



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

**UNIDAD DE GENÓMICA AVANZADA**

**Análisis filogenético de familias de enzimas que utilizan tRNA, indicios para  
el descubrimiento de productos naturales ocultos en *Actinobacteria***

Tesis presentada por

**IBQ. Paulina Mayell Mejía Ponce**

Para obtener el grado de

**Maestra en Ciencias en Biología Integrativa**

Dirección de tesis

**Dr. Francisco Barona Gómez**

Irapuato, Guanajuato

Mayo 2017

Este trabajo se realizó en el Centro de Investigación y de Estudios Avanzados del  
IPN – Unidad Irapuato, Unidad de Genómica Avanzada, LANGEBIO en el  
Laboratorio de Evolución de la Diversidad Metabólica bajo la dirección del Dr.  
Francisco Barona Gómez, gracias al apoyo de beca proporcionado por el Consejo  
Nacional de Ciencia y Tecnología (CONACyT).

Este trabajo se llevó a cabo del 1 de septiembre de 2014 al 3 de marzo de 2017.

Miembros del comité de asesores:

Dra. Laila Pamela Partida Martínez  
Dr. Cei Leander Gaston Abreu Goodger

Fecha de examen de grado: 9 de junio de 2017

## Agradecimientos

En primer lugar, agradezco al Dr. Francisco Barona Gómez no solo por darme la oportunidad de ser parte de su equipo de trabajo, también por haberme brindado un excelente entrenamiento científico para desarrollar un proyecto multidisciplinario el cual integra distintas áreas del conocimiento como la microbiología, química biológica, bioinformática y evolución. Gracias Dr. Barona por haberme transmitido tanta pasión y visión científica sin fronteras.

A mi comité tutorial, la Dra. Laila Pamela Partida Martínez y el Dr. Ceí Abreu por todas sus observaciones, asesorías, buenas críticas y recomendaciones para el presente proyecto.

Al Dr. Kapil Tahlan, profesor investigador de la Universidad Memorial de Newfoundland, Canadá, por haberme permitido realizar una estancia de investigación en su laboratorio durante la cual discutimos grandes temas relacionados al presente trabajo de investigación.

A mis compañeros del galerón 7 (grupo del Dr. Francisco Barona y grupo del Dr. Rafael Montiel, 2014-2017), por haber formado parte de una de las mejores etapas de mi vida. En especial, gracias a los auxiliares Hilda E., Romina B. y Christian E., por su apoyo incondicional para que este proyecto fuera posible. Gracias Pablito Cruz por haber sido mi más grande apoyo y amigo durante este tiempo. Gracias Nelly S. y César A. por ayudarme en todos los aspectos bioinformáticos y experimentales de mi proyecto. Y gracias Pablito Cruz (otra vez), Romina Berni, Cuauhtémoc Licona, Miguel Vallebuena y Ricardo Gamboa por su amistad, sin ella simplemente no hubiera sobrevivido.

Gracias a mi familia por creer en mí, por apoyarme y ser mi más grande motivación para ser una mejor persona.

Finalmente, pero no menos importante, agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT) por otorgarme el apoyo de beca durante el periodo de realización del presente proyecto.

A mis padres Guillermo y Fani

# ÍNDICE

Abstract .....	1
Resumen .....	2
<b>Capítulo I. Introducción .....</b>	<b>3</b>
1.    Productos naturales .....	3
1.1.    ¿Qué son los productos naturales? .....	3
1.1.1.    Función biológica de los NPs .....	5
1.1.2.    Función antropocéntrica de los NPs .....	7
1.2.    Problemática actual de los NPs .....	9
1.3. <i>Actinobacteria</i> .....	13
1.3.1.    Generalidades de las actinobacterias .....	13
1.3.2.    Ciclo de vida de los <i>Streptomyces</i> .....	14
1.3.3.    Correlación entre el ciclo de vida y la producción de NPs .....	17
1.4.    Estrategias para el descubrimiento de NPs en bacterias .....	18
1.4.1.    Metodologías empíricas .....	18
1.4.2.    Tecnologías -ómicas .....	20
1.5.    Minería genómica de NPs en bacterias .....	23
1.5.1.    ¿Qué es la minería genómica de NPs? .....	23
1.5.2.    Estructura general de un clúster de genes biosintéticos .....	24

1.5.3. Lógica biosintética de las principales clases de BGCs .....	25
PKSs .....	26
NRPSs .....	29
Terpenos .....	32
RIPPs.....	33
1.5.4. Fundamento y clasificación de las principales herramientas de minería genómica de NPs .....	36
<b>Capítulo II. Antecedentes.....</b>	<b>40</b>
2.1. EvoMining .....	42
2.1.1. Fundamentos básicos de EvoMining .....	42
2.2. El tRNA y enzimas que utilizan tRNA (tRUEs) .....	46
<b>Capítulo III. Hipótesis y objetivos del proyecto .....</b>	<b>50</b>
3.1. Hipótesis .....	50
3.2. Objetivo general .....	50
3.2.1. Objetivos específicos .....	50
<b>Capítulo IV. Metodología.....</b>	<b>51</b>
4.1. Generalidades y funcionamiento del programa EvoMining .....	51
4.2. Construcción de las bases de datos de EvoMining.....	52
4.2.1. Base de datos del metabolismo central (CE-DB).....	52
4.2.2. Base de datos de enzimas de genomas (G-DB).....	53

4.2.3	Base de datos de rutas biosintéticas de NPs (BGCs-DB).....	54
4.2.4	Base de datos de G-DBanotados en antiSMASH (antiSMASH-DB) .....	55
4.3	Descripción de la tubería bioinformática de EvoMining.....	55
4.4	Análisis de resultados de EvoMining.....	61
4.4.1	Análisis del contexto genómico de las expansiones enzimáticas .....	61
4.4.2	Predicción de nuevas rutas biosintéticas de productos naturales.....	61
4.5	Caracterización experimental del azoxy-BGC.....	62
4.5.1	Estrategia experimental general .....	62
4.5.2	Material biológico .....	63
4.5.3	Generación de mutantes del azoxy-BGC.....	64
4.5.4	Comparación morfológica entre cepas silvestres y mutantes .....	66
4.5.5	Extracción de RNA.....	67
4.5.6	RT-PCR .....	72
4.5.7	Cromatografía líquida de alta eficiencia .....	74
<b>Capítulo V. Resultados y discusión.....</b>		<b>75</b>
5.1	Adaptación de EvoMining al análisis de las tRUEs en <i>Actinobacteria</i> .....	75
5.1.1.	Construcción de la base de datos CE-DB centrada en tRUEs.....	76
5.1.2.	Construcción de la base de datos G-DB.....	81
5.1.3.	Construcción de la base de datos antiSMASH-DB .....	83
5.1.4.	Construcción de la base de datos BGCs-DB .....	84

5.2. Análisis cuantitativo de EvoMining enfocado a tRUEs .....	86
5.2.1. Expansiones enzimáticas de las familias tRUEs.....	86
5.2.2 Eventos de reclutamiento de tRUEs en biosíntesis de NPs.....	91
5.3 Análisis cualitativo de EvoMining enfocado a tRUEs .....	98
5.3.1 Reconstrucciones filogenéticas de las familias tRUEs reclutadas en BGCs de NPs .....	98
5.4 Expresión del azoxy-BGC .....	103
5.4.1 Propuesta bioinformática .....	103
5.4.2. Caracterización experimental del azoxy-BGC.....	106
Construcción de mutantes.....	106
Análisis comparativo entre el fenotipo de mutantes vs silvestres.....	107
Análisis de HPLC de los extractos de fermentación.....	112
Análisis de RT-PCR del azoxy-BGC .....	114
<b>Capítulo VI. Conclusiones y perspectivas .....</b>	<b>119</b>
6.1. Conclusiones.....	119
6.2. Perspectivas.....	121
<b>Capítulo VII. Referencias y Apéndice .....</b>	<b>122</b>
Referencias .....	122
Apéndice. Reconstrucciones filogenéticas de tRUEs expandidas y reclutadas por BGCs de NPs.....	135

<i>Currículum vitae</i> .....	141
Declaración de independencia .....	143
Índice de figuras .....	144
Índice de tablas .....	146

## Abstract

tRNA-utilizing enzymes (tRUEs) are an enzyme group that performs multiple functions in central metabolism, *e.g.* protein synthesis, cell wall synthesis, lipid aminoacylation and protein degradation. tRUEs have the common characteristic of being tRNA-dependent enzymes. For many years it was thought that these kind of enzymes were exclusively involved in the central processes mentioned. However, recently it has been reported some cases of tRUEs recruited for specialized metabolism, *i.e.* biosynthesis of natural products (NPs) such as valanimycin, pacidamycin, cyclodipeptides, and nisin. Since only few cases of tRUEs recruitment are currently known, an extensive exploration of tRUEs families as specialized-metabolism elements may reveal novel biosynthetic gene clusters (BGCs) of NPs. Also, these results could contribute to complementing conventional genome mining tools with new patterns.

Here, we analyzed the potential of 78 tRUEs families in a context of NPs biosynthesis in actinobacterial genomes using a non-conventional genome mining approach that incorporates enzyme evolutionary components, called EvoMining. Reconstruction of the evolutionary history of these enzyme families resulted in detection of hundreds of tRUEs enzymatic expansions related with BGCs of NPs, mostly cryptic pathways. These *in silico* predictions and the experimental characterization of one of them demonstrated that the tRUEs families analyzed are present in a large number of putative biosynthetic pathways of NPs and have enough potential to be converted in an additional guideline for conventional genome mining tools.

These findings serve as a basis for apply ingenious strategies to express the cryptic BGCs uncovered, elucidate the encoding metabolites and finally, develop new drugs that may help to combating the current crisis for antibiotics.

## Resumen

Las enzimas que utilizan tRNA (tRUEs, por sus siglas en inglés) son un grupo de enzimas que realizan múltiples funciones en el metabolismo central, *e.g.* síntesis de proteínas, síntesis de pared celular, aminoacilación de lípidos y degradación de proteínas. Las tRUEs tienen la característica común de ser enzimas dependientes de moléculas de tRNA. Por muchos años se pensó que este tipo de enzimas estaban involucradas exclusivamente en los procesos centrales mencionados. Sin embargo, recientemente se han reportado algunos casos de tRUEs reclutadas por el metabolismo especializado, *i.e.* biosíntesis de productos naturales (NPs) tales como la valanimicina, pacidamicina, nisina y ciclodipéptidos. Puesto que actualmente solo se conocen pocos casos de reclutamiento de tRUEs, una exploración amplia de familias de tRUEs como elementos del metabolismo especializado podría revelar nuevos clústeres de genes biosintéticos (BGCs) de NPs. Adicionalmente, estos resultados podrían complementar las herramientas de minería genómica con nuevos patrones.

El presente trabajo de tesis consistió en el análisis de 78 familias de tRUEs bajo un contexto biosintético de NPs en actinobacterias mediante un enfoque de minería genómica no-convencional que incorpora componentes evolutivos de las enzimas, llamado EvoMining. La reconstrucción de la historia evolutiva de estas familias de enzimas resultó en la detección de cientos de expansiones enzimáticas de tRUEs relacionadas con BGCs de NPs, la mayoría rutas crípticas. Estas predicciones *in silico* y la caracterización experimental de una de ellas demostraron que las tRUEs analizadas están presentes en un gran número de rutas biosintéticas putativas de NPs y que tienen el potencial suficiente para convertirse en una nueva pauta para las herramientas de minería genómica convencionales.

Estos descubrimientos sirven como base para aplicar estrategias ingeniosas de expresión de los BGCs crípticos descubiertos, elucidar los metabolitos codificados y, finalmente, desarrollar nuevos fármacos que puedan ayudar a combatir la crisis por antibióticos actual.

# Capítulo I. Introducción

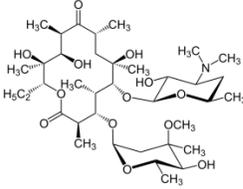
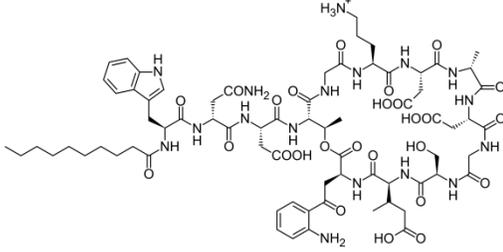
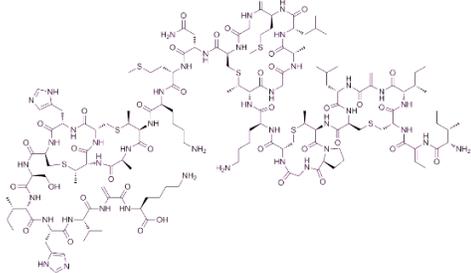
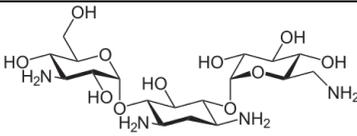
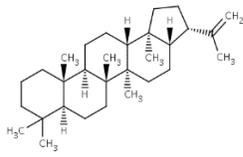
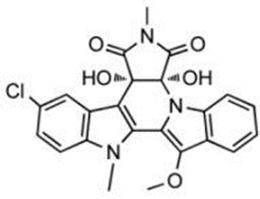
## 1. Productos naturales

### 1.1. ¿Qué son los productos naturales?

Los productos naturales (NPs) son moléculas orgánicas de bajo peso molecular que representan una familia muy amplia y diversa de entidades químicas con múltiples funciones biológicas. Estas moléculas pueden ser producidas por bacterias, hongos, plantas y animales. En bacterias y hongos, los NPs son metabolitos que por lo general se derivan del metabolismo especializado, también conocido como metabolismo secundario, por lo que no se consideran esenciales para el crecimiento, desarrollo y reproducción de la célula (Katz & Baltz, 2016; Jenke-Kodama, Müller, & Dittmann, 2008). Sin embargo, los NPs son fundamentales para enfrentar condiciones ambientales desfavorables que pueden afectar al organismo productor, e.g. escasez nutricional o presencia de otros organismos.

En la actualidad, se conocen más de 23,000 NPs, los cuales provienen de diferentes fuentes plantas, hongos, algunos animales, pero sobretodo bacterias del filo *Actinobacteria*. Estas moléculas han sido clasificados en 6 grandes grupos: policétidos (PKs), péptidos no-ribosomales (NRPs), péptidos sintetizados ribosomalmente y modificados pos-traduccionalmente (RiPPs), sacáridos, terpenos y alcaloides (**Tabla 1**). Los principales criterios considerados para esta clasificación son i) su estructura química, ii) el tipo de enzimas que los sintetizan, iii) el tipo de precursores o sustratos que utilizan las enzimas, y iv) el tipo de modificaciones finales de una molécula base.

Tabla 1. Clasificación de los principales productos naturales

Clase	Descripción de su estructura principal	Ejemplo
Policétidos (PKs)	Concatenación de un ácido esterificado con Coenzima A (e.g. Acetil-, Propionil-CoA) y varias unidades de ácidos malónicos (e.g. malonilo, metilmalonilo, etilmalonilo)	 <p>Eritromicina</p>
Péptidos no-ribosomales (NRPs)	Concatenación de residuos de aminoácidos mediante enlaces peptídicos fuera del ribosoma	 <p>Daptomicina</p>
Péptidos sintetizados ribosomalmente y modificados post-traduccionalmente (RIPPs)	Concatenación de residuos de aminoácidos mediante enlaces peptídicos por el ribosoma pero modificados post-traduccionalmente	 <p>Nisina</p>
Sacáridos	Concatenación de aminoglucósidos mediante enlaces glucosídicos	 <p>Kanamicina</p>
Terpenos	Concatenación de grupos isopreno (lineales o cíclicos)	 <p>Hopanoide</p>
Alcaloides	Estructuras complejas generalmente derivadas de aminoácidos y de carácter básico	 <p>BE-54017</p>

Los NPs cumplen funciones muy diversas, las cuales pueden definirse desde dos puntos de vista distintos: 1) función biológica *i.e.* el papel que juega el NP para el organismo productor y 2) función antropocéntrica *i.e.* la utilidad que le da el ser humano a un NP en beneficio propio, por ejemplo agente antibiótico, anticancerígeno, inmunosupresor, insecticida o herbicida.

### **1.1.1. Función biológica de los NPs**

La relevancia biológica de los NPs ha sido un tema ampliamente debatido desde el descubrimiento de las primeras moléculas. Las hipótesis más relevantes que han surgido desde entonces van desde asumir que los NPs son productos de desecho o desintoxicación del organismo productor (Malik, 1980); hasta sugerir que eran moléculas diversas no funcionales pero que le sirven al hospedero como metabolitos inventivos (Zähner & Anke , 1983). Otra hipótesis declaraba que los NPs solo eran el remanente de metabolitos más complejos dentro de la célula (Davies, 1990). Otros autores, exponían a los NPs como un conjunto de armas químico-biológicas que servían para competir contra otros organismos (Baba & Schneewind, 1998). Sin embargo, ninguna de estas especulaciones proporcionaba una descripción acertada del papel biológico de los NPs, ya que no lograban explicar la enorme diversidad estructural y funcional de los metabolitos conocidos o simplemente carecían de evidencias experimentales.

Uno de los argumentos mayormente aceptados hoy en día es aquel que sugiere que los NPs funcionan como moléculas de señalización, favoreciendo la comunicación entre comunidades complejas de organismos y su medio ambiente (Davies, 2006). En otras palabras, un organismo es capaz de producir y/o

detectar señales químicas que, de acuerdo a la naturaleza de la molécula, le permiten reaccionar positiva o negativamente ante dicho estímulo. Por ejemplo, se ha reportado que la bacteria *Bacillus subtilis* forma 'biofilms' (multicapa de bacterias) en respuesta a diversos NPs que causan pérdidas de potasio a través de la membrana citoplasmática, desencadenando la activación de una cinasa que regula la expresión de genes involucrados en la síntesis de la matriz extracelular (López , Fischbach, Chu, Losick, & Kolter , 2009).

Otra hipótesis fuertemente aceptada es aquella que menciona que los NPs confieren ventajas adaptativas al organismo productor. Por ejemplo, algunas bacterias del género *Streptomyces* producen moléculas de alta afinidad por  $Fe^{3+}$  llamados sideróforos, los cuales actúan como agentes quelantes secuestrando este metal y reduciéndolo a su forma biodisponible  $Fe^{2+}$ . Aquellos organismos que no producen sideróforos y que co-habitan un mismo nicho ecológico que los organismos productores, quedan en desventaja ante la escasez de este mineral esencial para la supervivencia de la célula (Barona-Gómez, y otros, 2006). También tenemos el caso de la rifampicina y la eritromicina, compuestos reconocidos por el hombre por su efecto antimicrobiano pero cuya función biológica es completamente distinta. Organismos productores de rifampicina y eritromicina emplean tales compuestos a concentraciones subinhibitorias para promover cambios transcripcionales globales (Goh, y otros, 2002). En condiciones fisiológicas normales, la concentración de antibióticos producidos por un organismo difícilmente se acerca a la dosis terapéutica necesaria en un humano para inhibir el crecimiento de patógenos (Yim, Wang, & Davies, 2007).

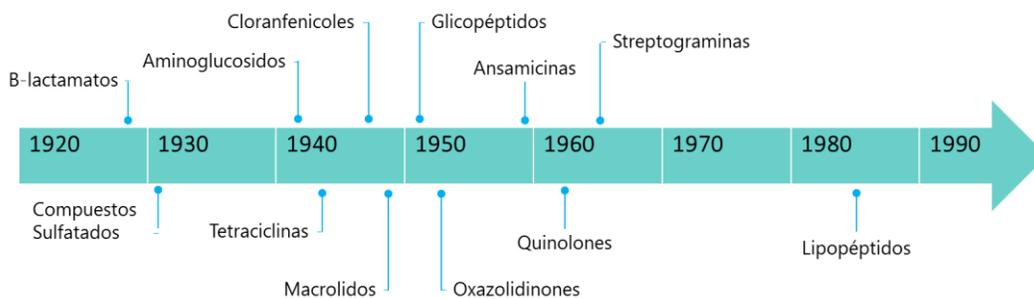
Por otro lado, tenemos el caso de la relación simbiótica entre el hongo *Rhizopus microsporus* y la bacteria *Burkholderia rhizoxinica*. Inicialmente se creía que *R. microsporus* era el organismo productor de la rizoxina, una fitotoxina que causa graves daños a la planta de arroz. Sin embargo, fue comprobado después que el verdadero productor de esta toxina era la bacteria *B. rhizoxinica*, quien vivía dentro de *R. microsporus* (Partida-Martínez & Hertweck, 2005). Otro ejemplo del papel de los NPs en la ecología microbiana y, en este caso, en las interacciones con otras especies.

### **1.1.2. Función antropocéntrica de los NPs**

Desde la antigüedad, los NPs han sido usados por el hombre para su propio beneficio, principalmente en el tratamiento de enfermedades. Los primeros registros del uso de NPs datan desde el año 2600 a.C. en Mesopotamia, quienes usaban aceites vegetales para tratar tos, resfriado e inflamación (Cragg & Newman, 2005). De igual manera egipcios, chinos, griegos, hindús, romanos y árabes obtenían extractos con propiedades medicinales derivadas de la herbolaria de su región mediante técnicas rudimentarias (Dias, Urban, & Roessner, 2012), sin conocer a ciencia cierta la química detrás de dichas infusiones. Sin embargo, el acontecimiento que marcó una gran diferencia en la historia de los NPs, fue el descubrimiento de la penicilina.

Una mañana de 1928 en el laboratorio del hospital de St. Mary, Londres, el microbiólogo Alexander Fleming, quien trabajaba estudiando cepas de *Staphylococcus* provenientes de pacientes enfermos, descubrió que uno de sus cultivos había sido contaminado por un hongo. Al examinar cuidadosamente esta

placa, Fleming notó que este hongo, identificado posteriormente como *Penicillium notatum*, inhibía el crecimiento del patógeno *Staphylococcus* que él mismo había inoculado una noche anterior. Después de varios experimentos y análisis, él y sus colaboradores descubrieron que *P. notatum* producía y secretaba una sustancia específica con actividad antibiótica responsable de la inhibición del crecimiento del patógeno, a la cual denominaron penicilina. Este histórico acontecimiento dio inicio a una revolución científica por encontrar nuevas moléculas con actividad antimicrobiana y, consecuentemente, la producción de fármacos. A esta época se le conoce como la era dorada de los antibióticos, comprendida entre 1930 y 1970. Durante este periodo, se descubrieron la mayoría de los NPs de uso terapéutico y agronómico hoy en día (**Fig. 1**).



**Figura 1. Línea del tiempo de las principales clases de NPs descubiertas durante la era dorada de los antibióticos.** Este periodo inició con el descubrimiento de la penicilina, y culminó cuando dejaron de invertir en investigación de NPs.

