



**Diseño de un constructo génico que potencie la expresión de lipopéptidos de la familia kurstakinas en *Bacillus thuringiensis* para la aplicación como biosurfactantes**

Erazo Román, Danny Paúl

Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Trabajo de titulación, previo a la obtención del título de Ingeniero en Biotecnología

Koch Kaiser, Alma Rosel Mgs.

08 de marzo de 2021

## Hoja de resultados de la herramienta Urkund



### Document Information

Analyzed document	TrabajoDeTitulacion_ErazoRomanDannyPaul_L00362474.docx (D97623773)
Submitted	3/8/2021 4:16:00 PM
Submitted by	
Submitter email	fjflores2@espe.edu.ec
Similarity	2%
Analysis address	fjflores2.espe@analysis.arkund.com

### Sources included in the report

SA	<b>Universidad de las Fuerzas Armadas ESPE / Armas_Arly_TRABAJO DE TITULACIÓN_Correcciones.docx</b>	2
	Document Armas_Arly_TRABAJO DE TITULACIÓN_Correcciones.docx (D96781542) Submitted by: fjflores2@espe.edu.ec Receiver: fjflores2.espe@analysis.arkund.com	
SA	<b>Universidad de las Fuerzas Armadas ESPE / tesis efecto de las NPHE sobre la poblacion bacteriana biblio.docx</b>	1
	Document tesis efecto de las NPHE sobre la poblacion bacteriana biblio.docx (D26700291) Submitted by: arizquierdo@espe.edu.ec Receiver: arizquierdo.espe@analysis.arkund.com	

ALMA  
ROSEL  
KOCH  
KAISE  
R

Firmado digitalmente  
por ALMA ROSEL  
KOCH KAISER  
DN: cn=ALMA  
ROSEL KOCH  
KAISER c=EC  
o=SECURITY DATA  
S.A. 2 su=ENTIDAD  
DE CERTIFICACION  
DE INFORMACION  
Motivo: Estoy  
aprobando este  
documento  
Ubicación:  
Fecha: 2021-03-08  
15:46:05:00

Firma:

**Koch Kaiser, Alma Rosel Mgs.**

**DIRECTOR**



**DEPARTAMENTO DE CIENCIAS DE LA VIDA Y DE LA AGRICULTURA**  
**CARRERA DE INGENIERÍA EN BIOTECNOLOGÍA**

**CERTIFICACIÓN**

Certifico que el trabajo de titulación, **“Diseño de un constructo génico que potencie la expresión de lipopéptidos de la familia kurstakinas en *Bacillus thuringiensis* para la aplicación como biosurfactantes”** fue realizado por el señor **Erazo Román, Danny Paúl** el cual ha sido revisado y analizado en su totalidad por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 08 de marzo de 2021

Firma:

ALMA  
ROSEL  
KOCH  
KAISER

Firmado digitalmente por ALMA  
ROSEL KOCH KAISER  
DN: cn=ALMA ROSEL KOCH  
KAISER, o=EC, ou=SEGURIDAD DE  
CERTIFICACION DE  
INFORMACION  
Método: Edificio de impresión de este  
documento  
Ubicación:  
Fecha: 2021-03-08 13:47:05-0500

**Koch Kaiser, Alma Rosel Mgs.**

C. C.: 170888079-2



DEPARTAMENTO DE CIENCIAS DE LA VIDA Y DE LA AGRICULTURA  
CARRERA DE INGENIERÍA EN BIOTECNOLOGÍA

RESPONSABILIDAD DE AUTORÍA

Yo **Erazo Román, Danny Paúl** con cédula de ciudadanía n° 040145657-9, declaro que el contenido, ideas y criterios del trabajo de titulación: **Diseño de un constructo génico que potencie la expresión de lipopéptidos de la familia kurstakinas en *Bacillus thuringiensis* para la aplicación como biosurfactantes** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 08 de marzo de 2021

Firma

**Erazo Román, Danny Paúl**

C.C.: 040145657-9



DEPARTAMENTO DE CIENCIAS DE LA VIDA Y DE LA AGRICULTURA  
CARRERA DE INGENIERÍA EN BIOTECNOLOGÍA

AUTORIZACIÓN DE PUBLICACIÓN

Yo, Erazo Román, Danny Paúl con cédula de ciudadanía n° 040145657-9, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **Diseño de un constructo génico que potencie la expresión de lipopéptidos de la familia kurstakinas en *Bacillus thuringiensis* para la aplicación como biosurfactantes en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.**

Sangolquí, 08 de marzo de 2021

Firma

Erazo Román, Danny Paúl

C.C.: 040145657-9

**Dedicatoria**

A mis padres Vilma y Bolívar,

por guiarme siempre con su amor, comprensión y apoyo absolutos.

A mis tías Judith y Tania,

por su compañía y ejemplo a lo largo de los años.

A mi primo Martín,

por ser el alma infantil que ilumina nuestras vidas.

A mis amigos más cercanos,

por permitirme encontrar en ellos a hermanos fuera de la familia.

A mi perrito Homero,

por haber sido un compañero incondicional durante los 16 años que vivió.

Todo lo que haga es y será por y para ustedes.

*Danny Paúl Erazo Román*

## **Agradecimiento**

A mi familia por ser siempre un soporte y una motivación para afrontar la vida y ser mejor cada día, en especial a mis padres Vilma y Bolívar por el total cariño, entrega y sacrificio con el que me han guiado desde la cuna. También a mis tías Judith y Tania y mi primo Martín por su comprensión y apoyo incondicional.

A mis profesores, todos quienes han participado en mi formación desde la infancia hasta este punto culminante de la educación superior. Lucía, Ghassan, Cecilia, María Augusta, Jorge, Alma y Francisco; sus enseñanzas me acompañan cada día.

A mis amigos por demostrarme que, sin importar el camino que elija en la vida, todo es mejor junto a las personas correctas, a quienes llegas a querer como una segunda familia. Erick, Carlos, Nicole, Daniela P., Daniela V., Nathaly F., Mauro, Michelle, Emilia, Jennifer, Stefanía, Camila, Alexander, Nathaly H., César, Verónica y Paúl; con ustedes he compartido los mejores y peores momentos y sé que el lazo que tengo con cada uno de ustedes trascenderá en el tiempo. Adicionalmente, a Geovanny, con quien compartí una breve temporada sin contar con que su vida se apagaría de forma repentina; te recordaré siempre amigo.

*Danny Paúl Erazo Román*

## Índice de contenido

Carátula

Hoja de resultados de la herramienta urkund

Certificación

Responsabilidad de autoría

Autorización de publicación

Dedicatoria

Agradecimiento

Índice de contenido

Listado de tablas

Listado de figuras

Listado de abreviaturas

Resumen

Abstract

Capítulo I: Introducción

Formulación del problema

Justificación del problema

Objetivos de la investigación

*Objetivo general*

*Objetivos específicos*

Hipótesis

Capítulo II: Marco teórico

Hidrocarburos de petróleo

*Contaminación por derrame de crudo*



*Estrategias de remediación*

*Bacillus* spp.

*Bacillus thuringiensis*

Lipopéptidos

*Kurstakinas*

*Actividad biosurfactante*

Bioinformática

*Alineamiento múltiple*

*Análisis filogenético*

*Análisis transcriptómico*

Constructo génico

*Vector de transporte*

*Promotor*

*Región 5'-UTR*

Recombinación homóloga

Capítulo III: Materiales y métodos

Revisión Bibliográfica

Identificación de la estructura consenso del locus *krs*

*Descubrimiento de los genes involucrados en genomas de microorganismos  
productores de kurstakinas*

*Caracterización de dominios y sustratos de los presuntos genes del locus *krs**

*Concatenación de secuencias, alineamiento múltiple y construcción de un  
árbol filogenético*

Anotación del plásmido BBa\_K802004

*Selección del plásmido*

*Búsqueda de ORFs y definición de genes presentes en el vector*

Análisis transcriptómico de RNA-seq proveniente de *Bacillus thuringiensis* serovar *chinensis* CT-43

Establecimiento del constructo génico

*Definición de una región potenciadora dentro de B. thuringiensis*

*PCR in silico de los sitios de interés*

*Digestión y ensamblaje del constructo*

Simulación del proceso de recombinación homóloga

Planteamiento del procedimiento a realizar en laboratorio

#### Capítulo IV: Resultados

Secuencia consenso del locus *krs*

Módulos y sustratos de las NRPS definidas para kurstakinas

Locus *krs* distribuido en el grupo *Bacillus cereus* (BLAST 2020) y árbol filogenético

pHT315-1, vector anotado

Recuento de lecturas y expresión diferencial en distintas fases de crecimiento de *Bacillus thuringiensis*

pHT315-1\_PrIpU\_krsE

Integración de la secuencia en el cromosoma bacteriano

Propuesta de protocolo experimental

#### Capítulo V: Discusión

#### Capítulo VI: Conclusiones

Capítulo VII: Recomendaciones

Capítulo VIII: Bibliografía

### Listado de tablas

- Tabla 1.** Taxonomía del género *Bacillus*.
- Tabla 2.** Programas utilizados para predecir dominios NRPS.
- Tabla 3.** Módulos hallados en KrsA, KrsB y KrsC.
- Tabla 4.** Aminoácidos usados como sustratos por NRPS.
- Tabla 5.** Distribución del locus *krs* en el grupo *Bacillus cereus*.
- Tabla 6.** Niveles de expresión.
- Tabla 7.** Resultados del análisis en DESeq2.
- Tabla 8.** Primers para amplificar PrlpU y *krsE*.

## Listado de figuras

**Figura 1.** Hipotético ensamblaje biosintético de kurstakinas.

**Figura 2.** Vector BBa\_K082004.

**Figura 3.** Crecimiento de *Bacillus thuringiensis*.

**Figura 4.** Locus *krs*.

**Figura 5.** Estructura de kurstakina.

**Figura 6.** Árbol filogenético.

**Figura 7.** Plásmido pHT315-1 anotado.

**Figura 8.** Número de lecturas mapeadas.

**Figura 9.** Gráficos generados por DESeq2.

**Figura 10.** Promotor del operón *krs*.

**Figura 11.** Promotores candidatos.

**Figura 12.** Vector pHT315-1\_PrlpU\_*krsE*.

**Figura 13.** Simulación de recombinación homóloga.

**Listado de abreviaturas**

<b>NRPS</b>	Sintetasa peptídica no ribosomal
<b>krs</b>	Kurstakina
<b>ADNc</b>	Ácido desoxirribonucleico complementario
<b>ARNm</b>	Ácido ribonucleico mensajero
<b>RNA-seq</b>	Secuencias de ADNc útiles para análisis transcriptómicos
<b>BLAST</b>	Herramienta de búsqueda de secuencias de ADN o proteicas
<b>NCBI</b>	Centro Nacional de Información Biotecnológica de Estados Unidos
<b>PCR</b>	Reacción en cadena de polimerasa
<b>SinR</b>	Factor de represión de transcripción génica
<b>SpoA</b>	Factor de activación de transcripción en una porción de células
<b>AbrB</b>	Factor de represión de transcripción génica
<b>ComK</b>	Factor de activación de transcripción génica
<b>mL</b>	Mililitros
<b>g</b>	Gramos
<b>µg</b>	Microgramos
<b>µL</b>	Microlitros
<b>h</b>	Horas
<b>°C</b>	Grados centígrados

## Resumen

Las bacterias del género *Bacillus*, junto a otros microorganismos como *Pseudomonas*, producen moléculas oligopeptídicas con versatilidad de aplicaciones, denominadas lipopéptidos. Durante el año 2000, Hathout y otros reportaron el hallazgo de un lipopéptido, la kurstakina; llamada así en honor a la variante *kurstaki* HD-1 de *Bacillus thuringiensis*. La biosíntesis es ejecutada por sintetasas peptídicas no ribosomales (NRPS, por sus siglas en inglés) distribuidas como operón en el genoma.

Estudios previos han propuesto su conformación y factores que regulan su expresión, en consecuencia, este trabajo pretende reunir los datos disponibles para explicar el sistema de *Bacillus thuringiensis* en la producción de kurstakinas.

Mediante el uso de distintos softwares bioinformáticos, se describen seis genes; además, un alineamiento múltiple y construcción de un árbol filogenético destaca la exclusividad del operón a especies del clado II dentro del grupo *Bacillus cereus*. Por otro lado, se tomó la parte BBa\_K802004 (disponible en el repositorio de iGEM) para su respectiva anotación y uso como vector, entonces, tras un análisis transcriptómico realizado en Galaxy, se propone el reemplazo del promotor de *krs* por la secuencia PrpIU.

Concluyendo, la potencialización de síntesis de kurstakinas es posible de acuerdo al procedimiento *in silico* presentado en este documento y se presenta una propuesta del protocolo a realizar en laboratorio, incluyendo el uso del lipopéptido como biosurfactante.

Palabras clave:

- **LIPOPÉPTIDO**
- **KURSTAKINA**
- **BIOSÍNTESIS**
- **CONSTRUCTO**
- **RENDIMIENTO**

## Abstract

Bacteria from *Bacillus* genus, among different microorganisms such as *Pseudomonas*, produce oligopeptide molecules with application versatility, referred as lipopeptides. In 2000, Hathout and colleagues found a brand new lipopeptide family, kurstakin, named after *Bacillus thuringiensis* serovar *kurstaki* HD-1. Biosynthesis is exerted by non-ribosomal peptide synthetases (NRPS), specialized enzymes gathered as operon within bacterial genomic DNA.

Previous research have purposed a description of its conformation and expression regulating factors, thereafter, this document aims to collect current available data in order to explain the mechanism used by *Bacillus thuringiensis* to synthesize kurstakins.

Thus, using a wide branch of bioinformatics software, six genes were described. Furthermore, multiple alignment and phylogenetic tree construction show how operon's distribution is exclusive for clade II in *Bacillus cereus* group. On the other hand, part BBa\_K082004 (from iGEM repository) was chosen to be annotated and perform the role of shuttle vector, then, after a transcriptomic analysis executed on Galaxy Project platform, it is suggested to replace *krs* promoter with PrlpU sequence.

To conclude, kurstakin production yield can be enhanced according to *in silico* procedures detailed in this thesis and a protocol plan is purposed for its implementation at laboratory level, including utilization of obtained lipopeptide as biosurfactant.

Keywords:

- LIPOPEPTIDE
- KURSTAKIN
- BIOSYNTHESIS
- CONSTRUCT
- YIELD



## Capítulo I: Introducción

### Formulación del problema

El Fondo Mundial para la Naturaleza o WWF – por su nombre en inglés World Wildlife Fund – resalta en un artículo sobre el petróleo cómo su extracción requiere de carreteras, oleoductos, torres de perforación y otras estructuras que, por sí mismas, representan una invasión en el hábitat de múltiples especies animales y, además, los eventos accidentales de corrosión de tuberías y derrames de crudo provocan daños en el entorno y a poblaciones humanas residentes en territorios aledaños (WWF, 2021).

En el caso particular de Ecuador, Becerra y colaboradores (2013) describen cómo la mayor parte de la economía del país tiene sus bases en actividades petrolíferas, a pesar de los múltiples impactos ambientales suscitados a lo largo del tiempo y donde resalta el caso de la empresa Texaco (ahora llamada Chevron) cuya descarga de desperdicios superó al gran derrame del Exxon Valdez en Alaska en 1989 y tiene secuelas ambientales, sociales y de salud pública hasta la actualidad.

Una de las propuestas de remediación en sitios contaminados recientemente generada en el país se llevó a cabo por el grupo conformado por Alexey, González, Enriquez y Mantilla en la Universidad Central del Ecuador (2017) donde se destaca la actividad de sustancias surfactantes en el proceso de recuperación mejorada; sin embargo, como el equipo menciona, diario El Comercio señaló en un artículo de 2016 que tal estrategia ha tenido bajos niveles de aplicación por su demanda monetaria.

Por otro lado, Baker y sus compañeros (2009) hacen hincapié en un contraste pues, pese a la capacidad de los tensoactivos para reducir la tensión superficial e interfacial favoreciendo al flujo y recuperación de componentes de petróleo, estos surfactantes representan un nuevo residuo en suelo y agua; por tanto, se precisa de la aplicación de agentes aptos para la remoción de compuestos contaminantes y su eventual degradación una vez culminado el proceso.

Así, los biosurfactantes – compuestos tensioactivos derivados de microorganismos – han surgido como una alternativa prometedora considerando su bajo nivel de toxicidad, tendencia a la biodegradabilidad, actividad en amplios rangos de pH y temperatura y su selectividad (Santos et al., 2016). La síntesis de biosurfactantes puede ser espontánea o inducida por las condiciones de crecimiento y en la mayoría de los casos presenta un rendimiento limitado; en consecuencia, el desafío planteado en el presente trabajo es proponer un mecanismo para incrementar el nivel de producción de un biosurfactante, específicamente kurstakina sintetizada por *Bacillus thuringiensis*, que sirva como un elemento de uso viable técnica y económicamente en la biorremediación, a la par de otros campos como la industria alimenticia, el desarrollo de nuevos antibióticos o el control de plagas.

### **Justificación del problema**

El Ministerio de Ambiente del Ecuador MAE informa en caso de eventuales derrames de crudo en el territorio nacional y participa con acciones pertinentes de remediación junto a las empresas públicas PetroEcuador EP y PetroAmazonas EP, fusionadas desde enero de 2021. Entre los desastres más recientes se destaca la rotura del oleoducto por la zona de San Rafael en la frontera entre las provincias de Napo y Sucumbíos, amenazando al suministro de agua y alimentos de calidad a los residentes de tal territorio del nororiente ecuatoriano en medio de la emergencia sanitaria por la pandemia de Covid-19 (Guerra, 2020). Situaciones similares, sumadas al impacto ambiental generado por Chevron-Exxon durante décadas en Sacha y Shushufindi (Fontaine, 2003), requieren de actividades conjuntas de remediación; por lo tanto, la biodegradación consta como una estrategia clave para eliminar los hidrocarburos del entorno y es ejecutada por microorganismos capaces de asimilar los residuos de hidrocarburos para degradarlos.

Esencialmente, la biorremediación constituye una estrategia de restauración de un entorno natural mediante la remoción de sustancias invasoras por parte de organismos vivos como plantas o microorganismos. Tomando en cuenta al último grupo, la mayor cantidad de bacterias reportadas en sitios afectados pertenecen al género *Pseudomonas*, asociadas comúnmente a miembros de divisiones taxonómicas como *Acinetobacter*, *Streptomyces*, *Enterobacter*, *Bacillus* y *Proteobacterium*, entre otras. Además, las prácticas recientes incluyen también organismos genéticamente modificados con el fin de mejorar el rendimiento de la técnica (Pande et al., 2020).

Entonces, para poder llevar a cabo el proceso es imprescindible la biodisponibilidad de los compuestos contaminantes, por ello se aplican tensoactivos capaces de favorecer a su solubilidad; sin embargo, tales productos se convierten en un problema ambiental adicional al concentrarse en el sitio. En consecuencia, se ha preferido el uso de biosurfactantes, sustancias de origen biológico con efecto tensoactivo útiles para facilitar a los consorcios microbianos del lugar la asimilación de hidrocarburos poco solubles (Li et al., 2018).

Con base en los antecedentes planteados y considerando los reportes de la existencia de microorganismos pertenecientes a la especie bacteriana *Bacillus thuringiensis* serovar *galleriae* strain HD-29 (identificada mediante secuenciación del fragmento ARNr 16S) en sitios contaminados por crudo en territorio ecuatoriano (Maddela et al., 2015), el presente estudio pretende incrementar la producción de los lipopéptidos de la familia kurstakinas, generados por *B. thuringiensis* (Hathout et al., 2000), mediante el diseño de un constructo génico basado en un promotor fuerte apto para potencializar la expresión de la sintetasa peptídica no ribosomal encargada de la función de biosíntesis.

