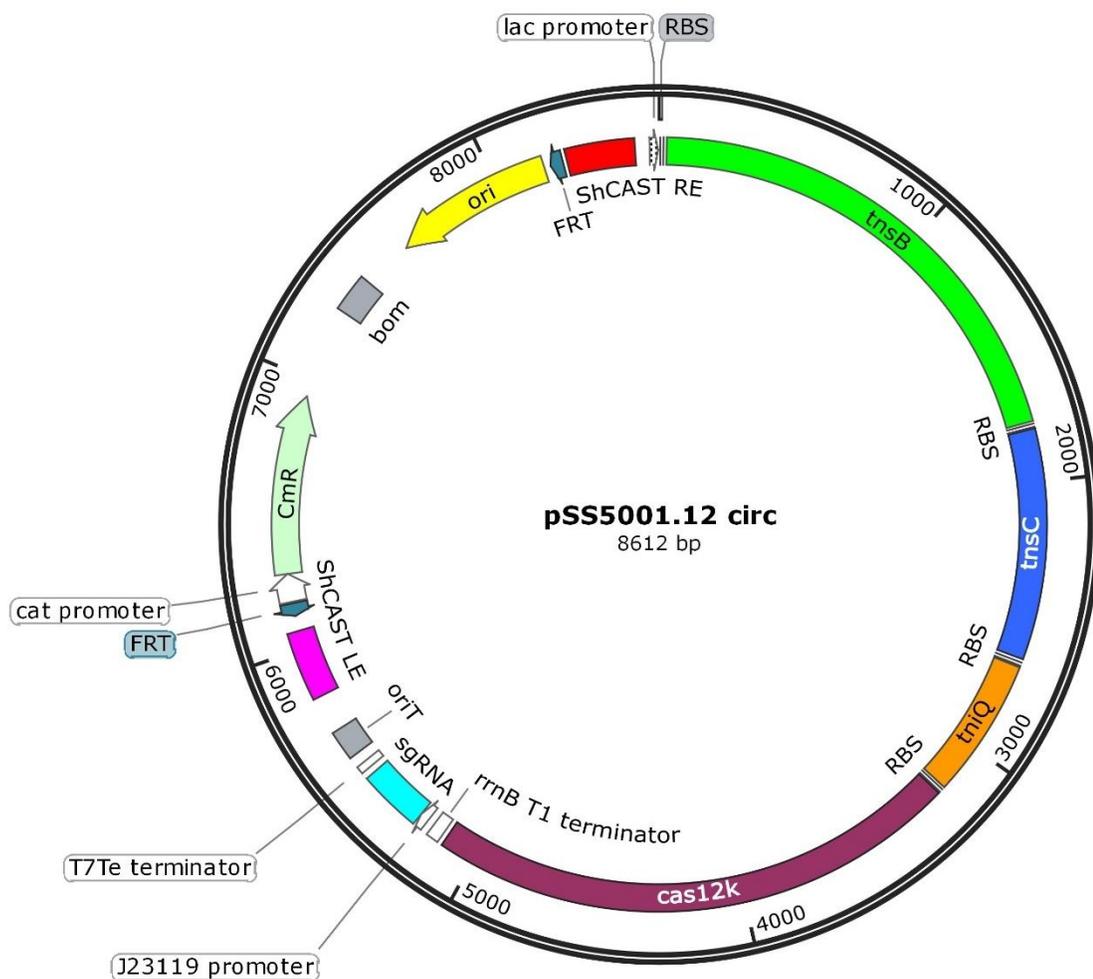
presentaran resistencia a ambos antibióticos. Sin embargo, para simplificar la estructura final del plásmido pSS5001, se consideró la segunda variante de diseño.



**Figura 11. Diseño inicial del plásmido pSS5001.** La construcción contempla el uso de dos marcadores de selección para rastrear coinserciones, los genes de resistencia a estreptomicina y kanamicina.

Cada uno de los genes del sistema CAST fue incluido en el plásmido pSS5001 (Figura 12) considerando la síntesis individual de las proteínas para las cuales codifican. Esta consideración se hizo debido a que los genes *tnsC* y *tniQ* se encuentran superpuestos en la región donde termina el gen *tnsC* y comienza *tniQ*. El sitio sgRNA fue incluido tomando también el promotor J23119 que participa en el inicio de la transcripción del RNA guía y el terminador T7Te. La inclusión del *oriT* tiene como finalidad proveer una estrategia de conjugación de *P.*

*chlororaphis*, por si se tuvieran dificultades con los sistemas de restricción bacterianos. Mientras que el cambio de marcador de selección fue planteado considerando que los genes que codifican para las proteínas que brindan resistencia a estreptomicina y kanamicina, contienen sitios de restricción redundantes a los seleccionados en las regiones que flanquean al sistema CAST. Los fragmentos FRT son sitios de reconocimiento de una recombinasa en *S. cerevisiae*, los cuales resultan útiles para la evaluación del sistema en levaduras.



**Figura 12. Diseño final del plásmido pSS5001.** La construcción fue optimizada para contener todos los elementos del sistema CAST, un marcador de selección (gen de resistencia a cloranfenicol) y el origen de replicación derivado de pBR322 para facilitar la observación de las inserciones genómicas.

El sitio de ensamble de los oligonucleótidos para la síntesis del sgRNA, está compuesto por dos regiones; la que se asocia a la proteína Cas (tracrRNA) y la región variable que es específica para un sitio diana dentro del genoma (crRNA). La secuencia del tracrRNA (Figura 13) está relativamente conservada para las proteínas Cas pertenecientes a las distintas familias. Sin embargo, en este diseño se utilizó sólo un fragmento del tracrRNA total contenido en el locus CAST debido a que se encontró que es la parte indispensable para llevar a cabo la formación del sgRNA y su asociación con la proteína Cas.

```
ATATTAATAGCGCCGCAATTCATGCTGCTTGCAGCCTCTGAATTTGTAAATGAGGGTTAGTTTGACT
GTATAAATACAGTCTTGCTTTCTGACCCTGGTAGCTGCTCACCTGATGCTGCTGTCAATAGACAGGAT
AGGTGCGCTCCCAGCAATAAGGGCGCGGATGTACTGCTGTAGTGGCTACTGAATCACCCCGATCAAGG
GGGAACCCTAAATGGGTTGAAAG
```

**Figura 13. Secuencia del tracrRNA en formato de DNA.** Permite el acoplamiento del sgRNA a la proteína Cas12k y sirve como andamio para la región variable o crRNA, que sitúa la maquinaria del sistema CAST a un sitio blanco.

### 8.2.3. Diseño de los oligonucleótidos para la amplificación de los fragmentos del plásmido pSS5001

Se propuso la amplificación de 11 fragmentos considerando los elementos seleccionados para la construcción del plásmido pSS5001, el orden que tienen dentro de los plásmidos originales y la estructura del diseño, por lo tanto, el diseño de 22 oligonucleótidos iniciadores utilizados en la técnica de PCR. Los iniciadores para la amplificación de los genes del sistema CAST fueron diseñados para agregar características específicas a cada fragmento como se muestra en la Tabla 6. Los oligonucleótidos que se diseñaron bajo esa estructura son los que amplifican los genes *tnsB*, *tnsC*, *tniQ* y *cas12k* con el sitio de ensamble del sgRNA.

**Tabla 6. Estructura de los oligonucleótidos para la amplificación de los genes del sistema CAST.**

Fragmento	Secuencia (5' → 3')	T <sub>m</sub> (°C)
tniQ_F	gtgaGGTCTCtgcgtaaagagagaaaacgctcatgatagaagcaccagatgtta	58
tniQ_R	gtgaGGTCTCgtactgctattcattggacaaccagatcaattca	58

En mayúsculas el sitio de reconocimiento de la enzima BsaI, en rojo los nucleótidos complementarios para el ensamble Golden Gate, en azul el RBS, en verde los nuevos sitios de restricción y en morado la región de alineamiento.

El resto de los oligonucleótidos (Tabla 7), que permiten la amplificación de los elementos restantes, no contienen características adicionales y sólo muestran el sitio de reconocimiento para la enzima BsaI y los nucleótidos complementarios para el ensamble Golden Gate. Las temperaturas de alineamiento son variables en virtud de la composición de la región de alineamiento, sin embargo, se cuenta con diversas estrategias para favorecer la amplificación específica de cada fragmento.

**Tabla 7. Oligonucleótidos para la amplificación de los 11 fragmentos del plásmido pSS5001.**

Oligo	Secuencia (5' → 3')	T <sub>m</sub> (°C)
tnsB_F	gtgaggtctcagtgtaaagaggagaaaactgtaggatctatgaacagtcagca	58°
tnsB_R	gtgaggtctccacgtgctattaaaacccatattcttcacga	58°
tnsC_F	gtgaggtctctacgtgaaagaggagaaaactgcagatgacagaagctcaggcgat	60°
tnsC_R	gtgaggtctcaacgcgtctatcatttgactccttagctact	60°
tniQ_F	gtgaggtctctgcgtaaagaggagaaaacgctcatgatagaagcaccagatgtta	60°
tniQ_R	gtgaggtctcggtaacctcattggacaaccagatcaattca	58°
cas12k_F	gtgaggtctcggtaaccaaagaggagaaaactgcagatgagtgcaataactattcaagct	62°
cas12k_R	gtgaggtctcatcgatgttcaccgacaaacaacagat	60°

**Tabla 12.** (Continuación)

sgRNA_F	gtg <b>agg</b> tctcatcgattgacagctagctcagtccta	60°
sgRNA_R	gtg <b>agg</b> tctctagcctaggggtccgcgcacatt	60°
OriT_F	gtg <b>agg</b> tctctggctttatcagagcttatcggcca	58°
OriT_R	gtg <b>agg</b> tctcagtccttgaagttcctattccgaagtt	60°
CmR_F	gtg <b>agg</b> tctctggactgatcggcacgtaagaggtt	60°
CmR_R	gtg <b>agg</b> tctcacgctagggcaccaataactgcctt	60°
Ori rep_F	gtg <b>agg</b> tctcgctacatgacggtgaaaacctctga	58°
Ori rep_R	gtg <b>agg</b> tctcaagcgaggatctaggtgaagatcct	60°
FRT_F	gtg <b>agg</b> tctctgtagagcgctctgaagttcctata	58°
FRT_R	gtg <b>agg</b> tctcagactatcttgaagtacctattccgaa	58°
ShCAST RE_F	gtg <b>agg</b> tctcgagtcaggcgacagtcaatttgta	58°
ShCAST RE_R	gtg <b>agg</b> tctcatccggttgactgagagtgccacat	58°
Lac prom_F	gtg <b>agg</b> tctctcggaCGatctcaagaagatcatctta	60°
Lac prom_R	gtg <b>agg</b> tctcatcacCaacatacGagccggaagcat	64°

En negritas se muestra la región de reconocimiento de la enzima Bsal.

#### **8.2.4. Diseño de los oligonucleótidos para la integración de los sgRNA**

Se seleccionaron diez regiones dentro del genoma de *P. chlororaphis* para probar la inserción sitio específico del sistema CAST; cinco en regiones codificantes y cinco en regiones no codificantes, como se muestra en la Tabla 8. El blanco de las regiones codificantes fueron genes implicados en la ruta de biosíntesis de aminoácidos, controlando su posible letalidad mediante el enriquecimiento del medio de cultivo con los posibles aminoácidos no sintetizados por la bacteria después de la transformación. La selección de sitios se hizo contemplando secuencias

PAM con estructura NGTN y considerando la distancia de inserción del sistema desde el sitio PAM, que es de alrededor de 60 – 66 nucleótidos.