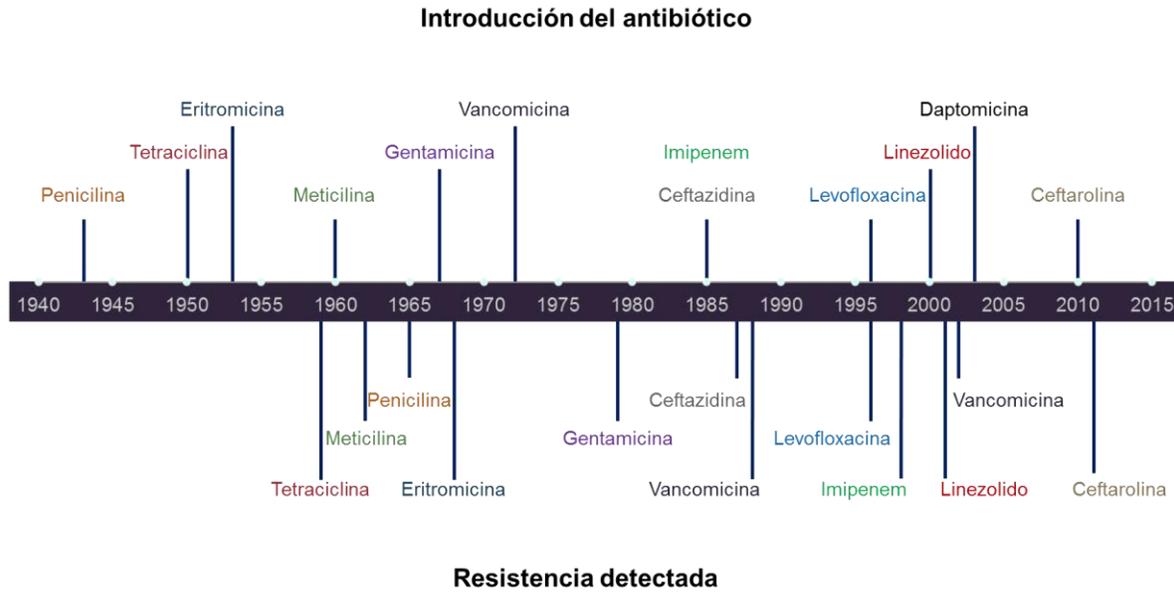
Durante la era dorada de los antibióticos, el ritmo de descubrimiento de NPs y su aprobación para uso terapéutico se mantuvo constante, de modo que hasta 3 nuevos fármacos eran liberados cada año. La vasta disponibilidad de medicamentos en esa época, propició el control efectivo de los organismos patógenos causantes de enfermedades infecciosas (OPEI), al grado de que

incluso algunas enfermedades mortales fueron erradicadas en su momento. Gracias a la revocación de la supremacía de muchos patógenos, fue posible desarrollar otros grandes avances médicos, e.g. intervenciones quirúrgicas, trasplantes de órganos y tratamiento de quemaduras, lo cual indudablemente mejoró la calidad y esperanza de vida de millones de personas alrededor del mundo.

### **1.2. Problemática actual de los NPs**

En un mundo ideal, los antibióticos empleados para el tratamiento de enfermedades infecciosas deberían mantener intacta su efectividad. Sin embargo, la realidad es todo lo contrario. Un fármaco aprobado y liberado para uso clínico pierde efectividad relativamente rápido (**Fig. 2**). Esto se debe a que, al igual que todos los seres vivos, los organismos patógenos evolucionan ante situaciones adversas contra su supervivencia y buscan la manera de sobrepasar dichos cambios perjudiciales.

Para comprender mejor este escenario, imaginemos que una bacteria patógena, e.g. *Mycobacterium tuberculosis*, infecta a un grupo de personas causándoles la enfermedad de tuberculosis. Los médicos tratarán a estas personas con antibióticos, como la rifampicina. En algunos casos este medicamento funcionará, siempre y cuando se siga el tratamiento adecuadamente. Sin embargo, la rifampicina, al igual que otros antibióticos, representa una amenaza para el patógeno *M. tuberculosis*, por lo que este organismo buscará la manera de resistir.



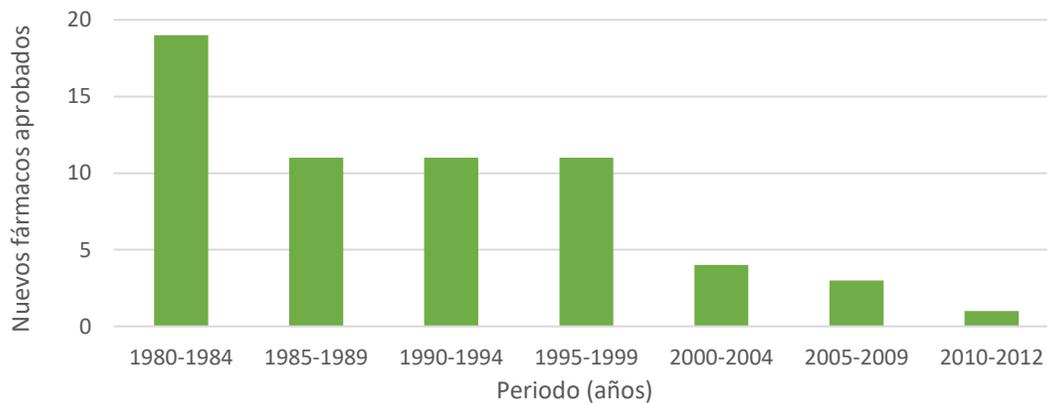
**Figura 2 . Cronología de los eventos clave de resistencia a antibióticos.** Modificado de (CDC, 2014)

Los organismos patógenos multi-resistentes adquieren resistencia mediante la transferencia y acumulación de genes codificantes para proteínas que modifican o degradan los antibióticos, provocando que estos fármacos no actúen sobre sus blancos y sean ineficientes contra la infección. Algunos ejemplos de estos genes de resistencia a antibióticos (MDR, por sus siglas en inglés) son: los genes *bla* que hidrolizan los anillos lactamatos de las penicilinas, los genes *tet* que expulsan las tetraciclinas de la célula, los genes *strA/B* que fosforilan la estreptomicina, o transportadores membranales como *Sul1/2*, *VanA*, *ermA/B*, *ABC*, *MFS*, *RND* y *MATE* los cuales expulsan a los antibióticos de las células (Asit, 2016).

La resistencia a antibióticos coexiste desde los primeros usos de antibióticos para el tratamiento de enfermedades, como podemos observar en la **figura 2**. Durante la era dorada de los antibióticos, este problema se solucionaba fácilmente

mediante el reemplazo del fármaco ineficiente por uno distinto. Esta estrategia fue factible por varias décadas debido a que los patógenos no contenían tantos genes MDR y además, se contaba con una amplia colección de fármacos eficientes disponibles. Sin embargo, el uso descontrolado de antibióticos a lo largo del tiempo ha provocado un brote desenfrenado de patógenos multi-resistentes que amenazan con derrumbar el sistema de salud actual. Esto supone que las bacterias han evolucionado a la par del cambio dramático de su hábitat, por lo que ahora adquieren resistencia más rápido.

Actualmente existen muy pocos antibióticos efectivos contra un número creciente de cepas multi-resistentes e incluso re-emergentes. Pasamos de una producción de 20-30 medicamentos por década, a 3-4 en los últimos años (**Fig. 3**).



**Figura 3. Decaimiento de la producción y aprobación de nuevos fármacos.** Modificado de (CDC, 2014)

Esta crisis por antibióticos se debe a tres causas principales (Bérdy, 2012): 1) factores humanos (mal uso de los antibióticos), 2) dificultades económicas y de

legislación (baja inversión, rigurosas regulaciones, licencias y patentes) y 3) errores científicos (limitaciones en la renovación de las técnicas de exploración). La solución a esta crisis requiere por un lado, promover el uso controlado de antibióticos desde hospitales, hogares, sector agrícola y ganadero. Por otro lado, retomar y formalizar el financiamiento de proyectos de investigación enfocados al descubrimiento y desarrollo de nuevos fármacos, así como una revisión y mejoramiento de las leyes que rigen su liberación al mercado. Y, en cuanto a la parte científica, nos corresponde innovar en técnicas de exploración, romper paradigmas y ser pioneros en el desarrollo de estrategias novedosas para encontrar moléculas con potencial farmacéutico.

En el presente trabajo de investigación, aspiramos a contribuir al mejoramiento de las técnicas de exploración y descubrimiento de NPs potenciales, mediante el uso de herramientas bioinformáticas no-convencionales y técnicas experimentales empleando como modelo a bacterias del filo *Actinobacteria*, los organismos más reconocidos en cuanto a la producción de NPs.

### 1.3. *Actinobacteria*

#### 1.3.1. Generalidades de las actinobacterias

Las actinobacterias constituyen uno de los 18 linajes más grandes del dominio *Bacteria*. Los ciclos de vida de algunas actinobacterias son muy peculiares, ya que son capaces de formar micelio multicelular y esporas, muy parecido al de los hongos (**Fig. 4**). Inicialmente, se consideraba que las actinobacterias eran formas transicionales entre bacterias y hongos debido a la similitud morfológica que compartían. Sin embargo, este parecido solo es superficial ya que, a nivel molecular, las actinobacterias poseen un único cromosoma, su pared celular está compuesta por moléculas de peptidoglicano (Gram-positivas) y por lo tanto, son susceptibles a compuestos antibacterianos.



*Streptomyces  
coelicolor*

*Nocardia  
cyriacigeorgica*

*Saccharopolyspora  
rectivirgula*

**Figura 4. Morfología de algunos ejemplos de actinobacterias.** Una de las características más sobresalientes de algunos miembros del filo *Actinobacteria*, es su capacidad de diferenciación en la formación de hifas aéreas y esporas como en los hongos.

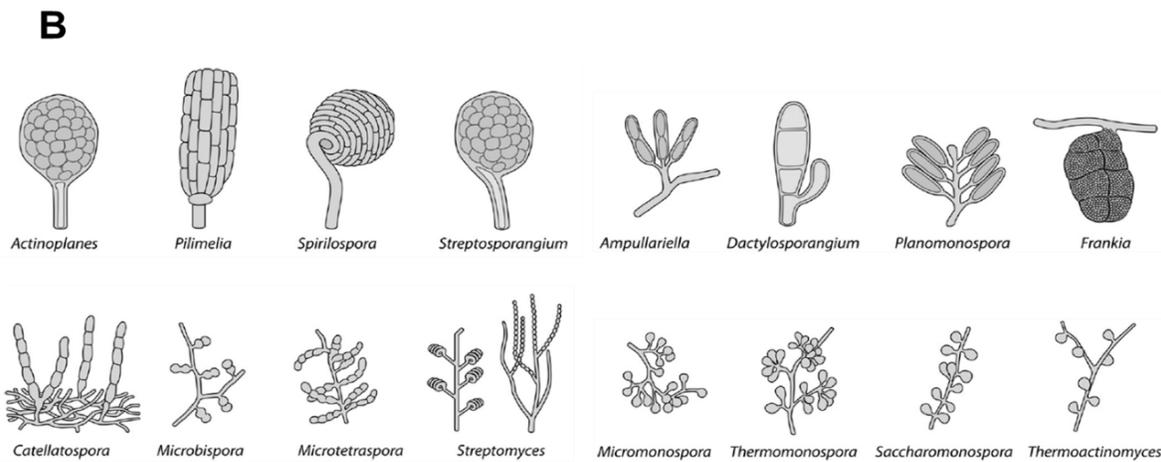
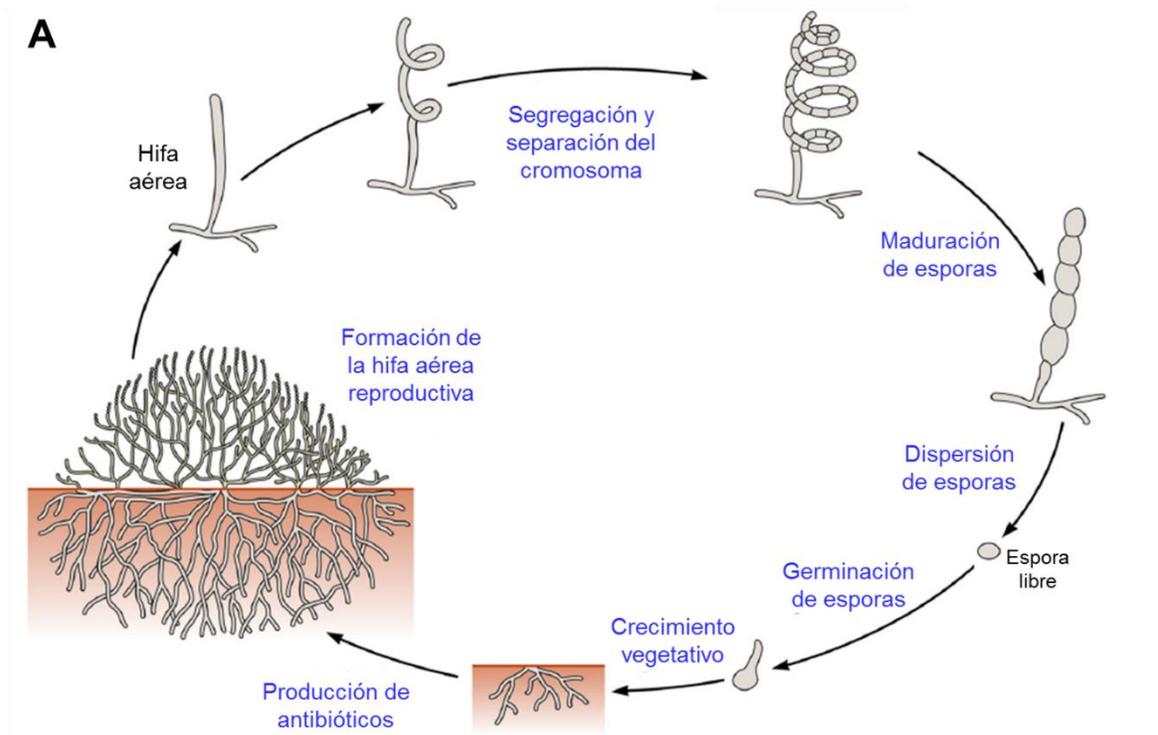
La mayoría de las actinobacterias son de vida libre y están ampliamente distribuidas tanto en ecosistemas terrestres como acuáticos (e.g. *Streptomyces*,

*Micromonospora*, *Rhodococcus*, *Salinispora*). Sin embargo, también existen algunas especies simbiotes de plantas (e.g. *Frankia* spp.), patógenos de plantas o animales (e.g. *Corynebacterium*, *Mycobacterium*, *Nocardia*) o comensales gastrointestinales (e.g. *Bifidobacterium* spp.) (Barka, y otros, 2015). Las condiciones ideales de crecimiento de estas bacterias dependen del estilo de vida que han adoptado, pero generalmente crecen a temperaturas de 25-30°C, un pH neutro y condiciones aerobias (con algunas excepciones e.g. *Thermobifida* y algunas *Streptomyces*).

Una de las características más relevantes de las actinobacterias es la diversidad de su metabolismo especializado. La mayoría de los miembros de este filo bacteriano, especialmente el género *Streptomyces*, son reconocidos por ser excelentes productores de NPs de interés farmacéutico.

### **1.3.2. Ciclo de vida de los *Streptomyces***

*Streptomyces* es el género más representativo del filo *Actinobacteria* por ser organismos ubicuos y ser los mayores productores de NPs. Juegan un papel importante en la ecología del suelo dado que tienen la capacidad de metabolizar diferentes fuentes de carbono y nitrógeno, así como mecanismos de adaptación a ambientes oligotróficos. Su ciclo vital es muy distinto al de otras bacterias, el cual se divide en 4 etapas principales: germinación, crecimiento vegetativo, crecimiento aéreo y esporulación (**Fig. 5-A**) (Barka, y otros, 2015).



**Figura 5. Representación del ciclo de vida del género *Streptomyces*.** A) Etapas del ciclo de vida. B) Diversidad morfológica de esporas producidas por diferentes géneros de actinobacterias. Modificado de (Barka, y otros, 2015)

La germinación de una espora ubicada en un sitio con las condiciones adecuadas para su desarrollo (*i.e.* humedad, temperatura, pH, disponibilidad de nutrientes) marca el inicio del ciclo de vida de *Streptomyces*. Esta primera espora crece y se reproduce exponencialmente formando hifas vegetativas, las cuales se ramifican e intercalan entre sí dando lugar al micelio vegetativo. La formación de este micelio vegetativo demanda una gran cantidad de nutrientes, por lo que es muy común que este suministro se agote rápidamente. Los *Streptomyces* son organismos sésiles que, para sobrevivir ante la escasez nutricional, acuden a la autólisis de una parte de su micelio vegetativo para poder solventar la siguiente etapa de desarrollo, el crecimiento aéreo. Este proceso autolítico ocurre mediante un mecanismo de muerte celular programada a partir del cual se obtienen carbohidratos, lípidos, aminoácidos y nucleótidos, necesarios para la formación del micelio aéreo.

El micelio aéreo se compone de hifas erectas esporogénicas derivadas de la diferenciación celular del micelio vegetativo. Al igual que las hifas vegetativas, las hifas aéreas crecen por extensión de la punta. Las diferencias principales entre ambos tipos de hifas es que las aéreas no forman tantas ramificaciones, miden el doble de ancho y crecen relativamente más rápido que las vegetativas. Además, las hifas aéreas están envueltas por una vaina hidrofóbica que posteriormente pasará a ser parte de la cubierta de las esporas (**Fig. 5-B**). En las hifas vegetativas, la división celular da origen a septos celulares irregulares con compartimentos multi-genómicos, mientras que en las hifas aéreas los septos se forman de manera simultánea, son altamente simétricos y contienen una única

copia de cromosoma. Cuando la cantidad de nutrientes vuelve a ser escasa, el crecimiento aéreo cesa e inicia la esporulación. Cada una de las hifas maduras aéreas dará origen a una cadena de esporas, donde cada septo formará una espora haploide mediante un evento único de división celular coordinado. Posteriormente, estas esporas son liberadas al ambiente donde podrán encontrar las condiciones adecuadas para iniciar el ciclo de vida nuevamente.

### **1.3.3. Correlación entre el ciclo de vida y la producción de NPs**

La etapa del ciclo de vida de los *Streptomyces* que se relaciona con la producción de NPs es el proceso de muerte celular programada (autólisis) del micelio vegetativo. El micelio lisado libera una gran cantidad de nutrientes que serán aprovechados por la misma bacteria para formar el micelio aéreo. Sin embargo, estos nutrientes liberados también quedan expuestos para otros posibles comensales cercanos. Es lógico pensar que, para proteger sus provisiones, los *Streptomyces* comienzan a producir antibióticos en esta etapa. La evidencia que soporta esta hipótesis surgió de observar que la *N*-acetilglucosamina (GlcNAc), uno de los principales componentes de la pared celular que se libera durante la autólisis, actúa como inductor global del desarrollo y la producción de antibióticos (Rigali, y otros, 2006). Esta señalización fue detectada en cultivos con composición mínima de nutrientes, donde el desarrollo y la producción de antibióticos se aceleraron al suplementar dicho medio con GlcNAc (Rigali, y otros, 2008).

Otra molécula que activa la expresión de genes involucrados en la diferenciación morfológica y la producción de antibióticos es la guanosina tetra-fosfato (ppGpp). La ppGpp es una pequeña molécula de señalización intracelular que detecta la ausencia de nutrientes y a su vez, desencadena una serie de respuestas adaptativas en una amplia variedad de bacterias. La síntesis de esta molécula se induce específicamente por insuficiencia de aminoácidos y, por ende, deficiencia de tRNAs aminoacilados para la síntesis de proteínas. Cuando el sitio A del ribosoma recibe un tRNA-no-aminoacilado, se activa la enzima ppGpp sintetasa codificada por el gen *relA*, la cual está unida al ribosoma y produce ppGpp a partir de GTP y ATP (Haseltine & Block, 1973). El gen *spoT* también produce ppGpp en respuesta a la escasez nutricional. El mecanismo de acción por el cual ppGpp regula varios genes aún se desconoce, sin embargo existen evidencias de que esta molécula causa cambios dramáticos en el desarrollo de los streptomycetes (Chakraburttu & Bibb, 1997; Hesketh, Chen, Ryding, Chang, & Bibb, 2007; Battesti & Bouveret, 2006). La regulación de la biosíntesis del antibiótico microbisporicina, un potente lantibiótico tipo I producido por *Microbispora corallina*, mediante la inducción de *relA* por deficiencia de nitrógeno es un claro ejemplo de la aplicación que se le ha dado a este mecanismo (Fernández-Martínez, Gomez-Escribano, & Bibb, 2015).

#### **1.4. Estrategias para el descubrimiento de NPs en bacterias**

##### **1.4.1. Metodologías empíricas**

Durante la era pre-genómica, antes de la secuenciación del primer genoma completo de bacteria (*Haemophilus influenza*, causante de la meningitis) en 1995

(Fleischmann, y otros, 1995), las estrategias para el descubrimiento de NPs consistían principalmente en cuatro etapas: 1) aislamiento de una gran variedad de microorganismos, especialmente bacterias y hongos, 2) fermentaciones a gran escala usando diferentes medios de cultivo, 3) ensayos de bioactividad tanto de extractos crudos como de extractos fraccionados provenientes de las fermentaciones, y 4) perfiles metabólicos por cromatografía para inferir las estructuras de los compuestos. Este tipo de estrategias fueron la base del éxito para descubrir NPs durante la era dorada de los antibióticos a pesar de su naturaleza empírica de “moler y encontrar”. Sin embargo, tenían grandes desventajas, e.g. bajo porcentaje de cepas cultivables en el laboratorio (menos del 1%) y bajo rendimiento de producción de NPs (1-3 por cepa). Las estrategias de innovación de esa época estaban limitadas a explorar nuevos nichos ecológicos, variar las técnicas de aislamiento de microorganismos, probar diferentes condiciones de fermentación y/o realizar distintos tipos de extracción con solventes. No obstante, el re-descubrimiento de NPs seguía siendo una constante, lo cual promovió la búsqueda de nuevas estrategias. Una de las apuestas más grandes por parte de la industria farmacéutica fue la inversión en proyectos de síntesis o modificación química de fármacos a finales de los años 70's. Durante esa época, se desarrolló un gran número de medicamentos totalmente sintéticos o semi-sintéticos considerando la química biológica de los NPs conocidos hasta el momento. Posteriormente, gracias al inicio de la era genómica y los avances en biología molecular, surgieron nuevas estrategias para el descubrimiento de NPs por medio de las tecnologías -ómicas.

### 1.4.2. Tecnologías -ómicas

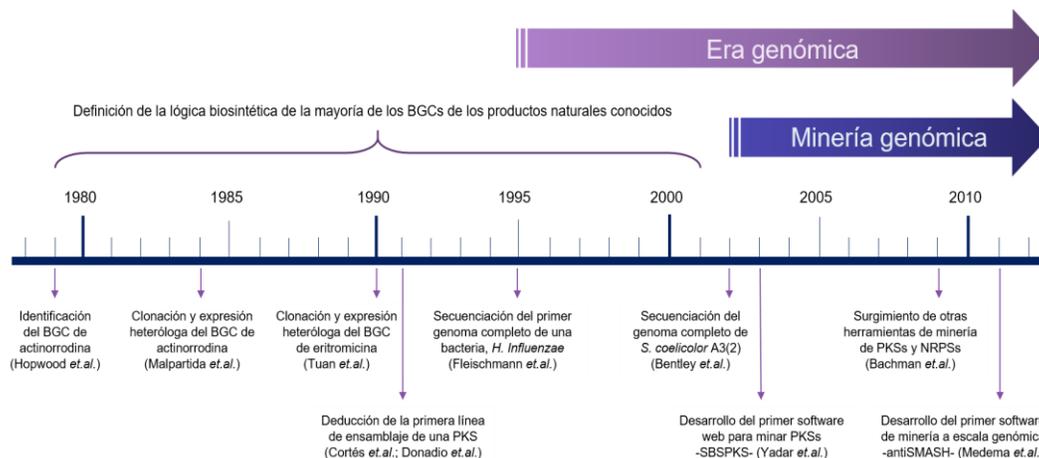
Cincuenta años después del descubrimiento serendípico de la penicilina, Rudd y Hopwood publicaron el primer estudio genético enfocado a la biosíntesis de NPs. En ese estudio, reportaron que los genes responsables de la biosíntesis de actinorrodina, un pigmento celular que funciona como indicador de pH y antibiótico, estaban localizados en un mismo *locus* del cromosoma de *S. coelicolor* A3 (2), sugiriendo que dichos genes formaban un clúster ininterrumpido de genes (Rudd & Hopwood, 1979). Este fue el origen del concepto “clúster de genes biosintéticos” (BGCs, por sus siglas en inglés). Cinco años más tarde, Malpartida y Hopwood lograron clonar y expresar este clúster biosintético en un hospedero diferente, *S. parvulus* (Malpartida & Hopwood, 1984), es decir, expresión heteróloga. Posteriormente, dos grupos de investigación independientes dieron a conocer la naturaleza y organización de los genes del BGC del antibiótico eritromicina, otro de los policétidos más representativos de la historia (Cortes, Haydock, Roberts, Bevitt, & Leadlay, 1990; Donadio, Staver, McAlpine, Swanson, & Katz, 1991). Y casi al mismo tiempo, se logró clonar y expresar de manera heteróloga este BGC de *Saccharopolyspora erythraea* (Tunan, y otros, 1990). El conjunto de estos trabajos derivó en definir por primera vez la lógica biosintética de un policétido, la cual es muy parecida a la síntesis de lípidos. Indiscutiblemente, estos trabajos marcaron el comienzo de una nueva etapa en la historia de los NPs, pues fueron la base para comprender el proceso de biosíntesis de una gran variedad de NPs en bacterias implementando estudios genéticos.