## **Objetivos de la investigación**

### ***Objetivo general***

Diseñar de un constructo génico que potencie la expresión de lipopéptidos de la familia kurstakinas en *Bacillus thuringiensis* para su aplicación como biosurfactantes.

### ***Objetivos específicos***

- Analizar secuencias promotoras y plásmidos para su aplicación en la modificación de *Bacillus thuringiensis*.
- Establecer la secuencia del constructo y su integración por recombinación homóloga mediante el uso de softwares bioinformáticos.
- Definir protocolos para ejecutar la modificación en laboratorio y la obtención de kurstakinas para aplicarlas en terrenos contaminados.

## **Hipótesis**

H<sub>0</sub>: La modificación del cluster génico *krs* mediante la integración de un constructo no es capaz de potenciar la expresión de sintetasa peptídica no ribosomal de kurstakinas en *Bacillus thuringiensis*.

H<sub>1</sub>: La modificación del cluster génico *krs* mediante la integración de un constructo es capaz de potenciar la expresión de sintetasa peptídica no ribosomal de kurstakinas en *Bacillus thuringiensis*.

## Capítulo II: Marco teórico

### Hidrocarburos de petróleo

El crudo de petróleo es una sustancia natural extraída desde formaciones geológicas ubicadas bajo la superficie de la Tierra, cuya composición específica es variada y se expresa en términos generales como un conjunto de hidrocarburos (Hanania et al., 2017). El término hidrocarburo hace referencia a moléculas constituidas por una combinación de átomos de carbono con átomos de hidrógeno, aunque algunas sustancias también pueden presentar azufre, oxígeno y nitrógeno; así, estos compuestos se manifiestan como líquidos claros con facultad de volatilizarse con facilidad o como líquidos oscuros y espesos difíciles de evaporar y pueden clasificarse en diferentes familias según su configuración de forma cíclica o en cadena y considerando si están o no saturados, de la siguiente manera:

#### Hidrocarburos alifáticos

- a) Saturados:** Moléculas formadas de acuerdo a la fórmula general  $C_nH_{2n+2}$  conformadas a manera de cadenas íntegras o ramificadas. Llamados también hidrocarburos parafínicos, metano, etano, propano y butano, se hallan en estado gaseoso a temperatura ambiente, los compuestos de 5 a 16 carbonos son líquidos y los de mayor peso molecular son sólidos estableciendo la parafina (Secretaría de Energía República de Argentina, 2003).
- b) No saturados:** Moléculas que presentan dobles o triples enlaces y también pueden hallarse en cadenas rectas o ramificadas. Llamados también olefínicos y no se encuentran en el petróleo crudo (Secretaría de Energía República de Argentina, 2003).

### **Hidrocarburos cíclicos**

La molécula forma ciclos que podrían contener varios átomos de carbono; sin embargo, la más común se forma de seis carbonos manteniendo un ángulo de 109° correspondientes a su distribución normal.

- a) Saturados:** Llamados también naftenos, mantienen una fórmula general  $C_nH_{2n}$  como isómeros de los hidrocarburos olefínicos y la condensación de dos núcleos naftenos genera como resultado un hidrocarburo complejo de elevado peso molecular (Secretaría de Energía República de Argentina, 2003).
- b) No saturados:** La forma más común es un anillo de seis carbonos y, a pesar de que la conformación puede incluir 1 o 2 dobles enlaces, la estructura es estable con tres dobles enlaces formando benceno. Los hidrocarburos bencénicos o aromáticos se hallan en el crudo de petróleo como un ciclo (benceno), dos ciclos (naftaleno), tres ciclos (antraceno) u otras moléculas incluyendo sus variaciones (Secretaría de Energía República de Argentina, 2003).

### ***Contaminación por derrame de crudo***

El petróleo, como se ha descrito previamente, se conforma de una mezcla de hidrocarburos y, en consecuencia, cuando existe un derrame se utiliza el término hidrocarburos totales de petróleo (TPH) como concepción general para medir su concentración en el sitio afectado; aunque, para tener una idea más específica de las sustancias contaminantes, se dividen los TPH en grupos de acuerdo a propiedades similares. La elevada demanda de productos derivados del petróleo exige la extensa explotación y comercialización de crudo, lo que a su vez genera una alta probabilidad de contaminación ambiental debida a roturas, falta de mantenimiento o errores operativos. Así, en caso de un eventual derrame, cuando el petróleo entra en contacto con el agua ocurre una separación de los TPH en fases dejando una fina capa en la superficie y el

resto tiende a sedimentar, afectando a peces y otros organismos vivos en el entorno; de igual manera, si el petróleo entra en contacto con el suelo, los TPH se movilizan hasta llegar al agua subterránea y los componentes pueden dividirse de acuerdo a sus propiedades químicas. Algunas de estas sustancias pueden evaporarse al aire y otros son disueltos en el agua; no obstante, algunos microorganismos (bacterias, hongos, entre otros) son capaces de degradarlos (Agencia para Sustancias Tóxicas y Registro de Enfermedades, 1999).

### ***Estrategias de remediación***

La contaminación ejercida por hidrocarburos de petróleo representa una amenaza a la salud humana y al mantenimiento del ambiente en general. Las estrategias de remediación son métodos de respuesta con el fin de restaurar la zona afectada, pero la aplicación de cualquiera de estos procesos en particular requiere estudios de lugar específico, la naturaleza de los contaminantes y las condiciones físicas, químicas y biológicas del sitio. Con respecto a la ejecución, debe tomarse en cuenta los mecanismos disponibles, el costo y tiempo que tomará la intervención; por tal motivo, existen distintos tipos de técnicas biológicas, químicas, fisicoquímicas, térmicas, eléctricas, electromagnéticas y de ultrasonido (Chukwunonso et al., 2019).

### **Tratamientos biológicos**

La aplicación de un tratamiento biológico está subordinada a varios factores limitantes y a la localización de los contaminantes; así mismo, debe considerarse la factibilidad de llevarla a cabo en el lugar o si se precisa una excavación, remoción y transporte del suelo y agua. Este tipo de estrategias de remediación se basan en el uso de organismos vivos (bacterias, hongos, algas, plantas) capaces de concentrar, e incluso, degradar hidrocarburos de petróleo para transformarlos en sustancias sencillas y menos tóxicas que no representan una amenaza para el ambiente (Chukwunonso et al., 2019).

### ***Biorremediación***

Los microorganismos son hábiles para utilizar los hidrocarburos de petróleo como fuente de energía, por lo que se disponen como una alternativa viable para el tratamiento y recuperación de zonas contaminadas. En primera instancia se permite el suceso de atenuación natural, que incluye transformación química, biodegradación, estabilización, volatilización, dispersión y dilución de las sustancias contaminantes y es ejecutada por la acción de los microorganismos y las condiciones propias del entorno. Sin embargo, con el objetivo de favorecer este proceso, algunas estrategias importantes incluyen la bioestimulación (adición de nutrientes en el medio para permitir el crecimiento de bacterias con capacidad degradadora) y el bioaumento (inoculación de microorganismos competentes y adaptados para el consumo de hidrocarburos), estableciendo un consorcio capaz de asegurar la recuperación de lugar (Naeem & Akhram, 2019).

### ***Biosurfactantes***

Los biosurfactantes son compuestos de naturaleza anfílica producidos por múltiples microorganismos a nivel de la superficie celular o secretado de manera extracelular y cuya principal función radica en disminuir la tensión superficial e interfacial. La manera en la que los biosurfactantes actúan para facilitar la remediación en terrenos afectados por hidrocarburos puede explicarse de dos formas: el incremento de la disponibilidad de las sustancias contaminantes para el consumo de los microorganismos o la interacción con la superficie celular para aumentar su hidrofobicidad; así, el sustrato hidrofóbico tiende a unirse a las células bacterianas (Li et al., 2018).



***Bacillus* spp.**

El género *Bacillus* hace referencia a un grupo de bacterias Gram positivas con morfología de bastón (Green, 2008). De acuerdo con el Manual de Bacteriología Sistemática de Bergey, pertenecen a la clasificación taxonómica que se aprecia en la Tabla 1:

**Tabla 1.**

Taxonomía del género *Bacillus*.

<u>Taxonomía</u>	
<u>Dominio</u>	Bacteria
<u>Filo</u>	Firmicutes
<u>Clase</u>	Bacilli
<u>Orden</u>	Bacilliales
<u>Familia</u>	Bacillaceae
<u>Género</u>	<i>Bacillus</i>

Descrito por primera vez por Cohn en 1872, el género *Bacillus* se caracteriza por tener especies aerobias o anaerobias facultativas y catalasa positivas en presencia de oxígeno. La identificación de especies se logra mediante observación morfológica y ensayos de carácter bioquímico; además, por la formación de endosporas, se reconoce que algunas especies son capaces de tolerar calor, radiación, desinfectantes y disecación. Con respecto a su distribución, se hallan principalmente en suelo, sedimentos marinos, alimentos y restos de animales muertos, debido a su naturaleza saprófita (Slepecky & Hemphill, 2006).

### ***Bacillus thuringiensis***

La especie *Bacillus thuringiensis* es una bacteria aerobia, Gram positiva y generadora de esporas que puede actuar como patógeno facultativo de insectos lepidópteros y coleópteros y es aislada fácilmente en un medio simple como agar nutriente a partir de varias fuentes como suelo, agua, superficies vegetales, polvo de granos, restos y heces de insectos (Sansinenea, 2012).

En Ecuador, el grupo de investigación de Maddela (2015) y colaboradores reportaron el aislamiento de dos especies bacterianas aisladas en terreno contaminado por hidrocarburos de petróleo en Lago Agrio. Tras un análisis filogenético de las secuencias de ADN ribosómico 16S se identificaron como miembros del género *Bacillus*, distinguiéndolas como *B. cereus* y *B. thuringiensis* por su porcentaje de identidad con las secuencias de Genbank con número de accesión EF582416.1 y CP010089.1, respectivamente; con capacidad de degradar algunos compuestos contaminantes, siendo más susceptibles los alcanos lineales y ramificados y algunas sustancias aromáticas pequeñas. El estudio se convierte en una ventana para la investigación enfocada en el reconocimiento de más microorganismos degradadores de la región Amazónica y la potencial optimización de su actividad (Maddela et al., 2015).

Con respecto a las aplicaciones biotecnológicas de *Bacillus thuringiensis*, la especie ha sido ampliamente sometida a ensayos de recombinación como la fusión del gen *P19* de *B. thuringiensis* subs. *israelensis* con un gen de chitinasa de *B. licheniformis* y se obtuvo un mejoramiento en la actividad insecticida de la bacteria. De igual manera, se ha promovido la producción de inclusiones cristalinas mediante la manipulación del gen *cry* con acción insecticida, sobre todo en coleópteros, y otras múltiples intervenciones (Abdelkefi-mesrati & Tounsi, 2012).

## **Lipopéptidos**

Los lipopéptidos son sustancias cíclicas de bajo peso molecular (1000-2000 Da) con potencial acción antimicrobiana producidos, principalmente, por bacterias de los géneros *Pseudomonas* spp. y *Bacillus* spp. La síntesis de estos compuestos está sujeta a un grupo de genes que forman una ruta de biosíntesis multienzimática dando lugar a complejos llamados sistetasas peptídicas no ribosomales o NRPS, por sus siglas en inglés provenientes de Non-Ribosomal Peptide Synthetases. La molécula presenta naturaleza anfifílica, pues está constituida por una cadena hidrofílica de oligopéptido (7-10 aminoácidos) enlazada a una cola hidrofóbica de ácido graso; además, tiene aplicaciones como antifúngico, antibacteriano, antitumoral y biosurfactante (Malviya et al., 2020).

### ***Kurstakinas***

El término *kurstakinas* hace referencia a una familia de lipopéptidos, reportados por primera vez a partir de un estudio en *Bacillus thuringiensis* subs. *kurstaki*. Su estructura es estrechamente similar a la de otros lipopéptidos generados por microorganismos del género *Bacillus* spp., hongos y levaduras, consta de los aminoácidos treonina (Thr), glicina (Gly), alanina (Ala), serina (Ser), histidina (His), glutamina (Gln) y glutamina; donde ocurre la formación de un enlace lactona entre el extremo C y el residuo de serina y cuenta con una cadena de ácido graso ligada al residuo de treonina (Hathout et al., 2000).

Las propiedades de las *kurstakinas* son similares a las de otros lipopéptidos, destacándose la capacidad antifúngica que se ha analizado con varias especies como *Stachybotrys charatum*; sin embargo, algunos estudios sugieren que esta correlación entre la aplicación de *kurstakinas* y la inhibición fúngica no existe o su acción es limitada, pues no tuvieron efecto sobre *Fusarium oxysporum*, *Aspergillus niger* y *Penicillium*

*roqueforti*, entre otros. Por otro lado, se ha reportado cómo estos lipopéptidos favorecen la supervivencia de *B. thuringiensis* en cadáveres de insectos o la manera en que su aplicación como biosurfactante facilita la asimilación de diésel de la bacteria *Acinetobacter haemolyticus* cepa 2SA al incrementar la hidrofobicidad de su membrana (Béchet et al., 2012; Diallo et al., 2019).

### **Biosíntesis de kurstakinas (sintetasas peptídicas no ribosomales)**

Los potenciales operones responsables de la síntesis de kurstakinas fueron identificados mediante análisis bioinformático utilizando dos enfoques: 1) proteómica basada en espectrometría de masas considerando el tamaño de otras NRPS y un marcador iónico único del cofactor fosfopantetenil común entre estas moléculas identificando los genes de formación de kurstakinas en *Bacillus* sp. NK2018 y *Bacillus cereus* AH1134; 2) PCR utilizando primers degenerados obtenidos a partir de secuencias intraoperón alineadas en los dominios de adenilación y tiolación de las enzimas implicadas en la biosíntesis de otros lipopéptidos, así dos pares de primers para identificación de bacilomicinas permitieron hallar los genes de biosíntesis de kurstakinas (Abderrahmani et al., 2011; Tapi et al., 2010).

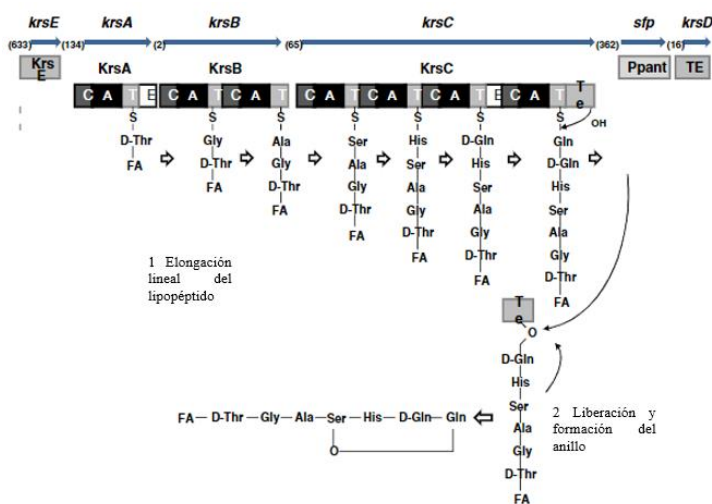
La conclusión de los acercamientos anteriormente descritos dio como resultado la identificación de un grupo de tres genes: *krsA*, *krsB* y *krsC*; cuya expresión da lugar a proteínas homónimas que forman la sintetasa completa. Así, la Figura 1 ilustra el desarrollo de la biosíntesis de kurstakinas iniciando en KrsA, donde se integra D-Thr como primer residuo a una cadena de ácido graso (C<sub>10</sub>-C<sub>14</sub>), luego atraviesa los respectivos módulos de KrsB y KrsC hasta completar el lipoheptapéptido (Béchet et al., 2012).

Existen tres genes adicionales participando en la biosíntesis de kurstakinas, tal como sucede con otras familias de lipopéptidos: 1) un gen *sfp* encargado de codificar una fosfopantetenil transferasa que adiciona el grupo fosfopantetenil de coenzima A a los

dominios tioesterasa (Mofid et al., 2004); 2) una segunda tioesterasa codificada por el gen *te* con un rol en la regeneración de proteínas portadoras de residuos peptídicos con error de acilación (Schwarzer et al., 2002); y 3) un gen *krsE*, ubicado aguas arriba de *krsA-C*, perteneciente al grupo de kurstakinas en *Bacillus thuringiensis* Bt407 e implicado en el direccionamiento efluente para la secreción del lipopéptido (Dubois et al., 2012).

### Figura 1.

*Hipotético ensamblaje biosintético de kurstakinas. Reproducido de (Béchet et al., 2012).*



Springer-Verlag © 2012

### Actividad biosurfactante

La concentración micelar crítica CMC es la concentración mínima requerida de una sustancia surfactante para reducir la tensión superficial del agua a un nivel máximo. Con base en el concepto establecido se evalúa la actividad de los biosurfactantes, por ejemplo, los lipopéptidos de las familias surfactina, iturina A y fengicina en estado puro han mostrado una CMC de 10, 20 y 11 mg/L, respectivamente. Tomando los valores mencionados como punto de partida, se ha logrado una reducción de tensión superficial significativa. Generalmente, los biosurfactantes presentan CMC notablemente menores que la de surfactantes químicos como el oleato de potasio con 350 mg/L o el éter sulfato de alcohol graso con 170 mg/L (Geissler et al., 2019).

Particularmente, la evaluación de actividad biosurfactante de la kurstakina se determinó en el estudio de Diallo y colaboradores (2019), en el que se midió la CMC tras la obtención del lipopéptido arrojando un valor de 144.504 mg/L (considerando 892 g/mol como el peso molecular de kurstakina, obtenido en el sitio web del NCBI), donde además se reportó cómo la kurstakina facilitaba la asimilación de diésel en muestras de suelo contaminado ejecutada por *Acinetobacter haemolyticus* al adherirse a la membrana de esta bacteria y permitiendo un incremento en su hidrofobicidad, favoreciendo la unión de los hidrocarburos hidrofóbicos con la célula (Diallo et al., 2019).

### **Bioinformática**

El proyecto genoma humano, junto a otros trabajos de investigación relacionados a la secuenciación y recopilación de datos biológicos, dieron lugar a una elevada demanda por herramientas de análisis e interpretación a principios del siglo XXI, así, la bioinformática nace como la aplicación de instrumentos computacionales enfocados en capturar e interpretar datos al integrar consideraciones interdisciplinarias que incluyen matemáticas, física y biología combinadas con programación. A través de los años, la bioinformática se ha venido aplicando con distintas perspectivas que tienen como base la comprensión funcional de las ciencias -ómicas (genómica, transcriptómica, proteómica, metabolómica, entre otras) y con vista hacia campos como ingeniería genética, sondeo de objetivos para fármacos, terapia génica, medicina personalizadas y un sin número de posibilidades (Bayat, 2002).

### ***Alineamiento múltiple***

El alineamiento múltiple de secuencias es un recurso de modelamiento desarrollado a partir de una serie de algoritmos enfocados en la alineación de secuencias evolutivamente relacionadas considerando eventos comunes ocasionados durante el proceso como mutaciones, inserciones, deleciones y rearrreglos. Estas estrategias

pueden llevarse a cabo en secuencias correspondientes a ADN, ARN y proteínas; y sus resultados son esenciales en posteriores análisis *in silico* como asociación de dominios, reconstrucción filogenética, búsqueda de motifs, determinación de elementos reguladores de expresión génica y una diversa rama de situaciones donde pueden emplearse. De tal forma, el principio en el que se basa el alineamiento de secuencias ha dado lugar a distintas propuestas incluidas Muscle, Mafft y Clustal $\Omega$  registrando un valor relacionado al posible evento evolutivo a través de una función de puntuación; por otro lado, también se cuenta con la herramienta en línea BLAST (Basic Local Alignment Search Tool), la cual enfrenta secuencias genómicas provenientes de una base de datos global con el fin de hallar regiones similares y destaca la significancia del análisis y el porcentaje de cobertura de un fragmento examinado (Chatzou et al., 2016).