**Tabla 8. Secuencias guía para la integración de los sgRNA.**

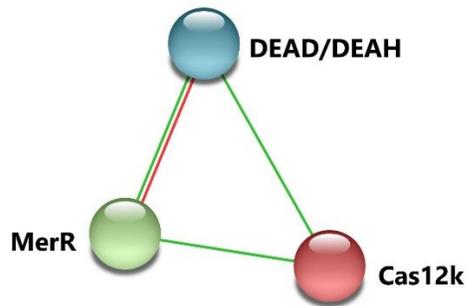
S	PAM	Secuencia guía (sgRNA)	Posición	Región
1	CGTT	TTACAACGGTTGCCCTTCCAGAA	23,327	No codificante
2	TGTG	TCAGGCAGAACGCTTCGCGGGCA	398,305	
3	AGTC	CCTTTCATATGGGTGCTCTCAGG	1,628,068	
4	AGTT	TTCCGCTGTTATTTTCAGGAGGA	1,745,450	
5	GGTT	GAGGAGGCTGGTAGTTATAGATG	3,042,215	
6	GGTT	ACCTTCGTCACCGCCGGCGACCC	1,050,327	Histidinol deshidrogenasa
7	TGTG	GTGCTACGACGCGCGGACCATCA	2,480,186	Diaminopimelato descarboxilasa
8	AGTT	GCTGCGCAATGCCCGCCCTGG	6,340,440	Shikimato deshidrogenasa
9	CGTT	CCTCGGCAATATCGAGCCGGAAC	6,784,499	Pirrolina – 5 – carboxilato reductasa
10	AGTC	CTGGGGCGGCCGCTTCAGTGAGC	6,945,665	Argininosuccinato liasa

S: sitio, PAM: protospacer adjacent motif.

### 8.2.5. Análisis de interacción entre los elementos del sistema CAST

Adicionalmente al diseño y las estrategias de construcción del plásmido pSS5001, se contempló un análisis bioinformático para observar posibles interacciones entre los elementos del sistema CAST y otros factores o proteínas dentro del entorno celular natural. Este análisis resulta relevante para una propuesta *in silico* y una posible validación experimental posterior, debido a que no existen reportes a la fecha que muestren las interacciones de los elementos de CAST más allá de las suposiciones basadas en sistemas similares. Con la predicción de interacciones para la proteína Cas12 (Figura 14), se encontró que existe una relación hipotética con al menos otras dos estructuras proteicas; el factor de transcripción de la familia de MerR

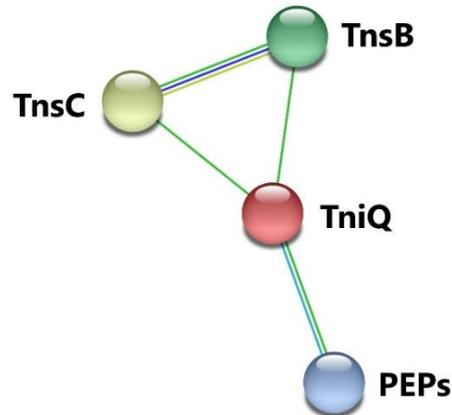
y una helicasa de la familia de las proteínas DEAD/DEAH box. Las interacciones que incluye la predicción del software String, están basadas en la presencia de los genes dentro de un mismo locus o al menos que se encuentran cercanos dentro del genoma.



**Figura 14. Interactoma de la proteína Cas12k.** Ilustra las interacciones directas del sistema, seleccionadas del interactoma completo obtenido.

Considerando las observaciones realizadas por Halpin – Healy *et al.* (2019), acerca de la naturaleza de interacción de la proteína TniQ y las proteínas del complejo Cascade, se planteó la hipótesis de que una interacción similar se lleva a cabo entre TniQ y Cas12k. Para validar esta hipótesis, se efectuó el mapeo del interactoma utilizando la secuencia de aminoácidos de la proteína TniQ como entrada. El interactoma (Figura 15) mostró que existe una interacción directa con TnsC, TnsB y una PEPs (fosfoenolpiruvato sintetasa), estas interacciones son por genes presentes en el mismo locus en el caso de todos los elementos. Sin embargo, destacan las interacciones de TnsC – TnsB y TniQ – PEPs debido a que están sustentadas en bases de datos para reportes experimentales de proteínas pertenecientes a la misma familia o proteínas homólogas. En primera instancia podemos observar que, a pesar de que los genes de estas

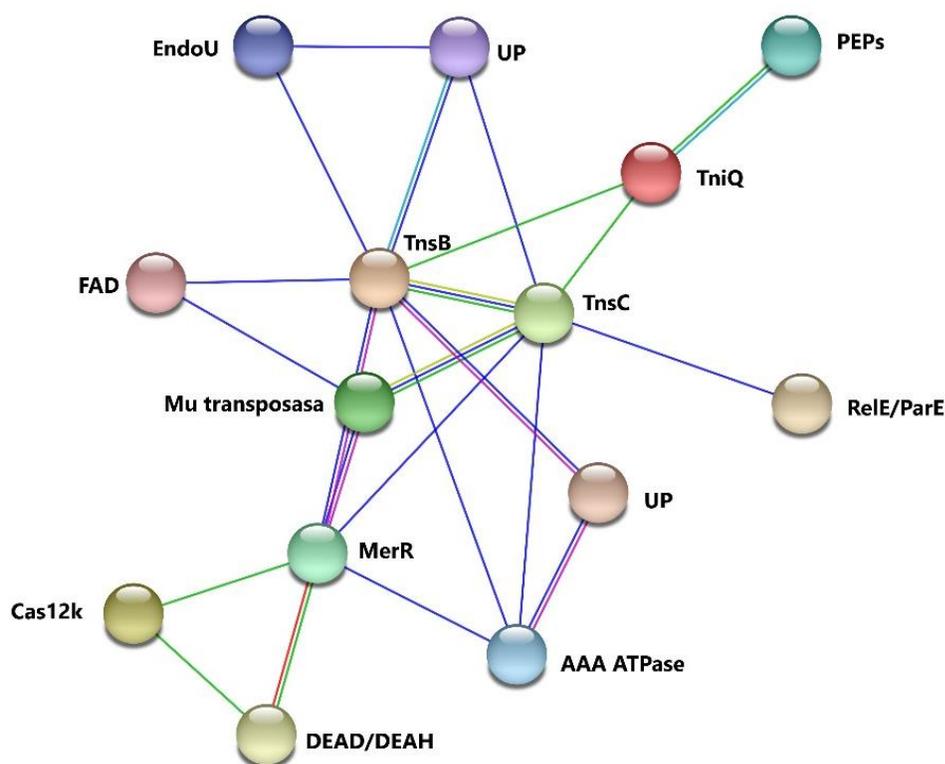
proteínas se encuentran en el mismo locus e incluso dentro del mismo transposón, no se reporta la ocurrencia de interacciones entre las proteínas de transposición y Cas12k.



**Figura 15. Interactoma de TniQ.** Las líneas verdes representan interacciones por genes en un mismo locus, la línea azul cielo incluye las interacciones reportadas determinadas experimentalmente, la línea azul rey y amarilla interacciones documentadas para proteínas homólogas.

Al no observarse una interacción hipotética entre las proteínas TniQ y Cas12k, se hizo el mapeo del interactoma completo empleando las secuencias de las cuatro proteínas que componen al sistema CAST. Las interacciones (Figura 16) fueron consistentes con los interactomas individuales, sin embargo, no se observó una predicción para una asociación entre la proteína Cas12k y algún elemento de la transposasa de Tn7. Se detectaron cuatro interacciones previamente reportadas; MerR – TnsB, MerR – Mu, TnsB – UP (proteína sin identificar) y UP – AAA ATPasa, y una interacción de proteínas homólogas contrastada de forma experimental (TnsB – UP). No se observaron interacciones hipotéticas entre los nodos de TniQ y Cas12k, aun cuando se incrementara el tamaño del interactoma agregando más elementos. De forma

interesante, se encontraron predicciones de asociación entre las subunidades de la transposasa y otras proteínas; TnsC – RelE/ParE (sistema toxina – antitoxina), TnsB – FAD (desaturasa de ácidos grasos), TnsB – EndoU (endoribonucleasa), TnsB – AAA ATPasa, TnsC – AAA ATPasa y la interacción de TnsB y TnsC con proteínas sin identificar.

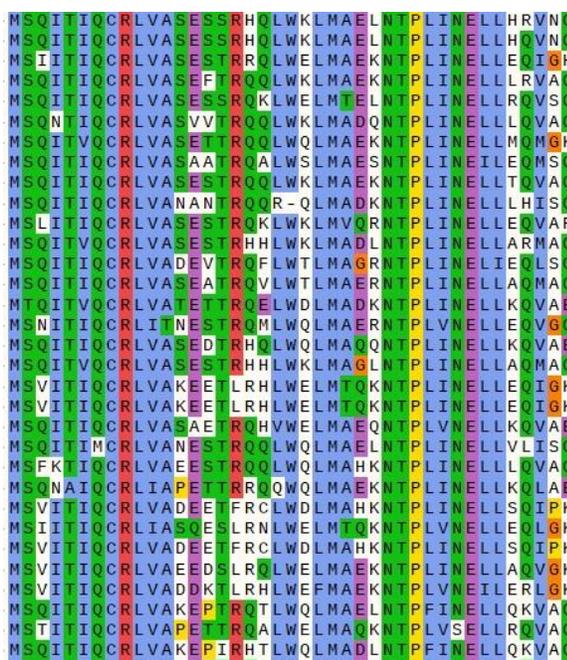


**Figura 16. Interactoma del sistema CAST.** Muestra las interacciones hipotéticas de los elementos del sistema CAST. UP: proteína no identificada.

### 8.2.6. Descripción de las relaciones filogenéticas y evolutivas

Al analizar los elementos individuales en el sistema CAST, surgió la duda acerca de si existían más sistemas con una estructura similar. La búsqueda de más sistemas que incluyeran la presencia de la proteína Cas12k dio 96 posibles locus CAST. Haciendo una depuración se

seleccionaron 45 locus que cumplieran con las características requeridas para una hipotética función de inserción génica sitio específico. Todos los locus hallados, aun los que no fueron seleccionados, se encontraron únicamente en cianobacterias, lo que podría representar que los sistemas CAST basados en la asociación de la proteína Cas12k y Tn7 son exclusivos de estos organismos. No obstante, la proteína Cas12k de *S. hofmannii* tuvo una importante divergencia con otras proteínas Cas12k, teniendo en su mayoría sustituciones aminoacídicas no sinónimas o pertenecientes a una clasificación diferente. El alineamiento de las 46 secuencias (45 seleccionadas y 1 consenso), de la proteína Cas12k (Figura 17) permitió establecer una comparación observable en la divergencia aminoacídica.