Cabe mencionar que cuando Hopwood y colaboradores descubrieron el BGC de la actinorrodina, no existían genomas completos secuenciados aún. El primer genoma completo que se secuenció fue el de la bacteria patógena *Haemophilus influenzae* (Fleischmann, y otros, 1995). Éste acontecimiento dio origen a la era genómica, durante la cual se han secuenciado ~24,000 genomas a la fecha. En un principio, la mayoría de los proyectos de secuenciación se centraron a organismos modelo tales como *Saccharomyces cerevisiae* (levadura), *Caenorhabditis elegans* (nemátodo), *Arabidopsis thaliana* (angiosperma). Sin embargo, el interés por secuenciar bacterias de importancia industrial surgió a partir de la secuenciación del genoma completo de *S. coelicolor* A3 (2) (Bentley, y otros, 2002). Para entonces ya se conocían algunos de los criterios para detectar BGCs, por lo que cuando estos investigadores analizaron el genoma, encontraron algo inesperado. La cepa modelo *S. coelicolor* A3 (2) fue –y sigue siendo- una de las bacterias predilectas para realizar experimentos, por lo que cuando se secuenció su genoma, se sabía que esta bacteria tenía la capacidad de producir 3 metabolitos secundarios en condiciones estándar de laboratorio: actinorrodina (Malpartida & Hopwood, 1984), prodiginina (Cerdeño, Bibb, & Challis, 2001) y CDA (calcium-dependent antibiotic, siglas en inglés) (Hojati, y otros, 2002). Sorprendentemente, el análisis genómico de *S. coelicolor* A3 (2) reveló que poseía “tesoros ocultos” en forma de genes biosintéticos, es decir, que su genoma albergaba muchos más BGCs de NPs que los 3 clústeres esperados (Bentley, y otros, 2002).

Para el año 2009, ya se habían secuenciado poco más de 70 genomas de actinobacterias tanto cepas patógenas, como cepas de importancia industrial y

ambiental. En un esfuerzo por dimensionar la diversidad metabólica de estas bacterias, los investigadores Nett, Ikeda y Moore realizaron una revisión extensa del contenido de BGCs en 11 actinobacterias representativas: *S. coelicolor* A3 (2), *Streptomyces avermitilis* MA-4680, *Streptomyces griseus* IFO 13350, *Saccharopolyspora erythraea* NRRL 2338, *Salinispora tropica* CNB-440, *Salinispora arenicola* CNS-205, *Frankia* sp. Ccl3, *Frankia alni* ACN14a, *Frankia* sp. EAN1pec, *Nocardia farcinica* IFM 10152, *Rhodococcus jostii* RHA1 (Nett, Ikeda, & Moore, 2009). Este estudio reveló que todas las cepas analizadas, contenían un mayor número de BGCs de lo esperado, tal como sucedió con *S. coelicolor* A3 (2) en 2002, posicionando a las actinobacterias como una de las principales fuentes de NPs de importancia industrial.

La acumulación de toda esta valiosa información, condujo al desarrollo de una nueva estrategia para revelar los “tesoros ocultos” en genomas secuenciados: la minería genómica (cuyos detalles se abordaran en la siguiente sección) (**Fig. 6**). La búsqueda de NPs se volvió más eficaz y más rápida gracias a estos avances de la era post-genómica. La innovación tecnológica actual en genómica, bioinformática, química analítica y biología sintética, también han impulsado la trascendencia de la investigación en NPs.



**Figura 6. Línea del tiempo que muestra el origen de la minería genómica.** La minería genómica surgió gracias a un cúmulo de información valiosa acerca de los principios de la biosíntesis de NPs y a la disponibilidad y avances de la secuenciación genómica.

## 1.5. Minería genómica de NPs en bacterias

### 1.5.1. ¿Qué es la minería genómica de NPs?

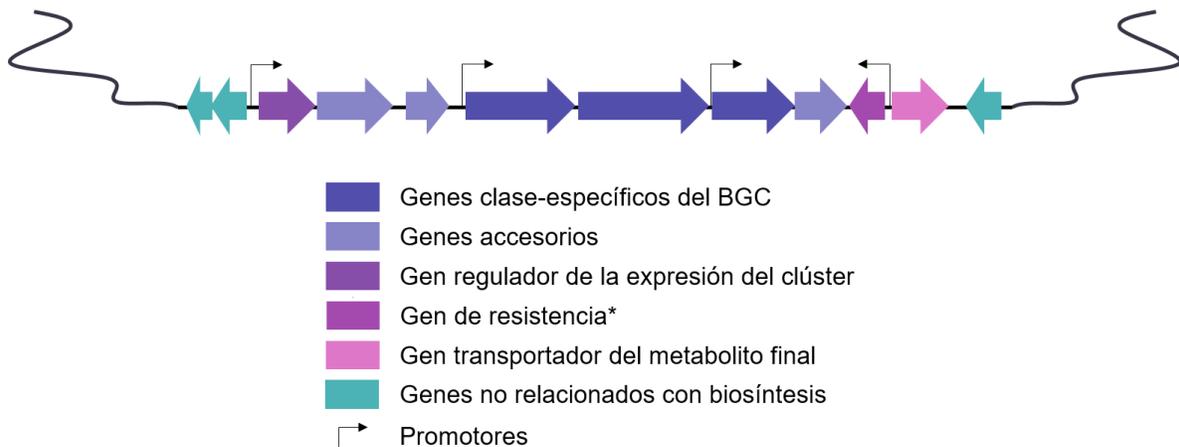
El término 'minería genómica' hace referencia a la fiebre del oro de California (California gold rush) de 1849 (Nett, 2014), periodo histórico en el que un gran número de inmigrantes llegaron a los alrededores de San Francisco, California, en busca de oro en los sedimentos de los ríos. De manera análoga, podemos considerar que dentro de los miles de genomas secuenciados disponibles (sedimento de ríos) existe una enorme riqueza en forma de genes productores de NPs (oro) que podemos aprovechar para obtener grandes beneficios.

En términos bioinformáticos modernos, la minería genómica se define como un análisis computacional de datos de secuencias nucleotídicas basado en la comparación y reconocimiento de patrones conservados (Nett, 2014). Bajo esta definición, cualquier método computacional que implique la búsqueda y predicción de propiedades fisiológicas o metabólicas se considera minería genómica. El

enfoque de minería genómica a NPs, se refiere específicamente a la identificación de clústeres de genes biosintéticos de NPs.

### 1.5.2. Estructura general de un clúster de genes biosintéticos

Un clúster de genes biosintéticos (BGC) se define como un conjunto de genes cuyas funciones están asociadas a la biosíntesis, regulación y transporte de un determinado metabolito. En bacterias, los genes que conforman un BGC generalmente están organizados de manera contigua en un mismo *locus* del cromosoma (**Fig. 7**).



**Figura 7. Estructura general de un clúster de genes biosintéticos (BGC) de NPs en bacterias.** La mayoría de los BGCs están compuestos por estos tipos de genes, por lo que la minería genómica se basa en la identificación de estos genes en un genoma. (\*) Presente solo en rutas de biosíntesis de antibióticos.

El conjunto de genes biosintéticos de un clúster funciona como una línea de ensamble, donde cada una de las enzimas codificadas realiza una función específica e.g. síntesis, ligación, deshidrogenación, oxido-reducción, transferencia, epimerización o ciclación, que van moldeando la estructura final de un metabolito determinado. Los BGCs se clasifican en base a las funciones de estos genes

biosintéticos, su combinación y la afinidad por su(s) sustrato(s). Si comparamos varios clústeres, podemos observar que ciertos genes siempre están presentes en los BGCs de una misma clase (genes-clase específicos). Estos genes determinan la estructura base del NP final. Los demás genes biosintéticos aledaños (genes accesorios) modifican esta estructura base añadiendo diferentes grupos funcionales o cambiando su conformación.

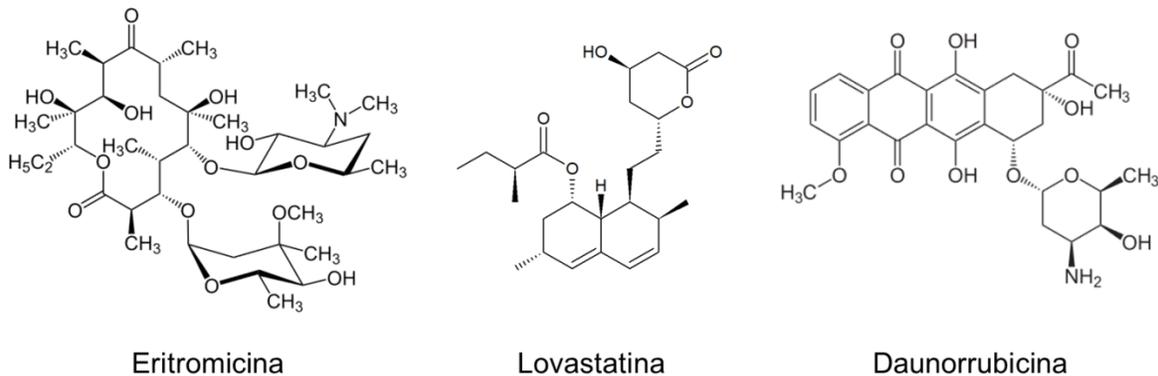
Los genes de regulación codifican para las proteínas que controlan, parcial o totalmente, la expresión del BGC. Estos genes suelen ser inducidos por estímulos ambientales o cambios en el desarrollo celular. Mientras que los genes de transporte, codifican para las proteínas encargadas de transportar el metabolito, ya sea dentro o fuera de la célula. Los promotores son regiones no-codificantes que funcionan como sitio de anclaje para la enzima RNA polimerasa y que consecuentemente ocurra el proceso de transcripción. La orientación y posición de los promotores dictan cuál de las dos cadenas del DNA servirá como molde para la síntesis del ARN mensajero.

### **1.5.3. Lógica biosintética de las principales clases de BGCs**

La minería genómica de NPs en bacterias consiste en la identificación de firmas nucleotídicas de genes asociados a biosíntesis de NPs conocidos y un análisis de vecindad genómica para delimitar los límites del clúster. Cada una de las clases de BGCs se caracteriza por el tipo de genes esenciales y accesorios que lo componen, lo cuales siguen una lógica biosintética específica al expresarse. A continuación se describe la lógica biosintética de las clases más predominantes de BGCs de NPs que se conocen en la actualidad: PKSs, NRPSs, RIPPps y terpenos.

## PKSs

La mayoría de los NPs conocidos producidos por microorganismos pertenecen a la clase de los policétidos, por ejemplo la eritromicina (antibiótico), la lovastatina (disminución de colesterol) y la daunorrubicina (quimioterapéutico) (**Fig. 8**). Los policétidos son moléculas orgánicas complejas cuya estructura base está compuesta por unidades de acetato y/o propionato. Las enzimas encargadas de ensamblar estos residuos son las policétido sintasas (PKSs, por sus siglas en inglés).

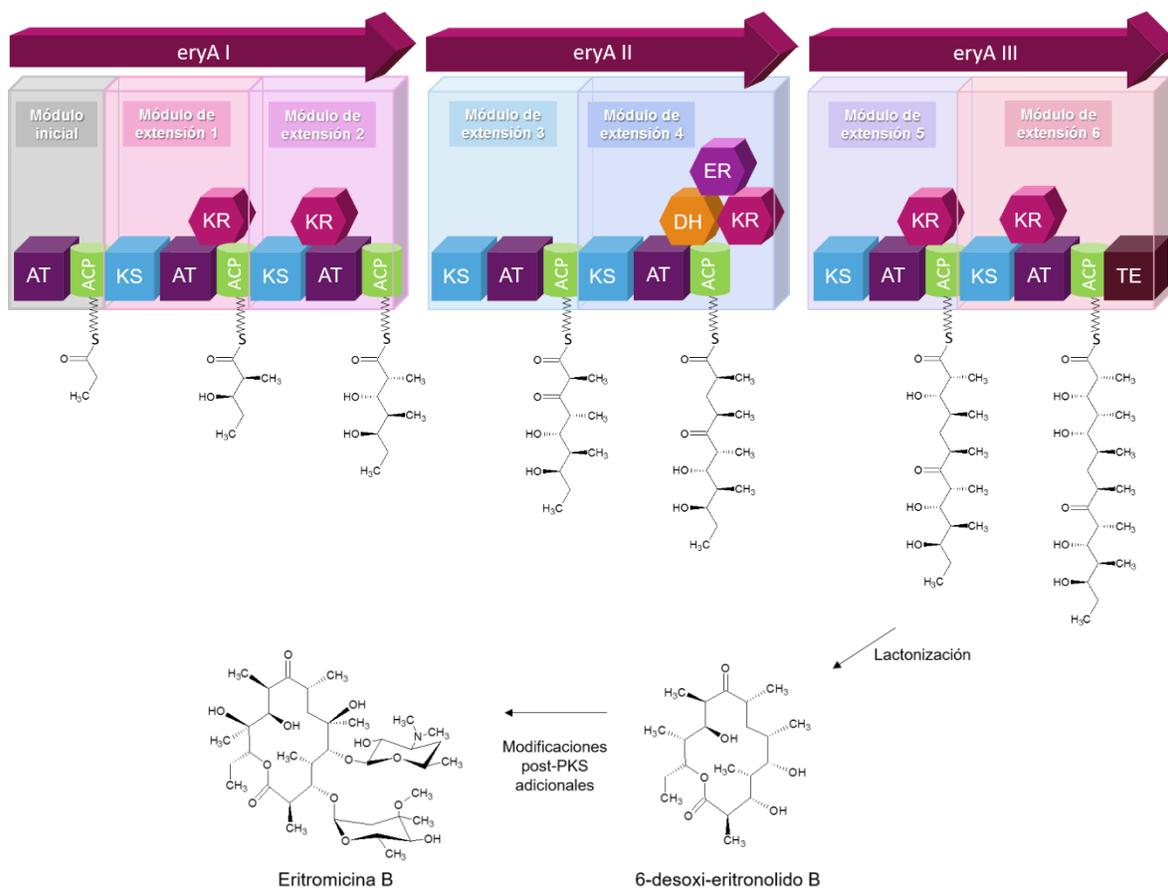


**Figura 8. Estructuras químicas de tres ejemplos de policétidos derivados de microorganismos.** 1) Eritromicina, funciona como antibiótico, producida por la bacteria *Saccharopolyspora erythraea*. 2) Lovastatina, derivado del hongo *Aspergillus terreus*, empleado para tratar altos niveles de colesterol. 3) Daunorrubicina, producto de la bacteria *Streptomyces peucetius*, anticancerígeno.

Las PKSs son proteínas enzimáticas de gran tamaño, cuya organización es muy peculiar. Estas enzimas se organizan en dos niveles: por módulos y por dominios. Una enzima PKS puede contener 2 o 3 módulos, y cada uno de estos módulos está compuesto por al menos 3 dominios esenciales: acil-transferasa (AT), proteína acarreadora de grupos acilo (ACP) y  $\beta$ -ceto-acil-sintasa (KS) (**Fig. 9**). Cada uno de estos dominios lleva a cabo una función específica para la elongación de la cadena del policétido, que es muy parecida a una línea de

ensamblaje. El dominio AT es el encargado de atraer los sustratos (monómeros) correctos que se incorporarán a la cadena. El dominio ACP funciona como el sitio donde ocurre el ensamblaje. Este dominio está unido a una molécula de 4-fosfopantetina, la cual enlaza covalentemente su grupo thiol con los monómeros que se irán cargando a la cadena. El dominio KS, es el encargado de catalizar la reacción de condensación tipo thiol-*Claisen* descarboxilativa de los sustratos provenientes del dominio AT y el precursor unido al dominio ACP.

Los dominios AT, ACP y KS son esenciales para la construcción de la cadena base del policétido. Sin embargo, la amplia diversidad de policétidos que existe en la naturaleza depende en gran medida de la presencia de dominios accesorios, los cuales modifican esta estructura básica. Estos dominios opcionales pueden ser de ceto-reducción (KR), deshidratación (DH) o enoil-reducción (ER). Al menos una PKS dentro del clúster debe contener un dominio de terminación (TE) o reducción (Red) que indican el final de la elongación del policétido correspondiente. El número de módulos y el tipo de dominios que contienen las PKSs, determinan la estructura final del metabolito.



**Figura 9. Estructura y organización general de las PKSs, ejemplo de la eritromicina.** La ruta de la eritromicina contiene tres PKS tipo I: eryAI, eryAII y eryAIII. Cada una de ellas se divide en módulos, y éstos a su vez en dominios. Cada módulo de extensión contiene al menos tres tipos de dominios esenciales: acil-transferasa (AT), proteína acarreadora de grupos acilo (ACP) y β-ceto-acil-sintasa (KS). Adicionalmente pueden contener dominios accesorios que modifican la estructura de los sustratos que se van añadiendo: ceto-reducción (KR), deshidratación (DH) o enoil-reducción (ER). Enzimas adicionales del clúster, decoran esta estructura basal hasta llegar a la eritromicina.

Hasta ahora hemos mencionado las características generales de las PKSs, sin embargo es importante aclarar que existen varios tipos, cuya clasificación se basa en el contenido y naturaleza de sus módulos, sus dominios y el tipo de sustrato que utilizan (Hertweck, 2009). La **tabla 2** resume esta información.

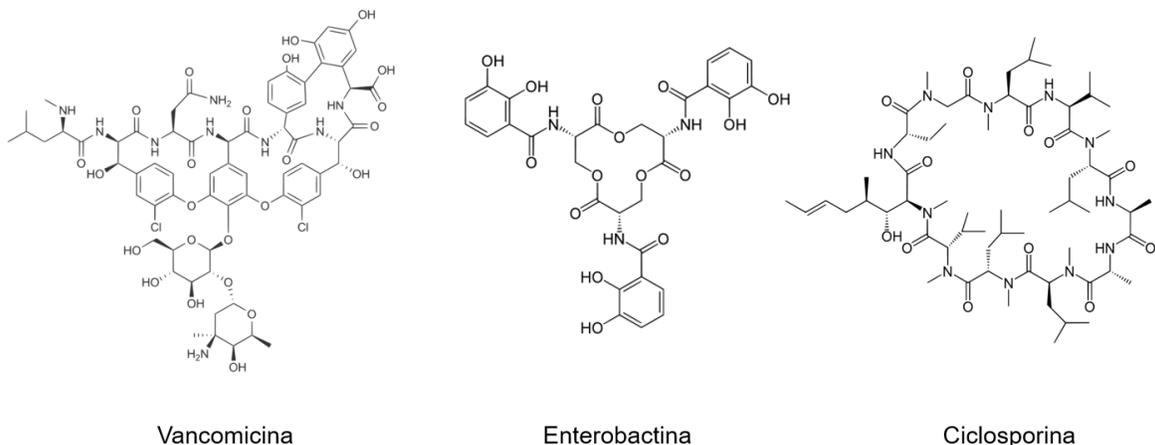
Tabla 2.Descripción general de los tipos de PKSs

Tipo de PKS	Sustratos precursores	Organismos
<b>Modular Tipo I (no iterativa)</b>	ACP, varias unidades de extensión	Bacterias, (protistas)
<b>Iterativa Tipo I</b>	ACP, solo unidades de malonil-CoA	Principalmente hongos, algunas bacterias
<b>(Iterativa) Tipo II</b>	ACP, solo unidades de malonil-CoA	Exclusivamente bacterias
<b>(Iterativa) Tipo III</b>	Acyl-CoA, solo unidades de malonil-CoA*	Principalmente plantas, algunas bacterias y hongos
<b>PKS-NRPS híbrido</b>	ACP, malonil-CoA, aminoácidos	Bacterias (modular) y hongos (iterativa)

\*Solo dos excepciones reportadas

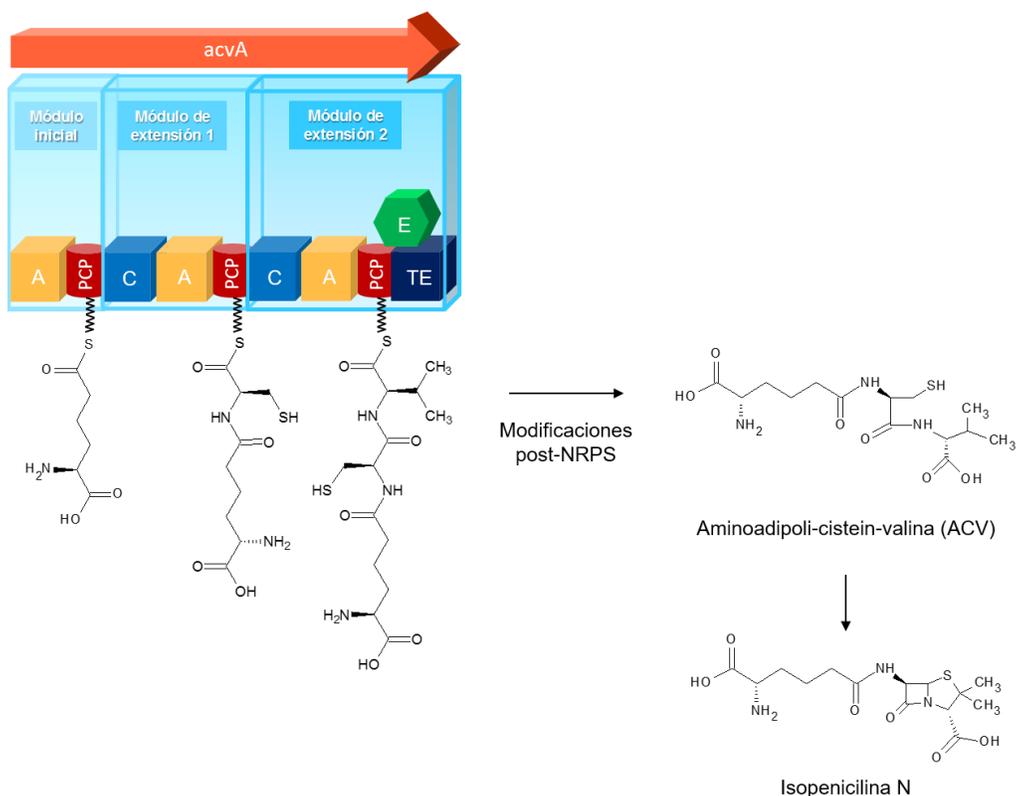
### **NRPSs**

Los péptidos no-ribosomales representan otra de las clases de metabolitos más abundantes y de mayor importancia en la naturaleza, por ejemplo la vancomicina (bactericida), enterobactina (sideróforo) y ciclosporina (inmunosupresor) (**Fig. 10**). La característica que tienen en común estas moléculas es que están formadas por la unión de residuos de aminoácidos o sus derivados. Sin embargo, como su nombre lo indica, estos péptidos no son sintetizados por los ribosomas. Las enzimas encargadas de ensamblar estas moléculas son las sintetisas de péptidos-no ribosomales (NRPSs, por sus siglas en inglés). Al igual que las PKS, las NRPSs son enzimas de gran tamaño y se organizan en módulos y dominios.