### ***Análisis filogenético***

La filogenética, también llamada ciencia de la sistemática, enfoca su interés en el estudio de la organización de la diversidad biológica; por lo tanto, un análisis filogenético radica en la integración de herramientas que permitan comparar información evolutiva de poblaciones, especies e individuos. Los datos procesados provienen de observaciones morfológicas, conductuales y moleculares para proceder a estimar la relación histórica existente entre géneros y especies; eventualmente, se organizan en diagramas ramificados conocidos como árboles filogenéticos (Hillis, 1997). La clasificación en distintos grupos se denomina taxonomía y está basada en la comparación de las similitudes descritas entre los organismos, tomando en cuenta tanto a vivos como extintos.

La construcción de árboles filogenéticos puede ser llevada a cabo considerando varios acercamientos incluyendo el método de distancia, parsimonia, máxima verosimilitud y bayesiano. Particularmente, la estrategia de máxima verosimilitud establece la probabilidad de la existencia de una secuencia basada en un modelo de

evolución definido previamente, así, mientras la secuencia propuesta sea más probable, el árbol generado será el elegido (Tennant, 2004).

### ***Análisis transcriptómico***

La secuenciación de nueva generación, a partir de su incursión en el mercado durante el año 2005, ha representado una plataforma alternativa para el estudio de la transcriptómica frente a la técnica de microarreglos. A manera de una sencilla explicación, la ciencia transcriptómica se entiende como la estimación cuantitativa del conjunto completo de transcritos en una célula; mientras, el término RNA-seq hace referencia al flujo de trabajo que abarca la obtención de ARN, su respectiva transformación a ADNc y la obtención de las secuencias con el uso de distintas plataformas de nueva generación (Illumina, Roche 454, Applied Biosystems SoLiD). De tal manera, los datos generados a través de RNA-seq sirven para establecer una medida de los niveles de expresión génica y, además, para otros propósitos como el hallazgo de nuevos patrones de splicing, uso de promotores alternativos, determinación de regiones 5' o 3' no traducidas (UTR) y el descubrimiento de eventos de transcripción antisentido, secuencias no codificantes y ARNs pequeños (Hoeijmakers et al., 2012).

### **Constructo génico**

Un constructo génico es una molécula de ADN que se desea insertar en un vector de clonación para lograr una modificación en una célula diana. Los objetivos de la integración de una secuencia de ADN en un organismo pueden variar desde una intervención o silenciamiento en un gen específico hasta el potenciamiento de la expresión de un gen o la incorporación de una secuencia de interés en un nuevo organismo. El vector de clonación es una molécula de ADN pequeña y capaz de auto replicarse, como un plásmido o ADN viral, al que se integra una secuencia de interés (Zaid et al., 1999).



### **Vector de transporte**

La manipulación genética en procariotas incluye un proceso de distribución de un nuevo gen o una modificación en un locus existente a través de plásmidos, en particular, *Escherichia coli* ha sido el organismo modelo para la aplicación de la técnica y por ende suelen generarse alteraciones en sus propias secuencias plasmídicas o sus células son utilizadas como chasis para la construcción de vectores enfocados en otras especie como *Bacillus subtilis*, otro microorganismo ampliamente estudiado (Sullivan et al., 1984).

En consecuencia, la clonación molecular y otras técnicas de ingeniería genética que ocurran en *Bacillus thuringiensis* precisan también de plásmidos capaces de replicarse tanto en la célula hospedera de origen como en *E. coli* y varios equipos de trabajo han desarrollado vectores de transporte enfocados en cumplir efectivamente con tal objetivo. Así, Ochoa y López (2012) reúnen en su trabajo varios plásmidos obtenidos a partir de la combinación de vectores de *B. thuringiensis* con otros convencionales provenientes de *E. coli* (pUC18 y pUC19), destacados por su número de copias por cromosoma o la longitud de secuencia que son capaces de soportar. Especialmente, destacan los vectores de transporte propuestos por Arantes y Lereclus (1991) pHT304, pHT315 y pHT370 que contienen el origen de replicación de los microorganismos en cuestión y genes de resistencia a ampicilina y eritromicina que permitan un aislamiento efectivo de las células transformadas.

### **Promotor**

Promotor es el término utilizado para definir a la secuencia de ADN hallada comúnmente aguas arriba con respecto a un gen y es necesaria para prender o apagar su expresión. Esta región contiene partes que permiten la unión de elementos regulatorios y de la enzima encargada del proceso de transcripción; sin embargo, su delimitación es ambigua, pues suelen incluirse en la clasificación de “promotor” a secuencias extensas

ubicadas antes del sitio de inicio de transcripción (TSS) que presenten características como nucleótidos consenso reconocidos por factores capaces de potenciar o reprimir la expresión del gen (NIH, 2021).

Cabe destacar que el conjunto de secuencias convencionalmente utilizadas por la ARN polimerasa de *Bacillus subtilis* (cercano a lo que sucede en *B. thuringiensis*) contiene un par de grupos formados por seis pares de bases ubicados en las posiciones -35 y -10 con respecto al sitio de inicio de transcripción, manteniendo consenso en al menos seis nucleótidos y corresponden a TTGACA y TATAAT, respectivamente. Tales sucesos de conservación en las regiones -35 y -10 se muestran como estrictos para promotores de *Bacillus*, en contraste a lo acontecido con *Escherichia coli* donde la distribución es ambigua, por lo tanto, este aspecto simplifica el hallazgo de secuencias promotoras en el microorganismo de interés (Moran et al., 1982).

### **Región 5'-UTR**

Estudios realizados en genes productores de proteínas insecticidas de *Bacillus thuringiensis* exhiben cómo la secuencia ubicada aguas arriba de los marcos de lectura abiertos, además de contener la región promotora, presentan un grupo clave en el extremo 3' del sitio de inicio de la transcripción. Este locus comprende varios pares de bases que, al ser transcritos en ARNm, permanecerían como regiones no traducidas (5'-UTR) compuestas por el sitio de unión del ribosoma y secciones que favorecen a la estabilidad del gen, como una secuencia Shine-Dalgarno consenso dentro del género *Bacillus* (Agaisse & Lereclus, 1996).

Considerando los propósitos de ingeniería genética que incluyen el mejoramiento en el rendimiento de producción de proteínas de *B. thuringiensis*, este conjunto de nucleótidos ubicados en la región no traducida es clave tanto en el entorno cromosómico

como plasmídico para potencializar los niveles de expresión, así como para mantener la estabilidad de unión del ribosoma con el ARNm de interés (Federici et al., 2010).

### **Recombinación homóloga**

La recombinación, como herramienta de modificación genética, se define como un procedimiento basado en la transferencia de una secuencia exógena dentro del genoma de un organismo de interés; sin embargo, cabe destacar la necesidad de ciertas condiciones favorables para su ejecución, caso contrario no causaría efecto pues no se trata de un evento espontáneo. Específicamente, la recombinación homóloga consiste en el alineamiento de secuencias complementarias de tamaño suficiente dar lugar a este fenómeno. Así, en *Bacillus thuringiensis* se han descrito estrategias para la recombinación homóloga considerando la estabilidad de segregación del vector de transporte o la especificidad de la secuencia diana a integrar durante la transformación del microorganismo (Didier Lereclus et al., 1992).

## Capítulo III: Materiales y métodos

### Revisión Bibliográfica

*Bacillus thuringiensis* es una bacteria entomoparásita cuyo uso se ha extendido sobre todo en el control de plagas de insectos en áreas de cultivo abierto y, en biotecnología, como fuente de la proteína Cry para la generación del maíz Bt, un producto transgénico resistente a invasión por insectos. No obstante, los libros y artículos científicos que destacan el uso de las células de este microorganismo en procedimientos de modificación genética son limitados y se centran principalmente en las mencionadas toxinas expresadas por los genes *cry*. Entonces, se procedió a realizar una exploración de material bibliográfico que contenga información sobre los vectores de transformación enfocados específicamente en *B. thuringiensis* y las estrategias de ingeniería genética implementadas para inducir cambios en esta bacteria.

Además, se indagó en publicaciones relacionadas a lipopéptidos, específicamente sobre las kurstakinas reportadas por primera vez en el año 2000, extrayendo esencialmente los datos asociados a su biosíntesis y la incursión en técnicas para potenciar el rendimiento de producción en moléculas similares como las surfactinas en *Bacillus subtilis*.

### Identificación de la estructura consenso del locus *krs*

Dubois y colaboradores (2012) destacan que, considerando la sección oligopeptídica de la molécula de kurstakina y la hipotética ruta biosintética que incluye tres subunidades de sintetasa peptídica no ribosomal, las enzimas correspondientes al locus *krs* serían correspondientes a las secuencias anotadas como BC\_2450 hasta BC\_2458 en el genoma de *Bacillus cereus* ATCC 14579.

Así, se procedió a realizar una búsqueda de cada una de estas secciones mediante el uso de la herramienta BLAST disponible en el sitio web del NCBI tomando

en cuenta varias consideraciones como la existencia de al menos 95% en los parámetros cobertura de query y porcentaje de identidad y la disposición de estos marcos de lectura abiertos de manera contigua en los genomas donde se determinó correspondencia.

### ***Descubrimiento de los genes involucrados en genomas de microorganismos productores de kurstakinas***

Los genes utilizados como objetivo de búsqueda se sometieron a la herramienta BLAST para compararse con los genomas de *Bacillus thuringiensis* serovar *kurstaki* HD-1 (Hathout et al., 2000) y *Bacillus thuringiensis* Bt407 strain 407 (Abderrahmani et al., 2011), en los que se ha reportado la producción de kurstakina. Los genes anotados como NRPS responsables de la síntesis de otros lipopéptidos, cuya distribución en el genoma de los microorganismos de interés fue caracterizada por sobre posición, se organizaron en un esquema para comparar si su longitud es consistente con la esperada en el biosíntesis de kurstakina y para ubicar los genes asociados tanto en el extremo 5' como en el 3'.

### ***Caracterización de dominios y sustratos de los presuntos genes del locus *krs****

Béchet y colaboradores (2012) proponen en su trabajo la disposición que tendrían los módulos de NRPS dentro de las proteínas principales del locus: *krsA*, *krsB* y *krsC*; además, coinciden con lo propuesto por Hathout y su equipo (2000), quienes reportaron la estructura de kurstakina como un oligopéptido conformado por los siguientes aminoácidos: D-Thr/Gly/Ala/Ser/His/D-Gln/Gln. Por lo tanto, tras la identificación de potenciales genes del locus *krs* obtenidos por BLAST se tomaron las secuencias de aminoácidos de sus respectivas proteínas para someterlas al análisis en los distintos softwares observados en la Tabla 2, enfocado en determinar la existencia de módulos de NRPS, sus dominios y posibles sustratos. La búsqueda incluye a las variantes HD-1 y Bt407, mencionadas previamente, y las NRPS de referencia en generación de kurstakinas

halladas en la base de datos NORINE (Flissi et al., 2020) pertenecientes al microorganismo *Bacillus thuringiensis* serovar *pondicheriensis* BSCG 4BA1.

**Tabla 2.**

*Programas bioinformáticos utilizados para predecir dominios NRPS.*

<b>Software</b>	<b>Descripción</b>	<b>Cita</b>
NRPSsp	Base de datos de módulos de NRPS elaborada mediante minería computacional utilizada para construir un predictor que funciona con base en modelos ocultos de Markov para hallar NRPS y sus potenciales sustratos.	(Prieto et al., 2012)
PKS/NRPS Analysis	Búsqueda BLAST que compara la secuencia con la NRPS de gramicidina A y utiliza métodos estadísticos refinados para establecer dos grupos de residuos críticos que definen cada dominio.	(Bachmann & Ravel, 2009)
Domain Search Program for NPRS-PKS	Compara las secuencias de 167 NRPS caracterizadas experimentalmente (aproximadamente 4400 dominios catalíticos) y otros dominios inusuales con una secuencia de búsqueda.	(Anand et al., 2010)
antiSMASH	Minería genómica que compara secuencias de NRPS convencionales y atípicas capaz de predecir sintetisas en genomas sin anotación.	(Blin et al., 2019)

### ***Concatenación de secuencias, alineamiento múltiple y construcción de un árbol filogenético***

El objetivo de verificar la existencia del locus *krs* en los genomas de otras especies del género *Bacillus* fue puesto en marcha por Gélis-Jeanvoine y colaboradores (2016) realizando una búsqueda con BLAST y utilizando su programa BLAStats para mantener las siguientes consideraciones: al menos 80% de identidad de la secuencia y 85% de

cobertura de query; esto con el fin de evitar el hallazgo de proteínas que no clasifiquen como NRPS. Entonces, se extendió la propuesta para incluir los genomas disponibles hasta 2020 (definiendo las condiciones en el propio interfaz de BLAST) y el ADN obtenido a partir del genoma de las especies afines al procedimiento anterior, correspondiente a cada gen de *krs*, se organizó para realizar un alineamiento múltiple de secuencias en el software MEGA7 a través del modelo ClustalΩ manteniendo los parámetros por defecto de 15 puntos de penalidad por apertura de gap y 6.6 puntos de penalidad por extensión de gap.

Una vez culminado el alineamiento múltiple se procedió a concatenar las secuencias de los seis genes involucrados y estimar el modelo evolutivo de máxima verosimilitud que mejor se acople a las variaciones halladas en las especies seleccionadas. Con base en esta información se llevó a cabo la curación del alineamiento con el software BMGE, asociado al flujo de trabajo que reposa en el sitio <https://ngphylogeny.fr> del Institut Pasteur, y la construcción de un árbol de máxima verosimilitud utilizando el programa RAxML-HPC Blackbox en los servidores de CIPRES.

#### **Anotación del plásmido BBa\_K802004**

##### ***Selección del plásmido***

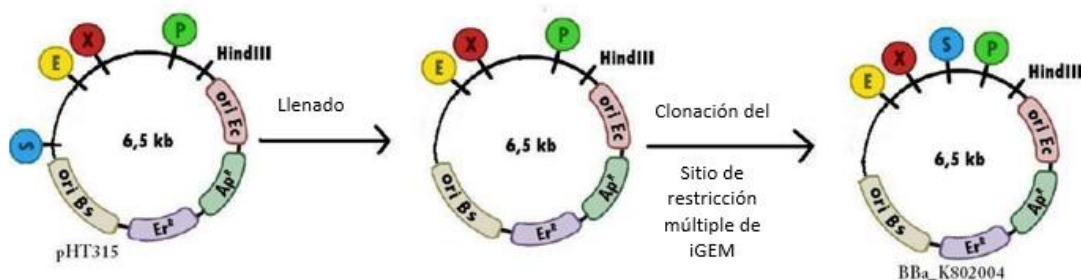
Los vectores de transformación de *Bacillus thuringiensis*, como lo destacan Ochoa-Zarzosa & López-Meza (2012), se originan principalmente a partir de la combinación de plásmidos: una parte proveniente del propio microorganismo y la otra de fragmentos originarios de *Escherichia coli* para favorecer al establecimiento de la secuencia deseada en esta última bacteria que actúa como chasis. Así, del grupo de vectores mencionados en tal trabajo se han seleccionado los elaborados por Arantes & Lereclus (1991) pHT304, pHT315 o pHT370; entonces, se efectuó una búsqueda del

mapa de la secuencia de cualquiera de estos plásmidos sin éxito en repositorios académicos o comerciales.

La base de datos de partes en el sitio web de iGEM pone a disposición la secuencia cruda del vector BBa\_K802004, creado por loana del equipo iGEM12\_Lyon\_INSA (2012), que es un plásmido derivado de pHT315 con una modificación que traslada el sitio de restricción de la enzima SpeI dentro del polylinker como se observa en la Figura 2, sin presentar alteraciones en los orígenes de replicación o los genes de resistencia a antibióticos.

### Figura 2.

Vector BBa\_K802004 que cuenta con una secuencia similar a pHT315, plásmido para clonación en *Bacillus thuringiensis* (loana, 2012).



### Búsqueda de ORFs y definición de genes presentes en el vector

Posteriormente, se introdujo la secuencia de BBa\_K802004 (al que a partir de ahora se denominará pHT315-1 en este documento) en el software Benchling donde se estableció el mapa del esqueleto del vector, entonces, se procedió a realizar su respectiva anotación para constatar que efectivamente cuenta con los genes necesarios para ejercer la transformación molecular.

La anotación fue lograda a partir del uso de distintas herramientas bioinformáticas. En primer lugar, se llevó a cabo una predicción de marcos de lectura abiertos a través de



ORF Finder del NCBI y a la vez se corrió el programa GeneMarks2 utilizando la secuencia completa como input para predecir genes con codones de inicio convencionales y atípicos con el fin de evitar la omisión de algún elemento importante.

Además, el origen de replicación de *B. thuringiensis* se encontró mediante alineamiento múltiple con las secuencias de referencia obtenidas de la base de datos DoriC (Luo & Gao, 2019), mientras que el correspondiente a *E. coli* se determinó por alineación con la región ORI del plásmido pUC19 obtenido del repositorio comercial Addgene.

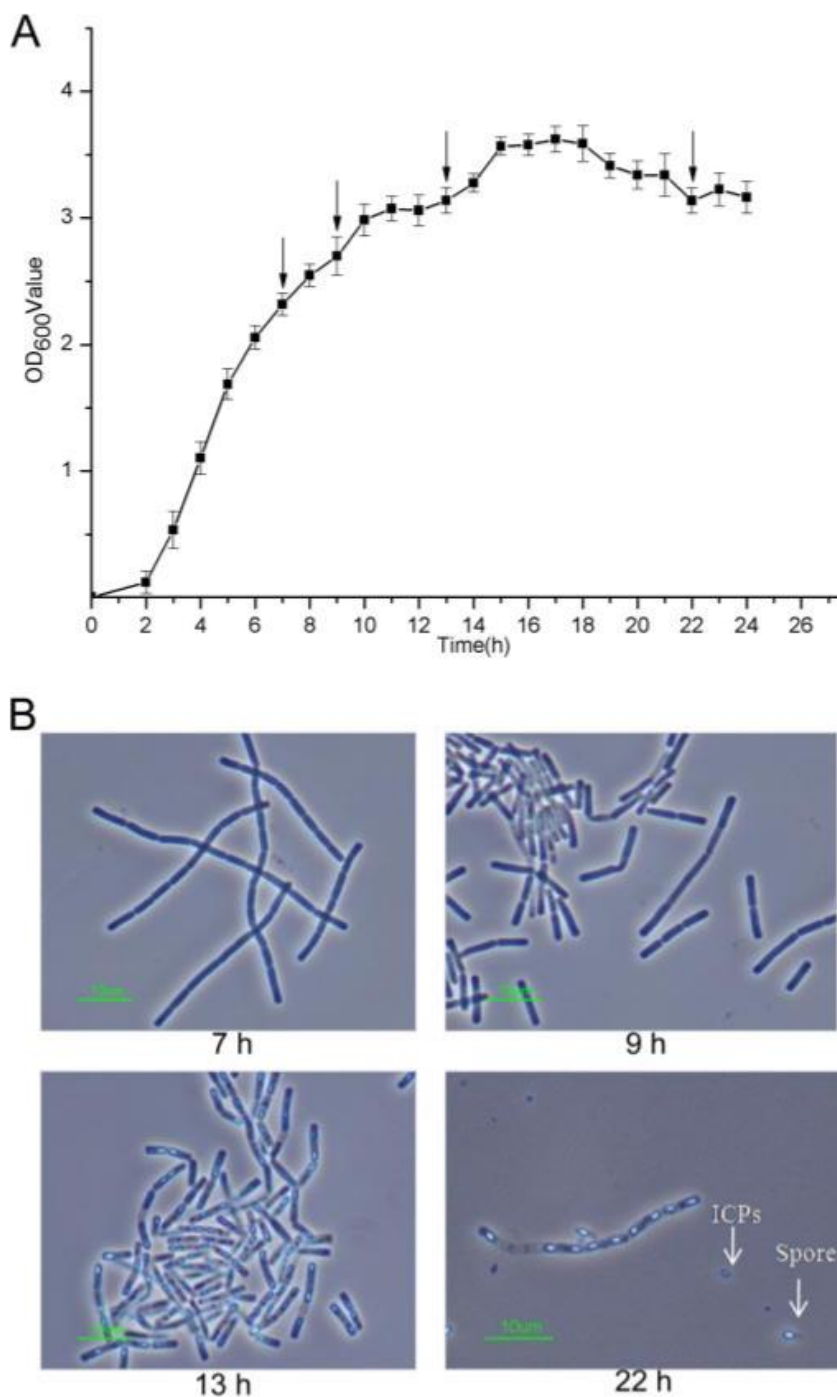
Finalmente, los genes y regiones de interés identificados se sometieron a una búsqueda en BLAST para comprobar su correspondencia a los genes buscados y se concluyó con las respectivas anotaciones en Benchling.