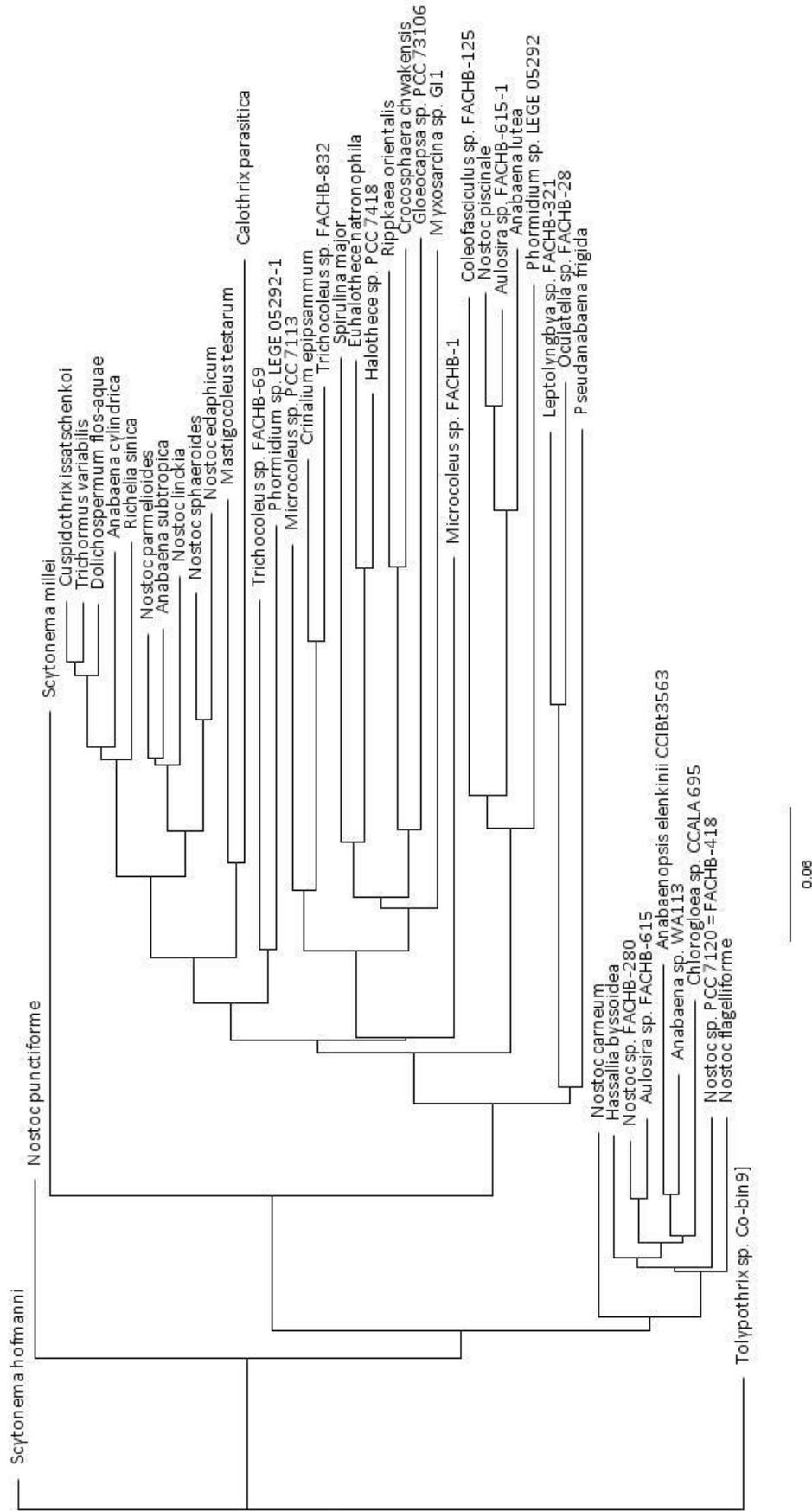


**Figura 17. Alineamiento de las secuencias aminoacídicas de Cas12k.** Se muestra el alineamiento parcial en la posición 1 – 41, contiene la secuencia de *S. hofmannii* como consenso y muestra la divergencia cualitativa entre secuencias.

Una vez alineadas las secuencias de la proteína Cas12k se construyó el árbol filogenético (Figura 18) para observar las distancias evolutivas dada la divergencia aminoacídica observada.

Es posible observar que la secuencia de Cas12k de *S. hofmannii* se agrupa en un clado único, aun cuando genéticamente se encuentra relacionada con otros organismos del orden Nostocales. Estas variaciones podrían estar relacionadas con la naturaleza efectora de los sistemas CAST encontrados, por lo que se analizó individualmente la estructura de cada locus para observar la presencia o ausencia de elementos. En todos los locus se encontró la presencia de los genes *tnsB* y *tnsC*, sin embargo, existió una variación en algunos casos debido a la presencia de duplicaciones génicas lo cual es relativamente común. En el caso del gen *tniQ*, no fue hallado en todos los locus y en cambio, se identificaron genes *tnsD* que tienen una estructura muy parecida, pero modifica ligeramente el mecanismo de inserción. Algunos de los locus mostraron la presencia de genes *tnsA*, que en algunos operones de la transposasa de Tn7 se encuentran fusionados a *tnsB*. Al comparar el análisis individual de los locus, fue posible agruparlos de acuerdo con su estructura y distancia evolutiva como se muestra en la Figura 18.

Adicionalmente, con el análisis de los locus CAST se encontró la presencia, en la mayoría de los casos, de los genes que componen al sistema RM (restricción/modificación) y *merR*, sin embargo, sólo con la proteína MerR se encontró una interacción directa de acuerdo con el interactoma. De forma interesante, fue posible observar la presencia de duplicaciones en el gen de la proteína Cas12k, con algunas modificaciones en la secuencia de nucleótidos y su distribución en distintos sitios a lo largo del locus, así como una variedad importante de genes cargo que codifican para diversas proteínas, la mayoría factores transcripcionales y enzimas. Los arreglos CRISPR en los locus CAST, tienen una diversidad notable pero como en el resto de los sistemas CRISPR – Cas, se encuentra en la cercanía de la proteína Cas12k.

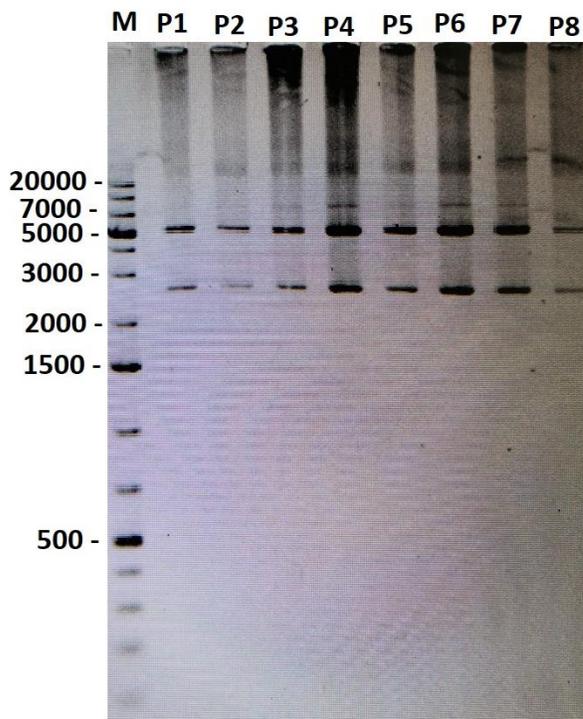


**Figura 18. Árbol filogenético de los locus identificados.** Se muestra el agrupamiento con base en las distancias evolutivas y los elementos que componen a los locus CAST hallados.

### 8.3. Obtención del material genético

#### 8.3.1. pHelper: Plásmido con los genes del sistema CAST

El plásmido pHelper con un tamaño de 7712 pb, fue sujeto a un análisis en cada una de sus regiones mediante digestión enzimática, para verificar su integridad y su identidad de acuerdo con la secuencia reportada por Strecker *et al.* (2019). Como ejemplo de este análisis, la Figura 19 que muestra un gel de electroforesis con los productos de la digestión enzimática del plásmido, empleando las enzimas XhoI y BamHI. Es posible observar la banda que corresponde al plásmido sin digestión como de un tamaño de fragmento de alrededor de 7000 y bandas de un tamaño aproximado de 2500 y 5200 pb, que corresponden a la digestión enzimática del plásmido Figura 19.



**Figura 19. Gel de electroforesis de la digestión enzimática del plásmido pHelper.** Contiene las enzimas XhoI y BamHI. M: marcador de tamaño de fragmento, P1 – P8: diversas digestiones del plásmido pHelper.

### **8.3.2. pDonor: Plásmido con el transposón de Tn7**

Se intentó recuperar el plásmido y propagarlo con los mismos procedimientos utilizados para el plásmido pHelper, sin embargo, al contener un origen de replicación R6K no fue posible propagarlo debido a que requiere una cepa de *E. coli* Pir1+, con la que no se cuenta dentro del laboratorio de trabajo. Para evitar una modificación mayor al plásmido se realizaron pruebas de amplificación de los fragmentos deseados, directamente sobre el plásmido recuperado del papel filtro obteniendo buenos resultados. Sin embargo, se plantea en un futuro emplear una técnica de propagación *in vitro* empleando un protocolo (Anexo 11) mixto que incluye la replicación del plásmido en cadena sencilla utilizando la enzima DNA polimerasa del fago  $\phi$ 29, la síntesis de la cadena complementaria mediante el fragmento largo de Klenow y la ligación final para obtener el plásmido completo. Esta técnica se encuentra en fase de desarrollo en el grupo de trabajo.

### **8.3.3. pTarget: Plásmido Diana (Control +)**

Este plásmido contiene una región de alineamiento para el sgRNA PSP1, que permite evaluar la capacidad del sistema de insertar un fragmento y rastrear fácilmente la eficiencia de inserción y la integridad del fragmento insertado. El plásmido pTarget fue recuperado, propagado y analizado de la misma forma que el plásmido pHelper, teniendo un tamaño de 2899 pb. Este plásmido será utilizado para validar la eficiencia de inserción del sistema CAST, al realizar la adición de un sgRNA con objetivo de reconocer el sitio PSP1.

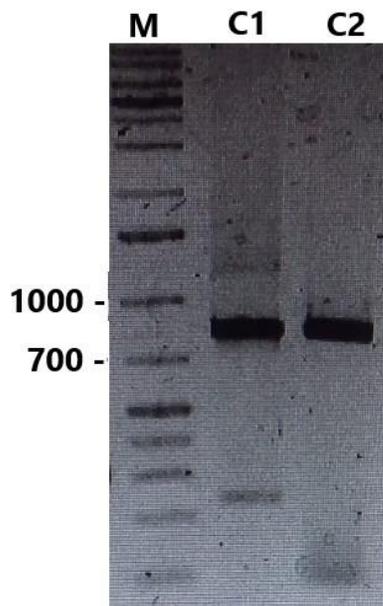
#### **8.3.4. Oligonucleótidos para la amplificación de los elementos del plásmido pSS5001**

Al efectuar las pruebas iniciales y la estandarización de las condiciones óptimas para la PCR, fue necesario hacer la corrección de los iniciadores que amplifican el origen de replicación y el fragmento que contiene el sitio de ensamble del sgRNA, debido a que el par de iniciadores no amplificaban de forma completa el origen de replicación y que existían discrepancias entre la secuenciación reportada por Addgene y el autor original, en el plásmido pHelper. Las pruebas de amplificación con el nuevo par de iniciadores resultaron en la obtención del fragmento deseado.

#### **8.4. Construcción del plásmido pSS5001**

##### **8.4.1. Amplificación de los elementos del plásmido pSS5001**

En la estandarización de las condiciones de amplificación de cada uno de los fragmentos mediante PCR, el cálculo manual de la temperatura de alineamiento ( $T_m$ ) resultó ser eficiente y no sólo un dato de referencia. A pesar de que algunos pares de iniciadores tienen temperaturas de alineamiento diferente, se encontró que la amplificación más eficiente se da considerando la temperatura mayor dentro del par de iniciadores. Inicialmente, todas las reacciones se hicieron en presencia de DMSO, sin embargo, las pruebas sugieren que, para los fragmentos amplificados y las condiciones incluidas en el programa, no es necesario añadirlo y se obtiene un mejor resultado en su ausencia. La Figura 20 muestra un ejemplo de amplificación de uno de los fragmentos. Todas las reacciones se realizaron con éxito, optimizando las condiciones para obtener una única banda de amplificación en el gel de agarosa a través de diversas estrategias.



**Figura 20. Gel de electroforesis con el fragmento CMR.** Muestra la amplificación del fragmento CMR que tiene una longitud de alrededor de 750 nt. M: marcador de tamaño de fragmento, C1 y C2: amplificaciones del fragmento CMR bajo dos condiciones distintas, en presencia y ausencia de DMSO, respectivamente.

#### **8.4.2. Purificación y cuantificación de los amplicones**

Los fragmentos purificados conservaron una buena concentración de material genético y mostraron una mayor limpieza de contaminantes, los cuales son agregados en el proceso de amplificación. A pesar de que se obtuvo una buena limpieza de los amplicones, está pendiente comparar este método con una purificación por diálisis y elegir la metodología más adecuada. La cuantificación de la concentración de cada amplicón se hizo para preparar una mezcla equimolecular en el proceso de ligación de los fragmentos, como está planteado de acuerdo con las recomendaciones en el protocolo de digestión con la enzima Bsal.