**Figura 10. Estructuras químicas de tres ejemplos de péptidos no-ribosomales.** 1) La vancomicina es un antibiótico producido por la bacteria *Amycolatopsis orientalis*. 2) Enterobactina, es un sideróforo altamente afín a hierro que se encuentra principalmente en bacterias Gram-negativas como *Escherichia coli* y *Salmonella typhimurium*. 3) Ciclosporina es un inmunosupresor, comúnmente usado en trasplante de órganos, producto del hongo *Tolypocladium inflatum* Gams.

Las funciones de los dominios de las NRPSs son equiparables a las de las PKSs. Sus dominios esenciales son: adenilación (A), proteína acarreadora de peptidilos (PCP) y condensación (C) (**Fig. 11**) El dominio A se encarga de la selección y activación de los sustratos mediante una reacción ATP-dependiente para producir acil-adenilatos. El dominio PCP permite la unión covalente de los residuos precursores que van formando la cadena peptídica a través de una molécula de fosfopanteteína, al igual que el dominio ACP de las PKS. El dominio C es el encargado de catalizar la reacción de condensación entre los residuos mediante enlaces amida, o en ocasiones especiales por enlaces éster.

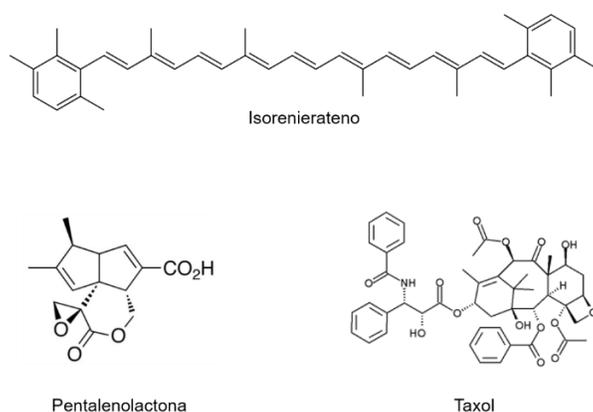


**Figura 11. Estructura y organización general de las NRPs, ejemplo de la isopenicilina.** La ruta de la isopenicilina contiene una enzima NRPS, la cual se divide en 3 módulos, uno inicial y dos de extensión. Al igual que las PKSs, cada uno de estos módulos de extensión contiene al menos tres dominios esenciales: adenilación (A), proteína acarreadora de grupos peptidil (PCP) y condensación (C). Adicionalmente pueden contener dominios accesorios que modifican la estructura de los aminoácidos que se van incorporando, como en este caso el dominio de epimerización (E). Las demás enzimas que están dentro del clúster se encargan de decorar esta estructura basal hasta llegar a la isopenicilina.

Los dominios accesorios de las NRPSs pueden ser de: epimerización (E), invierten la configuración de los residuos; metil-transferasa dependiente de S-adenosil-metionina (MT), que responden a residuos N-metilados; de ciclación (Cy) y/u oxidación (Ox) cuando se trata de unir residuos de cisteína, serina o treonina ciclados; y un dominio de terminación (TE), que marca el fin de la elongación de la cadena (Challis & Naismith, 2004). Al igual que en las PKSs, la cantidad de módulos así como el tipo y organización de los dominios determinan la estructura del péptido final.

## Terpenos

Los terpenos son moléculas derivadas principalmente de plantas, hongos y algunas bacterias, formadas por unidades de isopentenil-difosfato (IPP) y dimetilalil difosfato (DMAPP). Algunos ejemplos importantes de terpenos son: el isorenierateno (carotenoide), pentalenolactona (antibiótico) y taxol (antitumoral) (Fig. 12). En bacterias, los precursores IPP y DAMPP se derivan de la ruta del metil-eritritol fosfato.



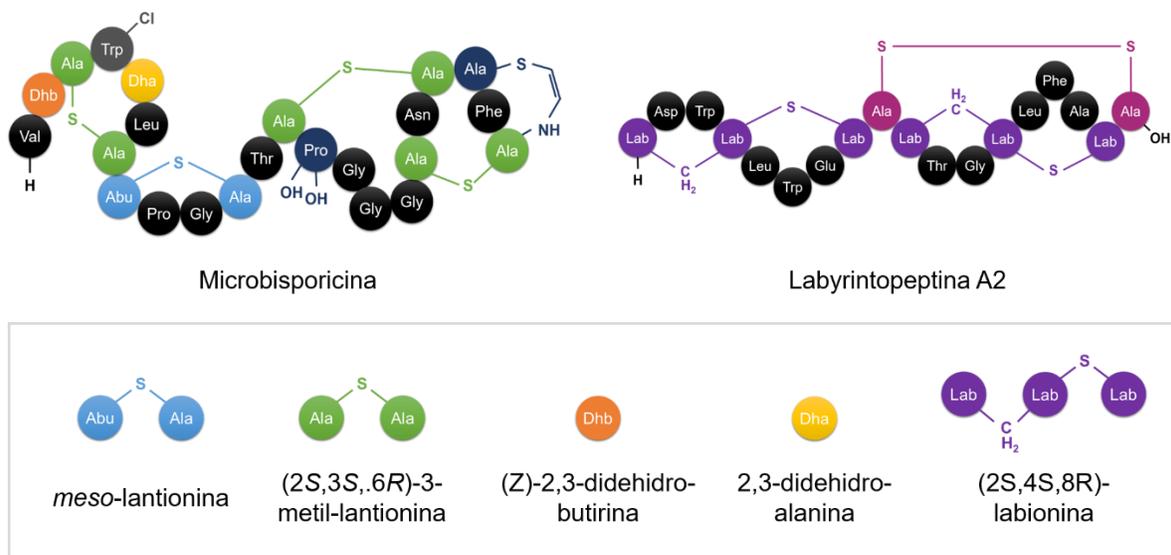
**Figura 12. Estructuras químicas de tres ejemplos de terpenos.** 1) El isorenierateno es un carotenoide-pigmento recolector de luz utilizado como biomarcador en estudios del ciclo del carbono en ciencias marítimas, bacterias verdes de azufre *Chlorobium*. 2) Pentalenolactona, es un antibiótico producido por la bacteria *Streptomyces avermitilis* MA-4680. 3) Taxol es un antitumoral, obtenido de las cortezas de árboles perenes de la especie *Taxus brevifolia*.

La síntesis de terpenos comienza con la formación de una cadena lineal de IPP y DMAPP. Posteriormente, la mayoría de los terpenos se someten a una serie de ciclaciones que pueden conllevar rearrreglos en su conformación o incluso eliminación de algunos carbonos. El BGCs de un terpeno también puede contener enzimas de metilación, hidroxilación o glicosilación, las cuales modifican su

estructura y aumentan la diversidad de este tipo de moléculas (Nett, Ikeda, & Moore, 2009).

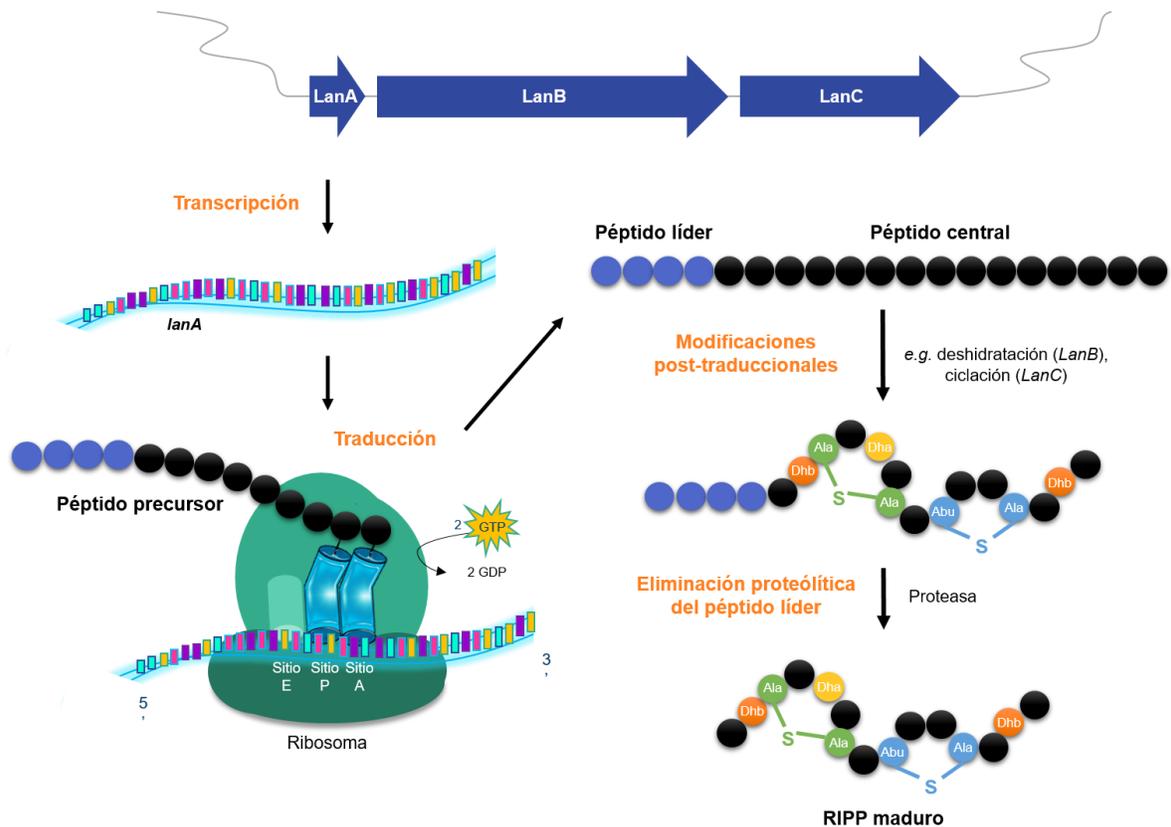
### ***RIPPs***

Los péptidos ribosomalmente sintetizados y modificados post-traduccionalmente (RIPPs, por sus siglas en inglés) son NPs cuya característica estructural consiste en una cadena peptídica policíclica donde los aminoácidos se entrecruzan por medio de enlaces tioéter formando uniones tipo *meso*-lantionina (Lan) y (2S, 3S, 6R)-3-metil-lantionina (MeLan), principalmente (**Fig. 13**). Los RIPPs más representativos son los lantipéptidos (Knerr & van der Donk, 2012), los cuales están distribuidos a lo largo del árbol de la vida. Entre ellos, los más relevantes son aquellos que tienen actividad antibiótica, mejor conocidos como lantibióticos, e.g. la nisina, empleada para la preservación de alimentos (Lubelski, Rink, Khusainov, Moll, & Kuipers, 2008); la actagardina usada para tratar infecciones de *Clostridium difficile* (Sedgwick & Dawson, 2009), la duramicina en el tratamiento de la fibrosis quística (Grasemann, y otros, 2007), y bacteriocinas que se usan en la agricultura (Cotter, Hill, & Ross, 2005).



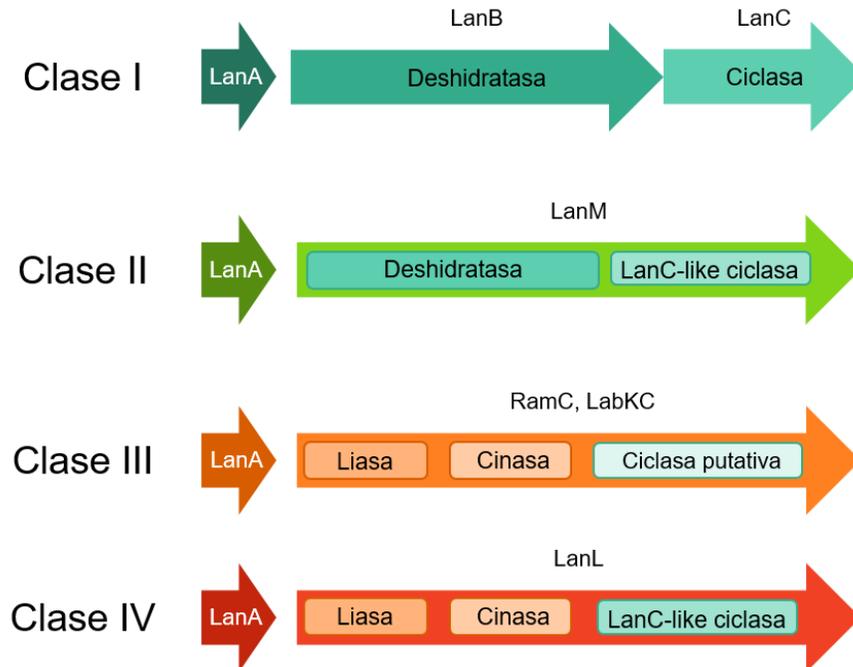
**Figura 13. Estructura química característica de los lantipéptidos.** La microbisporicina y labyrinthopeptina A2 son potentes antibióticos producidos por bacterias del género *Microbispora* y *Actinomadura* sp., respectivamente. En el recuadro se muestran las modificaciones y tipos de entrecruzamiento entre aminoácidos que pueden contener los lantipéptidos maduros. [Abreviaturas: Ala = alanina, Abu = ácido 2-aminobutírico, Asn = asparagina, Asp = ácido aspártico, Dha = 2,3-didehidroalanina, Dhb = (Z)-2,3-didehidrobutirina, Gly = glicina, Lab = labionina, Leu = leucina, Phe = fenilalanina, Pro = prolina, Thr = treonina, Trp = triptófano, Val = valina].

La biosíntesis de cualquier lantipéptido comienza con la transcripción y traducción ribosomal de un gen estructural (LanA), a partir del cual se obtiene un péptido precursor lineal (**Fig. 14**). Este péptido precursor se divide en dos partes, un péptido líder y un péptido central. El péptido líder es la porción N-terminal del péptido precursor que no sufre de modificaciones post-traduccionales (PTMs) ni va unido al producto final, pero que juega un papel importante en la correcta ejecución de las PTMs del péptido central y su transporte dentro de la célula (Knerr & van der Donk, 2012). El péptido central es la porción C-terminal del péptido precursor que se convierte en el lantipéptido final después de las PTMs y la eliminación del péptido líder (Knerr & van der Donk, 2012).



**Figura 14. Esquema general de la biosíntesis de los RIPP.** La biosíntesis de los RIPP comienza con la transcripción y traducción ribosomal del gen estructural (*LanA*) del cual se obtiene un péptido precursor lineal compuesto por un péptido líder y un péptido central. El péptido líder participa en la instalación de las PTMs, exportación e inmunidad del RIPP final. El péptido central se convierte en el RIPP maduro después de las PTMs y la eliminación del péptido líder. [Abreviaturas: *LanA*, gen estructural; *LanB*, deshidratasa; *LanC*, ciclasa; GTP=guanosina trifosfato; GDP=guanosina difosfato].

Dado que la composición basal de los lantipéptidos está restringida a los 20 aminoácidos proteínogénicos, la diversidad de estas moléculas depende en gran medida de las PTMs. Las PTMs más comunes son hetero- o macro-ciclación, deshidratación, acilación, glicosilación, halogenación, prenilación y epimerización (McIntosh, Donia, & Schmit, 2009; Nolan & Walsh, 2009). Los BGCs de RIPP se clasifican en cuatro clases de acuerdo al tipo de enzimas de modificación post-traduccionales que contienen (**Fig. 15**).



**Figura 15. Clasificación de lantipéptidos de acuerdo al tipo de enzimas de modificación post-traduccionales que conforman el BGC.** Todos los BGCs de lantipéptidos contienen un gen estructural (representado como LanA). La diferencia entre clases radica en que los BGCs de los lantipéptidos de la clase I contienen una deshidratasa y una ciclasa separadas individualmente, mientras que los lantipéptidos de las clases II, III y IV contienen enzimas multi-dominio.

#### 1.5.4. Fundamento y clasificación de las principales herramientas de minería genómica de NPs

El principal objetivo de la minería genómica de NPs es identificar clústeres de genes biosintéticos en un genoma determinado y predecir las estructuras químicas de sus productos. Si bien este proceso puede realizarse manualmente, *e. g.* usando BLAST (Basic Local Alignment Search Tool) (Altschul, Gish, Miller, Meyers, & Lipman, 1990) y/o HMMer (programa de análisis de secuencias que usa modelos ocultos de Markov) (Finn, Clements, & Eddy, 2011), actualmente existen diversas herramientas computacionales que han automatizado el proceso. Recientemente, estas herramientas de minería genómica han sido clasificadas en

dos contrastantes categorías: 1) Confiabilidad-alta/Novedad-baja y 2) Confiabilidad-baja/Novedad-alta (Medema & Fischbach, 2015).

La primera categoría incluye a la mayoría de las herramientas de minería genómica más empleadas, e.g. antiSMASH (Medema, y otros, 2011), NRPS-PKS (Ansari, Yadav, Gokhale, & Mohanty, 2004), ClustScan (Starcevic, y otros, 2008), NP.searcher (Li, Ung, Zajkowski, Garneau-Tsodikova, & Sherman, 2009), SMURF (Khaldi, y otros, 2010), BAGEL2 (de Jong, van Hijum, Bijlsma, Kok, & Kuipers, 2010), NaPDos (Ziemert, y otros, 2012), CLUSEAN (Weber, y otros, 2009). En general, estas herramientas persiguen una misma lógica computacional, que es comparar la secuencia nucleotídica o aminoacídica de un genoma contra bases de datos que almacenan información relacionada con BGCs conocidos. Específicamente, estas bases de datos contienen firmas o patrones conservados de los genes esenciales que conforman los BGCs, e.g. patrones conservados de los dominios AT (acyl-transferasa) y KS ( $\beta$ -ceto-acil-sintasa) de las PKSs; firmas de los dominios A (adenilación) y C (condensación) de las NRPSs; firmas de las prenil-transferasas y terpeno-ciclasas de los terpenos; o el gen estructural (bajo un contexto genómico *ad hoc*) de los RIPPs (Nett, 2014). En la práctica, estas herramientas son muy útiles para visualizar el potencial biosintético de un genoma de manera general, o para realizar búsquedas enfocadas a alguna clase de NP en particular. Sin embargo, debido a su alta especificidad por las firmas de BGCs conocidos, las predicciones no van más allá de lo establecido en sus bases de datos por lo que la probabilidad de descubrir nuevos BGCs es muy baja.

La segunda categoría Confiabilidad-baja/Novedad-alta, hasta el momento incluye tres estrategias vanguardistas que han quebrantado la barrera del redescubrimiento, *i.e.* han encontrado nuevas clases de BGCs: ClusterFinder, genómica comparativa seguido por análisis de sintenia y EvoMining. A diferencia de las herramientas de la primera categoría, estos programas no se basan en firmas de secuencias conservadas de BGCs conocidos. ClusterFinder detecta dominios proteicos contiguos cuya función esté relacionada con biosíntesis de NPs. Para esto, traduce un genoma a una cadena larga de dominios Pfam sobre la cual aplica un Modelo Oculto de Markov (HMM, por sus siglas en inglés) para identificar regiones que se asemejen a un BGC (Cimermancic, y otros, 2014). Este HMM fluctúa entre dos estados, BGC y no-BGC. La probabilidad de que un dominio Pfam de la cadena pertenezca a un BGC se designa en base a la frecuencia que tiene ese dominio en bases de datos de BGCs conocidos. Por ejemplo, la probabilidad de un dominio que pertenezca a familias de proteínas como las oxidorreductasas, metil-transferasas o ligasas, es alta, ya que estas proteínas son muy frecuentes en diferentes tipos de BGCs. En un análisis extenso de genomas microbianos, ClusterFinder logró identificar un gran número de nuevos BGCs, entre ellos aril-polienos desconocidos.

La segunda estrategia consiste en dos partes: 1) genómica comparativa y 2) un análisis de sintenia para encontrar nuevos BGCs (Takeda, Umemura, Koike, Asai, & Machida, 2014). Primero se comparan dos genomas distintos y se identifican alineamientos de genes homólogos contiguos (clústeres). Los límites de estos clústeres se determinan por el puntaje de similitud de los genes adyacentes. Por

último, se seleccionan solo los clústeres que se encuentren en regiones no sinténicas al genoma central (“core genome”) taxonómico de las especies comparadas. Aplicando esta metodología en genomas de *Aspergillus*, se descubrió el BGC del ácido kójico (Takeda, Umemura, Koike, Asai, & Machida, 2014).

La tercera estrategia, EvoMining, es una herramienta bioinformática diseñada en nuestro grupo para el descubrimiento de nuevos BGCs, cuya peculiaridad radica en la incorporación de conceptos evolutivos al proceso de minería genómica convencional (Cruz-Morales, y otros, 2016). EvoMining reconstruye la historia evolutiva de familias de enzimas que pertenecen al metabolismo central, mediante la cual puede detectar expansiones enzimáticas relacionadas con biosíntesis de NPs (Véase Sección 2.1 para más detalles). En la primera prueba de concepto de esta herramienta, se logró identificar y caracterizar un nuevo clúster en la cepa modelo *S. coelicolor* que produce un arseno-compuesto (Cruz-Morales, y otros, 2016), el primer caso de este tipo en una de las cepas más minadas de NPs. Estos resultados abrieron las puertas para ampliar nuestros análisis en la búsqueda de nuevos NPs ocultos en actinobacterias, y así contribuir al descubrimiento de nuevos compuestos bioactivos para combatir la actual crisis de antibióticos.

## Capítulo II. Antecedentes

Uno de los principales problemas que enfrentamos en la búsqueda de nuevos NPs es el redescubrimiento de moléculas. A pesar de que se han logrado grandes avances en la investigación en NPs, en las últimas décadas se han descubierto muy pocas nuevas clases de moléculas. Esto se debe principalmente a una falta de innovación en las estrategias utilizadas, tanto en los métodos experimentales como en las herramientas bioinformáticas. En el caso de las estrategias experimentales, algunos grupos han afrontado el redescubrimiento explorando nuevas fuentes de antibióticos (bacterias incultivables, nichos remotos) mediante metagenómica (Oulas, y otros, 2015), rescatando librerías existentes en busca de antibióticos potenciales olvidados o subestimados (Herrmann, y otros, 2016), o mediante ingeniería de metabolitos (Khosla & Keasling, 2003). Sin embargo, descubrir nuevas clases de NPs por este camino implica grandes retos y una gran inversión de tiempo, dinero y esfuerzo, lo cual es contraproducente para la urgencia de encontrar nuevas entidades químicas.

Bajo esta premisa, la minería genómica de NPs ofrece mejores ventajas en cuanto a la detección masiva y relativamente rápida de BGCs codificados en los genomas. Las primeras herramientas de minería genómica surgieron gracias a la acumulación de una gran cantidad de información en torno a la biosíntesis de NPs y a la creciente disponibilidad de secuencias genómicas. Sin embargo, la estrategia que implementan estos métodos, centrada en búsquedas por similitud de firmas conservadas en las secuencias, arroja predicciones completamente sesgadas a clases conocidas de BGCs, e.g. policétido sintasas (PKSs) o

sintetasas de péptidos no-ribosomales (NRPSs) (Medema & Fischbach, 2015). El uso de estas herramientas convencionales es útil para tener una visión general del contenido de BGCs conocidos en un genoma. Sin embargo, para detectar nuevas clases de BGCs es necesario cambiar esta estrategia.

antiSMASH, una de las herramientas de minería genómica de mayor renombre en el área, recientemente creó la versión antiSMASH 3.0 de su plataforma, la cual integra al algoritmo ClusterFinder (Weber, y otros, 2015). Como mencionamos anteriormente, ClusterFinder predice BGCs basándose en un análisis de dominios Pfam relacionados con biosíntesis de NPs (Cimermancic, y otros, 2014). La fusión de antiSMASH y ClusterFinder permite la detección tanto de clústeres biosintéticos de todas las clases conocidas como de clústeres putativos de nuevas clases de NPs, aumentando la posibilidad de encontrar nuevas clases de BGCs. Sin embargo, este tipo de búsquedas siguen siendo simples comparaciones de secuencias, ya sea de secuencias nucleotídicas o dominios proteicos asociados con biosíntesis de NPs. Sin quitarles mérito, su tasa de redescubrimiento de moléculas sigue siendo alta. Si el objetivo es encontrar nuevas clases de NPs, debe emplearse otro tipo de estrategias.