### **Análisis transcriptómico de RNA-seq proveniente de *Bacillus thuringiensis* serovar *chinensis* CT-43**

El análisis transcriptómico, como lo mencionan Hoeijmakers y colaboradores (2012), se presenta como una alternativa frente a los microarreglos para la estimación de los niveles de expresión de múltiples genes brindando un campo de visualización más extenso en vista de que no requiere de un conocimiento previo de las secuencias de los genes y hace posible, entre otras distintas aplicaciones, la selección de promotores y demás regiones reguladoras de expresión para potencializar o reprimir la producción de una proteína de interés. Por lo tanto, se tomaron los productos de la secuenciación de ARNm→ADNc proporcionados por Wang y su grupo de investigación (2013) pertenecientes a muestras de un cultivo de *Bacillus thuringiensis* subsp. *chinensis* CT-43 tomadas en diferentes etapas de crecimiento del microorganismo (Figura 3) a las 7 h, 9 h, 13 h y 22 h, correspondientes a las fases exponencial media, estacionaria temprana, estacionaria media y esporulación media, respectivamente.

**Figura 3.**

Crecimiento de *Bacillus thuringiensis* serovar *chinensis* CT-43. A: Curva de crecimiento de la bacteria obtenido por densidad óptica a 600nm. B: Imágenes de las bacterias en el punto de tiempo al que fueron muestreadas. Replicado de (Wang, Mei, et al., 2013).



Los archivos provenientes de la secuenciación (RNA-seq) fueron obtenidos de la base de datos European Nucleotide Archive – ENA bajo los números de acceso de SRR524792 hasta SRR524795 como secuencias crudas y fueron transferidas a una historia de trabajo nueva en los servidores del proyecto Galaxy, que permiten llevar a cabo análisis bioinformáticos de manera integral en una sola plataforma.

Los documentos RNA-seq, en formato .fastq, fueron sometidos a un análisis de calidad utilizando el programa FASTQC, donde fue posible deducir la presencia de adaptadores, el contenido GC y valoraciones de calidad en la secuencia. Inmediatamente, se procedió a remover los adaptadores propios del procedimiento de secuenciación Illumina por medio del software Trimmomatic para tener a disposición únicamente las lecturas que serán alineadas.

Las lecturas de todas las muestras (7H, 9H, 13H y 22H), considerando su pertenencia a *B. thuringiensis* CT-43 cuyo genoma está ensamblado y disponible en repositorios de acceso libre, pueden saltarse el procedimiento de ensamblaje *de novo* y pasar a un mapeo enfrentándose a un genoma de referencia. Los programas TopHat2, HiSat2 o STAR mapper se destacan por identificar potenciales exones para asociarlos con las lecturas de RNA-seq provenientes de organismos eucariotas; sin embargo, el presente estudio consideró la información proporcionada por el tutorial de uso de Galaxy Project y las recomendaciones impartidas en el foro de entrenamiento en bioinformática BioStars (2018) y se recurrió al manejo del software Bowtie2 (Langmead & Salzberg, 2012) que permite omitir la condición de splicing y facilita el trabajo con bacterias. El programa precisa de un segundo input, el genoma de referencia, que fue descargado en formato .fasta desde Genbank del NCBI con el número de acceso CP001907.1.

Los output del mapeo de lecturas se arrojan en formato SAM/BAM que a continuación deben ser sometidos al conteo de transcritos, entonces, a pesar de la existencia de programas como Cufflinks (descontinuado) o StringTie que brindan

estimaciones de nivel de expresión en términos de RPKM (lecturas por kilobase de exones por millón de lecturas mapeadas) o FPKM (fragmentos por kilobase por millón de lecturas mapeadas); se prefirió el uso de recuento de genes basado en lecturas – en vista de la carencia de eventos de splicing – con el software HTseq-count (Anders et al., 2015). Los inputs necesario por HTseq-count incluyeron al archivo BAM de cada muestra, la anotación del genoma de *B. thuringiensis* CT-43 (obtenido en formato GFF3 desde Genbank) y la dirección del gen (positiva o negativa) que se marcó como unknown para ser determinada durante la ejecución, además, el resto de parámetros se mantuvieron como default (Nekrutenko, 2020).

Consecuentemente, se pasó a un análisis de normalización de cuentas y el cómputo de expresión diferencial mediante la aplicación de la secuencia de comandos del programa DESeq2. Cabe destacar que la interfaz permite la incorporación de varios factores a comparar en distintos niveles determinados por el usuario; así, se colocó un único factor denominado GrowthStages (inglés para etapas de crecimiento) y se compararon dos niveles definidos como las dos fases principales del desarrollo de *Bacillus thuringiensis* que son la crecimiento vegetativo (denotado como “exp”) y la esporulación (denotado como “stat”) (Nekrutenko, 2020).

Para finalizar, se enfrentaron los outputs del recuento con HTseq-count y de expresión diferencial con DESeq2 (Love et al., 2014) junto al documento de anotación del genoma en el software Annotate DESeq2/DEXSeq buscando contar con datos como la posición de inicio y fin del gen, el nombre con el que cada uno se encuentra anotado en Genbank y su producto proteínico correspondiente (Nekrutenko, 2020).

Los resultados del análisis transcriptómico se utilizaron como fuentes primarias de información sobre potenciales regiones capaces de brindar regulación positiva a la expresión de kurstakinas considerando la cantidad de lecturas mapeadas, la variación de expresión entre las etapas del crecimiento y una actividad constitutiva.

## **Establecimiento del constructo génico**

### ***Definición de una región potenciadora dentro de B. thuringiensis***

La secuencia intergénica aguas arriba del gen que mostró el mejor perfil de transcripción se extrajo desde el genoma CP001907.1 en Genbank y se continuó con una serie de análisis en diferentes programas enfocados en la predicción de promotores y sitios de unión de elementos reguladores.

Para empezar, se llevó a cabo el análisis de la secuencia seleccionada en BProm, un predictor de promotores que arroja la posición más probable del sitio de inicio de transcripción (TSS) y sus regiones -35 y -10 asociadas (Solovyev et al., 2011). Al mismo tiempo, se sometió la secuencia a la ejecución de los programas iProEP, Berkeley Promoter Predictor e iPro70 generando los siguientes efectos: 1) evaluación de grupos de 81 pares de bases tomados de la secuencia input calculando la probabilidad de que sea un promotor; 2) el sitio más probable donde se halla una región promotora incluyendo su posición de inicio y final; y, 3) determinación de la probabilidad de unión del factor sigma 70 en la secuencia en estudio; respectivamente.

Finalmente, el sistema iPSW fue usado para verificar si el sitio seleccionado pertenece a un promotor fuerte o débil mediante la comparación con una base de datos, mientras al inquirir en el software DBTBS se comprobó la posible existencia de lugares para la unión de distintos factores sigma u otros elementos regulatorios.

### ***PCR in silico de los sitios de interés***

La región seleccionada para el análisis de promotor y otras características reguladoras fue, en efecto, elegida como un potencial elemento útil para mejorar el rendimiento en la producción de kurstakina. De este modo, el siguiente paso consistió en cargar al sistema Benchling el genoma de *Bacillus thuringiensis* serovar *galleriae* strain HD-29, reportada en sitios contaminados por compuestos de petróleo en Ecuador por

Maddela y sus colaboradores (2015), para diseñar primers específicos y simular una PCR *in silico* para determinar la viabilidad su uso.

Este procedimiento se realizó tanto para la región potenciadora así como para el gen *krsE* con el fin de integrar ambas secuencias al vector de transformación y permitir el reemplazo del sitio regulador a través de recombinación homóloga.

### ***Digestión y ensamblaje del constructo***

La observación del polylinker perteneciente al vector pHT315-1 considerando enzimas de un solo corte en la secuencia completa del plásmido condujo a la selección de cuatro enzimas de restricción cuyo resultado brinda un extremo pegajoso o 'sticky end': HindIII, SpeI, XbaI y EcoRI.

Los primers utilizados para la amplificación del promotor incluyeron en su secuencia satélite los sitios de restricción de HindIII (AAGCTT) y SpeI (ACTAGT); mientras que los aplicados en la PCR del gen *krsE* contenían los sitios de las enzimas XbaI (TCTAGA) y EcoRI (GAATTC). Enseguida, se procedió a digerir tanto a los amplicones como al vector y se ensamblaron los elementos de interés.

### **Simulación del proceso de recombinación homóloga**

El propósito en este punto radica en reemplazar la secuencia ubicada aguas arriba del locus *krs* para favorecer a la potencialización de rendimiento de expresión, en consecuencia, la inclusión del gen *krsE* permitiría que una vez ejercida la transformación de las células de *Bacillus thuringiensis* con el plásmido pHT315-1\_PrlpU\_*krsE* sea posible la integración de este constructo en el cromosoma bacteriano mediante recombinación homóloga.

Por lo tanto, se continuó con la utilización del software Molecular Cloning Designer Simulator MCDS que, además de mostrar cómo se desempeña clonación molecular *in silico*, incluye algoritmos para mostrar combinación de secuencias, ensamblaje Gibson,

CRISPR y recombinación (Shi & Vickers, 2016); así se suministraron el fragmento de genoma que incluye el locus *krs* con un par de kilobases extra en ambos extremos y el vector ensamblado y se seleccionaron los parámetros para establecer el escenario de una recombinación homóloga.

### **Planteamiento del procedimiento a realizar en laboratorio**

El contenido teórico revisado a lo largo de este trabajo fue combinado para sintetizar el protocolo ejecutable en laboratorio, concibiendo los procedimientos que abarcan desde el aislamiento de *Bacillus thuringiensis* hasta las pruebas de rendimiento en la producción de kurstakina con proyección a su aplicación en sitios contaminados por compuestos derivados de petróleo como elemento en un consorcio bacteriano.

## Capítulo IV: Resultados

### Secuencia consenso del locus *krs*

La búsqueda en BLAST por separado de los genes de *Bacillus cereus* ATCC 14579 que van desde BC\_2450 hasta BC\_2458 arrojó coincidencias con regiones correspondientes a NRPSs que intervienen en la producción de diferentes lipopéptidos como iturina, bacitracina, gramicidina, entre otros. Después, al limitarse la investigación con las variantes HD-1 y Bt407 de *Bacillus thuringiensis* se hallaron seis genes dispuestos a manera de operón que coinciden con las NRPS propuestas para la biosíntesis de kurstakinas.

Las secuencias encontradas pueden visualizarse en el esquema presentado en la Figura 4, donde se aprecia una anotación indefinida, pero que mantienen una organización propicia: una proteína permeasa que facilite la salida del péptido a sintetizar, tres sintetetas peptídicas no ribosomales distribuidas de acuerdo a su cantidad de módulos como 1 módulo NRPS, 2 módulos NRPS y 4 módulos NRPS, respectivamente; también, se halla una enzima fosfopanteteinil transferasa que activa los dominios de acetilación de cada módulo para la integración de aminoácidos por medio de fosforilación y, por último, una tioesterasa que hace posible la liberación del lipopéptido.

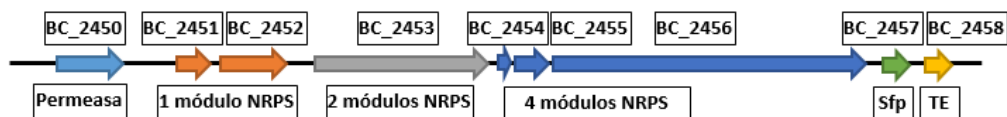
El fragmento completo con una longitud de alrededor de 29kb, mediante una nueva corrida en BLAST permitió identificar la presencia del locus en bacterias del grupo *Bacillus cereus*, principalmente en *B. cereus* y *B. thuringiensis*. La deducción pertinente al hallazgo es que, efectivamente, los seis genes constituyentes del operón corresponden a *krsE*, *krsA*, *krsB*, *krsC*, *sfp* y *te*; protagonistas en la biosíntesis de kurstakinas, como se describe en estudios previos.



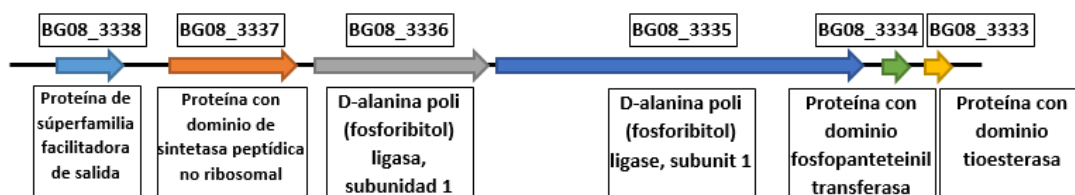
**Figura 4.**

*Locus krs identificado en el genoma de Bacillus thuringiensis, destacando una anotación ambigua en cada gen que conforma el operón. Basado en lo propuesto por (Dubois et al., 2012).*

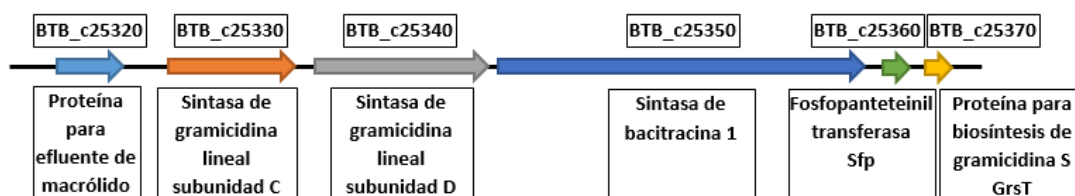
***Bacillus cereus* ATCC 14579**



***Bacillus thuringiensis* serovar *kurstaki* HD-1**



***Bacillus thuringiensis* Bt407 strain 407**



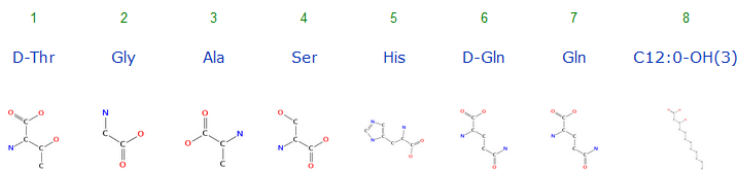
**Módulos y sustratos de las NRPS definidas para kurstakinas**

La secuencia de aminoácidos obtenida a partir de los genes previamente identificados (*krsA*, *krsB* y *krsC*) fue analizada para determinar sus propiedades para la formación del lipopéptido deseado, cuya estructura puede apreciarse en la Figura 5 obtenida de la base de datos NORINE y donde se ven los siete aminoácidos unidos a una cadena lipídica de doce carbonos y la formación de un anillo entre los residuos de histidina y serina.

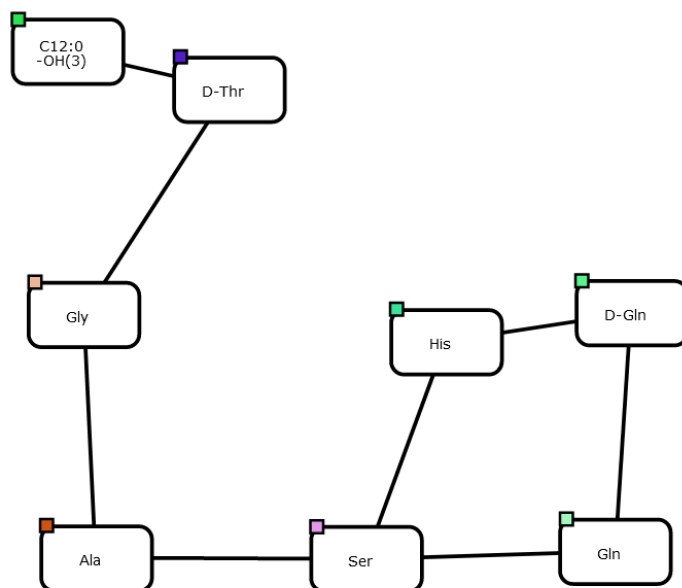
**Figura 5.**

*Estructura de kurstakina generada en la base de datos NORINE.*

▪ **Monomeric composition :**



▪ **Graph representation:** D-Thr,Gly,Ala,Ser,His,D-Gln,Gln,C12:0-OH(3) @1,7 @0,2 @1,3 @2,4,6 @3,5 @4,6 @3,5 @0



El resultado de la ejecución de los distintos softwares de predicción de NRPS arrojaron los datos mostrados en la Tabla 3. Con respecto a los dominios de cada módulo se obtuvo que todos contienen un sitio de condensación (reconocimiento de un aminoácido o la cadena lipídica), adenilación (identificación e integración del aminoácido específico a la molécula) y tiolación (liberación de la cadena favoreciendo su avance al siguiente módulo); además, en las regiones 1 y 6 se halla un dominio de epimerización (cambio de sentido de enantiómeros de levógiro a dextrógiro) y la región final cuenta con un dominio de tioesterasa (liberación de la molécula).

**Tabla 3.**

*Módulos hallados en las potenciales enzimas KrsA, KrsB y KrsC mediante análisis bioinformático.*

Enzima	KrsA			KrsB			KrsC														
Módulos	Módulo 1	Módulo 2	Módulo 3	Módulo 4	Módulo 5	Módulo 6	Módulo 7														
<b><i>Bacillus thuringiensis</i> serovar <i>kurstaki</i> HD-1</b>																					
Resultado	C	A	T	E	C	A	T	C	A	T	C	A	T	C	A	T	E	C	A	T	TE
<b><i>Bacillus thuringiensis</i> Bt407 strain 407</b>																					
Resultado	C	A	T	E	C	A	T	C	A	T	C	A	T	C	A	T	E	C	A	T	TE
<b><i>Bacillus thuringiensis</i> serovar <i>pondicheriensis</i> BSCG 4BA1 (UniProt accession numbers: C3GUG5, C3GUG6 &amp; C3GUG7)</b>																					
Resultado	C	A	T	E	C	A	T	C	A	T	C	A	T	C	A	T	E	C	A	T	TE

Nota: Basado en los resultados obtenidos desde PKS/NRPS Analysis, Domain Search Program for NRPS & antiSMASH. Abreviaturas: C-Condensation, A-Adenylation, T-Thiolation, E-Epimerization & TE-Thioesterase.