### **8.5. Nota aclaratoria**

Los resultados de las siguientes etapas se encuentran pendientes por efectuarse en el tiempo correspondiente a los estudios de doctorado. Esto considerando los objetivos cumplidos en el periodo ordinario de los estudios de maestría y la situación vivida durante 2020 como resultado de la pandemia por Covid – 19. Sin embargo, se incluye la propuesta metodológica para la etapa final del proyecto.

## 9. Discusión

El sistema CAST luce como una atractiva herramienta con la capacidad de inserción para regiones génicas de más de 1 kb, esto en consideración con el tamaño de los transposones presentes en la naturaleza. Este sistema posee la ventaja frente al sistema INTEGRATE de requerir un gasto energético menor en su síntesis, en virtud de un tamaño menor en todos sus elementos lo cual representa también una hipotética expresión mayor. Mientras que, la evidencia sugiere que tiene la misma eficiencia si se realiza la transformación del organismo con los vectores y con las proteínas purificadas, así como una inserción genómica y en DNA plasmídico. En contraste, el sistema INTEGRATE tiene reportes de pruebas exitosas en otros organismos además de *E. coli* y la ventaja de que su estructura de interacción ya se encuentra determinada a través de criomicroscopía electrónica.

Con el análisis de los elementos se encontró que existe la probabilidad de que el sistema CAST esté implicado en una función dentro de la inmunidad adaptativa de *S. hofmannii*, debido a que contiene elementos del sistema R – M, sitios blanco para secuencias víricas y una proteína que contribuiría en el funcionamiento de un elemento de R – M. No obstante, algunos de los espaciadores contienen sitios blanco hipotéticos de inserción dentro del genoma de la cianobacteria en regiones codificantes. Se requiere una validación experimental para determinar si estos sitios son funcionales, sin embargo, podrían también representar una forma de regulación génica o simplemente permiten la propagación el transposón. La hipótesis de la regulación génica estaría acompañada por la presencia de genes que codifican para CHAT, considerando que la expresión de estas caspasas es crucial para la formación de heterocistos, es decir, que participa en el desarrollo morfológico de la cianobacteria.

La información disponible acerca de los sistemas CAST e INTEGRATE indica que ambos requieren de consideraciones para un mejor funcionamiento. Particularmente, el sistema CAST requiere considerar que tiene una probabilidad de coinsertarse, es decir, la inserción no sólo del transposón sino del vector completo que lo contiene. Además, la evidencia experimental propone que se debe poner atención al genotipo del organismo donde se pretende emplear esta herramienta, ya que las deleciones en determinados genes afectan directamente la frecuencia de inserción. En el caso de la adaptación en este trabajo, es probable que se requiera una transformación bacteriana con el plásmido lineal y emplear diversas estrategias, en consideración a una posible interferencia de los sistemas de restricción bacterianos.

A la fecha se cuentan con múltiples reportes de predicciones de sistemas CAST, casposasas o asociaciones hipotéticas de proteínas Cas con transposones con capacidad de llevar a cabo inserción génica sitio específico, sin embargo, sólo han sido exploradas una mínima parte de ellas. Estas predicciones han abierto la posibilidad de realizar aportaciones experimentales debido a que es probable que en un futuro se describan sistemas más eficientes o se desarrollen herramientas basadas en la fusión de proteínas Cas con sistemas de transposición o diversas integrasas. Al inicio de este proyecto se planteó la fusión de una proteína dCas (dominio nucleasa inhibido) con la transposasa de Tn5. Chen & Wang (2019), y Strecker et al. (2019), reportaron fusiones de proteínas dCas con sistemas de transposición, encontrando una inserción sitio específico relativamente eficiente. Con base en estas observaciones, se podría pensar de que el sistema inicial habría operado con una buena eficiencia.

El diseño *in silico* del plásmido pSS5001 contempló todos los elementos necesarios para el funcionamiento del sistema CAST, pero también fueron incluidos elementos extra como el

*oriT*, para facilitar la transferencia del plásmido en caso de enfrentar problemas con la transformación bacteriana. Además, el diseño de los iniciadores obedeció a los requerimientos de la técnica de ensamble múltiple de fragmentos basada en Golden Gate, utilizando incluso los nucleótidos con una mayor afinidad de unión. La revisión de toda la información *in silico* brindó la certidumbre de poder obtener los fragmentos necesarios para ensamblar el plásmido, sin embargo, es probable que se requiera hacer un replanteamiento de algunos iniciadores debido a que amplifican regiones repetidas como los fragmentos FRT y podrían afectar en el orden de los elementos, a pesar de que difieren en la secuencia complementaria que permiten el ensamble de fragmentos. Además, con las pruebas de amplificación efectuadas se encontró en algunos casos la presencia de fragmentos con tamaños diferentes, aún cuando la revisión *in silico* mostrara que no existían otras regiones en los plásmidos donde pudieran alinear. Una explicación podría ser lo que sucedió con el plásmido pDonor que tenía dos secuencias diferentes reportadas; la de los autores y la de Addgene, debido a que se probaron distintas condiciones de reacción y ninguna permitió obtener una amplificación deseada, hasta que se hizo una variante de PCR anidada. Convendría probar quizá con un diseño de oligonucleótidos distinto con el objetivo de favorecer a una construcción más íntegra.

El mapeo de las interacciones de los elementos de CAST permitió identificar asociaciones hipotéticas interesantes, por ejemplo, la predicción de interacción entre Cas12k y MerR, debido a que podría ser clave para la asociación de Cas12k con alguna de las subunidades de la transposasa. No ha sido documentada de forma experimental la estructura de interacción de CAST y la estructura de INTEGRATE que podría representar un indicio de que Cas12k se

asocia con TniQ, sin embargo, con la predicción realizada en este trabajo no se observaron indicios de esta asociación. Una hipótesis es que se requieren elementos extra para la formación del complejo y al encontrar una interacción de MerR y TnsB, la propuesta es que MerR tiene un papel importante en la asociación de Cas12k y la transposasa de Tn7. Otra interacción que resulta interesante es la predicción de TnsB y EndoU, recientemente Michalska *et al.* (2018), publicaron un reporte de cómo EndoU está considerada como una enzima endoribonucleasa que participa en la competencia bacteriana por supervivencia. Sin embargo, este tipo de enzimas no sólo son encontradas en bacterias y podrían contribuir a la eliminación de ácidos nucleicos virales en un evento de respuesta inmune adaptativa. El resto de las interacciones encontradas sugieren que son factores o elementos que participan directa o indirectamente en la formación del complejo.

El análisis filogenético permitió inferir que el sistema CAST de *S. hofmannii* tiene particularidades que lo definen como único. El alineamiento mostró que la proteína Cas12k tienen una divergencia importante con el resto de las secuencias halladas en los diversos locus CAST identificados, no obstante, de manera global se esperaría que cumplieran con una función similar, pero tendrían que ser validados experimentalmente. Además, el análisis individual de cada locus CAST indicó que tienen una composición muy variable de elementos génicos, pero se mantienen similitudes como los genes del sistema R – M o la presencia de *merR*. Es interesante observar que conforme aumentan las distancias evolutivas en el árbol filogenético, también se generan variaciones en la estructura de la transposasa de Tn7. Una perspectiva acerca de este análisis sería la construcción de un árbol filogenético basado en las

distancias evolutivas entre los organismos seleccionados y ver cómo varía la estructura de los elementos del sistema CAST, para comparar si se obtiene un resultado similar.

## **10. Conclusión**

El planteamiento actual del proyecto sugiere que tiene el potencial para llevar a cabo su objetivo, que es la inserción génica sitio específico en *P. chlororaphis*. Tomando en cuenta las consideraciones aportadas por diversos autores, se plantean algunas modificaciones que le brinden una mayor eficiencia y la capacidad de ser universal. Este sistema de forma natural estaría implicado en la respuesta inmune adaptativa de la cianobacteria debido a los elementos que contiene y la disposición del arreglo CRISPR. Aún no se sabe con certeza cuál es el beneficio real de esta asociación, pero con base en la información hallada, podría ser que beneficia a la maquinaria de transposición al brindar una distribución más controlada del transposón. A pesar de que en los estudios experimentales del sistema CAST e INTEGRATE se propone que no se requieren elementos adicionales para su funcionamiento más que los que conforman a esos complejos, las interacciones encontradas en la predicción sugieren que su actividad y/o conformación pueda estar mediada por otras proteínas. Por otra parte, el análisis filogenético brinda un indicio de cómo se encuentran las distancias evolutivas entre los sistemas CAST con respecto a las especies donde están presentes, indicando que el caso de *S. hofmannii* es único. Finalmente, las perspectivas de este proyecto son llevar a cabo la construcción del sistema, su evaluación en distintas líneas celulares y su adecuación para hallar una herramienta eficiente de edición genética de carácter universal.

## **11. Recomendación**

Los resultados de mostrados en este escrito incluyen el cumplimiento parcial de los objetivos planteados en el proyecto original. Considerando la situación que se vive en torno a la pandemia global por Covid – 19, se hizo el replanteamiento de los objetivos y la delimitación del proyecto para cumplir con los estudios de maestría. Sin embargo, la propuesta gira entorno a la conclusión del proyecto durante los estudios de doctorado, complementándolo con aportaciones proyectadas en torno al tema, durante el análisis de la información disponible a la fecha. Este proyecto tiene múltiples características que podrían representar publicaciones potenciales, en virtud de que no existe mucha información acerca del sistema CAST. La recomendación general estriba en la conclusión del proyecto y su enriquecimiento con base en la información que podría ser aportada y que representaría información novedosa.

## 12. Bibliografía

- Anderson, J. A., Staley, J., Challender, M., & Heuton, J. (2018). Safety of *Pseudomonas chlororaphis* as a gene source for genetically modified crops. *Transgenic Research*, 27(1), 103–113. <https://doi.org/10.1007/s11248-018-0061-6>
- Birnboim, H., & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, 7(6), 1513–1523.
- Bondy-denomy, J., Pawluk, A., Maxwell, K. L., & Davidson, A. R. (2016). Bacteriophage genes that inactivate the CRISPR / Cas bacterial immune system. *Nature*, 493(7432), 429–432. <https://doi.org/10.1038/nature11723>. Bacteriophage
- Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Rik, J. H., Snijders, A. P. L., ... Eugene, V. (2018). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, 321(5891), 960–964. <https://doi.org/10.1126/science.1159689>. Small
- Chen, S. P., & Wang, H. H. (2019). An Engineered Cas-Transposon System for Programmable and Site-Directed DNA Transpositions. *The CRISPR Journal*, 2(6), 376–394. <https://doi.org/10.1089/crispr.2019.0030>
- Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A., & Yildiz, A. (2017). A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Science Advances*, (August).
- Deltcheva, E., Chylinski, K., Sharma, C. M., & Gonzales, K. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), 602–607.

<https://doi.org/10.1038/nature09886>.CRISPR

Faure, G., Scott, D. A., & Peters, J. E. (2019). CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nature Reviews Microbiology*, 17(August).  
<https://doi.org/10.1038/s41579-019-0204-7>

Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A., & Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*.  
<https://doi.org/10.1038/nature17945>

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9 – crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS*, 109(39), 2579–2586. <https://doi.org/10.1073/pnas.1208507109>

Gringauz, E., Orle, K. A., Waddell, C. S., & Craig, N. L. (1988). Recognition of Escherichia coli attTn7 by Transposon Tn7 : Lack of Specific Sequence Requirements at the Point of Tn7 Insertion. *Journal of Bacteriology*, 170(6), 2832–2840.