Recientemente, nuestro equipo de trabajo –el Laboratorio de Evolución de la Diversidad Metabólica-, desarrolló una herramienta bioinformática para descubrir BGCs de NPs ocultos en genomas bacterianos, llamada EvoMining (Cruz-Morales, y otros, 2016). Esta herramienta es única en su tipo ya que incorpora principios evolutivos de las enzimas y las rutas biosintéticas de NPs. A continuación, se detallan sus fundamentos.

## **2.1. EvoMining**

### **2.1.1. Fundamentos básicos de EvoMining**

EvoMining es una herramienta de minería genómica creada por el Laboratorio de Evolución de la Diversidad Metabólica que aborda el problema de redescubrimiento de moléculas implementando una estrategia no-convencional fundamentada en tres principios evolutivos. El primer principio asume que una enzima evoluciona funcionalmente reteniendo su mecanismo de reacción pero ampliando su especificidad por sustrato (Gerlt & Babbitt, 2001). El segundo, se refiere a que la evolución de las vías metabólicas contemporáneas ocurre frecuentemente mediante el reclutamiento de familias existentes para realizar nuevas funciones metabólicas (Caetano-Anollés, y otros, 2009). Y el tercero, indica que los BGCs son sistemas metabólicos que evolucionan rápidamente y que pudieran tener su origen en el metabolismo central (Vining, 1992; Medema, Cimermancic, Sali, Takano, & Fischbach, 2014).

Uno de los ejemplos canónicos que soporta estos principios evolutivos es el caso de la relación que existe entre las sintetasas de ácidos grasos (FASs, por sus siglas en inglés) y las PKSs. Las FASs son enzimas del metabolismo central reconocidas por su papel en la síntesis de lípidos. Recientemente se dedujo que las FASs son los ancestros de las PKSs, enzimas del metabolismo especializado encargadas de biosintetizar policétidos, dado que comparten una misma lógica biosintética pero diferente especificidad por sustrato (Jenke-Kodama, Sandmann, Müller, & Dittmann, 2005). A este vínculo entre enzimas se le conoce como expansión enzimática y, para fines del presente trabajo, representa una

correlación ancestral entre enzimas del metabolismo central y enzimas del metabolismo especializado asociadas con biosíntesis de NPs.

Las herramientas de minería genómica convencionales detectan BGCs en genomas enfocando sus búsquedas a firmas conservadas de los genes esenciales de cada una de las distintas clases de BGCs que se conocen en la actualidad. En cambio, EvoMining reconstruye la historia evolutiva de familias de enzimas fijas del metabolismo central con la finalidad de detectar eventos de expansiones enzimáticas que sean significativamente similares a genes altamente probables en BGCs. Con EvoMining es posible detectar nuevos clústeres biosintéticos dado que su búsqueda no está enfocada a genes característicos de las clases conocidas, sino que examina el potencial de enzimas centrales inexploradas en el contexto de biosíntesis de NPs.

A grandes rasgos, la tubería bioinformática de EvoMining comienza con la identificación de todas las proteínas homólogas, tanto ortólogos como parálogos, de una familia de enzimas fija en el metabolismo central (e.g. síntesis de proteínas, síntesis de componentes celulares) en un grupo de genomas seleccionados. Acto seguido, mediante una estimación heurística del número 'normal' de homólogos detectados en un conjunto de genomas analizados, EvoMining determina cuáles familias de enzimas centrales han tenido eventos de expansiones enzimáticas durante su historia evolutiva. Posteriormente, se comparan estas expansiones enzimáticas contra una base de datos que contiene información curada de todos los genes implicados en biosíntesis de NPs actualmente conocidos –repositorio MIBiG (Medema, y otros, 2015)-. Una vez que

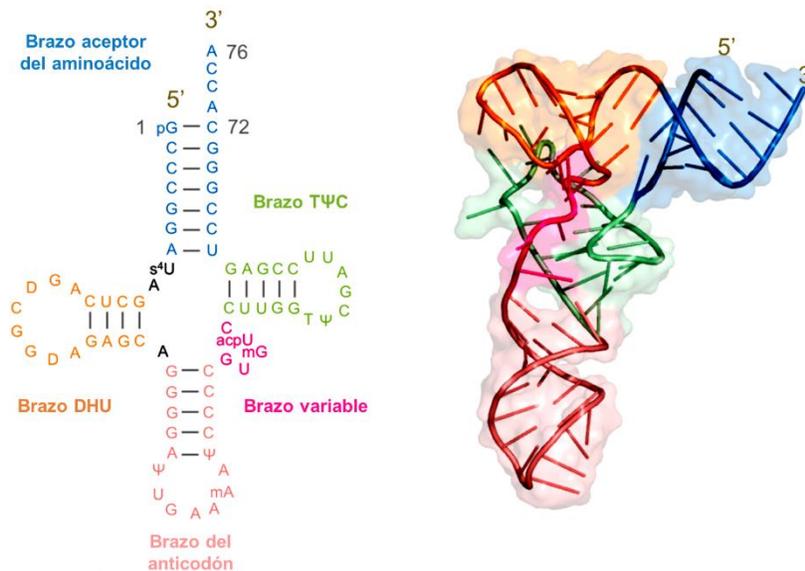
se han detectado asociaciones entre enzimas expandidas y genes de biosíntesis de NPs curados, se reconstruye una filogenia por cada familia de enzimas centrales analizada. La reconstrucción de estos árboles filogenéticos es la mejor manera de visualizar las expansiones enzimáticas. En este caso, los árboles incluyen todos los homólogos detectados en los genomas y los genes homólogos rescatados de la base de datos MIBiG. Por lo general, estos árboles mantienen una misma topología que se caracteriza por contener clados visiblemente divergentes, los cuales representan las expansiones enzimáticas de esa familia. Si el clado divergente además contiene la enzima rescatada de la base de datos MIBiG, quiere decir que las enzimas expandidas de este clado podrían estar formando parte de un BGC.

Durante la prueba de concepto del programa EvoMining se descubrió un compuesto arseno-orgánico nunca antes visto en *S. coelicolor* A3 (2) y *S. lividans* (Cruz-Morales, y otros, 2016), dos de las cepas modelo más estudiadas en el área. Esta predicción derivó de la reconstrucción de la historia evolutiva de la familia de enzimas *AroA* (5-enolpiruvilshikimato-3-fosfato sintasa), involucrada en la biosíntesis de aminoácidos aromáticos (fenilalanina, tirosina y triptófano) por medio de la ruta de shikimato. Esta enzima central nunca antes había sido explorada en el contexto de biosíntesis de NPs, por lo que no existían patrones o firmas conservadas conocidas que sirvieran de guía para detectar el clúster por medio de minería genómica convencional. El compuesto arseno-orgánico derivado de este BGC pertenece a una nueva clase de NPs, lo cual demuestra el poder predictivo de EvoMining.

En base a estos resultados, surgió la inquietud por investigar más familias de enzimas centrales inexploradas que pudieran revelarnos tesoros ocultos en actinobacterias. Una de las propuestas más interesantes en nuestro laboratorio, fue explorar familias de enzimas que utilizan tRNAs (tRNA-utilizing enzymes, tRUES). Como su nombre lo indica, las tRUEs son enzimas dependientes de tRNA, moléculas esenciales en diversos procesos del metabolismo central *e.g.* la síntesis de proteínas. El antecedente particular que motivó hacia el estudio de las tRUEs fue que en la última década se habían reportado 4 casos especiales de rutas biosintéticas de NPs que contienen este tipo de enzimas en sus BGCs: la ruta de la valanimicina (Grag, Qian, Alemany, Moran, & Parry, 2008), pacidamicina (Zhang, Ostash, & Walsh, 2010), ciclodipéptidos (Belin, y otros, 2012) y lantipéptidos (Ortega, y otros, 2015). Esto nos dio la premisa de que este tipo de enzimas dependientes de tRNA tenían buen potencial biosintético no solo en el metabolismo central, sino también en el especializado. Por lo tanto, diseñamos una estrategia para explorar este grupo de familias de tRUEs usando la herramienta bioinformática EvoMining.

## 2.2. El tRNA y enzimas que utilizan tRNA (tRUEs)

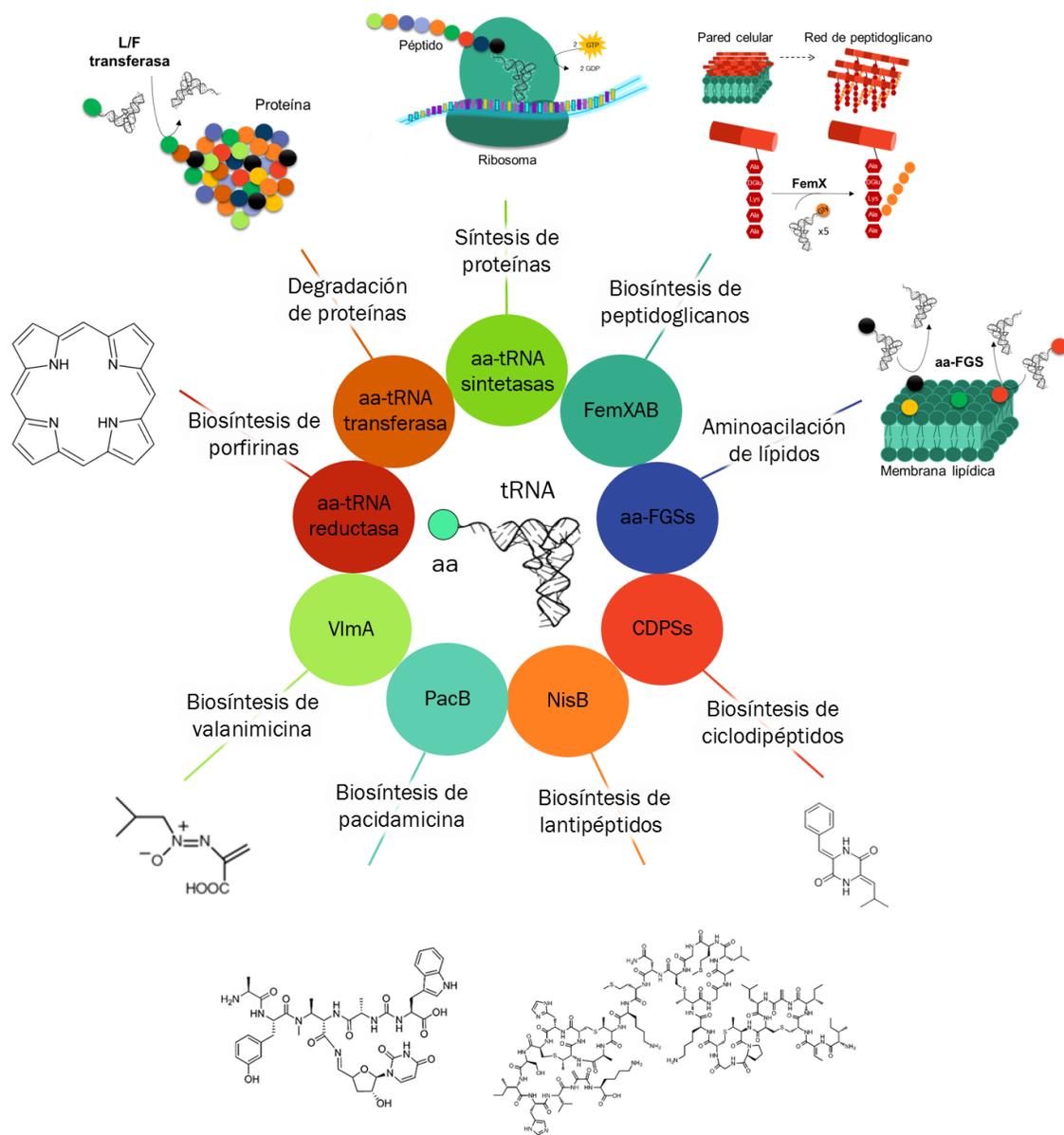
El ácido ribonucleico de transferencia (tRNA, por sus siglas en inglés) es una molécula de RNA formada por 70-100 nucleótidos, cuya naturaleza nucleotídica determina su estructura característica 2D en forma de 'trébol' y 3D en 'L' (**Fig. 16**) (Holley, y otros, 1965; Kim, y otros, 1973; Giegé, y otros, 2012). Los tRNAs están distribuidos y relativamente conservados en los tres dominios de la vida *Bacteria*, *Archaea* y *Eucarya* (Fujishima & Kanai, 2014).



**Figura 16. Estructura secundaria (modelo 'trébol') y terciaria (modelo 'L') del tRNA<sup>Ala</sup>.** La estructura en forma de 'trébol' se caracteriza por poseer 5 brazos formados por series de secuencias complementarias dentro de un mismo tRNA. El primer brazo se denomina brazo aceptor del aminoácido y está formado por los dos extremos de la molécula de tRNA, del cual sobresale una secuencia sencilla CCA en el extremo 3'. El segundo es el brazo TΨC, que recibe este nombre por la ocurrencia de la secuencia T-Psi-C (Psi, pseudouridina, un uracilo modificado) en la mayoría de los tRNAs. De manera adyacente, puede o no existir un tercer brazo pequeño de 4-21 nucleótidos, el cual difiere ampliamente en su composición entre especies. El cuarto brazo posee los tres nucleótidos que forman el anticodón que empalmará con la secuencia codón del mRNA. El quinto brazo se conoce como DHU, el cual contiene dos o más residuos de DHU (dihidrouridina) en diferentes posiciones. Una vez que la molécula de tRNA adopta la conformación 'L', uno de los extremos estará en contacto directo con el aminoácido y el otro extremo contendrá la secuencia anticodón que interactuará con el mRNA durante el proceso de la traducción de proteínas.

Enfocándonos al metabolismo central, los tRNAs llevan a cabo diferentes funciones esenciales para la célula (Francklyn & Minajigi, 2010). Entre ellas, la más destacada es el transporte de los aminoácidos hacia el ribosoma para la decodificación del ácido ribonucleico mensajero (mRNA) a proteínas. En la naturaleza existe al menos una isoforma de tRNA por cada uno de los 20 aminoácidos proteínogénicos. Cada célula contiene tanto tRNAs como tRNAs aminoacilados (tRNAs unidos a su aminoácido cognado, *e.g.* Ala-tRNA<sup>Ala</sup>), por lo que este tipo de moléculas son muy diversas y abundantes dentro de la célula. Dicho esto, es lógico imaginar que los tRNAs y los aminoacil-tRNAs pueden tener diferentes destinos dentro de la célula.

Los tRNAs como tal, pueden asimismo cumplir funciones regulatorias en los procesos de transcripción, traducción y recambio proteico (Hurto, 2011). Por su parte, los aminoacil-tRNAs son capaces de seguir diferentes directrices dentro de la célula. El principal destino celular es el ribosoma, donde ocurre la síntesis de proteínas. Sin embargo, los aminoácidos cargados por los aminoacil-tRNAs también pueden ser entregados a otras enzimas del metabolismo central, como las FemXAB para la síntesis de pared celular (Schepred & Ibba, 2013); las L/F transferasas involucradas en la degradación dirigida de proteínas (Bachmair, Finley, & Varshavsky, 1986); las Glu-tRNA reductasas que participan en la biosíntesis de porfirinas (Jahn, Verkamp, & Söll, 1992) o las aminoacil-tRNA fosfatidilglicerol sintasas (aa-FGSs) implicadas en mecanismos de resistencia a agentes catiónicos (Arendt, Hebecker, Jäger, Nimtz, & Moser, 2012) (**Fig. 17, parte superior**).



**Figura 17. Directrices de los aminoacil-tRNAs dentro de la célula.** Los aminoacil-tRNAs son elementos vitales para la célula dado que están involucrados en diversas rutas biosintéticas del metabolismo central como la síntesis de proteínas, la síntesis de peptidoglicanos, la síntesis de porfirinas, la degradación dirigida de proteínas y la aminoacilación de lípidos. Sin embargo, también se ha reportado recientemente que este tipo de moléculas participan en la biosíntesis de metabolitos especializados como la valanimicina, pacidamicina, lantipéptidos y ciclodipéptidos.

Recientemente se ha reportado que algunos aminoacil-tRNAs también han sido reclutados por enzimas del metabolismo especializado, específicamente en rutas de biosíntesis de antibióticos como la valanimicina (Grag, Ma, Hoyt, & Parry, 2002) ciclodipéptidos (Gondry, y otros, 2009), pacidamicina (Zhang, Ostash, & Walsh, 2010) y lantibióticos (Ortega, y otros, 2015) (**Fig. 17, parte inferior**). La función principal de las enzimas dependientes de tRNA que forman parte de la ruta de la valanimicina, la pacidamicina y los ciclodipéptidos (VlmL, PacB, CDPSs, respectivamente), es transferir el residuo de aminoácido correspondiente que proviene de moléculas aminoacil-tRNAs hacia la línea de ensamblaje del antibiótico. En el caso de los lantibióticos, las enzimas dependientes de tRNA (e.g. NisB, en la ruta de la nisina) intervienen en las modificaciones post-traduccionales del péptido central, concretamente en el proceso de deshidratación, activando los residuos Ser/Thr por glutaminación. Estos casos recientemente caracterizados han puesto en manifiesto que las tRUEs constituyen una familia de enzimas con gran potencial para el descubrimiento de nuevas clases de BGCs de NPs.

El presente trabajo está enfocado a la búsqueda de BGCs que contengan tRUEs distribuidos por el filo *Actinobacteria*, los organismos más reconocidos en la producción de NPs. Para lo cual, empleamos la herramienta EvoMining, que nos permitirá vincular evolutivamente dichas enzimas del metabolismo central con posibles expansiones asociadas a BGCs de NPs nunca antes vistos.

## Capítulo III. Hipótesis y objetivos del proyecto

### 3.1. Hipótesis

La reconstrucción de la historia evolutiva de familias de enzimas que utilizan tRNA (tRUEs) en *Actinobacteria* impulsa el descubrimiento de nuevas rutas biosintéticas de NPs.

### 3.2. Objetivo general

Explorar el potencial de las enzimas que utilizan tRNA en *Actinobacteria* como pauta para la minería genómica convencional mediante un análisis filogenético de sus expansiones enzimáticas asociadas con rutas biosintéticas de NPs.

#### 3.2.1. Objetivos específicos

1. Recolectar y clasificar las familias de enzimas que utilizan tRNA centrales para la identificación de sus respectivas expansiones enzimáticas en genomas del filo *Actinobacteria*.
2. Determinar si las expansiones enzimáticas pertenecen a rutas biosintéticas de NPs mediante el análisis de su contexto genómico y, en su caso, ubicar patrones que puedan ser incorporados a la minería genómica convencional.
3. Caracterizar experimentalmente una de las predicciones derivadas del análisis tipo EvoMining enfocado al análisis de familias de tRUEs.

## Capítulo IV. Metodología

### 4.1. Generalidades y funcionamiento del programa EvoMining

EvoMining es una herramienta bioinformática que consiste en un conjunto de subprogramas (codificados en Perl) y software de terceros concatenados. El uso de otros programas implica frecuentes cambios y actualizaciones, lo cual podría provocar graves problemas en cuanto a su ejecución y reproducibilidad. Para evitar estos problemas, EvoMining se encuentra empaquetado en Docker (<https://www.docker.com>), un motor de código abierto que automatiza el despliegue de aplicaciones en contenedores aislados y autónomos. El contenedor de EvoMining almacena diversos programas (BLAST, antiSMASH, Muscle, G-Blocks, QuickTree, Newick utilities) y archivos necesarios (código, herramientas del sistema, bibliotecas de Perl, etc.) para su despliegue. Esto facilita su distribución y ejecución de manera portátil. Para ejecutar EvoMining, el usuario debe crear 4 bases de datos iniciales sobre las cuales se centrará el análisis (**Tabla 3**), cuyos detalles se abordan a continuación.

Tabla 3. Bases de datos iniciales necesarias para ejecutar EvoMining

Base de datos	ID	Contenido	Formato
<b>Enzimas centrales</b>	CE-DB	Secuencias aminoacídicas de enzimas pertenecientes al metabolismo central	Fasta de aminoácidos
<b>Genomas bacterianos</b>	G-DB	Secuencias completas de los genomas bacterianos por minar	GenBank
<b>NPs-BGCs</b>	BGC-DB	Secuencias de todos los BGCs conocidos, previamente caracterizados y recopilados en la base de datos MIBiG	GenBank
<b>antiSMASH</b>	antiSMASH-DB	BGCs detectados por el programa antiSMASH sobre los mismos genomas bacterianos que se usarán para el análisis	GenBank

## 4.2. Construcción de las bases de datos de EvoMining

### 4.2.1. Base de datos del metabolismo central (CE-DB)

La construcción de la base de datos de enzimas del metabolismo central (CE-DB) se llevó a cabo en 3 etapas. La primera etapa consistió en la búsqueda y recopilación de todas aquellas enzimas cuyo nombre o función descrita en su anotación sugiriera una asociación con moléculas de tRNA, es decir, enzimas que utilizan tRNAs (tRUEs). Para esto, se aplicaron dos estrategias: 1) búsqueda de palabras clave ('tRNA', 't-RNA', 'transfer RNA', 'transfer-RNA', 'transfer ribonucleic acid') en los archivos GenBank de 4 genomas actinobacterianos representativos (*Streptomyces coelicolor* A3 (2), *Saccharopolyspora erythraea* NRRL 2338, *Mycobacterium tuberculosis* H37Rv y *Corynebacterium glutamicum* R); y 2) búsqueda profunda de tRUEs en literatura, esto para evitar perder enzimas o genes cuyos nombres no reflejaran asociación con tRNAs. Una vez que se tenían los identificadores proteicos de estas enzimas, se utilizó un script *ad hoc* (desarrollado por miembros del grupo, disponible en [github/nselem](#)) para descargar las secuencias aminoacídicas correspondientes en formato fasta desde la base de datos NCBI.

La segunda etapa consistió en la clasificación por familias de las tRUEs recopiladas y la ruta metabólica central a la que pertenecen. Esta información es importante para dar seguimiento a las enzimas centrales dentro del proceso de EvoMining y para una interpretación limpia y ordenada de los resultados finales. Para esto, lo que hicimos fue modificar y limpiar (eliminar caracteres especiales) los identificadores de las enzimas recopiladas, siguiendo el siguiente formato:

>RutaCentral|NúmeroFamiliaEnzima|NombreEnzima\_NúmeroEnzima|Organismo

Ejemplos de encabezados:

>BiosíntesisPeptidoglicano|1|FemX\_1|Scoe

>BiosíntesisAA\_tRNA|1|Ala\_tRNAsintetasa\_1|Sery

>BiosíntesisAA\_tRNA|2|Glu\_tRNAsintetasa\_1|Mtub

>AA\_transferasas|1|L\_F\_transferasa\_1|Scoe

>AA\_transferasas|1|L\_F\_transferasa\_2|Mtub

Para nuestro análisis, la base de datos CE-DB está compuesta por 254 enzimas, las cuales se dividen en 78 familias, que a su vez pertenecen a 9 rutas del metabolismo central.