Simultáneamente, en la Tabla 4 se exhibe cómo se logró pronosticar los posibles sustratos del dominio de adenilación en cada módulo, es decir, los aminoácidos específicos correspondientes a la estructura del lipopéptido. Cabe destacar la existencia de cierta inconsistencia en las tres últimas posiciones, sin embargo, se mantiene el patrón de generación del lipoheptapéptido esperado en la estructura de kurstakina.

De manera adicional, es importante mencionar el papel del programa antiSMASH, pues realiza un sondeo integral a la secuencia del genoma y por tal motivo se ratifica que el locus seleccionado es el único consistente con la descripción de biosíntesis de kurstakinas.

**Tabla 4.**

*Aminoácidos que serían utilizados como sustratos específicos por los módulos de las NRPS de kurstakina en Bacillus thuringiensis.*

Software	Módulos NRPS						
	Módulo 1	Módulo 2	Módulo 3	Módulo 4	Módulo 5	Módulo 6	Módulo 7
<b><i>Bacillus thuringiensis</i> serovar <i>kurstaki</i> HD-1</b>							
NRPSsp	D-Thr	Ser	Orn	Ser	Leu	D-Orn	Orn
PKS/NRPS Analysis	D-Thr	Gln	Gly	Ser	No Hit	No Hit	Gln
AntiSMASH	D-Thr	No Hit	Ala	Ser	Leu	D-Gln	Gln
<b><i>Bacillus thuringiensis</i> Bt407 strain 407</b>							
NRPSsp	D-Thr	Ser	Orn	Ser	Leu	D-Orn	Orn
PKS/NRPS Analysis	D-Thr	Gln	Gly	Ser	No Hit	D-Gln	Gln
AntiSMASH	D-Thr	No Hit	Ala	Ser	Leu	D-Gln	Gln
<b><i>Bacillus thuringiensis</i> serovar <i>pondicheriensis</i> BSCG 4BA1 (NORINE)</b>							
NRPSsp	D-Thr	Ser	Orn	Ser	Leu	D-Orn	Orn
PKS/NRPS Analysis	D-Thr	Gln	Gly	Ser	No Hit	D-Gln	Gln
AntiSMASH	D-Thr	No Hit	Ala	Ser	Leu	D-Gln	Gln

Nota: Basado en los resultados obtenidos desde NRPSsp, PKS/NRPS Analysis & antiSMASH. Abreviaturas de aminoácidos: Thr-Treonina, Ser-Serina, Orn-Ornitina, Leu-Leucina, Gln-Glutamina, Gly-Glicina, Ala-Alanina. La descripción "No Hit" hace referencia a que el software no detectó un sitio activo asociado a la integración de un aminoácido específico. La descripción (D-) previa a la abreviatura de algunos aminoácidos hace referencia al cambio de conformación, de levógiro a dextrógiro, ejercida por el dominio de epimerización en los módulos 1 y 6 de la NRPS.

#### **Locus *krs* distribuido en el grupo *Bacillus cereus* (BLAST 2020) y árbol filogenético**

La herramienta BLAST permite especificar entre sus parámetros de búsqueda la inclusión o exclusión de grupos taxonómicos señalados por el usuario. Así, tras la ausencia de un porcentaje de identidad significativo con especies ajenas al grupo *Bacillus*

*cereus*, la Tabla 5 destaca los resultados alcanzados al indagar sobre la existencia del operón *krs* en los microorganismos pertenecientes a la agrupación taxonómica, de manera que los datos confirman como las NRPS relacionadas a la biosíntesis de kurstakina se distribuyen arbitrariamente solo en ciertas especies.

**Tabla 5.**

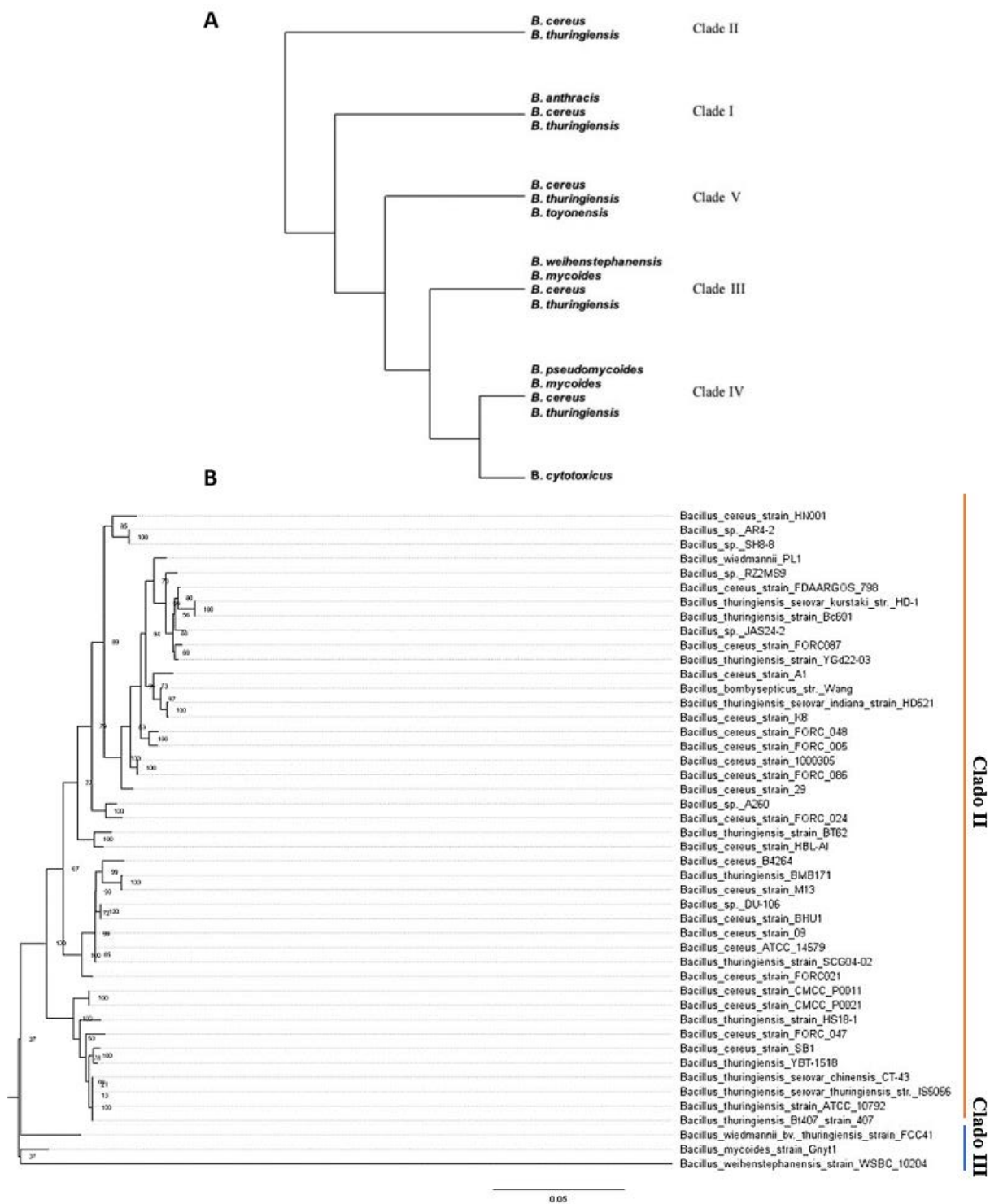
*Distribución del locus krs en las especies del grupo Bacillus cereus. Basado en el procedimiento realizado con el programa BLASTats por (Gélis-Jeanvoine et al., 2016).*

<b>Especie</b>	<b>Presente/Ausente</b>
<i>Bacillus anthracis</i>	Ausente
<i>Bacillus cereus</i>	Presente (≥ 30 genomas)
<i>Bacillus cytotoxicus</i>	Ausente
<i>Bacillus mycoides</i>	Presente (1 genoma y reportado en varios plásmidos)
<i>Bacillus pseudomycooides</i>	Presente (2 genomas, * < 85% de cobertura de query)
<i>Bacillus thuringiensis</i>	Presente (≥ 30 genomas)
<i>Bacillus toyonensis</i>	Ausente
<i>Bacillus weihenstephanensis</i>	Presente (1 genoma y reportado en varios plásmidos)

Las variantes de las especies *B. cereus*, *B. thuringiensis*, *B. mycoides* y *B. weihenstephanensis* donde se encontró el locus *krs* fueron tomadas en cuenta para construir el árbol filogenético expuesto en la Figura 6. El programa Mega7 estimó que el modelo que mejor se adecúa a la evolución del locus es GTR+G+I (incluyendo distribución gamma y proporción de sitios invariables) luego de haber concatenado los alineamientos múltiples de cada gen. Consecuentemente, a través del software RAXML-HPC Blackbox ejecutado en la plataforma CIPRES, se obtuvo un árbol de máxima verosimilitud organizado en la plataforma FigTreev1.4.4 y que muestra la conservación del operón dentro del grupo *B. cereus*, específicamente limitado al clado II de la clasificación propuesta por Ehling-Schulz y colaboradores (2019), considerando que los outgroups *B. mycoides* y *B. weihenstephanensis* son los únicos del clado III y teniendo en cuenta que la presencia del conjunto de genes ha sido reportada en plásmidos se podría esperar que lo hayan adquirido en eventos de transmisión horizontal.

Figura 6.

Árbol filogenético.



Nota: A) Cinco clados principales en la división filogenética del grupo *Bacillus cereus* propuesta por (Ehling-Schulz et al., 2019), las especies están íntimamente asociadas y

su clasificación se basa en parámetros morfológicos, fisiológicos y genómicos. B) Árbol de máxima verosimilitud construido con base en el alineamiento múltiple del operón *krs*, donde se aprecia la exclusividad de la NRPS de kurstakina en especies del clado II, mientras las especies del outgroup se hallan en el clado III. Basado en la propuesta previa de (Gélis-jeanvoine et al., 2016).

### **pHT315-1, vector anotado**

La Figura 7 indica el mapa de pHT315-1 anotado en Benchling con las partes inferidas de los procesos que se describen a continuación.

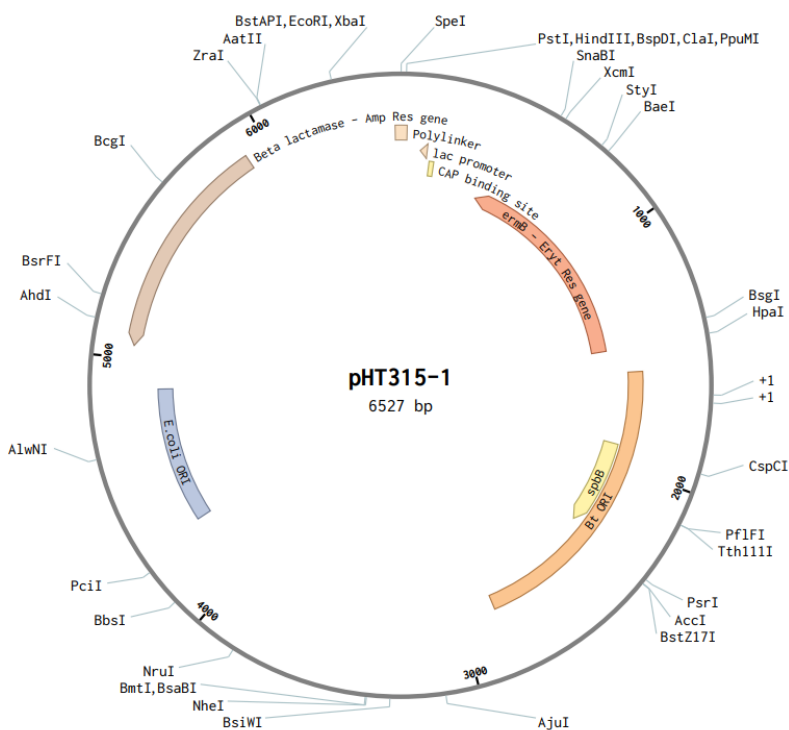
El paso de la secuencia completa de la parte BBa\_K082004 de iGEM por la herramienta ORF Finder arrojó 62 predicciones de marco de lectura basándose en la identificación de codones de inicio; sin embargo, el resultado principal fue un fragmento de 861bp, el cual fue sometido a búsqueda en BLAST y coincidió con el gen de  $\beta$ -lactamasa encargado de brindar resistencia a ampicilina.

Por otro lado, la prospección efectuada con ayuda del programa GeneMarks2 tuvo dos frutos importantes: el gen de resistencia a eritromicina *ermB* (erythromycin resistance methylase B) y el gen *spbB* que es parte del origen de replicación de *B. thuringiensis* y favorece a la estabilidad segregacional del plásmido.

En última instancia, se compararon las secuencias de pHT315-1 y pUC19 en SnapGene obteniendo varios elementos como el origen de replicación de *E. coli*, el sitio de unión de la proteína activadora de catabolitos CAP, promotor y operador del operón *lac* y el sitio de restricción múltiple o polylinker. También, el alineamiento con las secuencias de ori obtenidas de la base de datos DoriC hizo posible definir el ori de *B. thuringiensis*.

**Figura 7.**

*Plásmido pHT315-1 anotado.*



### **Recuento de lecturas y expresión diferencial en distintas fases de crecimiento de *Bacillus thuringiensis***

La existencia del locus *krs* se ha establecido en el genoma de las especies *B. thuringiensis* y *B. cereus*; por lo tanto, *Bacillus thuringiensis* serovar *chinensis* CT-43 se reconoce como microorganismo productor de kurstakinas. Así, se utilizaron los archivos de secuenciación RNA-seq proporcionados por Wang y colaboradores (2013) para llevar a cabo un análisis transcriptómico y, en la Tabla 6, se destacan algunos de los genes que tuvieron elevados niveles de expresión en los distintos puntos de muestreo incluyendo una comparación con el gen *krsC* que, con una longitud de 14kb, presentara la cantidad más alta en el recuento de lecturas mapeadas entre los elementos del operón de interés.



**Tabla 6.**

Niveles de expresión en base al recuento de lecturas mapeadas en HTSeq-count.

Comparación de los mejores perfiles para reemplazo de promotor.

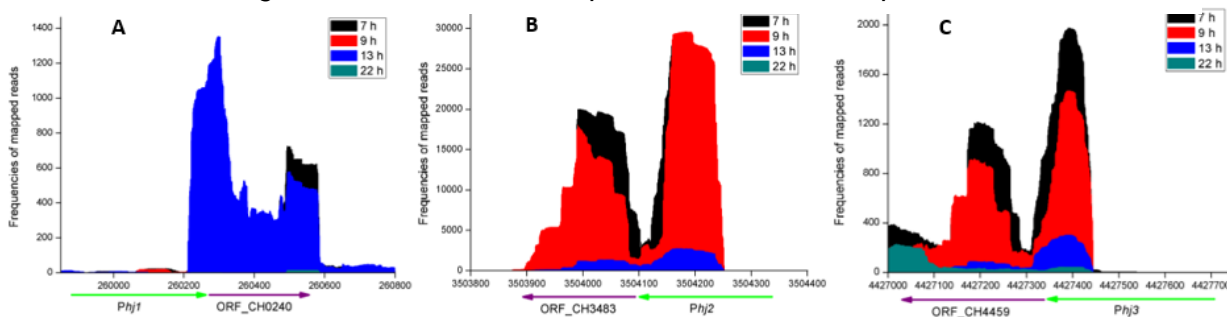
Gen	Recuento de lecturas mapeadas (HTSeq-count)				Producto
	7 h	9 h	13 h	22 h	
<i>krsC</i>	16	92	74	33	NRPS: KrsC
<i>rplU</i>	2258	1705	291	282	50S ribosomal protein L21
<i>cspB2</i>	22241	21324	1723	9	Cold shock protein
<i>groES</i>	1707	418	2242	16	Chaperonin GroES

El nivel de expresión de *krsC* fue relativamente bajo en comparación a los obtenidos por los otros genes elegidos, por tal motivo, se pretende utilizar la secuencia promotora de una de estas opciones para reemplazar la correspondiente al operón *krs* para potencializar el rendimiento de producción de kurstakinas. A pesar de que *cspB2* parece tener buen desempeño de expresión, es importante seleccionar una región reguladora capaz de mantener la generación de NRPS de manera estable a lo largo del crecimiento de *Bacillus thuringiensis* con el fin de lograr un rendimiento de kurstakina sostenido a lo largo de su aplicación en un consorcio bacteriano, así como en un posible escalamiento para la producción industrial en un biorreactor; entonces, basándose en la Tabla 6 y la Figura 8, el promotor perteneciente al gen *rplU* se presenta como una opción para la generación estable de NRPS de kurstakina.

**Figura 8.**

Número de lecturas mapeadas a través de las fases de crecimiento (Wang et al., 2013).

A. Promotor de *groES*. B. Promotor de *cspB2*. C. Promotor de *rplU*.



El crecimiento de *B. thuringiensis* está definido por dos etapas esenciales: vegetativa y esporulación; motivo por el que se consideró este factor para llevar a cabo un análisis de expresión diferencial utilizando DESeq2 en Galaxy. La Tabla 7 resume los datos que evalúan la variación en la expresión de los genes cuyos promotores se han tomado en cuenta como candidatos para reemplazar a la secuencia reguladora del operón *krs*.

**Tabla 7.**

*Resultados del análisis de expresión diferencial en DESeq2. Únicamente se incluyen los genes cuyo promotor se considera apto para su uso en el constructo.*

<b>Gen</b>	<b>Media</b>	<b>Log<sub>2</sub>(FC)</b>	<b>StdErr</b>	<b>Wald-Stats</b>	<b>P-value</b>	<b>P-value adj</b>
<i>rlpU</i>	907.714	-1.192	1.465	-0.813	0.416	0.655
<i>cspB2</i>	7496.389	-3.749	1.659	-2.26	0.024	0.105
<i>groES</i>	839.634	0.474	1.669	0.284	0.776	0.899

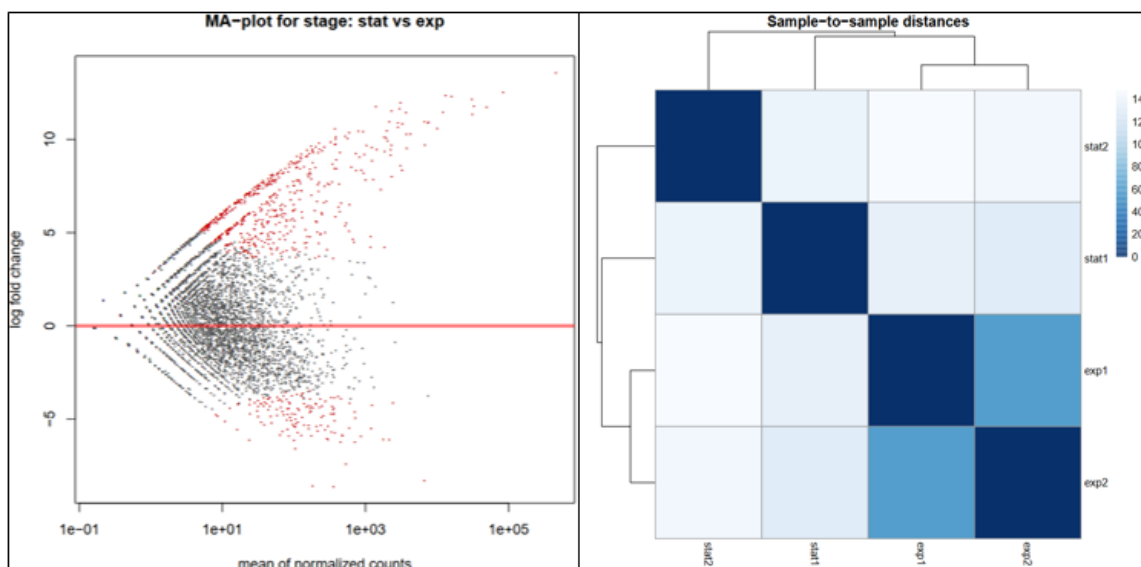
En resumen, la información expuesta en la tabla anterior está sujeta a las siguientes consideraciones: la media se calculó considerando las lecturas mapeadas en cada punto de muestreo, el logaritmo en base 2 expresa la magnitud del cambio en los niveles de expresión existente entre las etapas de crecimiento vegetativo y esporulación, la prueba estadística de Wald determina qué tan extrema es la diferencia entre un valor esperado y el observado (funciona como una prueba  $\chi^2$  con 1 grado de libertad), el valor p estima la significancia del cambio generado y el valor p ajustado a un proceso de varios análisis controla la tasa false discovery rate (FDR). Con fundamento en los detalles planteados, el promotor de *rlpU* sigue siendo la mejor opción en vista de que su factor de cambio (variación de expresión en las etapas de crecimiento con respecto a la media) es bajo manteniendo un rendimiento promedio alto, la prueba Wald no muestra una variación

importante o extrema y el valor p de significancia – aunque no es el mejor – permanece aceptable para el propósito de este estudio.