Grissa, I., Vergnaud, G., & Pourcel, C. (2007). CRISPRFinder : a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research*, 35(19), 52–57.  
<https://doi.org/10.1093/nar/gkm360>

Halpin-healy, T. S., Klompe, S. E., Sternberg, S. H., & Tnsa, W. (2019). Structural basis of DNA targeting by a transposon-encoded CRISPR – Cas system. *Nature*, (July).  
<https://doi.org/10.1038/s41586-019-1849-0>

Hickman, A. B., Kailasan, S., Genzor, P., Haase, A. D., & Dyda, F. (2020). Casposase structure

- and the mechanistic link between DNA transposition and spacer acquisition by CRISPR-Cas. *ELife*, 9, 1–26. <https://doi.org/10.7554/eLife.50004>
- Holder, J. W., & Craig, N. L. (2011). Architecture of the Tn7 Post-Transposition Complex: an Elaborate Nucleoprotein Structure. *Journal of Molecular Biology*, 401(2), 167–181. <https://doi.org/10.1016/j.jmb.2010.06.003>.Architecture
- Hudaiberdiev, S., Shmakov, S., Wolf, Y. I., Terns, M. P., Makarova, K. S., & Koonin, E. V. (2017). Phylogenomics of Cas4 family nucleases. *BMC Evolutionary Biology*, 1–14. <https://doi.org/10.1186/s12862-017-1081-1>
- Hur, J., Olovnikov, I., & Aravin, A. (2015). Prokaryotic Argonautes defend genomes against invasive DNA. *Trends in Biochemistry*, 39(6), 257–259. <https://doi.org/10.1016/j.tibs.2014.04.006>.Prokaryotic
- Iwasaki, K., Uchiyama, H., Yagi, O., Kurabayashi, T., Ishizuka, K., & Takamura, Y. (1994). Transformation of *Pseudomonas putida* by Electroporation. *Bioscience, Biotechnology, and Biochemistry*, 58(5), 851–854. <https://doi.org/10.1271/bbb.58.851>
- Jansen, R., Van Embden, J. D. A., Gastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, 43(6), 1565–1575. <https://doi.org/10.1046/j.1365-2958.2002.02839.x>
- Jiang, F., Zhou, K., Gressel, S., & Doudna, J. A. (2015). A Cas9 – guide RNA complex preorganized for target DNA recognition. *Structural Biology*, 348(6242), 1477–1482.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., & Jennifer, A. (2012). A programmable dual RNA-

- guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096), 816–821.
- Jore, M. M., Lundgren, M., Duijn, E. Van, Bultema, J. B., Westra, E. R., Waghmare, S. P., ... Brouns, S. J. J. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Publishing Group*, 18(5), 529–536. <https://doi.org/10.1038/nsmb.2019>
- Kapitonov, V. V, & Jurka, J. (2006). Self-synthesizing DNA transposons in eukaryotes. *PNAS*, 103(12).
- Klompe, S. E., Vo, P. L. H., Halpin-healy, T. S., & Sternberg, S. H. (n.d.). Transposon-encoded CRISPR – Cas systems. *Nature*, 21. <https://doi.org/10.1038/s41586-019-1323-z>
- Komor, A. C., Badran, A. H., & Liu, D. R. (2018). CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell*, 168, 20–36. <https://doi.org/10.1016/j.cell.2016.10.044>.CRISPR-based
- Koonin, E. V, & Krupovic, M. (2018). Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Natural Reviews in Genetics*, 16(3), 184–192. <https://doi.org/10.1038/nrg3859>.Evolution
- Koonin, E. V, Makarova, K. S., & Koonin, E. V. (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions B*.
- Krupovic, M., Béguin, P., & Koonin, E. (2017). Casposons : mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Current Opinion in Microbiology*. <https://doi.org/10.1016/j.mib.2017.04.004>
- Krupovic, M., & Koonin, E. V. (2016). Self-synthesizing transposons : unexpected key players in

- the evolution of viruses and defense systems. *Current Opinion in Microbiology*, 31, 25–33. <https://doi.org/10.1016/j.mib.2016.01.006>
- Kupferschmied, P., Maurhofer, M., & Keel, C. (2013). Promise for plant pest control : root-associated pseudomonads with insecticidal activities. *Frontiers in Plant Science*, 4(July), 1–17. <https://doi.org/10.3389/fpls.2013.00287>
- Luo, W., Galvan, D. L., Woodard, L. E., Dorset, D., Levy, S., & Wilson, M. H. (2017). Comparative analysis of chimeric ZFP- , TALE- and Cas9-piggyBac transposases for integration into a single locus in human cells. *Nucleic Acids Research*, 45(14), 8411–8422. <https://doi.org/10.1093/nar/gkx572>
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., ... Koonin, E. V. (2015). An updated evolutionary classification. *Nature Reviews Microbiology*, (September), 1–15. <https://doi.org/10.1038/nrmicro3569>
- Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2009). Comprehensive comparative-genomic analysis of Type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *BioMed Central*, 38. <https://doi.org/10.1186/1745-6150-4-19>
- Mendes, R. (2013). Deciphering the Rhizosphere Microbiome for Disease - Suppressive Bacteria. *Science*, 1097(2011). <https://doi.org/10.1126/science.1203980>
- Mulepati, S., Bailey, S., & Directorate, S. (2015). Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science*, 345(6203), 1479–1484. <https://doi.org/10.1126/science.1256996>.Crystal

- Murray, N. E. (2000). Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle). *Microbiology and Molecular Biology Reviews*, 64(2), 412–434.  
<https://doi.org/10.1128/membr.64.2.412-434.2000>
- Peters, J. E. (2014). Tn7. *Microbiology Spectrum*, 1–20.  
<https://doi.org/10.1128/microbiolspec.MDNA3-0010-2014.f1>
- Peters, J. E., Makarova, K. S., Shmakov, S., & Koonin, E. V. (2017). Recruitment of CRISPR-Cas systems by Tn7-like transposons. *PNAS*, 7358–7366.  
<https://doi.org/10.1073/pnas.1709035114>
- Pourcel, C., Salvignol, G., & 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*, 653–663. <https://doi.org/10.1099/mic.0.27437-0>
- Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., & Bhat, P. (2015). Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR-Cas system. *Cell*, 163(4), 854–865.
- Reznikojf, S. (1993). The tn5 transposon. *Annual Reviews Microbiology*, 945–963.
- Richter, F., Fonfara, I., Gelfert, R., Nack, J., Charpentier, E., & Mo, A. (2017). Switchable Cas9. *Current Opinion in Microbiology*.
- Risser, D. D., & Callahan, S. M. (2008). HetF and PatA control levels of HetR in *Anabaena* sp. strain PCC 7120. *Journal of Bacteriology*, 190(23), 7645–7654.  
<https://doi.org/10.1128/JB.01110-08>

- Sarnovsky, R. J., May, E. W., & Craig, N. L. (1996). The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *The EMBO Journal*, *15*(22), 6348–6361.
- Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Jonathan, S., Semenova, E., ... Koonin, E. V. (2016). CRISPR-Cas systems. *Molecular Cell*, *60*(3), 385–397. <https://doi.org/10.1016/j.molcel.2015.10.008>.Discovery
- Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., ... Koonin, E. (2018). Diversity and evolution of class 2 CRISPR–Cas systems. *Natural Reviews in Microbiology*, *15*(3), 169–182. <https://doi.org/10.1038/nrmicro.2016.184>.Diversity
- Simon, R., Prieffer, U., & Hüler, A. (1983). A broad host range mobilization system for in vivo genetic engineering: Transposon mutagenesis in gram negative bacteria. *Biotechnology*, (1), 784–790.
- Sternberg, S. H., Lafrance, B., Kaplan, M., & Doudna, J. A. (2015). Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature*, 1–14. <https://doi.org/10.1038/nature15544>
- Tanner, W. D., Atkinson, R. M., Goel, R. K., Toleman, M. A., Benson, L. S., Porucznik, C. A., & VanDerslice, J. A. (2017). Horizontal transfer of the bla<sub>NDM-1</sub> gene to *Pseudomonas aeruginosa* and *Acinetobacter baumannii* in biofilms. *FEMS Microbiology Letters*, *364*(8), 22–25. <https://doi.org/10.1093/femsle/fnx048>
- Velivelli, S. L. S., Vos, P. De, Kromann, P., Declerck, S., & Prestwich, B. D. (2014). Biological control agents : from field to market , problems , and challenges. *Trends in Biotechnology*,

32(10), 493–496. <https://doi.org/10.1016/j.tibtech.2014.07.002>

Wang, T. Y., Wang, L., Zhang, J. H., & Dong, W. H. (2011). A simplified universal genomic DNA extraction protocol suitable for PCR. *Genetics and Molecular Research : GMR*, 10(1), 519–525. <https://doi.org/10.4238/vol10-1gmr1055>

Wolkow, C., Deboy, R. T., & Craig, N. L. (1996). Conjugating plasmids are preferred targets for Tn7. *Genes and Development*, 2145–2157.

Wright, A. V., & Doudna, J. A. (2016). Protecting genome integrity during CRISPR immune adaptation. *Nature Structural & Molecular Biology*, (September), 1–9. <https://doi.org/10.1038/nsmb.3289>

Xiao, Y., Ding, F., Erp, P. B. G. Van, Rajashankar, K., Bailey, S., & Wiedenheft, B. (2016). Structural basis for promiscuous PAM recognition in type I–E Cascade from *E. coli*. *Nature*, 1–16. <https://doi.org/10.1038/nature16995>

Xiao, Y., Luo, M., Hayes, R. P., Kim, J., Ng, S., Ding, F., ... Ke, A. (2018). Structure basis for directional R-loop formation and substrate handover mechanisms in Type I CRISPR-Cas system. *Cell*, 170(1), 48–60. <https://doi.org/10.1016/j.cell.2017.06.012>.Structure

Yamada, M., Watanabe, Y., Gootenberg, J. S., Yamada, M., Watanabe, Y., Gootenberg, J. S., ... Nakane, T. (2017). Crystal Structure of the Minimal Cas9 from *Campylobacter jejuni* Reveals the Molecular Diversity in the CRISPR-Cas9 *Molecular Cell*, 65(6), 1109–1121.e3. <https://doi.org/10.1016/j.molcel.2017.02.007>

Yamaguchi, Y., & Inouye, M. (2009). *Sequence - Specific Endoribonucleases from the Toxin –*

*Antitoxin Systems I. Introduction A. Toxin – Antitoxin ( TA ) Systems in Bacteria. Progress in Molecular Biology and Translational Science* (1st ed., Vol. 85). Elsevier Inc.  
[https://doi.org/10.1016/S0079-6603\(08\)00812-X](https://doi.org/10.1016/S0079-6603(08)00812-X)

Yutin, N., Bäckström, D., Ettema, T. J. G., Krupovic, M., & Koonin, E. V. (2018). Vast diversity of prokaryotic virus genomes encoding double jelly-roll major capsid proteins uncovered by genomic and metagenomic sequence analysis. *Virology Journal*, 1–17.

Zhang, Y., Sun, X., Wang, Q., Xu, J., Dong, F., Yang, S., ... Yang, S. (2020). Multicopy Chromosomal Integration Using CRISPR-Associated Transposases. *ACS Synthetic Biology*, 9(8), 1998–2008. <https://doi.org/10.1021/acssynbio.0c00073>

## **13. Anexos**

### **13.1. Anexo 1**

#### **Recuperación de DNA a partir de un liofilizado en papel filtro**

1. Tomar el papel filtro con unas pinzas estériles y recortar la sección que contiene el liofilizado con unas tijeras o bisturí también estériles. Trabajar en mechero.
2. Colocar el papel en un tubo con un volumen máximo de 50  $\mu$ l de agua inyectable estéril y realizar un pipeteo en forma de lavado.
3. Incubar el tubo durante 15 – 20 minutos a 65°C en agitación de 650 rpm.
4. Hacer un pipeteo en forma de lavado durante 30 segundos.
4. Centrifugar en minicentrífuga para separar el papel y utilizar la fase líquida para transformar células competentes.
5. Conservar el tubo en congelación a -20°C hasta por seis meses según la recomendación de Addgene.