Dado que EvoMining tiene la finalidad de detectar expansiones enzimáticas de enzimas que pertenecen al metabolismo central que estén relacionadas con biosíntesis de NPs, es sumamente importante asegurarnos que nuestra CE-DB contiene únicamente enzimas del metabolismo central. De lo contrario, nuestros resultados no estarían contando la misma historia.

#### **4.2.2. Base de datos de enzimas de genomas (G-DB)**

Para la construcción de la base de datos de genomas (G-DB), se seleccionó un total de 1244 genomas bacterianos pertenecientes al filo *Actinobacteria*. Estos genomas fueron descargados de la base de datos NCBI (Junio 2016) procurando elegir solamente uno o dos organismos representativos de cada género o, en el caso de los géneros más extensos *e.g. Streptomyces*, uno o dos organismos de cada especie, optando por los de mejor ensamblado (mayor número de bases

pero menor número de contigs). Dado que cada uno de los genomas tiene su propio estilo de anotación, fue necesario re-anotar todos ellos usando la plataforma RAST (Aziz, y otros, The RAST Server: rapid annotations using subsystems technology, 2008) para estandarizar la nomenclatura de los genes. La descarga y re-anotación de estos genomas se realizó de manera automática usando scripts *ad hoc* (desarrollado por miembros del grupo, disponible en [github/nselem](https://github.com/nselem)).

#### **4.2.3 Base de datos de rutas biosintéticas de NPs (BGCs-DB)**

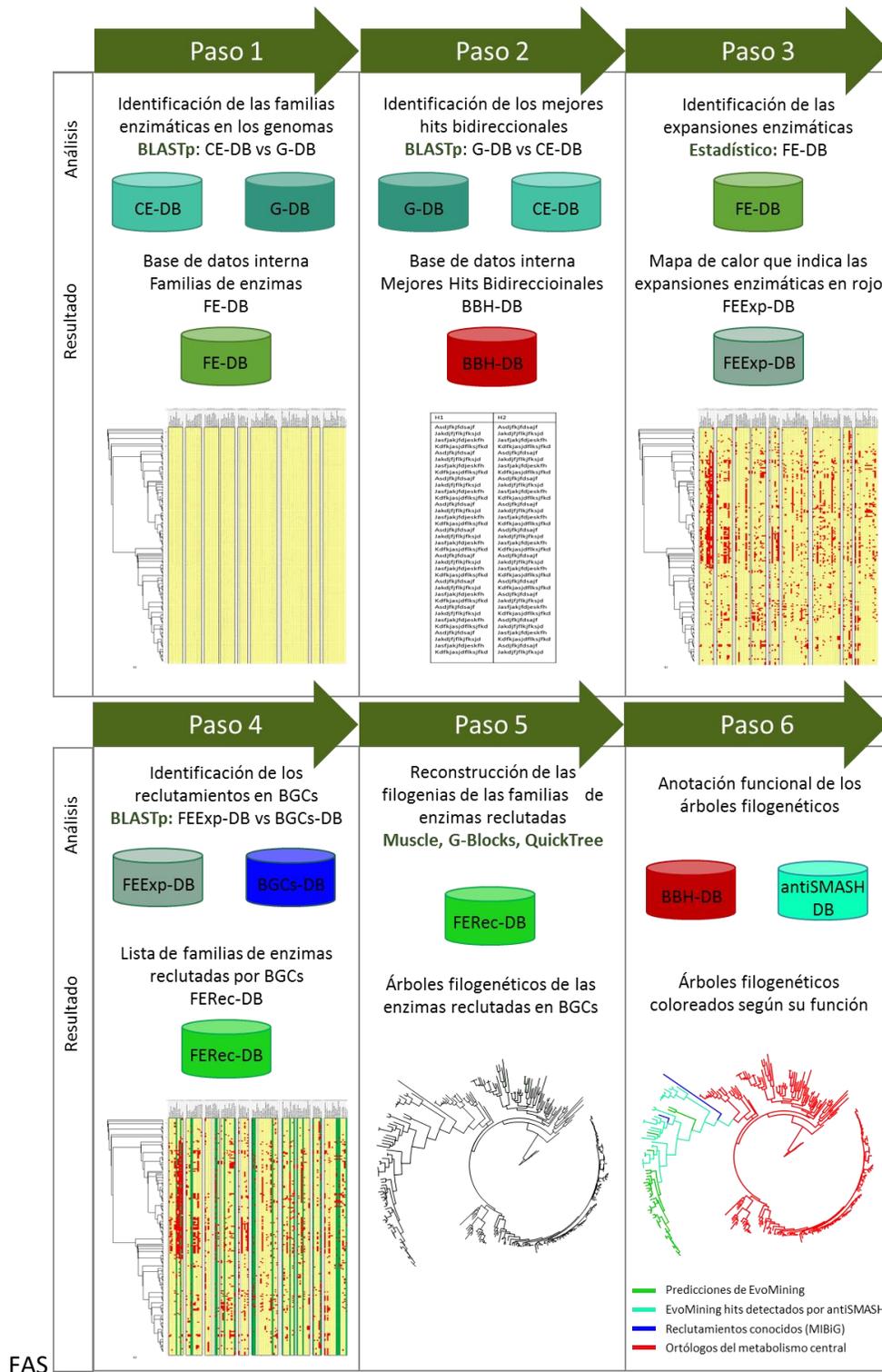
La construcción de la base de datos de clústeres de genes biosintéticos de NPs (BGCs-DB) para EvoMining, se realizó a partir de una compilación de las rutas biosintéticas de NPs en formato GenBank depositadas en MIBiG (<http://mibig.secondarymetabolites.org/>). MIBiG es un repositorio que contiene un total de 1294 BGCs de origen bacteriano, fúngico y vegetal, pertenecientes a las principales clases de NPs conocidos *i.e.* PKs, NRPs, terpenos, alcaloides, RIPPs, sacáridos, nucléosidos y clústeres híbridos (Agosto 2016). Este repositorio, fue creado por una comunidad global de científicos dedicados a la investigación en NPs, quienes se han caracterizado diversas clases de BGCs. De tal manera que MIBiG representa la base de datos de BGCs curada por expertos más completa en la actualidad, donde cualquier persona tiene acceso a dicha información (clase y estructura del metabolito, organismo productor, anotación del BGC, referencias bibliográficas) (Medema, y otros, 2015).

#### **4.2.4 Base de datos de G-DB anotados en antiSMASH (antiSMASH-DB)**

Para validar la originalidad de las predicciones de EvoMining, es necesario comparar dichos resultados con predicciones obtenidas mediante un enfoque distinto. Para esto, se realiza la anotación funcional de NPs-BGCs de cada uno de los genomas de nuestra base de datos G-DB usando el enfoque antiSMASH (Medema, y otros, 2011), la herramienta de minería genómica convencional de mayor renombre para la detección de BGCs conocidos en un genoma determinado. Los BGCs detectados con antiSMASH se descargan y se compilan para crear antiSMASH-DB. Esta información será utilizada por EvoMining para validar sus predicciones.

#### **4.3 Descripción de la tubería bioinformática de EvoMining**

Una vez que se construyeron las bases de datos iniciales (G-DB, CE-DB, BGCs-DB y antiSMASH-DB), se ejecuta un script preliminar que formatea y coloca dichas bases de datos en las carpetas correspondientes dentro de la tubería. De esta manera, EvoMining podrá reconocer los datos de entrada sobre los cuales hará el análisis completo. La tubería bioinformática de EvoMining puede dividirse en seis pasos principales: 1) identificación de las familias de enzimas, 2) identificación de los mejores hits bidireccionales (BBH), 3) identificación de las expansiones enzimáticas, 4) identificación de los reclutamientos en BGCs de NPs, 5) reconstrucción de los árboles filogenéticos y 6) anotación funcional de los árboles filogenéticos (**Fig. 18**).



**Figura 18. Tubería bioinformática de EvoMining.** El proceso bioinformático de EvoMining se puede dividir en seis pasos principales de acuerdo al tipo de análisis general que se lleva a cabo. CE-DB= Base de datos (DB) de enzimas centrales, G-DB= DB de genomas, BGCs-DB= DB de clústeres de genes biosintéticos de NPs, FE-DB= DB de familias enzimáticas, BBH-DB= DB de los mejores hits bidireccionales, FEExp-DB= DB de familias enzimáticas expandidas, FERec-DB= DB de Familias enzimáticas reclutadas, antiSMASH-DB= DB de los genomas anotados con antiSMASH.

### ***Paso 1***

El primer paso consiste en una comparación de las bases de datos CE-DB vs G-DB usando BLAST (Altschul, Gish, Miller, Mayers, & Lipman, 1990) (e-value < 0.0001 y bit score > 100), para identificar todas las familias de enzimas tRUEs homólogas dentro de los genomas de actinobacterias. Los resultados de este primer análisis dan lugar a una base de datos interna de familias enzimáticas que denominamos FE-DB. Usando la información de FE-DB, se crea una tabla FE (**Panel superior izquierdo, Fig. 18**) donde las filas representan los genomas, las columnas representan las familias de enzimas tRUEs y en cada celda se indica el número de tRUEs homólogas encontradas.

### ***Paso 2***

Una vez identificados los homólogos de las tRUEs en los genomas, es importante determinar cuáles de estas enzimas pertenecen al metabolismo central. En teoría, nuestra base de datos CE-DB fue debidamente curada y contiene sólo enzimas centrales, por lo cual puede usarse como referencia para un análisis BBH (en inglés, Bidirectional Best Hits), el segundo paso de EvoMining (**Panel superior central, Fig.18**). Este análisis consiste en hacer un segundo BLAST pero esta vez comparando G-DB vs CE-DB, *i.e.* el inverso del primer BLAST que realizamos. Posteriormente se comparan los resultados de ambos BLAST, para encontrar pares de enzimas que comparten la puntuación más alta (mejor score) y que por lo tanto, representan los mejores hits en ambas direcciones. Esta información se almacena en una segunda base de datos interna (BBH-DB) que servirá como referencia durante el proceso de anotación de EvoMining.

### *Paso 3*

Regresando a nuestra tabla FE, el siguiente paso consiste en la identificación de las expansiones enzimáticas. De manera heurística, se determina un valor numérico que represente los eventos de expansión significativos por cada familia de enzimas. Este valor se obtiene de sumar al promedio del número de homólogos de cada familia una desviación estándar, por lo tanto tendremos tantos valores numéricos como número de columnas de nuestra tabla. Todas aquellas celdas que sean mayores al valor numérico de expansión, se marcarán en rojo (**Panel superior derecho, Fig. 18**). Nuestra tabla se convierte pues en un mapa de calor con la finalidad de visualizar rápidamente las expansiones enzimáticas y nuestra FE-DB ahora contiene esta información (FEExp-DB).

### *Paso 4*

El cuarto paso consiste en identificar cuáles de las familias enzimáticas con expansiones están relacionadas con biosíntesis de NPs, para lo cual se realiza una comparación de las bases de datos FEExp-DB vs BGCs-DB mediante BLASTp (e-value < 0.0001 y bit score > 100). Encontrar pares de homólogos entre estas dos bases de datos, significa haber detectado expansiones de enzimas tRUEs centrales reclutadas en biosíntesis de NPs (**columnas verdes, panel inferior izquierdo, Fig. 18**), dado que tenemos la certeza de que las enzimas contenidas en BGCs-DB han sido caracterizadas experimentalmente. Internamente, se crea una base de datos de las familias de enzimas reclutadas (FERec-DB), a la cual también se añaden los genes homólogos encontrados en los BGCs de la BGCs-DB.

Adicionalmente, se crea una tabla interfaz que consta de tres columnas que indican: 1) nombre de la familia de enzimas, 2) genes homólogos encontrados en los BGCs y 3) un botón para la construcción del árbol filogenético. Este botón es controlado por el usuario, el cual puede elegir el orden y/o número de árboles por reconstruir.

### ***Paso 5***

El quinto paso es la reconstrucción de los árboles filogenéticos que sean seleccionados por el usuario a través del botón (**Panel inferior central, Fig. 18**). Este proceso implica 1) el alineamiento con Muscle (versión 3.8.31) de todas las secuencias de la familia de enzimas más los genes homólogos detectados en BGCs-DB, 2) el rasurado de los alineamientos con GBlocks y 3) la reconstrucción de la filogenia con QuickTree (método Neighbour-Joining).

### ***Paso 6***

A grandes rasgos, el último paso de EvoMining se refiere a una anotación funcional del árbol filogenético (**Panel inferior derecho, Fig. 18**). Cada una de las ramas de un árbol filogenético representa una enzima tRUE o, en su caso, el homólogo detectado en un BGCs. Para la anotación funcional, se compara cada una de estas ramas contra las bases de datos BBH-DB y antiSMASH-DB. Cuando la rama se encuentra dentro de la BBH-DB, quiere decir que ésta enzima pertenece al metabolismo central dado que es uno de los mejores hits bidireccionales de la CE-DB y, por tanto se colorea de rojo (usando el programa de edición de filogenias Newick Utilities). En caso de que la rama se encuentre dentro de la antiSMASH-DB, indica que esta enzima también ha sido detectada

por una de las herramientas más usadas de minería genómica de NPs (control positivo de EvoMining), y esta rama se pintará de color cian. Los genes homólogos de BGCs se señalan en color azul oscuro. El resto de las ramas se pintarán de color verde y representan las predicciones propias de EvoMining, es decir, los nuevos genes detectados a formar parte de BGCs nunca antes vistos ni detectados por otros métodos (*e.g.* antiSMASH y Cluster Finder).

EvoMining despliega estos árboles filogenéticos en una interfaz gráfica que permite explorar y visualizar detalladamente su topología. Cada rama del árbol está etiquetada con el identificador del gen y el organismo al que pertenece la enzima. La mayoría de las topologías de los árboles mostrarán clados divergentes muy marcados. Estas divergencias representan los eventos de expansión de las familias de enzimas, ya que cada árbol simboliza una historia evolutiva. Es de esperarse que estos clados divergentes contengan las enzimas homólogas más cercanas a los BGCs de NPs, por lo que generalmente encontraremos dentro de estos clados las predicciones propias de EvoMining (color verde), los controles positivos detectados por antiSMASH (color cian) y los genes extraídos de la base de datos BGCs-DB (color azul oscuro). Los clados restantes que no muestran divergencia, representan las enzimas no expandidas que pertenecen al metabolismo central (color rojo).

Esta interfaz, también contiene un complemento que permite visualizar el contexto genómico de las enzimas que el usuario desee analizar, el cual se explica a continuación.

## **4.4 Análisis de resultados de EvoMining**

### **4.4.1 Análisis del contexto genómico de las expansiones enzimáticas**

Una vez que obtenemos un árbol filogenético que representa la historia evolutiva de una familia tRUE dentro del contexto de biosíntesis de NPs, es momento de analizar su filogenia. Para esto, la interfaz final de EvoMining que despliega la topología del árbol resultante cuenta con un complemento para visualizar el contexto genómico de cualquier enzima que el usuario elija. El procedimiento consiste en dar clic sobre las enzimas (ramas del árbol) que deseamos analizar, preferentemente del clado divergente. El sub-programa que rige este complemento enlista estas enzimas seleccionadas y las busca dentro de la base de datos de genomas (G-DB) de EvoMining. Una vez que las localiza, extrae el contexto genómico alrededor de éstas, es decir, 50 kpb de secuencia río arriba y 50 kpb río abajo. Posteriormente, convierte esta información a una gráfica que representa el contexto genómico de cada enzima seleccionada (organización de los genes, nombre, dirección y tamaño correspondiente). A partir de estos gráficos, podemos hacer las primeras indagaciones de posibles nuevas rutas biosintéticas examinando la anotación funcional de los genes del contexto genómico.

### **4.4.2 Predicción de nuevas rutas biosintéticas de productos naturales**

Actualmente no existe un método estandarizado o automático para determinar si las predicciones de EvoMining (enzimas en color verde de los clados divergentes de los árboles filogenéticos) son realmente parte de una nueva ruta biosintética. Hoy en día, este proceso se realiza de manera manual, donde el usuario distingue

aquellas enzimas que pueden ser parte de una ruta de biosíntesis de NPs y las que no. Por ejemplo, se puede reconocer un BGC si un contexto genómico contiene genes que codifican para enzimas deshidrogenasas, oxidasas, ciclasas, sintasas, sintetisas, transferasas. Esta curación 'manual' es la que se llevó a cabo en el presente trabajo para proponer nuevas rutas biosintéticas predichas por EvoMining.

El siguiente paso, después de obtener una gran diversidad de predicciones de BGCs con EvoMining, es elegir candidatos para su caracterización experimental. De acuerdo a la anotación de sus genes y al organismo al que pertenece, se diseña una estrategia para expresar dicho BGC y obtener el metabolito que produce. A continuación se describe la metodología que seguimos para la caracterización de uno de los BGCs predichos por EvoMining.

#### **4.5 Caracterización experimental del azoxy-BGC**

##### **4.5.1 Estrategia experimental general**

La estrategia experimental general que planteamos para expresar el azoxy-BGC predicho por EvoMining consistió en la construcción de mutantes de genes clave del BGC con la finalidad de realizar un escaneo fenotípico de comparación entre cepas silvestres y mutantes creciendo en diferentes medios de cultivo, esperando encontrar cambios en su morfología que estuvieran asociados con la expresión del clúster. Una vez que seleccionamos las condiciones de crecimiento que podrían estar induciendo la expresión del azoxy-BGC, se hicieron análisis de RT-PCR sobre genes puntuales del clúster para verificar su expresión y por otro lado, se

realizó un análisis de cromatografía líquida de alta eficiencia (HPLC) para obtener el extracto del metabolito del azoxy-BGC y su posterior identificación.

#### 4.5.2 Material biológico

Las cepas modelo empleadas para la construcción de mutantes y la realización del resto de los experimentos fueron *Streptomyces lividans* 66 y *Streptomyces coelicolor* M1154, provistas por el John Innes Center. En algunos experimentos también incluimos a la bacteria *Streptomyces rochei* 36, una cepa perteneciente a la colección de aislados ambientales de nuestro laboratorio, proveniente de la poza Churince (Cuatro Ciénegas, Coahuila), la cual es altamente sinténica a *S. coelicolor* y contiene el azoxy-BGC.

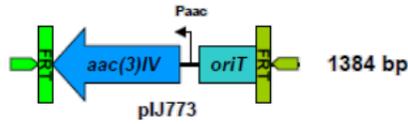
Tabla 4. Oligonucleótidos

Clave	Longitud	Secuencia	Nota
SCO1288_P1	59	GCG CCC CGG AAT CGT ACG CGC AAT CCA GGA GCG CCC GTG - ATT CCG GGG ATC CGT CGA CC	
SCO1280_P2	58	CGG ACG GTG CGT CCG GGG CCT GCG CGT GGC GCC GTG TCA - TGT AGG CTG GAG CTG CTT C	Oligonucleótidos para el casete de resistencia con el sitio <i>oriT</i>
SCO1287_P1	59	AAA GCA CCG CCG CGC GGA CTT CCG TGG AGG ACG CCG ATG - ATT CCG GGG ATC CGT CGA CC	
SCO1287_P2	58	ATC GGC GCG GCA CTC GGA CGG CCG TGG ACG CGG TGC TCA - TGT AGG CTG GAG CTG CTT C	
SCO1288_Scr_F	21	TCA ATC ATC CCC ACT TTA CGA	Oligonucleótidos para comprobación de las mutantes
SCO1280_Scr_R	21	GTA TCT GCC GCT GAT CGG CAC	
SCO1287_Scr_F	21	AGC CCT TCG GTG TCG CCT TCG	
SCO1287_Scr_R	21	TCG GGC AGC AGC TTC GCG AGG	
SCO1286_F	22	ATG AGA TTC CTA GAG CGT GAA C	Oligonucleótidos para análisis de expresión por RT-PCR
SCO1286_R	19	GCC GTA GCG CAC GTA CAA C	
SCO1287_F	20	GTC CGT ACT TGA CAC ACT CG	
SCO1287_R	19	TGA ACA TAC GGC TAG CAC C	
SCO1289_F	19	TGA CGG AGG CAT ACG AGC A	
SCO1289_R	19	GTG CCC AGT CGA GCA GTT C	
SCO1278_F	19	CTC GGT CCC TTC TAC GTG A	
SCO1278_R	20	CCA GCG CGA AAA GGA TCT TC	

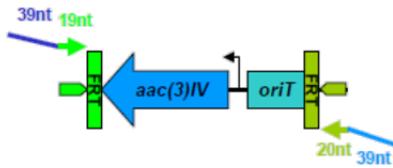
### 4.5.3 Generación de mutantes del azoxy-BGC

Las mutantes del azoxy-BGC se generaron mediante el sistema  $\lambda$ -Red (Gust, y otros, 2004). Este sistema de transformación es apto para la delección sencilla de genes o de BGCs completos en *Streptomyces* (Gust, Challis, Fowler, Kieser, & Chater, 2003). El sistema  $\lambda$ -Red, también conocido como REDIRECT-Technology, se divide en 6 pasos principales (**Fig. 19**). Los requerimientos más importantes para este método es el diseño de oligos especiales (**Tabla 4**), el uso de cepas de *E. coli* con características puntuales para la transformación y un cósmido de *S. coelicolor* que contenga la región que se desea mutar. En nuestro caso, se realizó una búsqueda del cósmido de *S. coelicolor* que contuviera el azoxy-BGC en el servidor StrepDB (<http://strepdb.streptomyces.org.uk/>). Encontramos que el cósmido 1E11 (~35,500 pb) de *S. coelicolor* contenía el total de los genes del azoxy-BGC. Debido a que los genomas de *S. coelicolor* y *S. lividans* son altamente sinténicos y ambos conservan el azoxy-BGC, el cósmido 1E11 y los oligos diseñados eran aptos para transformar a ambos organismos. Decidimos hacer dos tipos de mutantes en cada organismo: una delección sencilla sobre el gen hit de EvoMining (SCO1287 y SLI1569, respectivamente) y una delección de 9 genes del clúster (SCO1280-SCO1288 y SLI1560-SLI1570, respectivamente). En el caso de *S. lividans*, utilizamos la cepa *S. lividans* 66 (Sli66). Para el caso de *S. coelicolor*, elegimos trabajar con la cepa M1154 (ScoM1154), diseñada especialmente para una caracterización más limpia de BGCs.