Finalmente, en la Figura 9 se aprecia como la diferencia de expresión entre etapas es extensa (MA-plot), motivo por el que se toman únicamente los tres genes con mejor proyección de estabilidad; por otro lado, se evidencia cómo existen variaciones entre las muestras de fase de esporulación (heatmap), lo que hace aún más trascendental la selección de un promotor cuya actividad sea independiente del periodo de crecimiento de la bacteria.

**Figura 9.**

Gráficos generados por DESeq2. MA-plot muestra la relación entre el logaritmo de cambio de expresión y la media normalizada de los recuentos (para el presente objetivo se prefiere una media elevada y un logaritmo de cambio cercano a cero). Heatmap indica una matriz que resalta la semejanzas y disparidades entre las muestras utilizada (es particularmente llamativa la diferencia entre stat1 y stat2).



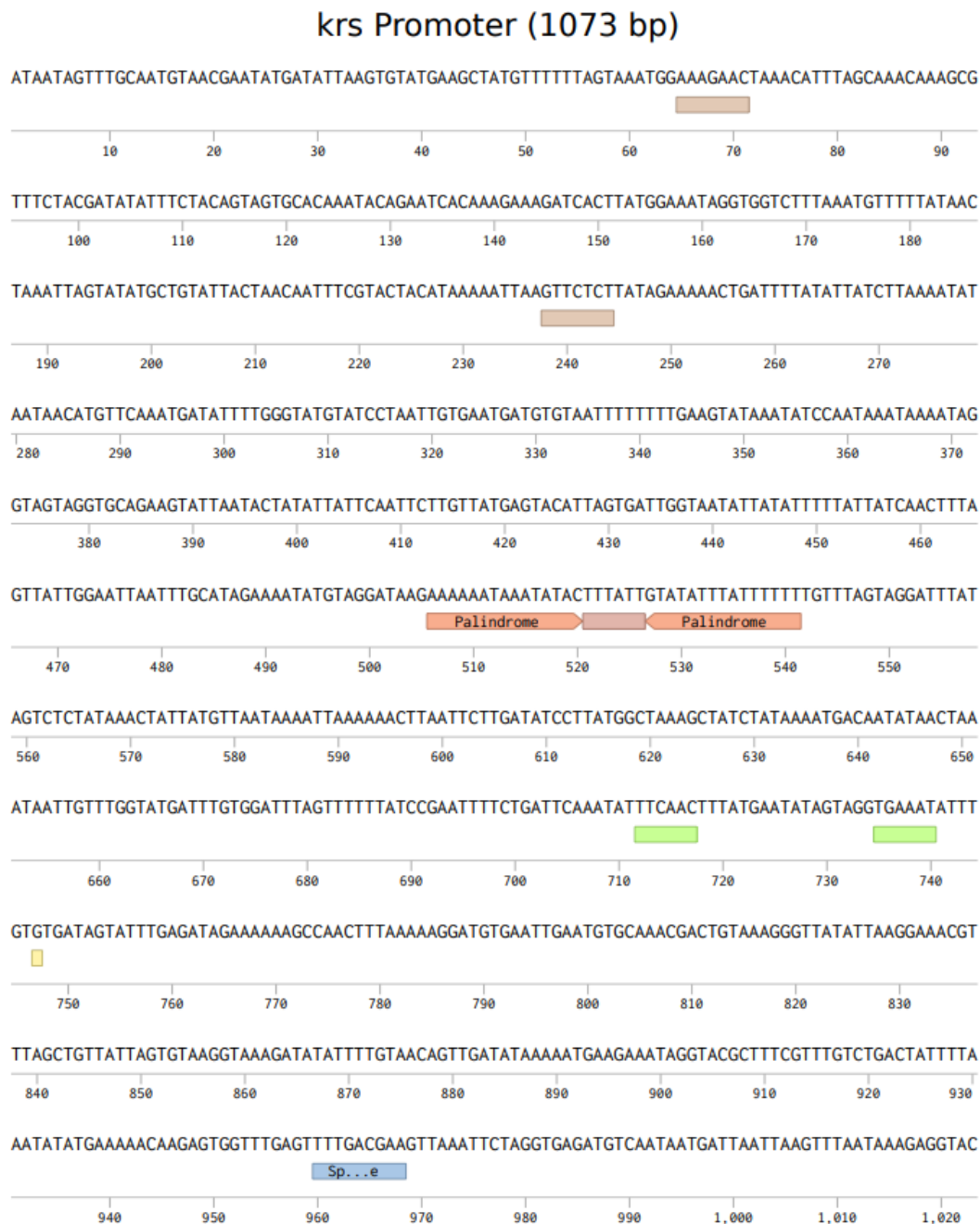
### pHT315-1\_PrlpU\_krsE

El promotor del operón *krs* descrito por Gélis-Jeanvoine y su equipo de trabajo (2016) se presenta en la Figura 10 donde se resaltan los elementos regulatorios. La secuencia abarca regiones de unión de factores como SinR, Spo0A y un palíndromo (previamente desconocido) conservados en los genomas de *B. thuringiensis* y con un efecto negativo enfocado en reprimir la expresión de las NRPS; además, el sitio de inicio de la transcripción TSS, de acuerdo a un análisis en BProm y DBTBS, tendría afinidad con el factor sigma F, el cual es activado principalmente durante la etapa de esporulación, hecho consistente con los niveles de expresión obtenidos en el análisis transcriptómico.

En consecuencia, se procedió a analizar también las secuencias de los dos mejores promotores elegidos como candidatos para potencializar la expresión del locus de interés, *r1pU* y *cspB2*. Los resultados de la observación en softwares relacionados a la predicción de promotores y sus componentes, presentados en la Figura 11, marcan una importante diferencia, pues el promotor de *r1pU* consiste en una región compuesta estructurada por tres potenciales inicios de transcripción: dos de ellos funcionarían con el factor sigma A cuya actividad es constitutiva y otro trabajaría con el factor sigma F durante la esporulación. Por otro lado, el promotor de *cspB2* incluye únicamente un sitio para el factor sigma X, activado tras la etapa vegetativa del crecimiento, y cuenta con lugares de unión de los factores AbrB y AraR de regulación negativa y ComK de regulación positiva de acuerdo al programa DBTBS.

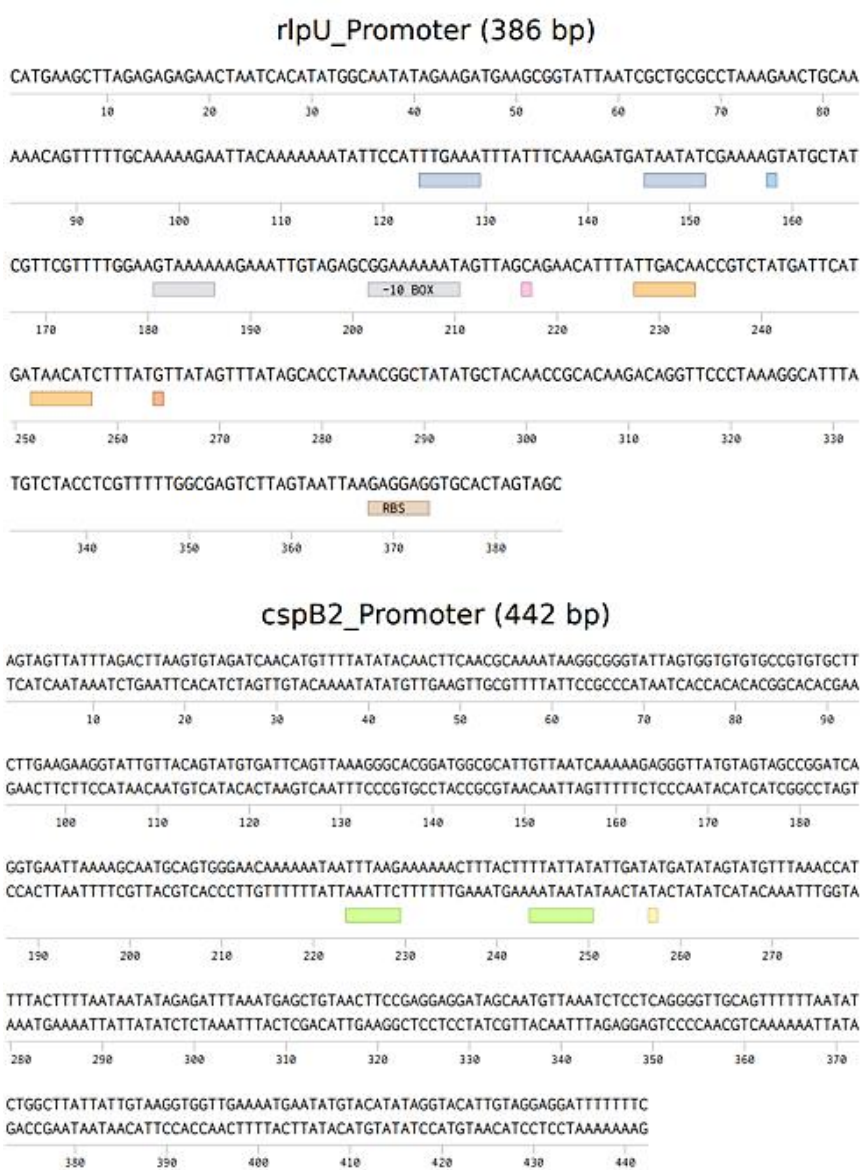
**Figura 10.**

*Promotor del operón krs (Gélis-jeanvoine et al., 2016). Incluye: sitios de unión del factor SinR (posiciones 64 y 237), palíndromo de regulación negativa (posición 505-542), región -35 (posición 711), región -10 (posición 735), sitio de inicio de transcripción TSS (posición 747) y potencial lugar de unión del factor Spo0A (posición 959).*



**Figura 11.**

Promotores candidatos para potencializar la expresión del operón *krs*. P<sub>rlpU</sub> contiene: regiones -35 y -10 con el sitio de unión de sigmaA (posición 123-158), regiones -35 y -10 con el sitio para el factor sigmaF (posición 181-216), regiones -35 y -10 con el TSS principal (posición 227-264) y el potencial sitio de unión del ribosoma (posición 367). P<sub>mspB2</sub> contiene: regiones -35 y -10 con sitio de unión del factor sigmaX (posición 224-257).



Entonces, se llevó a cabo el ensamblaje *in silico* del vector de transformación compuesto por el plásmido pHT315-1 junto a las secuencias del promotor PrlpU y el gen *krsE* que facilite la recombinación homóloga. La Tabla 8 exhibe los primers diseñados y utilizados para la amplificación de los fragmentos de interés mediante PCR y contienen en la región sobresaliente (overhang) 5' los sitios de restricción útiles para la clonación molecular.

**Tabla 8.**

*Oligonucleótidos (primers) utilizados para amplificar los fragmentos de interés PrlpU y krsE en Bacillus thuringiensis serovar galleriae strain HD-29.*

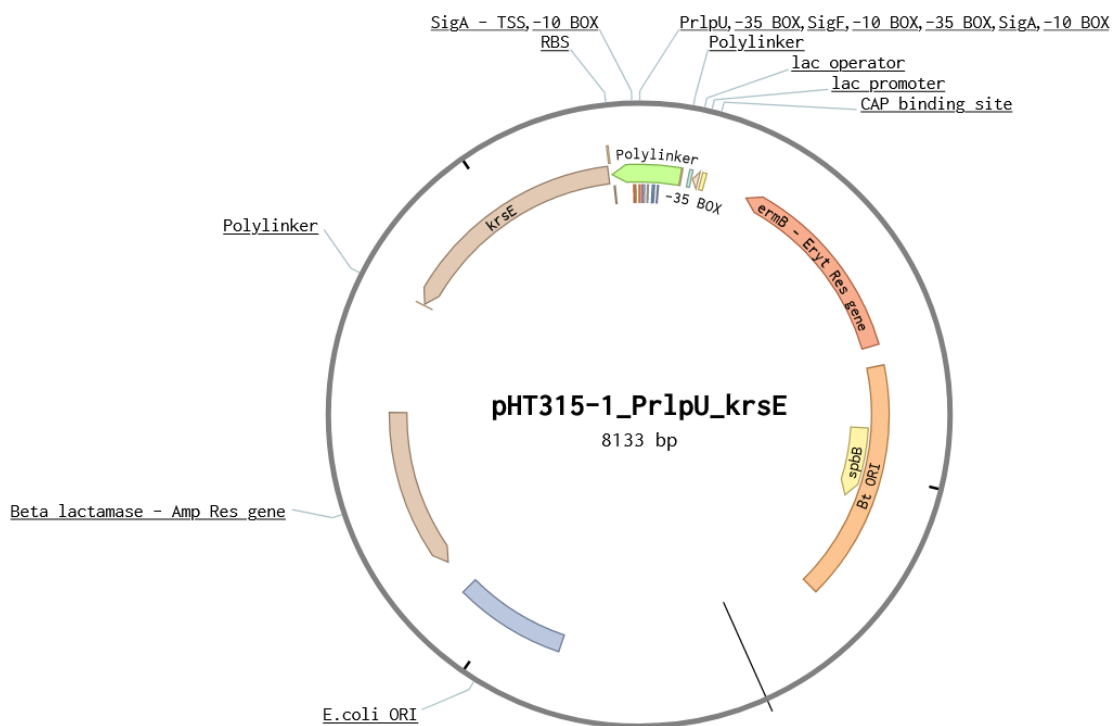
<b>Primers</b>	<b>Secuencia (5'-3')</b>	<b>Enzima de restricción</b>
PrlpU_Fwd	CATGA <b>AAGCTT</b> AGAGAGAGAACTAATCAC	HindIII
PrlpU_Rev	GCT <b>ACTAGT</b> GCACCTCCTCTTAATTAC	SpeI
krsE_Fwd	CTAG <b>TCTAGAG</b> CCTTTTATGAAA <b>ACTATTGATTC</b>	XbaI
krsE_Rev	CTGAG <b>GAATTC</b> CTTCAAAAATTATTTGCTCTT	EcoRI

La respectiva digestión *in silico* se desarrolló con todas las enzimas de restricción mencionadas tanto en el plásmido como en los fragmentos obtenidos. Como tal, la Figura 12 muestra el vector final a utilizar en la transformación que tendría lugar en las células de *B. thuringiensis*.



## Figura 12.

Vector pHT315-5 ensamblado con las secuencias PrlpU y krsE para reemplazar la secuencia de *Bacillus thuringiensis*.



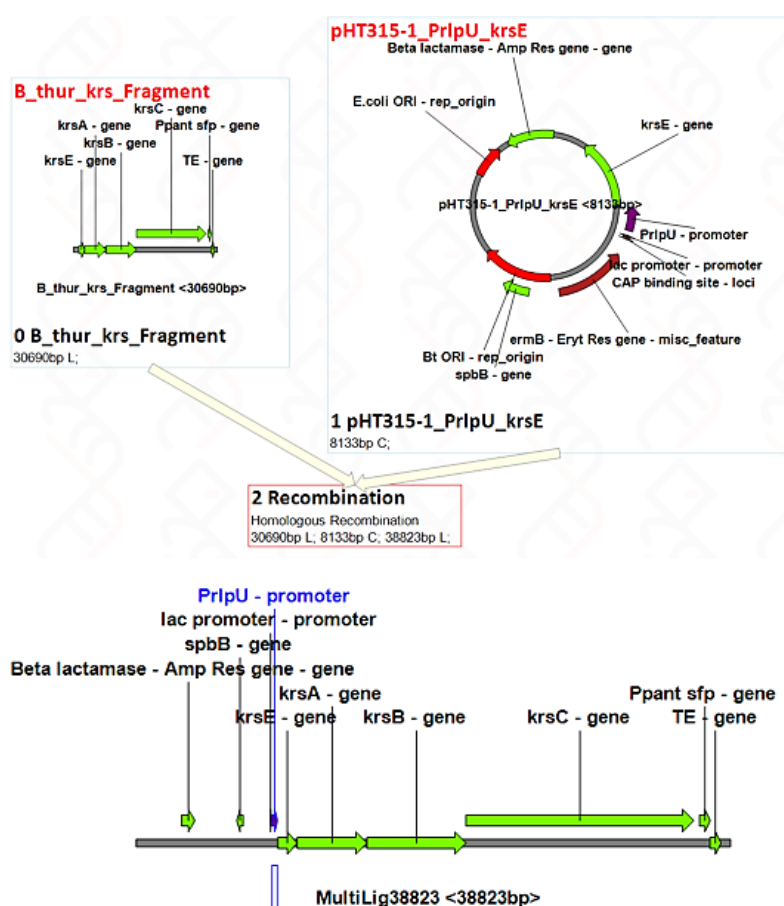
## Integración de la secuencia en el cromosoma bacteriano

Una vez definido el constructo génico para transformar el cromosoma bacteriano y aspirar a un mejor rendimiento en la producción de kurstakinas, procede simular el acontecimiento de recombinación homóloga. Así, el software MCDS (Shi & Vickers, 2016), diseñado con algoritmos que consideran los factores necesarios para dar lugar a eventos de recombinación homóloga, recombinación lambda red, unión de extremos no homólogos NHEJ y otros; fue empleado para introducir al vector pHT315-1\_Pr1pU\_krsE en el segmento aguas arriba al locus *krs* del genoma de *Bacillus thuringiensis* serovar *galleriae* strain HD-29. Consecuentemente, la Figura 13 exhibe el resultado arrojado por el sistema, donde puede notarse cómo, efectivamente, el nuevo promotor se integraría al

locus *krs* haciendo posible esperar un mayor rendimiento de kurstakinas y, sobre todo, que sea constitutivo (a lo largo de todo el ciclo de crecimiento) para favorecer activamente al consorcio bacteriano del cual formaría parte en futuras aplicaciones.

### Figura 13.

Simulación de recombinación homóloga entre *pHT315-1\_PrlpU\_krsE* y el locus *krs* de *Bacillus thuringiensis*.



### Propuesta de protocolo experimental

El procedimiento que se propone a continuación concibe la cadena de actividades necesarias para elevar la producción de kurstakinas en *Bacillus thuringiensis*, con el fin de aplicar el lipopéptido en distintos campos considerando sus propiedades y sobre todo

en suelos contaminados con sustancias derivadas de petróleo como se ha planteado en el desarrollo de este documento.