\*Protocolo desarrollado en el laboratorio 50

## 13.2. Anexo 2

### **Transformación química de células competentes por choque térmico**

1. Tomar un vial de células competentes conservadas en ultracongelación y mantener la cadena de frío, colocando el tubo en hielo.
2. Añadir un volumen de 2 a 5  $\mu\text{l}$  de DNA plasmídico, dependiendo de la concentración podría ser necesaria una dilución.
3. Incubar en hielo durante 5 – 30 minutos, procurando que el tubo se encuentre completamente inmerso.
4. Transcurrido el periodo de incubación, inmediatamente incubar durante 30 segundos a 42°C.
5. Regresar el tubo al hielo, teniendo cuidado con la cadena de frío.
6. Añadir 200 – 500  $\mu\text{l}$  de medio LB y realizar un pipeteo suave.
7. Incubar el tubo durante 1 h a 37 °C en agitación de 200 rpm.
8. Sembrar en placa tomando tres volúmenes diferentes: 20  $\mu\text{l}$ , 50  $\mu\text{l}$  y el resto del tubo, para tener colonias transformante visibles y fácilmente seleccionables.
9. Observar el crecimiento y seleccionar colonias transformantes transcurridas 12 h, para efectuar una resiembra en placa.

### 13.3. Anexo 3

#### Extracción de DNA plasmídico mediante técnica modificada de Birnboim y Doly

<b>Solución I BD<sup>a</sup>:</b>	volumen (ml)	concentración final
Tris·HCl (pH 8.0) 1 M	1.25	25 mM
EDTA 0.25 M	2.00	10 mM
Glucosa 1 M	2.50	50 mM
Agua destilada c.b.p.	50.00	
RNasa (10 mg/ml) <sup>b</sup>	0.5	0.1 mg/ml

<sup>a</sup> **BD**= Birnboim y Doly.

<sup>b</sup> Añadir la RNasa sólo para hacer las preparaciones en pequeña escala (Minipreps).  
Guardar en refrigeración.

#### **Solución II BD:**

NaOH 1 N	5.00	0.2 N
SDS al 10 %	2.50	1.0 %
Agua destilada c.b.p.	25.00	

Guardar a temperatura ambiente no más de dos meses, si se ve floculada, la solución ya no sirve.

### **Solución III BD:**

Acetato de potasio 5 M	30.00
Acido acético glacial	5.75
Agua destilada c.b.p.	50.00

Guardar en refrigeración no más de seis meses, el pH debe permanecer aproximadamente en 4.6, cerrar perfectamente el frasco para evitar que se eleve el pH.

### **Procedimientos:**

#### **1. Preparación a pequeña escala.**

- a. Obtener un paquete celular de la bacteria con el plásmido que se quiere purificar y resuspenderlo en 150  $\mu$ l de solución I BD:
  - a partir de medio líquido, colocar en un tubo para microcentrífuga, 1.5 ml de un cultivo de toda la noche en el medio selectivo apropiado, centrifugar 1 min en una microcentrífuga, eliminar el sobrenadante, agregar 150  $\mu$ l de solución I BD y agitar hasta que se resuspenda el paquete celular.
  - a partir de un cultivo de toda la noche en medio selectivo sólido, tomar una asada gruesa y resuspenderla en 150  $\mu$ l de solución I BD, en un tubo para microcentrífuga.
- b. Agregar 300  $\mu$ l de la solución II BD y agitar en agitador orbital por 1 ó 2 seg y dejar reposar 5 min en un baño de hielo.
- c. Agregar 300  $\mu$ l de la solución III BD, agitar por inversión unas 100 veces para mezclar

perfectamente y dejar reposar en el baño de hielo por 10 min.

- d.** Centrifugar por 10 min (de preferencia a 4°C) y transferir el sobrenadante a un tubo para microcentrífuga nuevo y estéril.
- e.** Agregar 700 µl de isopropanol, agitar bruscamente por inversión y dejar reposar por 10 a 15 min en el baño de hielo o bien, por 20 min a -20°C.
- f.** Centrifugar por 15 min (de preferencia a 4°C) y eliminar el sobrenadante.
- g.** Para eliminar el exceso de sales agregar 1 ml de etanol al 70 %, agitar en agitador orbital por 30 seg y centrifugar nuevamente por 5 min, eliminar el sobrenadante.
- h.** Secar la pastilla calentando a 56 ó 65°C el tiempo requerido y resuspenderla con 40 µl de agua bidestilada estéril.

El rendimiento es entre 2 y 40 µg de plásmido por preparación, dependiendo del plásmido.

\*Protocolo modificado de Birnboim & Doly (1979).

#### 13.4. Anexo 4

### Protocolo de PCR para la DNA polimerasa de alta fidelidad del kit Phusion™ de Thermo Scientific

Componente	Volumen
DNA polimerasa Phusion™, 2U/μl	100 U – 50 μl
5X Phusion™ HF Buffer	2 x 1.5 ml
5X Phusion™ GC Buffer	1.5 ml
Solución de MgCl <sub>2</sub> 50 mM	1.5 ml
DMSO	500 μl

1. Prepare la mezcla de reacción añadiendo los siguientes componentes en el orden que se muestra en la siguiente Tabla. Es importante mencionar que la DNA polimerasa debe ser el último componente añadido a la mezcla debido a que la enzima exhibe actividad exonucleasa 3' → 5' y podría degradar los iniciadores en ausencia de dNTPs.

<i>Componente</i>	<i>Reacción típica de 20 μl</i>	<i>Concentración final</i>
<i>H<sub>2</sub>O</i>	12.4 μl	
<i>5X Phusion™ HF Buffer</i>	4 μl	1X
<i>10 mM dNTPs</i>	0,4 μl	200 μM
<i>Iniciador forward</i>	1 μl	0.5 μM
<i>Iniciador reverse</i>	1 μl	0.5 μM
<i>DNA muestra</i>	1 μl	
<i>DNA polimerasa Phusion™</i>	0.2 μl	0.02U/ μl

2. Programar el termociclador de acuerdo con los siguientes parámetros

ETAPA	PARÁMETROS		CICLOS
	Temperatura	Tiempo	
DESNATURALIZACIÓN INICIAL	98°C	30 seg	1
DESNATURALIZACIÓN ALINEAMIENTO	98°C	5 – 10 seg	25 – 35
EXTENSIÓN	X°C	10 – 30 seg	
EXTENSIÓN FINAL	72°C	15 – 30 s/kb	1
ALMACENAMIENTO	72°C	5 – 10 min	
	4°C	Mantener	

### 13.5. Anexo 5

#### Diálisis para remoción de sales en muestras de DNA

1. Tomar un filtro Millipore™ de 0.024  $\mu\text{m}$  de diámetro de poro previamente esterilizado con unas pinzas limpias y estériles, trabajar junto al mechero.
2. Colocar el filtro Millipore™ en una caja Petri estéril con agua inyectable estéril, con la cara brillante hacia arriba y cuidando que quede en la superficie.
3. Con una pipeta, colocar el DNA en el centro del papel filtro y esperar 5 - 20 min.
4. Recuperar la muestra y colocarla en un tubo limpio y estéril.
5. Evaluar el DNA cualitativamente a través de electroforesis en gel de agarosa.

\*Protocolo desarrollado en el laboratorio 50.

### 13.6. Anexo 6

#### Purificación de DNA mediante kit de Zymo™ Research

Componente	Cantidad	Almacenamiento
DNA Binding Buffer	50 ml	
DNA Wash Buffer	6 ml	
DNA Elution Buffer	1 ml	Temperatura ambiente
Zymo-Spin™ Columns	50	
Collection Tubes	50	

1. Agrega 2 – 7 volúmenes de DNA Binding Buffer a cada muestra de DNA. Mezcla por vórtex.
2. Transfiere la mezcla a una columna Zymo – Spin™ ensamblada con el tubo de recolección.
3. Centrifuga durante 30 segundos y descarta el líquido colectado.
4. Agrega 200 µl DNA Wash Buffer a la columna. Centrifuga durante 30 segundos. Repita el paso de lavado.
5. Agrega al menos 6 µl de DNA Elution Buffer o agua, directamente a la columna e incuba durante un minuto a temperatura ambiente. Transfiere la columna a un tubo de 1.5 ml y centrifuga durante 30 segundos para eluir el DNA.
6. Analiza la pureza del DNA por espectrofotometría y mediante electroforesis en gel de agarosa.

### 13.7. Anexo 7

#### Ensamble de fragmentos por técnica de Golden Gate de NEB™

1. Cuantificar y purificar los amplicones para calcular su concentración relativa. Hacer una dilución de ser necesario, para facilitar su mezclado.
2. En un tubo apto para PCR, hacer la mezcla de los amplicones en concentraciones equimoleculares.
3. Añadir los siguientes componentes.

<b>Componentes</b>	<b>Volumen</b>
<i>Agua inyectable estéril</i>	Ajustar a 50 µl
<i>Bsal</i>	3 µl
<i>Ligasa T4</i>	2.5 µl
<i>Buffer Cutsmart</i>	4 µl
<i>ATP</i>	2 µl

4. Colocar la mezcla de reacción en el termociclador e introducir el siguiente programa: 134 ciclos alternando 5 min a 37°C y 5 min a 28°C.
5. Transformar células competentes de *E. coli* y seleccionar colonias.
6. Extracción y observar el DNA plasmídico por electroforesis en gel de agarosa.

\*Protocolo adaptado de forma experimental en el laboratorio 50, considerando las recomendaciones de NEB™

## 13.8. Anexo 8

### Electroporación en células de *Pseudomonas*

#### Preparación de células competentes

1. Crecer las células en medio LB hasta alcanzar una  $A_{600} = 0.26 - 0.38$ . Colocar las células en hielo y centrifugar para eliminar el medio.
2. Lavar dos veces con 300 – 500  $\mu\text{l}$  de una solución estéril de sacarosa 300 mM. Resuspender las células en la misma solución y conservar en congelación.

#### Electroporación

1. Colocar 100  $\mu\text{l}$  de células en una celda estéril de electroporación y agregar el DNA plasmídico (de acuerdo con la concentración añadir entre 2 – 5  $\mu\text{l}$ ). Incubar la celda en hielo durante 10 min.
3. Colocar la celda en el electroporador y ajustar los siguientes parámetros: voltaje 2.5 kV (12.5kV/cm), capacitor 25  $\mu\text{F}$  y el controlador de pulso 200  $\Omega$ .
4. Inmediatamente después de la descarga, agregar 900  $\mu\text{l}$  de medio LB. Transferir el contenido de la celda a un tubo de 1.5 ml e incubar durante 2 hr a 28 °C en agitación a 140 rpm.
5. Sembrar en placa a distintos volúmenes de células transformadas (50  $\mu\text{l}$ , 100  $\mu\text{l}$  y el resto).
6. Seleccionar colonias y realizar la resiembra.

\*Protocolo adaptado de Iwasaki *et al.* (1994).

### 13.9. Anexo 9

#### Transformación bacteriana por conjugación

1. Cultivar la cepa donante (\*) con el plásmido a transferir y la cepa receptora (*P. chlororaphis*), en medio LB durante toda la noche, utilizando un marcador de selección apropiado.
2. Tomar 1 ml cada cultivo y centrifugar para eliminar el medio.
3. Lavar dos veces con PBS, para eliminar el medio residual y el agente selectivo. Resuspender en 500 µl de PBS.
4. En un tubo limpio y estéril, colocar 50 µl de cada cepa (donadora y receptora) y mezclar por inversión.
5. Sembrar en placa por cuadrantes 25 µl de la mezcla sin esparcir en la superficie. Utilizar medio LB sin marcadores de selección. Incubar durante toda la noche a 30°C.
6. Tomar con un asa todo el paquete celular de cada cuadrante y resuspender en 1 ml de PBS. Lavar dos veces con PBS y resuspender en 1 ml de PBS.
7. Sembrar en una placa con medio selectivo 50 µl de distintas diluciones (1:1, 1:10 y 1:100) de cada uno de los tubos.
8. Observar el crecimiento y monitorear la transferencia del plásmido a través de PCR colony.