1 Purificación del templado para clonar el casete de resistencia con *oriT* por PCR



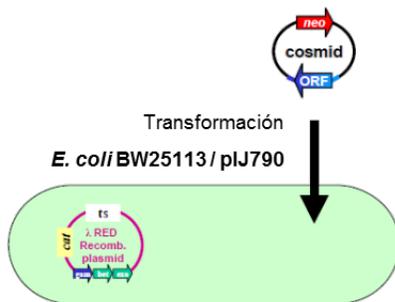
2 Diseño de oligos para la amplificación del casete con 39 nt adicionales que sean homólogos a la region adyacente por inactivar en *Streptomyces*



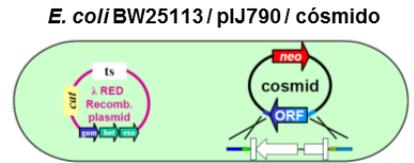
3 Amplificación del casete de interrupción por PCR



4 Introducción del cósmido 1E11 de *S. coelicolor* a *E. coli* BW25113/pIJ790 que contiene el sistema  $\lambda$ -Red por electroporación

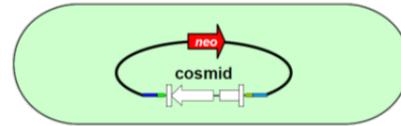


5 PCR-targeting del cósmido de *S. coelicolor*



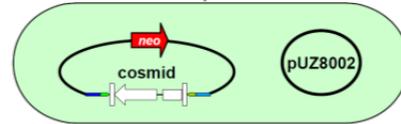
30°C  
37°C  
Pérdida del plásmido de recombinación  $\lambda$ -Red sensible a temperatura (37°C ON)

*E. coli* BW25113/cósmido

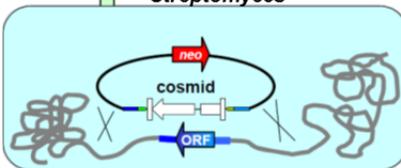
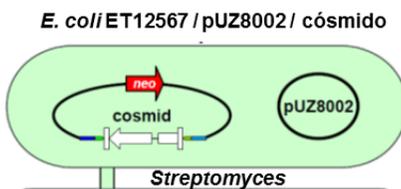


Transformación de la cepa ET12567/pUZ8002 para recombinación

*E. coli* ET12567/pUZ8002/cósmido



6 Transferencia por conjugación y detección de doble entrecruzamiento de las mutantes



Mutante de interrupción de *Streptomyces*

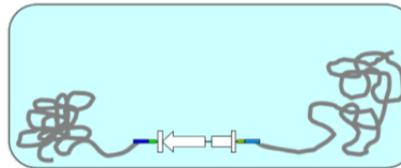
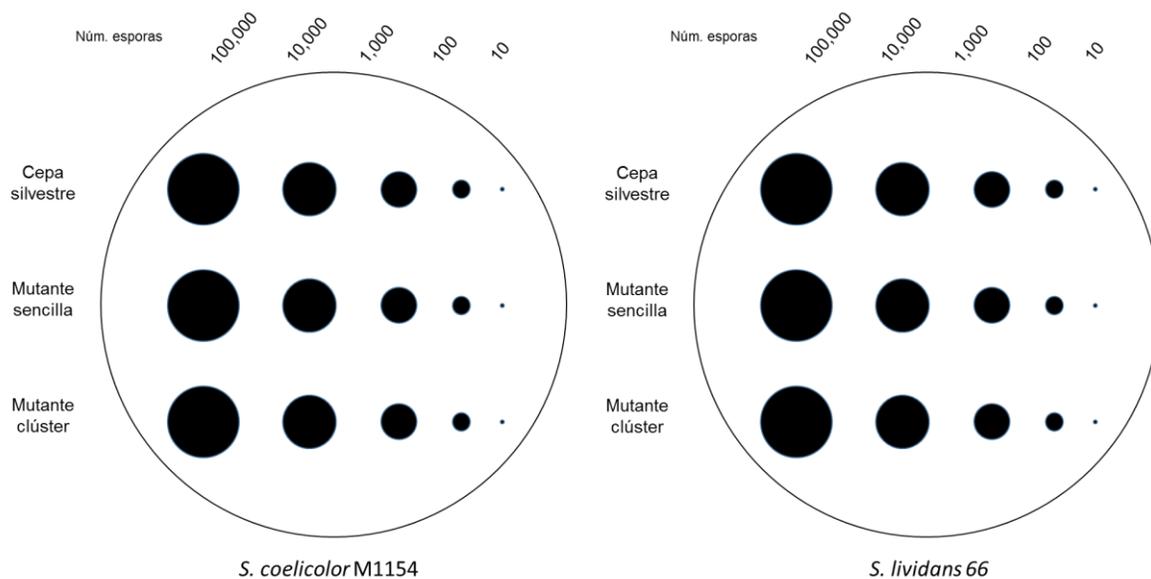


Figura 19. Sistema de transformación  $\lambda$ -Red (REDIRECT-Technology). Modificado de (Gust, Kieser, & Chater, 2002)

*S. coelicolor* M1154 es una mutante derivada de la cepa M145 que contiene deleciones en los clústeres biosintéticos de la figura para reducir el potencial competitivo del suministro de carbono y nitrógeno, así como mutaciones puntuales en los genes *rpoB* ([C128T]) y *rpsL* ([A262G]) para incrementar pleiotrópicamente la producción de NPs (Gomez-Escribano & Bibb, 2011). Para todas las mutantes, se utilizaron los marcadores de selección de kanamicina (para la selección de transformantes en *E. coli*) y apramicina (*aac(3)IV*) (para seleccionar las transformantes en *Streptomyces*).

#### **4.5.4 Comparación morfológica entre cepas silvestres y mutantes**

Una vez que se obtuvieron y comprobaron las mutantes por PCR, se realizó un análisis del fenotipo que adquirirían cepas silvestres en comparación con cepas mutantes bajo distintas condiciones de crecimiento en el laboratorio. La finalidad de realizar estas comparaciones, fue detectar cambios en la morfología de las cepas que pudieran estar asociadas a la expresión/delección del azoxy-BGC. Para esto, se probaron diferentes concentraciones de esporas en placas de agar (**Fig. 20**) con diferentes composiciones del medio. Los medios probados fueron 5 medios de cultivo sólidos ricos en nutrientes (YEME, R2YE, TSB, 2XYT y SFM) y 2 medios de cultivo sólidos de composición mínima de nutrientes con la variante de la presencia/ausencia de fósforo (SMMS +P, SMMS -P, SMM +P y SMM -P). Estas placas crecieron a una temperatura de 30°C por un tiempo de 4-5 días. Durante este periodo, se monitoreo el crecimiento de las cepas silvestres y mutantes para tratar de detectar diferencias morfológicas entre las cepas.



**Figura 20. Estrategia de siembra en placa para el análisis de fenotipo.** La razón por la que elegimos esta estrategia fue para evaluar cambios en el fenotipo a nivel micro (colonias únicas) y semi-macro de las células (conjunto de colonias).

#### 4.5.5 Extracción de RNA

Después de identificar las condiciones de crecimiento en las que podíamos notar cambios en la morfología de mutantes y silvestres, debíamos verificar si realmente los genes del azoxy-BGC se estaban expresando. Para esto planteamos hacer un análisis de RT-PCR, para lo cual primero es necesario extraer RNA total de las cepas silvestres. Debido a la morfología tan peculiar de los *Streptomyces*, se probaron distintos métodos de extracción de RNA. El mejor protocolo para la extracción de RNA de nuestras *Streptomyces* es una combinación de los diferentes probados, el cual se describe a continuación:

### ***Preparación del material***

- Preparar una solución de agua-DEPC al 0.1%. Agregar 1mL de DEPC a 999mL de H<sub>2</sub>OmilliQ y agitar por 30 min con agitador magnético.  
ALTAMENTE TÓXICO, INDISPENSABLE USAR GUANTES.
- Sumergir todo el material que estará en contacto con la muestra en dicha solución (tubos Falcon, morteros y pistilos, tubos eppendorf, espátulas, puntas) y dejar incubando a temperatura ambiente una noche completa.
- Escurrir el material y colocarlo en recipientes aptos para esterilización. Los frascos de vidrio deberán hornearse previamente a 250°C, por lo menos 4 horas. Las cajas para puntas deberán estar limpias.
- Para eliminar las trazas de DEPC, esterilizar el material a 121°C por 15 min.
- Mantener el material en un área exclusiva para RNA y siempre usar guantes.

### ***Preparación de soluciones***

- Preparar una solución de agua-DEPC al 0.1%.
- Dejar que el tratamiento actúe por una noche completa
- Para eliminar las trazas de DEPC, esterilizar el material a 121°C por 15 min.
- Usar esta agua tratada con DEPC para preparar las soluciones que se usarán en la extracción (amortiguador de lisis, etanol 75%, agua de elución).

### ***Precauciones generales***

- Siempre usar guantes cuando se trabaje con RNA, ya que la piel está llena de RNAsas y otras bacterias que podrían contaminar la muestra. Además

de que la mayoría de los reactivos usados en este protocolo son altamente tóxicos.

- Mantener el área de trabajo libre de RNAsas. Usar soluciones descontaminantes (e.g. RNA-Zap) para limpiar la mesa, las pipetas, las gradillas y, sobretodo, los guates. Evitar tocar equipo, manijas de puertas o cualquier otra superficie con los mismos guantes que se maneja la muestra. Recomendable trabajar en un lugar separado a donde se trabaja con DNA.
- El procedimiento de extracción debe realizarse lo más rápido posible para evitar la degradación del RNA, por lo que es deseable tener todo el material y soluciones listas para usarse.
- Usar puntas con filtro para evitar la contaminación cruzada.

### ***Procedimiento de extracción de RNA***

A partir de un cultivo celular de *Streptomyces* (30°C, 200rpm, 5-6 días):

1. Recolectar la biomasa
  - a. Vaciar 50 ml del cultivo celular a un tubo Falcon.
  - b. Centrifugar a 10,000rpm por 2 min a 4°C.
  - c. Recuperar el sobrenadante en otro tubo Falcon para su posterior análisis por HPLC. Guardar a -80°C.
2. Macerar el micelio

Importante: Para evitar en gran medida la degradación del RNA, es necesario mantener la biomasa congelada en todo momento y macerar lo más rápido posible.

- a. Transferir el pellet celular a un mortero previamente congelado a -80°C.

- b. Agregar N<sub>2</sub> líquido para congelar las células.
- c. Macerar hasta reducir el tamaño del pellet. NO DEJAR QUE EL PELLETT LLEGUE A TEMPERATURA AMBIENTE. Agregar más N<sub>2</sub> cuando sea necesario.
- d. Congelar una espátula y un tubo eppendorf con N<sub>2</sub>.
- e. Usando esta espátula, transferir la biomasa macerada al tubo congelado. La cantidad de biomasa no debe sobrepasar el 10% del volumen del tubo.

Nota: Si queda biomasa en el mortero, puede guardarse como reserva. Para ello, transferir la biomasa a un tubo congelado y resuspender con 1mL de TRIzol. Almacenar a -80°C.

### 3. Lisis celular

- a. Agregar 200uL de amortiguador de lisis (10mM Tris-HCl, 1mM EDTA, pH8 + 2mg/ml de lisozima) y mezclar por inversión.
- b. Incubar esta reacción de lisis a 37°C por 15min.
- c. Agregar 1mL de TRIzol y mezclar por inversión.
- d. Incubar 5 min a temperatura ambiente.

Nota: Si después de la reacción de lisis quedan restos celulares insolubles en la solución, centrifugar a 12,000 x g por 10min a 4°C y transferir solo la fase acuosa a un tubo nuevo.

### 4. Separar las fases

- a. Añadir 200uL de cloroformo, cerrar los tubos perfectamente y mezclar vigorosamente por 15 segundos. No vórtex.

- b. Incubar a temperatura ambiente por 2-3min.
  - c. Centrifugar a 12,000 x g por 15 min a 4°C.
  - d. Transferir la fase acuosa (parte superior, transparente) a un tubo nuevo con el extremo cuidado de no tocar la interfase, mucho menos la fase fenólica (parte inferior, color rosa).
5. Precipitar el RNA
- a. Añadir 500uL de isopropanol y mezclar por inversión.
  - b. Incubar a temperatura ambiente por 10min.
  - c. Centrifugar a 12,000 x g por 10min a 4°C. Se formará un pellet transparente con aspecto de gel.
  - d. Remover el sobrenadante cuidadosamente.
6. Lavar el RNA
- a. Añadir 1mL de etanol al 75% y mezclar por inversión.
  - b. Centrifugar a no más de 7500 x g por 5min a 4°C.
  - c. Decantar el sobrenadante y dejar secar al aire por 10min. No dejar que seque por completo ya que la solubilidad disminuiría en gran medida. No secar por centrifugación a vacío, nunca.
7. Disolver el RNA
- a. Agregar 50uL de agua tratada con DEPC y mezclar por pipeteo lentamente.
  - b. Dividir la muestra de RNA en alícuotas (10uL, 20uL, 20uL).
  - c. Almacenar inmediatamente a -80°C.
8. Verificar la calidad y cantidad del RNA

Usando la alícuota de 10uL:

- a. Cuantificar el RNA en el NanoDrop. Idealmente  $A_{260}/A_{280} \geq 1.8$ .
- b. Correr un gel de electroforesis ajustando la cantidad a 1ug de RNA.

En la fotografía deberán verse dos bandas bien definidas que representan el RNA ribosomal así como una banda más pequeña que indique la presencia de sRNA. La presencia de una banda de alto peso molecular (más de 10,000 kb) indica presencia de DNA, el cual debe eliminarse con DNase I antes de proceder con el ensayo de RT-PCR.

#### **4.5.6 RT-PCR**

Antes de realizar la RT-PCR, se llevó a cabo un tratamiento con DNase I de todas las muestras de RNA obtenidas, siguiendo el protocolo de INVITROGEN. Para evitar la degradación de la muestra, no usamos EDTA ni calentamos la muestra para detener la reacción de la DNase I como lo indicaba el protocolo. Lo que hicimos fue inactivar la reacción en hielo únicamente. Inmediatamente después se procedió a realizar la RT-PCR usando el kit OneStep RT-PCR de QIAGEN, el cual permite la generación de cDNA y la reacción de PCR en un mismo paso (un solo tubo, una sola reacción). El procedimiento consiste en preparar la reacción de RT-PCR mezclando agua destilada libre de RNasas, un buffer de reacción, una mezcla de dNTPs, un aditivo que permite la amplificación eficiente de templados ricos en 'GC' (Q-solution) y un coctel de enzimas reversa-transcriptasas (Sensiscript y Omniscript) y DNA polimerasa (HotStarTaq). A esta mezcla se le añade nuestra muestra de RNA de interés y los oligonucleótidos correspondientes al gen que queremos evaluar. El programa del termociclador es crucial para la funcionalidad de este protocolo:

Transcripción reversa (Obtención del cDNA)	50°C , 30 min				
Activación de la DNA polimerasa y separación de la cadena	95°C , 15 min				
Amplificación (30 ciclos)	<table> <tr> <td rowspan="3">}</td> <td>95°C , 30 seg</td> </tr> <tr> <td>60°C , 30 seg</td> </tr> <tr> <td>72°C , 90 seg</td> </tr> </table>	}	95°C , 30 seg	60°C , 30 seg	72°C , 90 seg
}	95°C , 30 seg				
	60°C , 30 seg				
	72°C , 90 seg				
Extensión final	72°C , 10 min				
Mantenimiento	12°C				

Es necesario incluir reacciones control en cada corrida de RT-PCR para asegurarnos de que los resultados son verídicos. Como control negativo, usamos una reacción de RT-PCR sin templado, este control nos indica que los reactivos empleados están libres de contaminación por RNA o DNA. Como control positivo, usamos otra reacción de RT-PCR con templado pero usando oligonucleótidos del algún gen constitutivo, esto nos indica que los reactivos del kit fueron viables. Además, el control posiblemente más importante es el control para verificar que no hay presencia de DNA genómico en nuestras muestras de RNA. Para este control, se prepara una reacción de RT-PCR con todos los componentes, con la diferencia de que esta reacción se mantiene en hielo durante el periodo de retrotranscripción y se coloca en el termociclador sólo hasta el periodo de la PCR.

#### **4.5.7 Cromatografía líquida de alta eficiencia**

Los análisis de cromatografía líquida de alta eficiencia (HPLC) se realizaron con un equipo Agilent con arreglo de diodos. Como muestras se usaron extractos crudos de los mismos cultivos celulares de los que se obtuvo la biomasa para extracción de RNA. De acuerdo con una predicción preliminar del compuesto derivado del azoxy-BGC, decidimos usar columnas C-18 (Zorbax), como fase móvil gradientes de acetonitrilo:agua (90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, 10:90) y leyendo longitudes de onda de 210nm, 230nm y 280nm.

## Capítulo V. Resultados y discusión

EvoMining es una herramienta de minería genómica no convencional en muchos sentidos. En primer lugar, es la única herramienta capaz de identificar nuevas clases de BGCs mediante análisis filogenéticos de familias de enzimas del metabolismo central. Su estrategia consiste en identificar un número aberrante de parálogos de enzimas centrales, ubicados en las ramas más divergentes de su filogenia, que residen dentro de BGCs. Por otro lado, las bases de datos iniciales de EvoMining (CE-DB, G-DB, BGC-DB y antiSMASH-DB) pueden ser adaptadas para ejecutar análisis independientes de acuerdo a los intereses del usuario. Dado que existe una gran variedad de enzimas del metabolismo central y un sinnúmero de genomas, las posibilidades son muy amplias.

En el presente proyecto, adaptamos EvoMining al estudio de un grupo de enzimas que utilizan tRNA (tRUEs) en un conjunto de genomas representativos del filo *Actinobacteria*. A continuación prestamos los resultados del enfoque de este análisis.

### 5.1 Adaptación de EvoMining al análisis de las tRUEs en *Actinobacteria*

La adaptación de EvoMining al estudio de las enzimas que utilizan tRNA (tRUEs) consistió en la construcción de las bases de datos iniciales (CE-DB, G-DB, BGC-DB y antiSMASH-DB) con las secuencias correspondientes (**Tabla 5**). En las siguientes secciones, se explica detalladamente la obtención de los datos.

**Tabla 5. Bases de datos iniciales del análisis EvoMining enfocado a tRUEs en Actinobacteria**

<b>Base de datos EvoMining</b>	<b>Naturaleza de las secuencias</b>	<b>Número de secuencias</b>	<b>Formato de las secuencias</b>
<b>CE-DB</b>	tRUEs del metabolismo central de bacterias	254	Fasta de aminoácidos
<b>G-DB</b>	Genomas completos de bacterias del filo <i>Actinobacteria</i>	1244	GenBank
<b>antiSMASH-DB</b>	Anotación funcional de G-DB con el programa de minería genómica antiSMASH*	1244	Predicciones de BGCs de NPs GenBank
<b>BGCs-DB</b>	BGCs de NPs depositados en MIBiG	1294*	BGCs de NPs GenBank

\* = <https://antismash.secondarymetabolites.org/>

ψ = <http://mibig.secondarymetabolites.org/repository.html>

### 5.1.1. Construcción de la base de datos CE-DB centrada en tRUEs

La CE-DB es posiblemente la base de datos más importante de un análisis tipo EvoMining. Esta base de datos debe contener las enzimas del metabolismo central de las cuales se identificarán sus expansiones enzimáticas que estén ubicadas en un contexto genómico asociado a biosíntesis de NPs. Por lo tanto, cuidar la calidad de la información que depositemos en ella es crucial. En nuestro caso, la construcción de CE-DB consistió en estrategias diferentes de búsqueda y en la clasificación por familias de las enzimas tRUEs recopiladas.

La primera estrategia de búsqueda de tRUEs que probamos consistió en la revisión de diversas bases de datos de proteínas disponibles en línea como NCBI-protein, UniProt, SwissProt, BioCyc, KEGG, PANTHER y GeneOntology. Sin embargo, observamos que estas plataformas contenían información incompleta, sesgada a ciertos organismos modelo (que no incluían actinobacterias) y/o las secuencias no estaban clasificadas de acuerdo a nuestros intereses. Por lo tanto,

seguimos una segunda estrategia basada en la identificación de palabras clave ('tRNA', 't-RNA', 'transfer RNA', 'transfer-RNA', 'transfer ribonucleic acid') en 4 de los genomas mejor anotados y más representativos de las actinobacterias: *Streptomyces coelicolor* A3(2) (Scoe), *Saccharopolyspora erythraea* NRRL 2338 (Sery), *Mycobacterium tuberculosis* H37Rv (Mtub) y *Corynebacterium glutamicum* R (Cglu). Por medio de esta estrategia, logramos identificar un total de 213 tRUEs (Scoe = 71, Sery = 57, Mtub = 42 y Cglu = 43). Adicionalmente, complementamos esta base de datos con enzimas tRUEs reportadas en la literatura cuya anotación no coincidía con nuestras palabras clave utilizadas en la búsqueda 'manual' (e.g. MprF, L/F transferasa, FemXAB) pero que sí fungían como tRUEs. Mediante esta estrategia de revisión bibliográfica pasamos de 213 a un total de 254 tRUEs. Cabe mencionar que algunas de estas enzimas adicionales pertenecían a otros géneros de bacterias, algunas incluso fuera del filo *Actinobacteria*. La intención de incluir enzimas tRUEs de otros géneros fue llevar a cabo un análisis EvoMining más amplio y relativamente más completo de estas enzimas. La profunda revisión bibliográfica también fue útil para verificar que las tRUEs identificadas realmente pertenecían al metabolismo central de bacterias. Posterior a esta curación manual de los datos, procedimos a descargar las 254 secuencias.

El siguiente paso para construir nuestra CE-DB, fue la clasificación por familias de las 254 tRUEs recopiladas. Para esto, utilizamos la herramienta BLASTClust (<https://www.ncbi.nlm.nih.gov/Web/Newsltr/Spring04/blastlab.html>), la cual forma clústeres de secuencias proteínicas o nucleotídicas que cumplen con ciertos rangos de similitud asignados por el usuario. De esta manera, todas las tRUEs de

una misma familia, e.g. las Ala-tRNA sintetetasas, quedan agrupadas en un mismo clúster. Con esta metodología, nuestra CE-DB adaptada a tRUEs quedó clasificada en 78 familias, las cuales a su vez pertenecen a 10 rutas del metabolismo central de bacterias (**Tabla 6**).