- **Muestreo de suelo**

El lugar de recolección debe ser reconocido y delimitado, de preferencia incluirá una imagen satelital o un mapa del sitio y sus respectivas coordenadas. Con base en antecedentes de reportes en Ecuador, *Bacillus thuringiensis* puede hallarse en compost proveniente de florícolas en las cercanías de Cayambe o Tabacundo (Narváez, 2015) y en sectores aledaños a Lago Agrio contaminados con productos de petróleo (Maddela et al., 2015). Entonces, se tomarán ahí muestras de 400 g en la superficie del suelo (cavando hasta 15 cm de profundidad) en cuatro puntos elegidos aleatoriamente. Las alícuotas de suelo deberán ser secadas al aire libre, tamizadas y conservadas a 4°C (Maddela et al., 2015).

- **Caracterización fisicoquímica**

En primer lugar, se determina el porcentaje de material grueso al pesar las partículas retenidas al pasar la muestra original por un tamiz con orificios de 2 mm. Después, se establecerá el pH introduciendo un potenciómetro en una mezcla suelo-agua de proporciones 1:1. Con respecto a la conductividad eléctrica, se ocupará un conductímetro introducido en un matraz que contenga 100 ml de agua al que se adicionará 1 g de suelo (Nagaraju et al., 2009).

- **Aislamiento de la bacteria**

Una porción de 1 g de la muestra de suelo se disolverá en 10 ml de agua destilada asegurándose de mezclar completamente y se procederá a calentar la solución hasta 80°C durante 10 minutos para deshacerse de células vegetativas y las bacterias carentes de esporas. Diluciones seriadas hasta  $10^{-10}$  serán realizadas y se tomarán alícuotas de cada dilución para esparcirlas en medio T3 (por cada litro: 3 g de triptona, 2 g de triptosa, 1.5 g de extracto de levadura, 0.05 M de buffer fosfato a pH 6.8 y 0.005 g de  $MnCl_2$ ;

(Shishir et al., 2012)). Las placas deberán ser incubadas a 37°C durante la noche para proseguir a observar las colonias con tinción Gram.

El medio T3 favorece a la esporulación del grupo *B. cereus* (Lout et al., 2020), posteriormente, para discernir entre *Bacillus cereus* y *Bacillus thuringiensis* se utilizará 450 ml de base para el medio MYP estéril mezclada con 50 ml de emulsión de yema de huevo y una solución de Polimixina B sulfato hasta lograr una concentración de 100 unidades por ml. Una colonia de cada aislado es suficiente para sembrar por estriado en el agar MYB incubándose a 30°C durante 24 horas. Al concluir, la diferenciación se logra teniendo en cuenta que *B. cereus* crece formando colonias rosadas de borde amarillo, entonces, las colonias que no presenten estas características pertenecerán a *Bacillus thuringiensis* (Rabha et al., 2017).

- **Medio de cultivo y crecimiento bacteriano**

Las bacterias a utilizar son *Bacillus thuringiensis* aislado de las muestras de suelo y *Escherichia coli* competente adquirida de manera comercial. Por otro lado, el medio de cultivo destinado para el crecimiento de ambos microorganismos es Luria-Bertani (también llamado LB, que contiene: 10 g/L de triptona, 5 g/L de extracto de levadura y 10 g/L de NaCl).

Particularmente, el primer paso de crecimiento de *B. thuringiensis* se lleva a cabo a 42°C durante la noche en constante agitación a 200 rpm para conseguir una cepa *cry*<sup>-</sup> o acristalífera (D Lereclus et al., 1989). A continuación, las células de *E. coli* se cultivarán en LB a 37°C en mezcla continua y *B. thuringiensis* a 30°C. En último lugar, y enfocado en la diferenciación de células transformadas, se añadirán en agar LB 100 µg/mL de ampicilina, para *E. coli*, y 25 µg/mL de eritromicina, para *B. thuringiensis* (Gélis-jeanvoine et al., 2016; Jiao et al., 2016).

La transformación de *B. thuringiensis* se desarrollará por electroporación tras preparar células competentes mediante la exposición a un pulso eléctrico de 12.5 kV/cm en un equipo Micropulser.

- **Manipulación de ADN**

El plásmido puede ser solicitado al autor (Arantes & Lereclus, 1991) o al equipo de iGEM encargado de la modificación mencionada para generar pHT315-1 (Ioana, 2012). Los primers, sintetizados por MacroGen® o ThermoFisher®, serán empleados en los procedimientos de PCR con una ADN polimerasa estándar como DreamTaq (ThermoFisher) o similares en un volumen de reacción ajustado a 25 µL. Las enzimas de restricción y la ADN ligasa T4 serán utilizadas de acuerdo a las indicaciones del fabricante.

- **Extracción y cuantificación de kurstakinas**

Las células transformadas de *B. thuringiensis* se cultivarán por triplicado en medio LB agar incubado a 37°C durante 24 horas y se conserva otros 6 días dentro de la incubadora a 27°C. Entonces, se procederá a tomar 100 mg de esporas secas para su respectiva suspensión en 10 mL de una solución de acetonitrilo/agua/ácido trifluoroacético (ATF) en proporciones V/V de 90:10:0.1 mezclando enérgicamente con vortex y centrifugando levemente a 3000 g durante 10 minutos. El lipopéptido se hallará en el sobrenadante, por lo que se secará utilizando una centrífuga de vacío y el extracto seco se disolverá en 4 mL de la solución acetonitrilo/agua/ATF ahora en proporción V/V 30:70:0.1 para pasar por una columna de cartucho empacada con 500 mg de sílica gel C18. Posteriormente, los elementos retenidos se lavan en dos pasos: primero con 10 mL de ATF al 0.1% y segundo con 10 mL de acetonitrilo/agua/ATF (30/70/0.1 V/V). En un paso final, la fracción de lipopéptido se remueve con 10 mL de la solución de elución acetonitrilo/agua/ATF en proporción V/V 90/10/0.1 y secada (Hathout et al., 2000).

Un siguiente procedimiento de purificación se efectuará a través de cromatografía líquida de alta eficacia HPLC ocupando una columna semi preparativa de tamaño de partícula 10  $\mu\text{m}$  y dimensiones de 250x10 mm conectada a una bomba de distribución. La muestra debe ser diluida en 500  $\mu\text{L}$  de la solución acetonitrilo/agua/ATF (30/70/0.1 V/V) para inyectarse en la columna. La fase móvil estará diseñada con un flujo fijo de 20  $\mu\text{L}/\text{min}$  formada por dos componentes: el solvente A que consiste en 0.1% de ácido trifluoroacético en agua y el solvente B de acetonitrilo suministrado en un gradiente que inicia en 30% durante un minuto y se eleva progresivamente hasta 70% hasta cumplir los 20 minutos del proceso. Los picos con distinto tiempo de retención serán recolectados y liofilizados; así, la fracción de lipopéptido puede ser pesada en una microbalanza (Diallo et al., 2019; Hathout et al., 2000).

- **Degradabilidad de diésel**

La metodología para evaluar la capacidad de degradación de diésel en presencia del lipopéptido extraído es la propuesta por Diallo y colaboradores (2019). En primer lugar, se prepararán 800 mL de medio Bushnell-Hass BH (por litro: 1 g de  $\text{KH}_2\text{PO}_4$ , 1 g de  $\text{K}_2\text{HPO}_4$ , 0.2 g de  $\text{MgSO}_4$ , 0.02 g de  $\text{CaCl}_2$ , 1 g de  $\text{NH}_4\text{NO}_3$ , 0.05 g de  $\text{FeCl}_3$ , 0.05 g de extracto de levadura) conteniendo diésel en concentración del 1%. El caldo nutriente se separará en dieciséis matraces conteniendo 50 mL cada uno, en ocho de los cuales se colocará el biosurfactante extraído anteriormente hasta lograr una concentración de 120 mg/L; mientras los otros matraces tendrán solamente medio. Entonces, el ensayo se basa en el cultivo de un microorganismo degradador de hidrocarburos, como *Acinetobacter haemolyticus*, en los envases de ambos grupos de estudio manteniendo agitación constante a 150 rpm durante 21 días a 30°C.

La determinación de resultados será llevada a cabo por métodos de análisis gravimétricos para establecer la tasa de degradación a los a, 10, 14 y 21 días. Así, el hidrocarburo remanente será extraído con una solución 1:1 V/V de hexano, exponiendo

la mezcla a vortex por 5 minutos y permitiendo la separación de la fase orgánica para su recolección. Consecuentemente, se extraerá dos veces la fase acuosa y el solvente será evaporado mediante el uso de un rotavapor. Cada muestreo se realizará por duplicado.

Entonces, la tasa de degradación se calculará utilizando la siguiente fórmula:

$$TDH (\%) = \frac{(PHR \text{ en cultivo control} - PHR \text{ en la muestra}) \times 100\%}{PHR \text{ en cultivo control}}$$

Considerando: TDH como tasa de degradación del hidrocarburo y PHR como peso de hidrocarburo remanente.

Además, el mejoramiento de la degradación del hidrocarburo con el uso del lipopéptido se determina con la siguiente fórmula:

$$MD (\%) = \frac{(TD \text{ con kurstakina} - TD \text{ sin kurstakina}) \times 100}{TD \text{ con kurstakina}}$$

Donde, MD se refiere a mejoramiento en la degradación y TD significa tasa de degradación.

También, se evaluará el crecimiento de la bacteria influenciado por la presencia de kurstakina, para lo que se usará el método de conteo en placas. Se tomarán alícuotas de 1 mL desde los cultivos en medio BH a los 4, 7, 10, 14, 17 y 21 para inocularlas en agar nutriente incubado a 30°C durante 24 horas. El resultado será reportado como un logaritmo en base 10 de las unidades formadoras de colonia por mL (UFC/mL).

## Capítulo V: Discusión

La familia de lipopéptidos kurstakina fue descrita por Hathout y su equipo de investigación (2000), quienes reportaron tres moléculas representantes de este grupo describiendo una estructura similar que consta de siete aminoácidos D-Thr/Gly/Ala/Ser/His/D-Gln/Gln y difieren únicamente por la longitud de la cadena de ácido graso que se une al residuo de treonina. De tal manera, con los conocimientos recopilados hasta ese entonces y durante las últimas dos décadas, este tipo de lipopéptidos ha pasado con bajo perfil en estudio indagatorios frente a otras moléculas similares sintetizadas por bacterias del género *Bacillus* como los pertenecientes a familias como surfactina, iturina y fengicina. Particularmente, las iturinas mantienen una relación estrecha con las kurstakinas en términos de biosíntesis en vista de que ambas cuentan con una estructura heptapeptídica cíclica y sus sintetetasas peptídicas no ribosomales NRPS de mayor longitud (ItuB y KrsC, respectivamente) están conformadas por cuatro módulos y, en algunos casos como la bacilomicina, coinciden en ciertos aminoácidos utilizados como sustrato (Geissler et al., 2019). Por lo tanto, con el antecedente de esta familia de compuestos sintetizados por *Bacillus subtilis*, cabe comprender que las herramientas informáticas de predicción y anotación de genes pasen por alto la descripción del operón *krs* y la información disponible en las bases de datos convencionales como NCBI y EMBL sea ambigua o nula.

Por otro lado, un mecanismo que puede ponerse en marcha con el fin de dilucidar si una NRPS forma parte del sistema de biosíntesis de un lipopéptido en específico es la identificación de los dominios que constituyen sus distintos módulos, pues existen motivos que sugieren una actividad específica como el caso del dominio de epimerización encargado de alterar la configuración espacial del aminoácido de levógiro a dextrógiro. A la par, la definición del sustrato específico del dominio de adenilación permite establecer



una predicción de la composición del péptido; sin embargo, algunas de las inconsistencias obtenidas durante este tipo de análisis en las NRPS de kurstakinas se deberían al patrón degenerado de selección exhibido por su tipo de enzimas, D-alanil-poli(fosforibitol) sintetasa, por lo que los modelos ocultos de Markov precisan de un input limitado a una secuencia reducida donde se encuentre tal dominio de adenilación o, en su defecto, el enriquecimiento de la base de datos experimentales de referencia para afinar la especificidad del software de predicción (Khayatt et al., 2013).

La distribución del operón *krs* dentro del grupo *Bacillus cereus* fue previamente establecida por Gélis-Jeanvoine y colaboradores (2016) ocupando los genomas secuenciados disponibles al momento de su investigación; así, la actualización del registro de especies portadoras (Tabla 5) permite corroborar que, en efecto, los genes de NRPS de kurstakinas son endémicos del grupo taxonómico hallándose exclusivamente en *B. cereus* y *B. thuringiensis* a nivel cromosómico y en *B. mycoides* y *B. weihenstephanensis* dentro de plásmidos; además, y el árbol filogenético de máxima verosimilitud ratifica lo expuesto por el grupo de investigación, pues el operón estaría conservado en especies específicamente del clado II, lo que sugiere un papel ecológico de kurstakina que facilitaría la invasión de *Bacillus thuringiensis* en larvas y cadáveres de insectos hospederos. En adición a la teoría planteada, la recopilación de información brindada por Ehling-Schulz y sus colegas (2019) serviría como sustento tomando en cuenta que especies como *B. anthracis* tienen un comportamiento marcado de infección en mamíferos y la tendencia de *B. cereus* y *B. thuringiensis*, aunque ambiguamente descrita hasta la fecha, sugiere una preferencia a establecerse en suelos agrícolas y remanentes de insectos del orden lepidóptera y coleóptera. Las kurstakinas favorecerían a la formación de biofilm y la estabilidad de la población de la bacteria (Dubois et al., 2012; Gélis-jeanvoine et al., 2016).

En contraste con el rol ecológico o fisiológico que puedan tener las kurstakinas en su microorganismo de origen, el hecho de ser un lipopéptido implica un potencial para aplicaciones biotecnológicas dentro de industrias como la farmacéutica, química, agrícola o de alimentos, así como en el tratamiento y remediación de sitios contaminados (Malviya et al., 2020). Para este fin, es necesario considerar que las kurstakinas son producidas durante la fase de esporulación del crecimiento de *B. thuringiensis*, una característica similar a la de las proteínas cristalíferas entomotóxicas cuya expresión ha sido extendida para llevarse a cabo de manera constitutiva mediando la integración de distintos promotores o fragmentos reguladores a través de sistemas de recombinación (Schnepf, 2012).

Por consiguiente, es importante revisar las herramientas actuales que faciliten el mejoramiento en el rendimiento de generación de metabolitos secundarios como los lipopéptidos. Así, Hu junto a su equipo (2019), recopilan estrategias para potenciar la producción de surfactina en *B. subtilis* y proyectan estos procedimientos para su ejecución con respecto a otras sustancias como las kurstakinas; entre estas, la ingeniería de promotor resalta como una alternativa viables teniendo en cuenta el panorama extenso que puede obtenerse a partir de un análisis transcriptómico.

Este tipo de observaciones han sido anteriormente realizadas en una canalización de programas configurada de la siguiente manera: TopHat→Cufflinks→Cuffmerge, enfocada en el mapeo de las lecturas de RNA-seq y una estimación de RPKM como un método para describir los niveles de expresión (Trapnell et al., 2013). No obstante, los software del grupo TopHat y Cufflinks han sido marcados como discontinuados en los servidores de Galaxy en vista de la disponibilidad de alternativas actualizadas que incluyen parámetros que permiten predecir eventos de splicing alternativo o *de novo* y, además, destacan como los valores de RPKM son en esencia descriptivos y se recomienda evitar su uso al definir niveles de expresión (Nekrutenko, 2020). Por lo tanto,

el alineamiento y mapeo elaborado con Bowtie2 fue satisfactorio tomando en cuenta su diseño moderno y su disponibilidad ante la voluntad del usuario a manipular ciertos parámetros, como los referentes a splicing, que destacan su fidelidad a los patrones de expresión esperados en procariontas. También, el acercamiento estadístico efectuado por el software de análisis de expresión diferencial DESeq2 proyecta una visión panorámica del comportamiento de los promotores, facilitando a la selección de una secuencia reguladora capaz de favorecer la expresión constitutiva y elevada de un gen de interés, como *PrlpU*, en lugar de elegir basándose únicamente en valores puntuales de alto rendimiento.

En adición a los elementos a considerar al momento de seleccionar la región promotora candidata para reemplazar a *Pkrs* (promotor del operón *krs*) se encuentran los factores de transcripción, pues la región propia del locus de interés presenta un patrón definido de represión con sitios de unión para SinR, Spo0A y una región palindrómica represora; en contraste, *PrlpU* cuenta con sitios para la potencial unión de proteínas como AbrB, ComK, GlnR, MntR, PurR, RocR y TnrA que son en su mayoría de regulación positiva o están encaminadas a mantener la homeostasis de elementos como manganeso y sodio en el entorno. Del mismo modo, fue determinante la observación de la estructura secundaria formada por el ARNm estimada mediante el uso del software UNAFold (Markham & Zuker, 2008), donde se evidencia en *Pkrs* la conformación de múltiples bucles que favorecen la exposición del sitio de unión de Spo0A; mientras, *PrlpU* da lugar a una estructura de un bucle principal a manera de vástago en el tallo de una planta y, sobre todo, cabe remarcar la manera en la cual favorece a la accesibilidad del sitio de unión del ribosoma para continuar con la traducción (Wang, Ai, et al., 2013).

Finalmente, el protocolo propuesto para la aplicación experimental de la información detallada en este documento se formó con base en el primer reporte de kurstakinas (Hathout et al., 2000), la investigación pionera en describir su región

promotora y su capacidad de sostener la formación de biofilm (Gélis-jeanvoine et al., 2016) y el estudio más reciente de kurstakinas encargado de evaluar su potencial aplicabilidad en lugares contaminados por hidrocarburos derivados de petróleo (Diallo et al., 2019); motivo por el cual se sostiene la viabilidad del procedimiento tanto en campo como en laboratorio.

Sin embargo, es importante traer a colación la técnica propuesta en el artículo de Dimkic y su grupo de investigación (2017), donde se utiliza como fuente de extracción de kurstakinas el sobrenadante de cultivo en medio LB y se somete a procesos de identificación y obtención como la espectrometría de masas MALDI-TOF y la cromatografía de capa fina de alto rendimiento (HPTLC). De acuerdo con el equipo de trabajo, la elución con acetato de etilo tendría mejores resultados para conseguir kurstakinas como compuestos puros y tiene un mayor espectro de alcance, pues es aplicable con otros lipopéptidos como iturinas, fengicinas y surfactinas.

## Capítulo VI: Conclusiones

Los genes que codifican las NRPS encargadas de la biosíntesis de kurstakinas están dispuestos en el genoma de *Bacillus thuringiensis* a manera de operón cuya expresión está dirigida por el promotor P<sub>krs</sub> de una longitud aproximadas de 1.1 kb y cuenta con sitios de unión de factores de transcripción que reprimen su actividad. Los productos de esta región son seis proteínas que se denotan como KrsE, KrsA, KrsB, KrsC, Krs D (fosfopanteteinil transferasa) y TE (tioesterasa).

La potencialización del rendimiento de expresión génica en este locus de interés es viable y requiere, de acuerdo a este caso de estudio, del reemplazo de su región promotora a través de la integración de un promotor candidato en el cromosoma bacteriano mediante recombinación homóloga. Para esta finalidad, el plásmido pHT315-1 puede actuar como un vector de transporte albergando un constructo que consiste en el promotor constitutivo fuerte PrpU (proveniente del gen de la proteína ribosomal 50S L21) y la secuencia del gen *krsE* que dirija y facilite el evento de recombinación.