\*Protocolo adaptado de Tanner *et al.* (2017).

### 13.10. Anexo 10

#### Extracción de DNA genómico

##### Buffer de extracción

1. Agrega 24 g de tioscianato de guanidina (GuSCN) en 20 ml de Tris·HCl 0.1 M, pH 6.4, en un tubo Falcon de 50 ml.
2. Calienta a 60°C para disolver el GuSCN.
3. Agrega 4.4 ml de EDTA 0.2 M, pH 8.0.
4. Agrega 0.5 ml de Triton X-100 y mezcla por inversión.

##### Buffer de unión

1. Mezcla 24 g de GuSCN y 20 ml de Tris·HCl 0.1 M, pH 6.4, en un tubo Falcon de 50 ml.
2. Calienta a 60°C para disolver el GuSCN.
3. Agrega 4.4 ml de EDTA 0.2 M, pH 8.0 y mezcla por inversión.

##### Buffer de lavado

Componente	Concentración final
EtOH	70%
Buffer NaCl TE	10 mM
Tris·HCl pH 8.0	10 mM
EDTA pH 8.0	1 mM

## Protocolo

1. Agrega 1 ml de buffer de extracción a  $10^5 - 10^8$  de células, agitar vigorosamente con la pipeta o vórtex.
2. Centrifuga por 10 min a 13,000 g, 4 – 25°C
3. Recupera el sobrenadante (fase superior viscosa) y coloca en un tubo limpio y estéril.
4. Añade 500  $\mu$ l de buffer de unión y mezcla por inversión.
5. Transfiere la mezcla a una columna de afinidad y centrifuga a 12,000 rpm, 30 seg. Descarta el líquido del tubo.
6. Agrega 600  $\mu$ l de buffer de lavado a la columna y centrifuga a 12,000 rpm, 30 seg. Descarta el líquido y repite el lavado.
7. Centrifuga durante 1 min a 12,000 rpm y transfiere la columna a un tubo limpio y estéril.
8. Agrega 10 – 200  $\mu$ l de buffer TE o agua inyectable estéril e incuba durante 1 min a temperatura ambiente.
9. Centrifuga a 12,000 g durante 1 min y conserva el líquido en el tubo.
10. Observa la concentración del DNA por electroforesis y espectrofotometría.

\*Protocolo adaptado de Wang *et al.* (2011).

### 13.11. Anexo 11

#### Replicación *in vitro* de plásmidos

##### Reactivos

Componente	Inactivación
Buffer Cutsmart	
DTT	
DNA polimerasa I (Klenow)	10 mM de EDTA, 20 min a 75 °C
DNA polimerasa de $\phi$ 29	10 min a 65 °C
dNTPs	
T4 DNA ligasa	10 min a 65 °C
ATP	
Oligonucleótido 1	
Oligonucleótido 2	

1. Mezclar los siguientes componentes:

	Volumen	Concentración final
Agua inyectable	Ajustar a 50 $\mu$ l	
Buffer Cutsmart	5 $\mu$ l	1x
DTT 10 mM	0.02 $\mu$ l	4 $\mu$ M
dNTPs 10 mM	0.165 $\mu$ l	33 $\mu$ M
DNA polimerasa de $\phi$ 29	1.5 $\mu$ l	1 U por $\mu$ g de DNA
DNA plasmídico	X	X
Oligonucleótido 1	2.5 $\mu$ l	0.5 $\mu$ M

2. Incubar durante toda la noche a 30°C o temperatura ambiente.

3. Inactivar la enzima a 65 °C durante 10 min.

4. Agregar 1.5  $\mu$ l de DNA polimerasa I (Klenow) y 2.5  $\mu$ l del oligonucleótido 2, e incubar por 2 h a 25°C.

5. Inactivar la enzima calentando a 65°C durante 20 min.

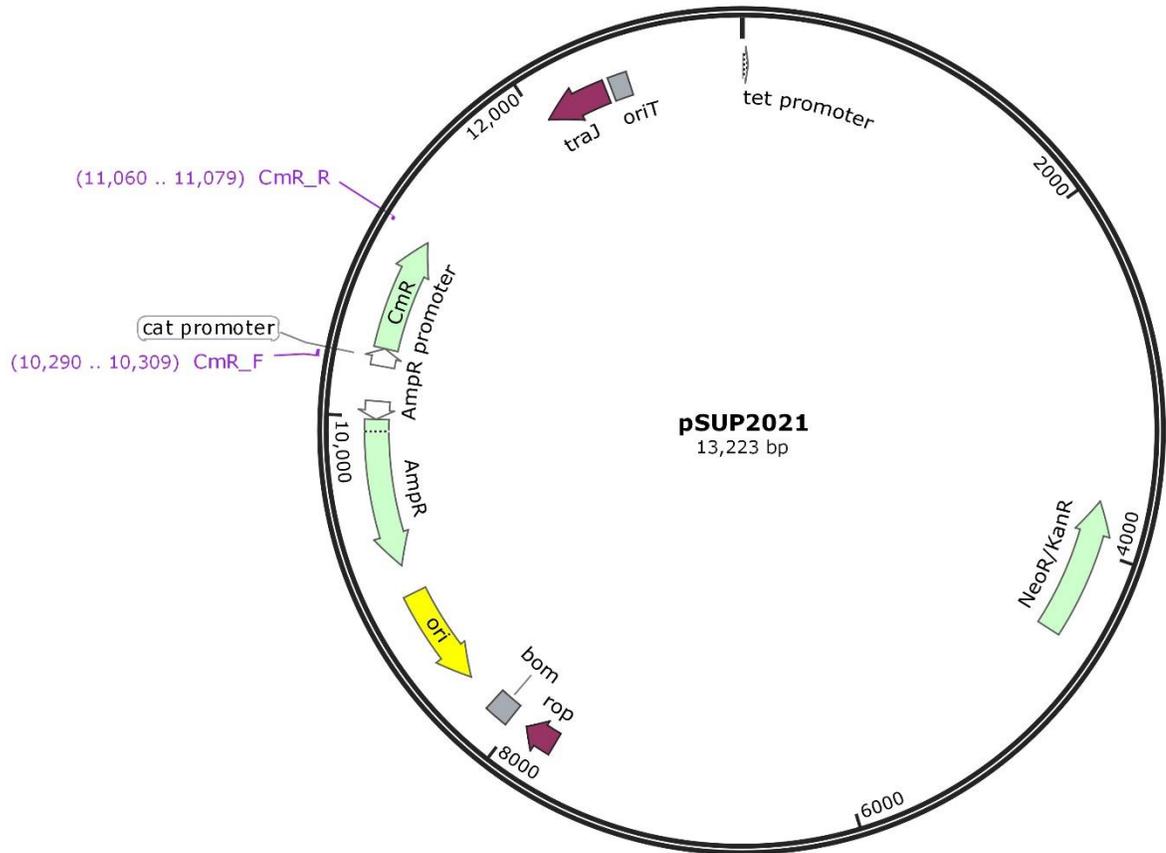
6. Agregar 1.5  $\mu$ l de ligasa T4 y 2  $\mu$ l de ATP. Incubar durante toda la noche a 37°C.

7. Inactivar la ligasa T4 a 65°C a durante 20 min. Agregar la enzima de restricción adecuada para generar un corte único en el plásmido y seguir las recomendaciones del fabricante. Opcional: si se requiere un cambio de buffer, se puede precipitar el DNA con cloroformo/isopropanol/etanol y realizar la mezcla de reacción en un tubo limpio y estéril.
8. Inactivar la enzima de restricción y agregar 2.5  $\mu$ l de ligasa T4 y 2  $\mu$ l de ATP. Incubar durante toda la noche a 37°C. Opcional: si se empleó una enzima de restricción que requirió un cambio de buffer, se requiere precipitar el DNA antes de agregar la ligasa T4 y ajustar las condiciones para la operación de esta enzima.
9. Evaluar la amplificación del plásmido mediante electroforesis y a través de un análisis de restricción.

\*Protocolo desarrollado en el Laboratorio 50.

13.12. Anexo 12

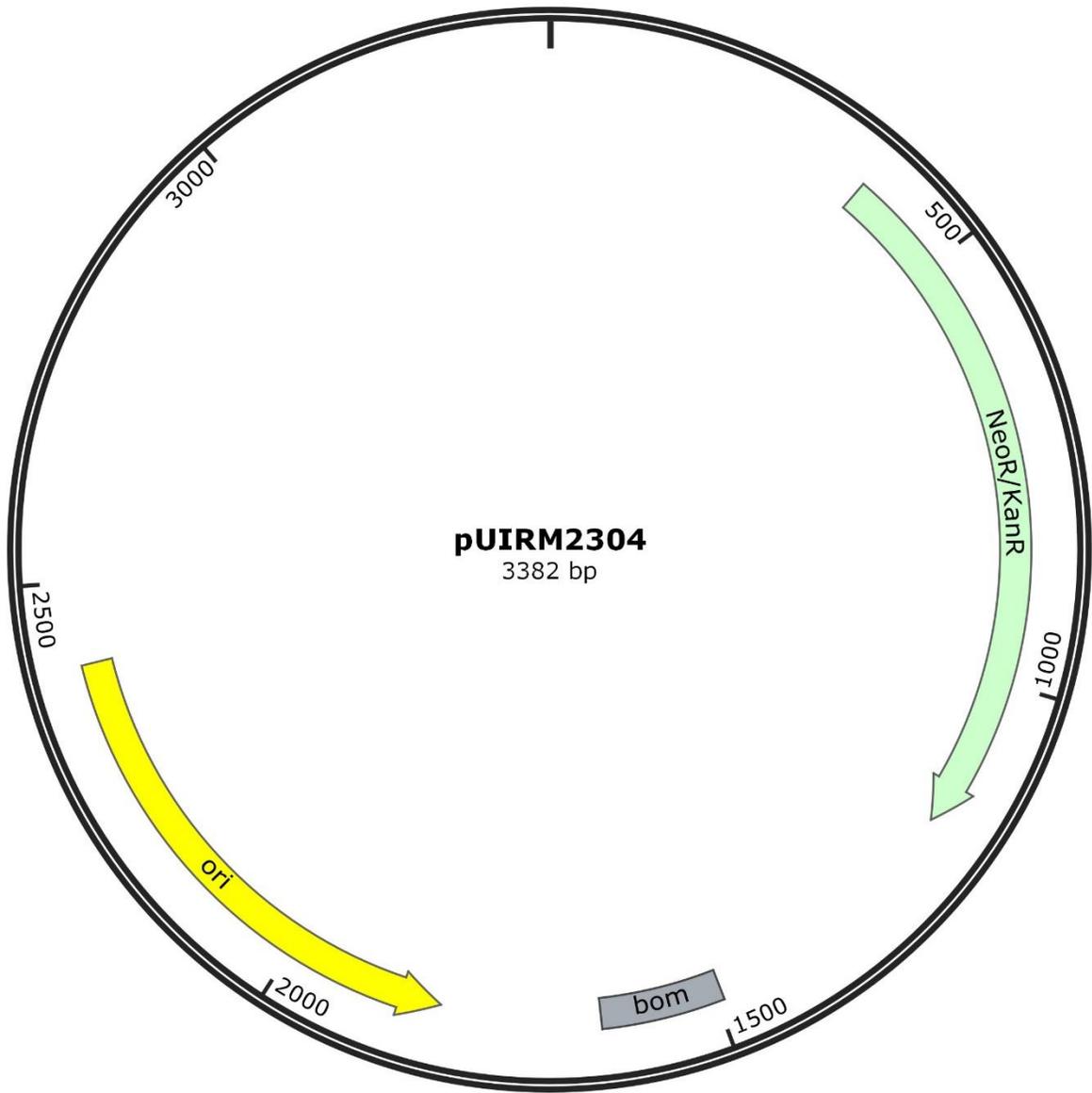
Mapa del plásmido pSUP2021



\*Desarrollado por Simon *et al.* (1983).