Este proceso de ingeniería genética tiene potencial de ser ejecutado en laboratorio con especies reportadas en Ecuador, como *Bacillus thuringiensis* serovar *galleriae* strain HD-29, y la combinación de técnicas que incluyen la reacción en cadena de polimerasa y digestión por enzimas de restricción, concebidas en el procedimiento general de clonación y transformación molecular. Por lo tanto, los lipopéptidos obtenidos pueden emplearse en diversas áreas, teniendo como protagonista en el presente documento a la remediación de sitios contaminados por hidrocarburos derivados del petróleo teniendo en cuenta la propiedad biosurfactante de la kurstakina.

## Capítulo VII: Recomendaciones

Este documento presenta el diseño de un constructo dentro del vector de transporte pHT315, un miembro del reducido grupo de plásmidos descritos para la transformación en *Bacillus thuringiensis*; sin embargo, se sugiere comparar durante el procedimiento experimental su efectividad frente a otros plásmidos, como pBta, cuya aplicación se ha reportado como diversificada dentro del género *Bacillus* y que, además, permitiría el control de su prevalencia dentro de la bacteria por su naturaleza termosensible.

Por otro lado, la extracción del lipopéptido se ha propuesto de acuerdo a la estrategia utilizada en el primer reporte de kurstakinas y que ha sido replicada en distintos artículos de investigación, pero es recomendable la puesta en marcha de la técnica generada por Dimkic y su equipo (2017) con el fin de sacar el mayor provecho al biosurfactante extrayéndolo tanto desde las esporas como del sobrenadante del medio de cultivo.

## Capítulo VIII: Bibliografía

- Abdelkefi-mesrati, L., & Tounsi, S. (2012). Recombination in *Bacillus thuringiensis*. *Bacillus Thuringiensis Biotechnology*, 201–214. <https://doi.org/10.1007/978-94-007-3021-2>
- Abderrahmani, A., Tapi, A., Nateche, F., Chollet, M., Leclère, V., Wathelet, B., Hacene, H., & Jacques, P. (2011). *Bioinformatics and molecular approaches to detect NRPS genes involved in the biosynthesis of kurstakin from Bacillus thuringiensis*. 571–581. <https://doi.org/10.1007/s00253-011-3453-6>
- Advisor, W. P. I., Mello, N. A., Baker, C., & Johnson, B. (2009). Oil Pollution in Ecuador : Remediation Approach. *Worcester Polytechnic Institute*.
- Agaisse, H., & Lereclus, D. (1996). STAB-SD : a Shine – Dalgarno sequence in the 5 ' untranslated region is a determinant of mRNA stability. *Molecular Microbiology*, 20, 633–643.
- Agencia para Sustancias Tóxicas y Registro de Enfermedades. (1999). ATSDR - Resumen de Salud Pública: Hidrocarburos totales de petróleo. CDC.
- Alexey, C., Rivadeneira, M., Wladimir, A., Antonio, M., & Aldás, M. (2017). *Influencia de surfactantes en la evaluación de la tensión interfacial para una emulsión agua petróleo relacionada al proceso de recuperación mejorada*. Escobar 2014.
- Anand, S., Prasad, M. V. R., Yadav, G., Kumar, N., Shehara, J., Ansari, Z., & Mohanty, D. (2010). SBSPKS : structure based sequence analysis of polyketide synthases. *Nucleic Acids Research*, 38(May), 487–496. <https://doi.org/10.1093/nar/gkq340>
- Anders, S., Pyl, P. T., & Huber, W. (2015). Genome analysis HTSeq — a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Arantes, O., & Lereclus, D. (1991). *Construction of cloning vectors for Bacillus thuringiensis*. 108, 115–119.

- Bachmann, B. O., & Ravel, J. (2009). Methods for In Silico Prediction of Microbial Polyketide and Nonribosomal Peptide Biosynthetic Pathways from DNA Sequence Data. In *Complex enzymes in microbial natural product biosynthesis* (1st ed., Vol. 458, Issue 09). Elsevier Inc. [https://doi.org/10.1016/S0076-6879\(09\)04808-3](https://doi.org/10.1016/S0076-6879(09)04808-3)
- Bayat, A. (2002). Science, medicine, and the future Bioinformatics: Clinical Review. *British Medical Journal*, 324(April), 1018–1022.
- Becerra, S., Paichard, E., Sturma, A., & Maurice, L. (2013). VIVIR CON LA CONTAMINACIÓN PETROLERA EN EL SANITARIO Y CAPACIDAD DE RESPUESTA Living with oil contamination in Ecuador : social perceptions of health risks and coping capacities también es conocido por ser el país que acoge la biodiversidad más densa del. *Revista Lider*, 23, 102–120.
- Béchet, M., Caradec, T., Hussein, W., Abderrahmani, A., Chollet, M., Leclère, V., Dubois, T., Lereclus, D., Pupin, M., & Jacques, P. (2012). Structure , biosynthesis , and properties of kurstakins , nonribosomal lipopeptides from *Bacillus* spp . *Applied Microbiology and Biotechnology*, 907, 593–600. <https://doi.org/10.1007/s00253-012-4181-2>
- BioStars. (2018). *Aligning with Bacterial Genome*. Bioinformatics Training. <https://www.biostars.org/p/193312/>
- Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, Y., Medema, M. H., & Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, 47(April), 81–87. <https://doi.org/10.1093/nar/gkz310>
- Chatzou, M., Magis, C., Chang, J., Kemena, C., Bussotti, G., Erb, I., & Notredame, C. (2016). Multiple sequence alignment modeling: methods and applications. *Briefings in Bioinformatics*, 17(October 2015), 1009–1023. <https://doi.org/10.1093/bib/bbv099>
- Chukwunonso, I., Ahmed, A., Hassan, A., & Shahul, F. (2019). Remediation of soil and



- water contaminated with petroleum hydrocarbon: A review. *Environmental Technology & Innovation*. <https://doi.org/10.1016/j.eti.2019.100526>
- Diallo, M. M., Vural, C., Şahar, U., & Ozdemir, G. (2019). Kurstakin molecules facilitate diesel oil assimilation by *Acinetobacter haemolyticus* strain 2SA through overexpression of alkane hydroxylase genes strain 2SA through overexpression of alkane hydroxylase genes. *Environmental Technology*, *0*(0), 1–15. <https://doi.org/10.1080/09593330.2019.1689301>
- Dimkic, I., Stankovic, S., Nisavic, M., Petkovic, M., Ristivojevic, P., Fira, D., & Beric, T. (2017). The Profile and Antimicrobial Activity of *Bacillus* Lipopeptide Extracts of Five Potential Biocontrol Strains. *Frontiers in Microbiology*, *8*(May), 1–12. <https://doi.org/10.3389/fmicb.2017.00925>
- Dubois, T., Lemy, C., Buisson, C., Faegri, K., Nielsen-Ieroux, C., Gohar, M., Jacques, P., Ramarao, N., & Kolstø, A. (2012). *Necrotrophism Is a Quorum-Sensing-Regulated Lifestyle in Bacillus thuringiensis*. *8*(4). <https://doi.org/10.1371/journal.ppat.1002629>
- Ehling-schulz, M., Koehler, T. M., & Lereclus, D. (2019). The *Bacillus cereus* Group: *Bacillus* species with Pathogenic Potential. *HHS Public Access*, *7*(3), 3–32. <https://doi.org/10.1128/microbiolspec.GPP3-0032-2018>.The
- Federici, B. A., Park, H., & Bideshi, D. K. (2010). Overview of the Basic Biology of *Bacillus thuringiensis* with Emphasis on Genetic Engineering of Bacterial Larvicides for Mosquito Control. *The Open Toxicology Journal*, *3*, 154–171.
- Flissi, A., Ricart, E., Chevalier, M., Dufresne, Y., Michalik, J., Jacques, P., Flahaut, C., & Pupin, M. (2020). Norine : update of the nonribosomal peptide resource Fr ed. *Nucleic Acids Research*, *48*(November 2019), 465–469. <https://doi.org/10.1093/nar/gkz1000>
- Fontaine, G. (2003). Más allá de Texaco, ¿se puede rescatar el nororiente ecuatoriano? *FLACSO*, 1–10.

- Geissler, M., Heravi, K. M., Henkel, M., & Hausmann, R. (2019). Lipopeptide Biosurfactants From *Bacillus* Species. In *Biobased Surfactants* (Second Edi). Elsevier Inc. <https://doi.org/10.1016/b978-0-12-812705-6.00006-x>
- Gélis-jeanvoine, S., Canette, A., Caradec, T., Lemy, C., & Jacques, P. (2016). Genetic and functional analyses of *krs*, a locus encoding kurstakin, a lipopeptide produced by *Bacillus thuringiensis*. *Research in Microbiology*, December 2017. <https://doi.org/10.1016/j.resmic.2016.06.002>
- Green, L. H. (2008). *MICROBIOLOGY* (Issue August). <https://doi.org/10.1201/9781420009330.ch24>
- Guerra, S. (2020). *La producción petrolera de Ecuador se ha reducido en 58%*. <https://www.primicias.ec/noticias/economia/produccion-petrolera-ecuador-reducido-mitad/>
- Hanania, J., Stenhouse, K., & Donev, J. (2017). *Crude Oil*. Energy Education Canada - University of Calgary.
- Hathout, Y., Ho, Y. P., Ryzhov, V., Demirev, P., & Fenselau, C. (2000). Kurstakins: A new class of lipopeptides isolated from *Bacillus thuringiensis*. *Journal of Natural Products*, 63(11), 1492–1496. <https://doi.org/10.1021/np000169q>
- Hillis, D. M. (1997). Phylogenetic analysis. *Current Biology: Magazine*, 7(3), 129–131.
- Hoeijmakers, W. A. M., Bártfai, R., & Stunnenberg, H. G. (2012). Chapter 15 Transcriptome Analysis Using RNA-Seq. *Methods in Molecular Biology*, 923(4), 221–239. <https://doi.org/10.1007/978-1-62703-026-7>
- Hu, F., Liu, Y., & Li, S. (2019). Rational strain improvement for surfactin production: enhancing the yield and generating novel structures. *Microbial Cell Factories*, 30, 1–13. <https://doi.org/10.1186/s12934-019-1089-x>
- Ioana, S. (2012). *Part BBa K802004 Design - parts*. Registry of Standard Biological Parts. [http://parts.igem.org/Part:BBa\\_K802004:Design](http://parts.igem.org/Part:BBa_K802004:Design)

- Jiao, S., Li, X., Yu, H., Yang, H., Li, X., & Shen, Z. (2016). In Situ Enhancement of Surfactin Biosynthesis in *Bacillus subtilis* Using Novel Artificial Inducible Promoters. *Biotechnology and Bioengineering*, 9999(xxx), 1–11. <https://doi.org/10.1002/bit.26197>
- Khayatt, B. I., Overmars, L., Siezen, R. J., & Francke, C. (2013). Classification of the Adenylation and Acyl-Transferase Activity of NRPS and PKS Systems Using Ensembles of Substrate Specific Hidden Markov Models. *PLoS ONE*, 8(4). <https://doi.org/10.1371/journal.pone.0062136>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–360. <https://doi.org/10.1038/nmeth.1923>
- Lereclus, D, Arantrs, O., Chaufaux, J., & Lecadet, M. (1989). Transformation and expression of a cloned delta-endotoxin gene in *Bacillus thuringiensis*. *FEMS Microbiology Letters*, 60, 211–217.
- Lereclus, Didier, Vallade, M., Chaufaux, J., Arantes, O., & Rambaud, S. (1992). Expansion of Insecticidal Host Range of *Bacillus thuringiensis* by in vivo Genetic Recombination. *Nature Bio/Technology*, 10(4), 418–421.
- Li, X., Fan, F., Zhang, B., Zhang, K., & Chen, B. (2018). Biosurfactant enhanced soil bioremediation of petroleum hydrocarbons : Design of experiments ( DOE ) based system optimization and phospholipid fatty acid ( PLFA ) based microbial community analysis. *International Biodeterioration and Biodegradation*, March, 0–1. <https://doi.org/10.1016/j.ibiod.2018.04.009>
- Lout, H., Pellen, F., Jeune, B. Le, Lteif, R., Kallassy, M., B, G. Le, & Abboud, M. (2020). Real-time monitoring of bacterial growth kinetics in suspensions using laser speckle imaging. *Scientific Reports*, 1–10. <https://doi.org/10.1038/s41598-019-57281-2>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 1–21.

<https://doi.org/10.1186/s13059-014-0550-8>

- Luo, H., & Gao, F. (2019). *DoriC 10 . 0: an updated database of replication origins in prokaryotic genomes including chromosomes and plasmids*. 47(October 2018), 74–77. <https://doi.org/10.1093/nar/gky1014>
- Maddela, N. R., Masabanda, M., & Leiva, M. (2015). Novel diesel-oil-degrading bacteria and fungi from the Ecuadorian Amazon rainforest. *Water Science and Technology*, 1554–1562. <https://doi.org/10.2166/wst.2015.142>
- Malviya, D., Sahu, P. K., Singh, U. B., & Paul, S. (2020). *Lesson from Ecotoxicity : Revisiting the Microbial Lipopeptides for the Management of Emerging Diseases for Crop Protection*.
- Markham, N. R., & Zuker, M. (2008). Chapter 1 Software for Nucleic Acid Folding and Hybridization. *Methods in Molecular Biology*, 453(1). <https://doi.org/10.1007/978-1-60327-429-6>
- Mofid, M. R., Finking, R., Essen, L. O., Marahiel, M. A., Biochemie, F. C., & Philipps-uni, V. (2004). *Structure-Based Mutational Analysis of the 4 ' -Phosphopantetheinyl Transferases Sfp from Bacillus subtilis : Carrier Protein Recognition and Reaction Mechanism †,‡*. 4128–4136.
- Moran, C. P., Lang, N., Legrice, S. F. J., Lee, G., Stephens, M., Sonenshein, A. L., Pero, J., & Losick, R. (1982). Nucleotide Sequences that Signal the Initiation of Transcription and Translation in *Bacillus subtilis*. *Molecular and General Genetics*, 186, 339–346.
- Naeem, U., & Akhram, M. (2019). Leading edges in bioremediation technologies for removal of petroleum hydrocarbons. *Environmental Science and Pollution Research, Boopathy 2000*.
- Nagaraju, M., Narasimha, G., & Rangaswamy, V. (2009). International Biodeterioration & Biodegradation Impact of sugar industry effluents on soil cellulase activity.

- International Biodeterioration & Biodegradation*, 63(8), 1088–1092.  
<https://doi.org/10.1016/j.ibiod.2009.09.006>
- Narváez, A. (2015). Caracterización morfológica y molecular de cepas de *Bacillus* nativas de suelos de la provincia de Pichincha para biocontrol del ácaro *Tetranychus urticae*. In *Universidad San Francisco de Quito*.
- Nekrutenko, A. (2020). *Reference-based RNAseq data analysis (long)*. Galaxy Tutorials. <https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/rb-rnaseq/tutorial.html>
- NIH. (2021). *Promoter*. National Human Genome Research Institute.
- Ochoa-zarzosa, A., & López-meza, J. E. (2012). *Shuttle Vectors of Bacillus thuringiensis*. 175–184. <https://doi.org/10.1007/978-94-007-3021-2>
- Pande, V., Pandey, S. C., Sati, D., Pande, V., & Samant, M. (2020). Bioremediation: an emerging effective approach towards environment restoration. *Environmental Sustainability*, 0123456789. <https://doi.org/10.1007/s42398-020-00099-w>
- Prieto, C., García-estrada, C., Lorenzana, D., & Martín, J. F. (2012). NRPSsp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics Oxford Academics*, 28(3), 426–427. <https://doi.org/10.1093/bioinformatics/btr659>
- Rabha, M., Sharma, S., Acharjee, S., & Sarmah, B. K. (2017). Isolation and characterization of *Bacillus thuringiensis* strains native to Assam soil of North East India. *3 Biotech*, 7(5), 1–9. <https://doi.org/10.1007/s13205-017-0935-y>
- Sansinenea, E. (2012). Discovery and Description of *Bacillus thuringiensis*. *Bacillus Thuringiensis Biotechnology*. <https://doi.org/10.1007/978-94-007-3021-2>
- Santos, D. K. F., Rufino, R. D., Luna, J. M., Santos, V. A., & Sarubbo, L. A. (2016). Biosurfactants: Multifunctional Biomolecules of the 21st Century. *International Journal of Molecular Sciences*, 17, 1–31. <https://doi.org/10.3390/ijms17030401>
- Schnepf, H. E. (2012). *Bacillus thuringiensis* Recombinant Insecticidal Protein Production.

- Bacillus Thuringiensis Biotechnology*, 259–281. <https://doi.org/10.1007/978-94-007-3021-2>
- Schwarzer, D., Mootz, H. D., Linne, U., & Marahiel, M. A. (2002). *Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases*. 99(22).
- Secretaría de Energía República de Argentina. (2003). Hidrocarburos. *Conceptos Sobre Hidrocarburos*.
- Shi, Z., & Vickers, C. E. (2016). Molecular Cloning Designer Simulator ( MCDS ): All-in-one molecular cloning and genetic engineering design , simulation and management software for complex synthetic biology and metabolic engineering projects. *Metabolic Engineering Communications*, 3, 173–186. <https://doi.org/10.1016/j.meteno.2016.05.003>
- Shishir, A., Akter, A., Hassan, H., Kibria, G., Ilias, M., & Nargis, S. (2012). Characterization of locally isolated *Bacillus thuringiensis* for the Development of Eco-friendly Biopesticides in Bangladesh. *Journal of Biopesticides*, 5, 216–222.
- Slepecky, R. A., & Hemphill, H. E. (2006). *The Genus Bacillus — Nonmedical*.
- Solovyev, V. V, Salamov, A., Berkeley, L., & Bachinsky, A. (2011). Automatic Annotation of Bacterial Community Sequences and Application to Infections Diagnostic. *Metagenomics and Its Applications in Agriculture*, January.
- Sullivan, M., Yasbin, R., & Young, F. (1984). New shuttle vectors for *Bacillus subtilis* and *Escherichia coli* which allow rapid detection of inserted fragments. *Elsevier Gene*, 29, 21–26.
- Tapi, A., Chollet-imberty, M., Scherens, B., & Jacques, P. (2010). *New approach for the detection of non-ribosomal peptide synthetase genes in Bacillus strains by polymerase chain reaction*. 1521–1531. <https://doi.org/10.1007/s00253-009-2176-4>
- Tennant, M. (2004). *Maximum Likelihood*. NCBI Phylogenetics Resources.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D., Pimentel, H., Saizberg,

- S., Rinn, J., & Pachter, L. (2013). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *NIH Public Access*, 7(3), 562–578. <https://doi.org/10.1038/nprot.2012.016>.Differential
- Wang, J., Ai, X., Mei, H., Fu, Y., Chen, B., Yu, Z., & He, J. (2013). High-Throughput Identification of Promoters and Screening of Highly Active Promoter-5'-UTR DNA Region with Different Characteristics from *Bacillus thuringiensis*. *PLoS ONE*, 8(5), 1–11. <https://doi.org/10.1371/journal.pone.0062960>
- Wang, J., Mei, H., Zheng, C., Qian, H., Cui, C., Fu, Y., Su, J., Liu, Z., Yu, Z., & He, J. (2013). The Metabolic Regulation of Sporulation and Parasporal Crystal Formation in *Bacillus thuringiensis* Revealed by Transcriptomics and Proteomics \* □. *Molecular and Cellular Proteomics*, 12(5), 1363–1376. <https://doi.org/10.1074/mcp.M112.023986>
- WWF. (2021). *Oil and Gas Development Threats WWF*. <https://www.worldwildlife.org/threats/oil-and-gas-development>
- Zaid, A., Hughes, A., Porceddu, E., & Nicholas, F. (1999). *Glossary of biotechnology and genetic engineering*. FAO.