13.13. Anexo 13

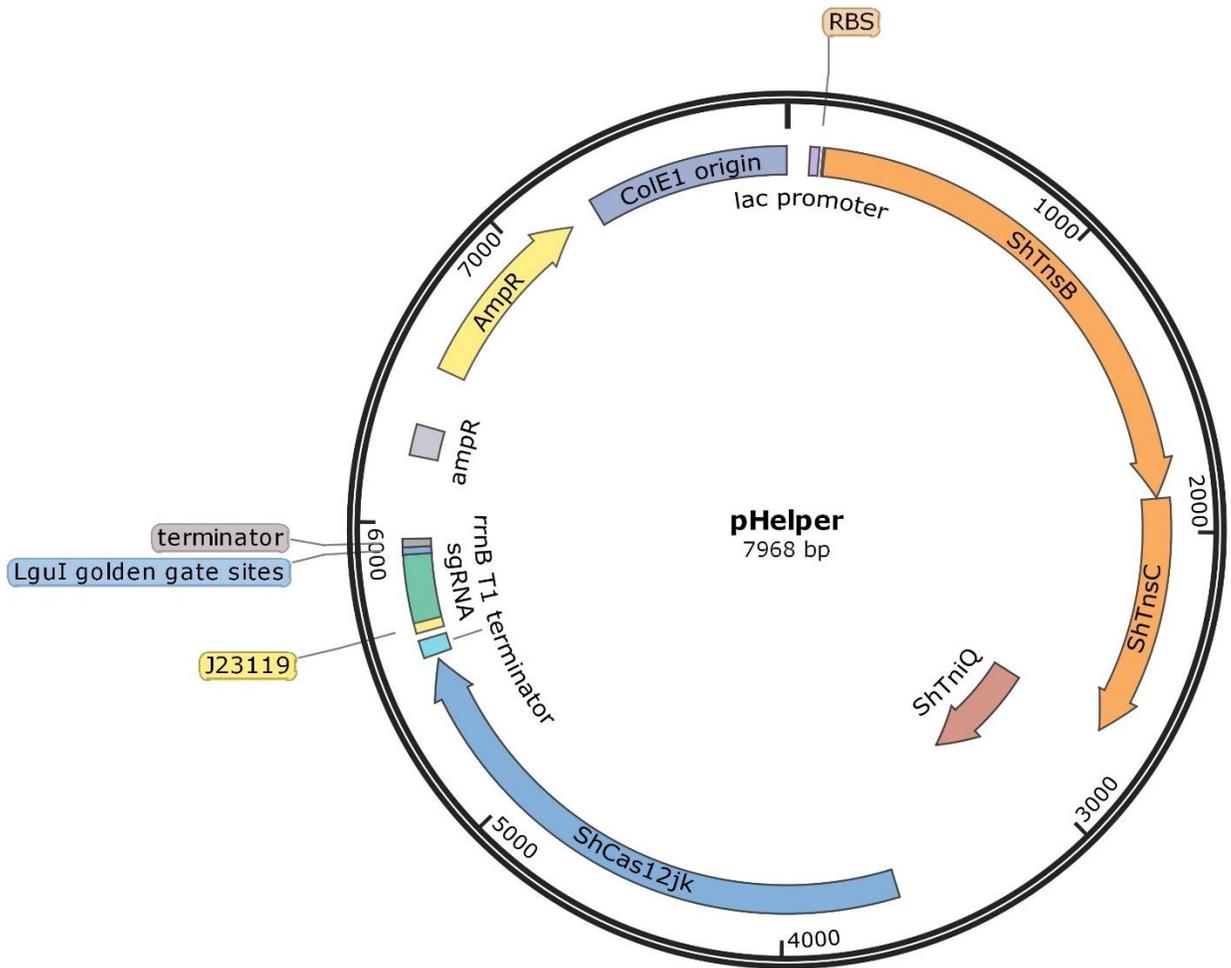
Mapa del plásmido pUIRM2304



\*Construido por el Dr. Rodolfo Marsch Moreno

13.14. Anexo 14

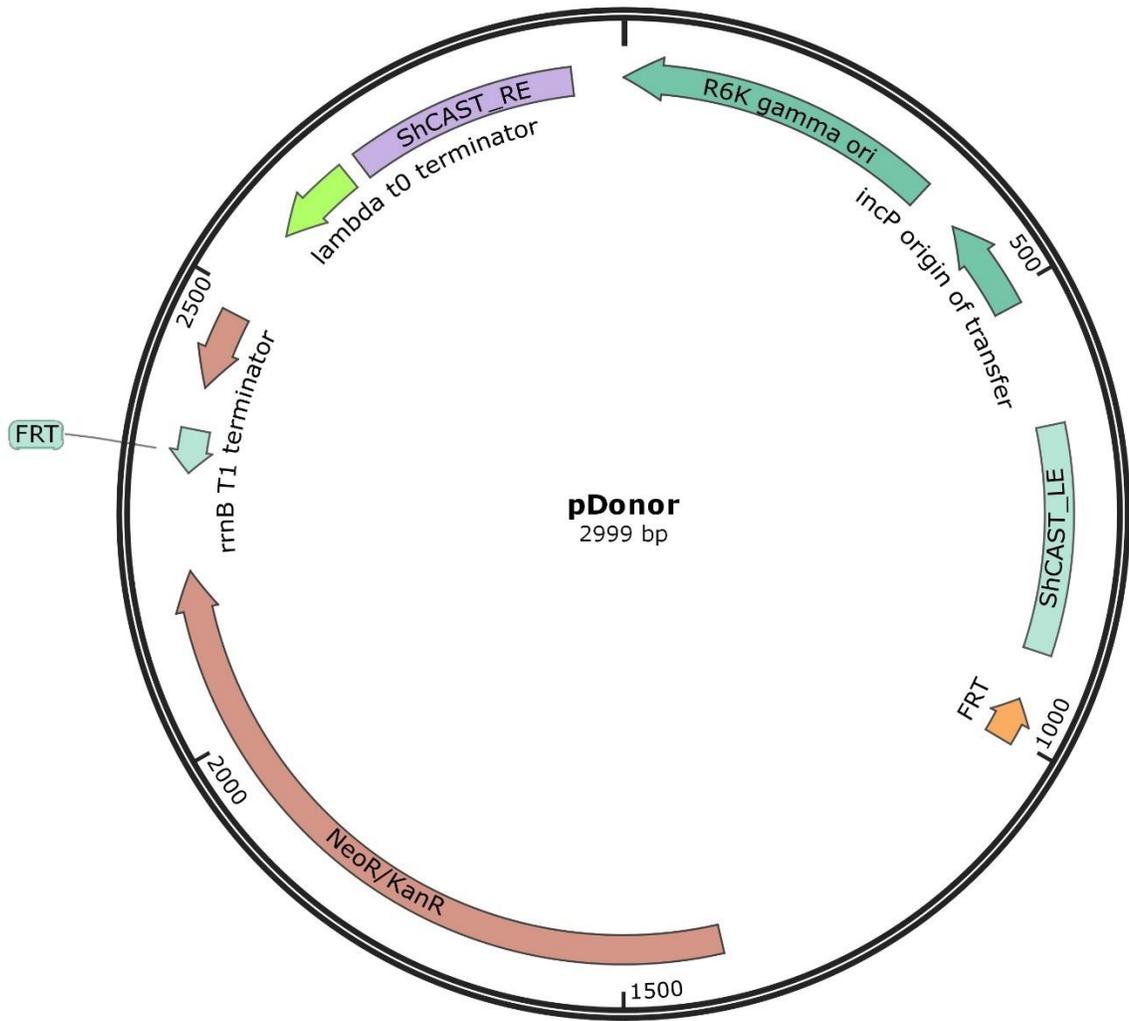
Mapa del plásmido pHelper



\* Construido por Strecker *et al.* (2019).

13.15. Anexo 15

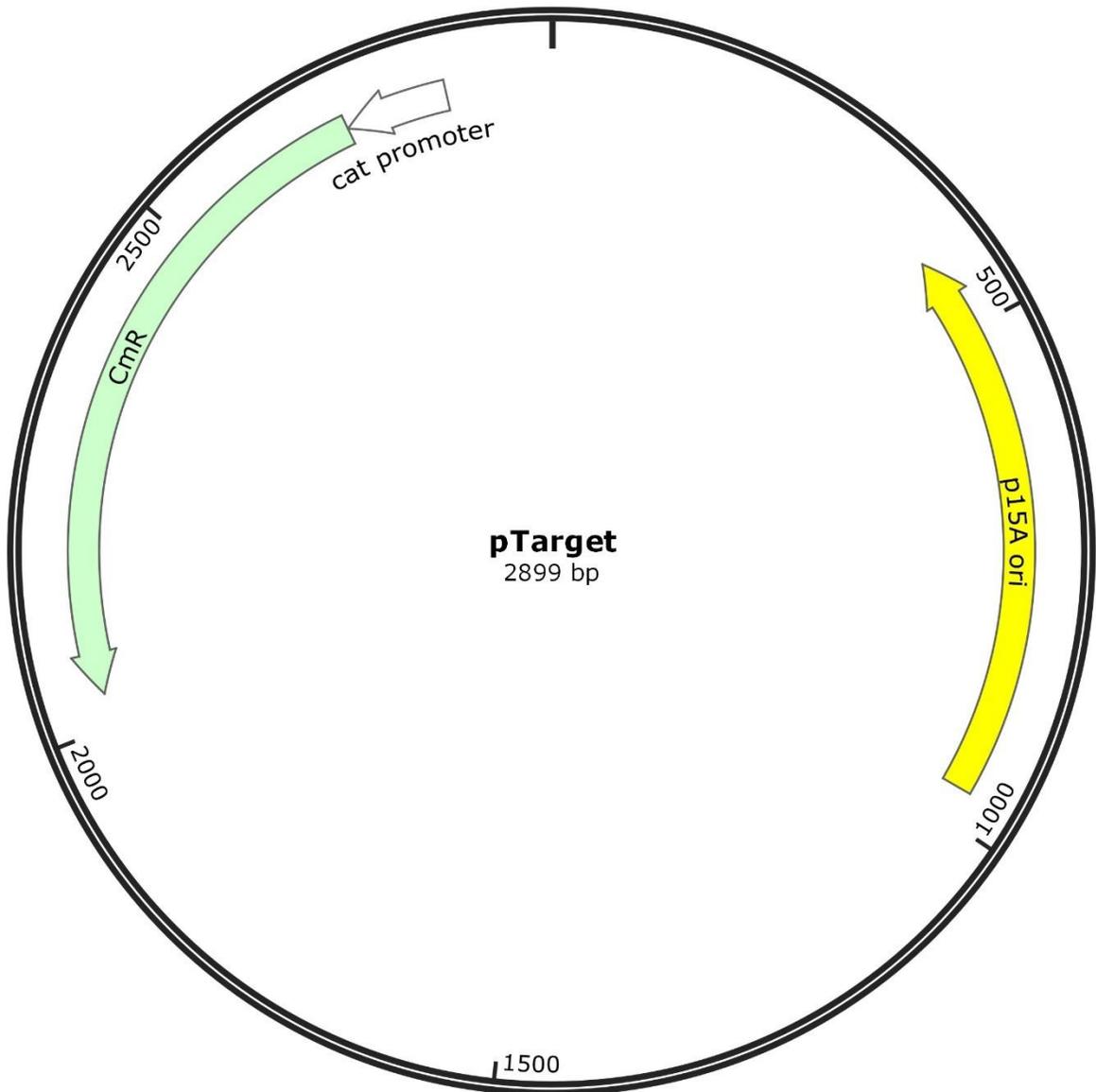
Mapa del plásmido pDonor



\* Construido por Strecker *et al.* (2019).

13.16. Anexo 16

Mapa del plásmido pTarget



\* Construido por Strecker *et al.* (2019).