**Tabla 6. Clasificación por familias de las tRUEs y ruta metabólica central al que pertenecen**

Familias enzimáticas	Ruta metabólica central	Número de secuencias por familia
<b>Sub-Familia: aminoacil-tRNA sintetetasas clase I</b>		
Arginil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	5
Cisteinil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	6
Glutaminil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	2
Glutamil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	5
Isoleucil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	4
Leucil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	4
Lisil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	6
Metionil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	4
Triptofanil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	6
Tirosil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	4
Valil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	4
Glutamil-tRNA sintetasa no-discriminante	Biosíntesis de aminoacil-tRNAs	2
<b>Sub-Familia: aminoacil-tRNA sintetetasas clase II</b>		
Alanil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	5
Asparaginil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	2
Aspartil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	4
Glicil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	4
Histidil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	4
Lisil-tRNA sintetasa integral de membrana	Biosíntesis de aminoacil-tRNAs	2
Aspartil-tRNA sintetasa no-discriminante	Biosíntesis de aminoacil-tRNAs	2
Fenilalanil-tRNA sintetasa subunidad A	Biosíntesis de aminoacil-tRNAs	4
Fenilalanil-tRNA sintetasa subunidad B	Biosíntesis de aminoacil-tRNAs	4
Protil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	6
Seril-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	4
Treonil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	6
Pirrolisil-tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	2
<b>Sub-Familia: Methionyl tRNA formyltransferase-like</b>		
Metionil-tRNA formiltransferasa	Biosíntesis de aminoacil-tRNAs	6
<b>Sub-Familia: GatB/GatE</b>		
Aspartil/glutamil-tRNA amidotransferasa subunidad A	Biosíntesis de aminoacil-tRNAs	6
Aspartil/glutamil-tRNA amidotransferasa subunidad B	Biosíntesis de aminoacil-tRNAs	4
Aspartil/glutamil-tRNA amidotransferasa subunidad C	Biosíntesis de aminoacil-tRNAs	4

Glutamil-tRNA amidotransferasa	Biosíntesis de aminoacil-tRNAs	2
--------------------------------	--------------------------------	---

**Sub-Familia: *SepSec C-like***

Selenocisteinil-tRNA sintetasa (Sela)	Biosíntesis de aminoacil-tRNAs	2
Sep-tRNACys -tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	2
Sep-tRNA <sup>Sec</sup> -tRNA sintetasa	Biosíntesis de aminoacil-tRNAs	3
O-fosfoseril-tRNACys sintetasa	Biosíntesis de aminoacil-tRNAs	2

**Sub-Familia: *Ser-tRNA<sup>Sec</sup> seleno transferasa***

L-seril-tRNA <sup>Sec</sup> seleno transferasa	Biosíntesis de aminoacil-tRNAs	1
Aminotransferasa dependiente de piridoxal-fosfato	Biosíntesis de aminoacil-tRNAs	1

**Sub-Familia: *Sec cinasa***

O-fosfoseril-tRNA <sup>Sec</sup> cinasa	Biosíntesis de aminoacil-tRNAs	2
---	--------------------------------	---

**Sub-Familia: *FemXAB***

FemA	Síntesis de pared celular	3
FemAB	Síntesis de pared celular	2
FemB	Síntesis de pared celular	2
FemX	Síntesis de pared celular	5
FemY	Síntesis de pared celular	1
MurM	Síntesis de pared celular	2
MurN	Síntesis de pared celular	2
Epr	Síntesis de pared celular	1
Lif (factor de inmunidad a lisostafina)	Síntesis de pared celular	1
VanK	Síntesis de pared celular	1

**Sub-Familia: *MprF***

Alanil-diacilglicerol sintetasa	Aminoacilación de lípidos	1
Alanil-fosfatidilglicerol sintetasa	Aminoacilación de lípidos	6
Lisil-fosfatidilglicerol sintetasa	Aminoacilación de lípidos	11
RAK-fosfatidilglicerol sintetasa	Aminoacilación de lípidos	2
MprF1	Aminoacilación de lípidos	1
MprF2	Aminoacilación de lípidos	1

**Sub-Familia: *LFTR-like***

Leucil/fenilalanil-tRNA transferasa	Degradación de proteínas	4
R transferasa	Degradación de proteínas	2

**Sub-Familia: *aminoacil-tRNA reductasas***

Glutamil-tRNA reductasa	Biosíntesis de porfirina	4
-------------------------	--------------------------	---

**Sub-Familia: *peptidil-tRNA hidrolasas***

Peptidil-tRNA hidrolasa	Síntesis de proteínas	7
-------------------------	-----------------------	---

**Sub-Familia: *Ribosomal proteins***

Proteína ribosomal 50s L19	Síntesis de proteínas	2
Factor de elongación P	Síntesis de proteínas	2
Ribonucleasa PH	Síntesis de proteínas	3

**Sub-Familia: *tRNA guanina transglicosilasas***

Queuina-tRNA ribosiltransferasa	Biosíntesis de tRNA-queuosina	1
<b>Sub-Familia: tRNA ribosyl transferase isomerase</b>		
S-adenosilmetionina-tRNA ribosiltransferasa isomerasa	Biosíntesis de tRNA-queuosina	1
<b>Sub-Familia: Amidasas</b>		
Amidasa	Amidación de péptidos	5
<b>Sub-Familia: DTD-like</b>		
D-tirosil-tRNA deacilasa	Catabolismo de aminoácidos	4
<b>Sub-Familia: tRNA/rRNA metiltransferasas</b>		
tRNA/rRNA-metiltransferasa	Procesamiento del tRNA	5
<b>Sub-Familia: tRNA pseudouridina sintasa</b>		
tRNA-pseudouridina sintasa A	Procesamiento del tRNA	4
tRNA-pseudouridina sintasa B	Procesamiento del tRNA	3
<b>Sub-Familia: tRNA Ile lisidina sintasas</b>		
tRNAIle-lisidina sintasa	Procesamiento del tRNA	3
<b>Sub-Familia: tRNA adenosina deaminasas</b>		
tRNA-adenosina deaminasa	Procesamiento del tRNA	1
<b>Sub-Familia: tRNA dihidrouridina sintasa</b>		
tRNA-dihidrouridina sintasa	Procesamiento del tRNA	3
<b>Sub-Familia: Proteínas G</b>		
tRNA-modificación GTPasa	Procesamiento del tRNA	1
<b>Sub-Familia: tRNA guanina metiltransferasa</b>		
tRNA-guaninaN1 metiltransferasa	Procesamiento del tRNA	4
tRNA-guaninaN7 metiltransferasa	Procesamiento del tRNA	4
<b>Sub-Familia: tRNA dimetilaliladenosina sintasa</b>		
tRNA-2-metilio-N6-dimetilaliladenosina sintasa	Procesamiento del tRNA	4
<b>Sub-Familia: tRNA adenina metiltransferasa</b>		
tRNA-adenina-5(8) N1-metiltransferasa	Procesamiento del tRNA	1
<b>Sub-Familia: tRNA isopentenilpirofosfato transferasa</b>		
tRNA-delta-2-isopentenilpirofosfato transferasa	Procesamiento del tRNA	2
<b>Sub-Familia: tRNA dimetilaliltransferasa</b>		
tRNA-dimetilaliltransferasa	Procesamiento del tRNA	2
<b>Sub-Familia: tRNA tiouridilasa</b>		
tRNA-especifico-2 tiouridilasa	Procesamiento del tRNA	4
<b>Total</b>		<b>254</b>

Una vez que clasificamos nuestras tRUEs por familias y ruta metabólica central, se modificaron los encabezados de cada una de las secuencias con tales datos. Este paso es importante, ya que EvoMining sólo reconocerá aquellas secuencias que cumplan con tal formato (ver sección 4.2.1 de Métodos).

### 5.1.2. Construcción de la base de datos G-DB

Las bacterias del filo *Actinobacteria* son los organismos más reconocidos en cuanto a producción de NPs. Por tal razón, construimos nuestra G-DB con los genomas más representativos y mejor ensamblados de este filo disponibles en la base de datos NCBI.

Al momento de construir nuestra G-DB, la base de datos NCBI contenía ~1600 genomas de *Actinobacteria*, sin embargo elegimos solo aquellos que cumplieran con un criterio de calidad de ensamblado establecido. Este criterio de calidad lo determinamos en base al menor número de contigs pero mayor tamaño entre dos genomas de la misma especie. Con este filtro logramos rescatar 1244 genomas (**Tabla 7**). El género más sobrerrepresentado fue *Streptomyces* (367 cepas), seguido de *Corynebacterium* (100), *Mycobacterium* (90), *Bifidobacterium* (63), *Actinomyces* (49), *Nocardia* (31), *Microbacterium* (25), *Rhodococcus* (25), *Gordonia* (23) y *Amycolatopsis* (20).

Tabla 7. Composición de la G-DB

Género(s)	Número de cepas por género
<i>Streptomyces</i>	367
<i>Corynebacterium</i>	100
<i>Mycobacterium</i>	90
<i>Bifidobacterium</i>	63
<i>Actinomyces</i>	49
<i>Nocardia</i>	31
<i>Microbacterium, Rhodococcus</i>	25
<i>Gordonia</i>	23
<i>Amycolatopsis</i>	20
<i>Arthrobacter, Nocardiosis</i>	16
<i>Frankia</i>	13
<i>Nocardioides</i>	12
<i>Clavibacter, Micromonospora, Propionibacterium</i>	11
<i>Kitasatospora, Saccharomonospora</i>	10
<i>Streptacidiphilus</i>	9
<i>Actinoplanes, Brevibacterium, Cellulomonas, Kocuria, Mobiluncus</i>	8
<i>Pseudonocardia</i>	7
<i>Atopobium, Rothia, Saccharothrix</i>	6
<i>Actinobaculum, Actinomadura, Brachybacterium, Candidatus, Leucobacter, Micrococcus, Saccharopolyspora, Salinispora</i>	5
<i>Actinoalloteichus, Actinopolyspora, Collinsella, Knoellia, Leifsonia, Nesterenkonia</i>	4
<i>Actinokineospora, Dietzia, Geodermatophilaceae, Glycomyces, Kutzneria, Lechevalieria, Lentzea, Microbispora, Olsenella, Slackia, Streptosporangium, Tetrasphaera</i>	3
<i>Aeromicrobium, Agrococcus, Agromyces, Arcanobacterium, Arsenicococcus, Blastococcus, Chloroflexus, Dermabacter, Dermacoccus, Enterorhabdus, Intrasporangium, Janibacter, Jonesia, Kribbella, Marmoricola, Mobilicoccus, Modestobacter, Nakamurella, Nonomurea, Rathayibacter, Rubrobacter, Sciscionella, Segniliparus, Serinicoccus, Thermobifida, Thermocrispum, Tropheryma, Trueperella, Tsukamurella, Varibaculum</i>	2

---

*Acaricomes, Acidimicrobium, Acidothermus, Actinobacteria, Actinocatenispora, Actinomycetospora, Actinopolymorpha, Actinospica, Actinosynnema, Actinotignum, Adlercreutzia, Aestuariimicrobium, Agreia, Alloactinosynnema, Allokutzneria, Alloscardovia, Amycollicoccus, Anaerolinea, Austwickia, Beutenbergia, Caldilinea, Catelliglobospora, Catenulispora, Catenuloplanes, Cellulosimicrobium, Cellvibrio, Citricoccus, Conexibacter, Coriobacteriaceae, Coriobacterium, Corynebacteriales, Cryobacterium, Cryocola, Cryptobacterium, Cryptosporangium, Dactylosporangium, Dehalococcoides, Dehalogenimonas, Demetria, Dermatophilus, Eggerthella, Enorma, Frigoribacterium, Gardnerella, Geodermatophilus, Glaciibacter, Gordonibacter, Granulicoccus, Gryllotalpicola, Gulosibacter, Haloglycomyces, Herbiconiux, Humibacter, Ilumatobacter, Intrasporangiaceae, Isopterocola, Jiangella, Kibdelosporangium, Kineococcus, Kineosphaera, Kineosporia, Ktedonobacter, Kytococcus, Longispora, Luteipulveratus, Lysinimicrobium, Microlunatus, Microtetrastroma, Mycetocola, Nitrolancea, Nocardioideae, Oerskovia, Ornithinimicrobium, Oscillochloris, Paraoerskovia, Parascardovia, Phycococcus, Pilimelia, Pimelobacter, Prauserella, Promicromonospora, Propionimicrobium, Pseudoclavibacter, Renibacterium, Roseiflexus, Ruania, Salinibacterium, Sanguibacter, Scardovia, Senegalimassilia, Sinomonas, Sphaerobacter, Sporichthya, Stackebrandtia, Streptomonospora, Terrabacter, Terracoccus, Thermobispora, Thermogemmatispora, Thermomicrobium, Thermomonospora, Timonella, Tomitella, Verrucosipora, Williamsia, Xylanimonas, Yaniella*

---

1

### 5.1.3. Construcción de la base de datos antiSMASH-DB

antiSMASH es la herramienta de minería genómica convencional de NPs más reconocida a nivel mundial. Este programa analiza el contenido de BGCs de NPs de un genoma bacteriano comparándolo contra una base de datos que contiene todas las firmas conservadas de las principales clases de NPs conocidos hasta la fecha. Anotar los genomas actinobacterianos de nuestra G-DB con antiSMASH nos sirve como ‘control positivo’ para las predicciones de BGCs de EvoMining. Por lo tanto, nuestra base de datos antiSMASH-DB se construyó con cada uno de los archivos resultantes (concatenado de BGCs detectados, formato GenBank) de un análisis tipo antiSMASH de cada uno de los 1244 genomas de G-DB.

### 5.1.4. Construcción de la base de datos BGCs-DB

La construcción de la base de datos BGCs-DB consistió en la recopilación de los BGCs (formato GenBank) depositados en el repositorio MIBiG. Al mes de agosto de 2016, esta base de datos contenía 1294 BGCs pertenecientes a 26 clases, o combinaciones de clases conocidas de NPs (**Tabla 8**).

Tabla 8. Clases de BGCs del repositorio MIBiG

Clases de BGC	Número de BGCs	Ejemplos de NPs de cada clase de BGC
<i>Clases únicas de BGCs</i>		
Policétidos	328	ciclohexamida, aflatoxina, avermectina
NRPs	209	CDA, valinomicina, daptomicina, lipstatina
RiPPs	178	nisina A, microbiosporicina, goadsporina
Otros	141	anfotericina, ácido clavulónico, arsenopolicétidos
Sacáridos	123	apramicina, kanamicina, estreptomina
Terpenos	72	geosmina, grixazona, hopeno, diazepinomicina
Alcaloides	21	scitonemina, violaceina, fumigaclavina C
Nucleosidos	1	tunicamicina
<i>Clases de BGCs híbridos</i>		
NRP / Policétido	149	undecilprodigiosina, rapamicina, lancacidina
Policétido / Sacárido	25	nistatina, clorotricina, arixantomina
Policétido / Terpeno	10	fumagillina, ácido micofenólico, piriropeno A
NRP / Otro	5	pacidamicina, fengycina, napsamicina
NRP / Policétido / Sacárido	6	estreptolidigina, talisomicina, zorbamicina
Alcaloide / Terpeno	4	xiamicina, ambigua, lolitrem
NRP / Sacárido	4	hassalidinas, ristomicina A, brabantamida
Alcaloide / NRP	3	fumitremorgina, malbrancheamida
Otro / Policétido / Sacárido	3	pactamicina, SF2575, simociclina
Otro / Sacárido	3	clorobiocina, coumermicina A1, novobiocina
Sacárido / Terpeno	3	fenalinolactona, astaxantina didesoxiglicósido
Alcaloide / NRP / Policétido	1	cilindrospermopsina
Alcaloide / NRP / Terpeno	1	notoamida/stefacidin
Alcaloide / Policétido	1	anatoxina
NRP / Otro / Policétido	1	zeamina
NRP / RiPP	1	feganomicina
NRP / Terpeno	1	telecidina B

Cabe mencionar que durante la revisión de la literatura cuando buscamos las tRUEs, encontramos varios reportes recientes de BGCs que no estaban registrados en el repositorio MIBiG. Por lo tanto, complementamos nuestra BGCs-DB con 4 BGCs adicionales que pertenecen a 4 casos de rutas biosintéticas que contienen enzimas tRUEs: valanimicina (componente azoxy), albonoursina, micociclosina y pulcherrimina (ciclodipéptidos). En total, BGC-DB estuvo conformada por 1298 BGCs de NPs conocidos.



En nuestro análisis, encontramos que 71 de las 78 familias de tRUEs analizadas contenían expansiones enzimáticas en al menos 2 genomas de *Actinobacteria* (**Tabla 9**). Las 7 familias de tRUEs en las que no se detectaron expansiones fueron: pirrolisil-tRNA sintetasa, sep-tRNA<sup>Cys</sup>-tRNA sintetasa, sep-tRNA<sup>Sec</sup>-tRNA sintetasa, O-fosfoseril-tRNA<sup>Cys</sup> sintetasa, O-fosfoseril-tRNA<sup>Sec</sup> cinasa, R-transferasa y aspartil/glutamil-tRNA amidotransferasa subunidad C. Cabe resaltar que no se detectaron expansiones enzimáticas en las primeras 6 familias de tRUEs antes mencionadas debido a que tampoco se encontraron sus homólogos en los genomas analizados. Al no haber homólogos, no hay expansiones. El hecho de no encontrar homólogos de estas familias de tRUEs particularmente en los genomas de actinobacterias se debe probablemente a la naturaleza de las secuencias de referencia que se usaron para el análisis, las cuales provenían de organismos ajenos al filo *Actinobacteria*. Además, las familias pirrolisil-tRNA sintetasa, sep-tRNA<sup>Cys</sup>-tRNA sintetasa, sep-tRNA<sup>Sec</sup>-tRNA sintetasa, O-fosfoseril-tRNA<sup>Cys</sup> sintetasa y O-fosfoseril-tRNA<sup>Sec</sup> cinasa están funcionalmente involucradas en la biosíntesis de pirrolisil-tRNAs y selenocisteinil-tRNAs, moléculas altamente conservadas en archeas (Sauerwald, y otros, 2005; Rother & Krzycki, 2010) mas no en actinobacterias, lo cual también podría explicar su ausencia en nuestro conjunto de genomas. Lo mismo ocurrió con la enzima R-transferasa, involucrada en la degradación dirigida de proteínas, cuyas secuencias de referencia provenían de dos bacterias del filo *Proteobacteria*. En el caso de la enzima aspartil/glutamil-tRNA amidotransferasa subunidad C, la cual sí está presente en *Actinobacteria*, hay homólogos pero sin expansiones. Sin embargo, las subunidades A y B de esta misma enzima sí presentan expansiones enzimáticas en algunos genomas.

**Tabla 9. Número de homólogos y familias de tRUEs expandidas detectadas por EvoMining**

Familias de tRUEs	Núm. de homólogos totales detectados en G-DB ( $\Sigma$ de cada columna)	Promedio del núm. de homólogos totales ( $\bar{X}$ de cada columna)	Desviación estándar del núm. de homólogos totales ( $\sigma$ de cada columna)	Valor estadístico para determinar expansión ( $\bar{X} + \sigma$ )	Núm. de genomas con tRUEs expandidas (celdas rojas)
Glutaminil-tRNA sintetasa	1112	0.8939	0.7716	1.6655	605
Glutamil-tRNA sintetasa	1746	1.4035	0.5053	1.9089	502
Triptofanil-tRNA sintetasa	1739	1.3979	0.5412	1.9391	488
Queuina-tRNA ribosiltransferasa	491	0.3947	0.5004	0.8951	485
Arginil-tRNA sintetasa	1691	1.3593	0.5279	1.8872	432
Metionil-tRNA formiltransferasa	1579	1.2693	0.4853	1.7546	345
tRNA-adenosina deaminasa	324	0.2605	0.4391	0.6995	324
S-adenosilmetionina-tRNA ribosiltransferasa isomerasa	322	0.2588	0.4436	0.7025	319
L-seril-tRNA <sup>Sec</sup> seleno transferasa	315	0.2532	0.4514	0.7046	307
Selenocisteinil-tRNA sintetasa	313	0.2516	0.4451	0.6967	307
MprF1	406	0.3264	0.6697	0.9961	296
Treonil-tRNA sintetasa	1527	1.2275	0.4436	1.6711	281
Metionil-tRNA sintetasa	1557	1.2516	0.5646	1.8162	273
Valil-tRNA sintetasa	2750	2.2106	0.5128	2.7234	260
RAK-fosfatidilglicerol sintetasa	313	0.2516	0.5278	0.7794	260
Prolil-tRNA sintetasa	2162	1.7379	0.8253	2.5632	252
Asparaginil-tRNA sintetasa	271	0.2178	0.4573	0.6752	247
Alanil-tRNA sintetasa	1465	1.1777	0.4571	1.6348	222
FemX	752	0.6045	0.8582	1.4627	221
Alanil-fosfatidilglicerol sintetasa	915	0.7355	0.9381	1.6737	220
FemY	710	0.5707	0.856	1.4267	215
VanK	698	0.5611	0.853	1.4141	214
Glutamil-tRNA amidotransferasa	5800	4.6624	3.2595	7.9219	201
Aspartil/glutamil-tRNA amidotransferasa subunidad A	5196	4.1768	3.1665	7.3433	170
Lisil-fosfatidilglicerol sintetasa	1405	1.1294	1.1232	2.2526	168
Aspartil-tRNA sintetasa	1405	1.1294	0.359	1.4884	165
Amidasa	5240	4.2122	3.1493	7.3615	162
Lisil-tRNA sintetasa integral de membrana	1251	1.0056	1.0989	2.1045	145
Lisil-tRNA sintetasa	2551	2.0506	1.1043	3.1549	145
Isoleucil-tRNA sintetasa	2616	2.1029	0.4272	2.5301	140
Leucil/fenilalanil-tRNA transferasa	137	0.1101	0.3233	0.4334	134
Aspartil-tRNA sintetasa no-discriminante	1350	1.0852	0.312	1.3972	116
MprF2	490	0.3939	0.6884	1.0823	114
Glutamil-tRNA sintetasa no-discriminante	110	0.0884	0.3162	0.4046	99
Seril-tRNA sintetasa	1342	1.0788	0.3237	1.4025	99
Alanil-diacilglicerol sintetasa	504	0.4051	0.6328	1.038	83

Aminotransferasa dependiente de piridoxal-fosfato	1880	1.5113	0.7313	2.2426	82
FemA	97	0.078	0.3581	0.4361	68
Cisteinil-tRNA sintetasa	2305	1.8529	0.5122	2.3651	66
FemAB	85	0.0683	0.3394	0.4077	58
FemB	59	0.0474	0.2895	0.337	39
Epr	59	0.0474	0.295	0.3425	38
tRNA-modificación GTPasa	541	0.4349	0.5893	1.0242	38
Lif	56	0.045	0.2913	0.3363	35
tRNA-2-metiltio-N6-dimetilaliladenosina sintasa	1911	1.5362	0.5697	2.1058	32
MurN	49	0.0394	0.2736	0.313	31
Tirosil-tRNA sintetasa	1257	1.0105	0.1857	1.1962	29
Leucil-tRNA sintetasa	1260	1.0129	0.1789	1.1918	27
MurM	26	0.0209	0.1431	0.164	27
Glutamil-tRNA reductasa	1124	0.9035	0.3547	1.2583	22
Peptidil-tRNA hidrolasa	2037	1.6375	0.5378	2.1752	22
Histidil-tRNA sintetasa	872	0.701	0.4935	1.1945	21
tRNA/rRNA-metiltransferasa	3041	2.4445	0.7489	3.1935	18
Fenilalanil-tRNA sintetasa sub A	1253	1.0072	0.1234	1.1307	14
tRNA-delta-2-isopentenilpirofosfato transferasa	1240	0.9968	0.1268	1.1236	9
tRNA-dimetilaliltransferasa	1240	0.9968	0.1268	1.1236	9
Fenilalanil-tRNA sintetasa sub B	1204	0.9678	0.2098	1.1776	8
tRNA-dihidouridina sintasa	1191	0.9574	0.2282	1.1856	8
tRNA-adenina-5(8) N1-metiltransferasa	1197	0.9622	0.2108	1.173	6
D-tirosil-tRNA deacilasa	1080	0.8682	0.3478	1.216	5
Factor de elongación P	1244	1	0.0802	1.0802	5
tRNA-guaninaN1 metiltransferasa	1234	0.992	0.1132	1.1051	4
Glicil-tRNA sintetasa	1186	0.9534	0.2221	1.1754	3
tRNA-guaninaN7 metiltransferasa	1089	0.8754	0.3352	1.2106	3
tRNAIle-lisidina sintasa	1133	0.9108	0.2908	1.2015	3
tRNA-pseudouridina sintasa B	1228	0.9871	0.1262	1.1133	3
tRNA-especifico-2 tiouridilase	1158	0.9309	0.26	1.1909	3
Aspartil/glutamil-tRNA amidotransferasa subunidad B	1234	0.992	0.1058	1.0978	2
Proteína ribosomal 50s L19	1238	0.9952	0.0801	1.0753	2
Ribonucleasa PH	1209	0.9719	0.1702	1.1421	2
tRNA-pseudouridina sintasa A	1223	0.9831	0.135	1.1181	2
Aspartil/glutamil-tRNA amidotransferasa subunidad C	986	0.7926	0.4056	1.1982	0
O-fosfoseril-tRNACys sintetasa	0	0	0	0	0
O-fosfoseril-tRNA Sec cinasa	0	0	0	0	0
Pirrolisil-tRNA sintetasa	0	0	0	0	0
Sep-tRNACys -tRNA sintetasa	0	0	0	0	0
Sep-tRNA Sec -tRNA sintetasa	0	0	0	0	0
R transferasa	0	0	0	0	0

De estos resultados preliminares, surgieron preguntas como: ¿cuál es la familia de tRUEs con el mayor número de expansiones en los genomas de *Actinobacteria*? ¿Y, cuál tiene el menor? ¿Existe alguna relación entre el número de homólogos totales de una familia y el número de expansiones en los genomas? Para responder estas dudas, únicamente ordenamos los datos y extrajimos la información.

**Tabla 10. Comparación de las tRUEs con el mayor y menor número de homólogos, promedio  $\bar{X}$ , desviación estándar  $\sigma$ , valor estadístico ( $\bar{X} + \sigma$ ) y número de genomas con expansiones enzimáticas**

<b>Característica</b>	<b>Familia de tRUEs (valor)</b>	
<b>Mayor núm. de homólogos</b>	Glutamil-tRNA amidotransferasa	(5800)
<b>Mayor <math>\bar{X}</math> del núm. de homólogos</b>	Glutamil-tRNA amidotransferasa	(4.6624)
<b>Mayor <math>\sigma</math> del núm. de homólogos</b>	Glutamil-tRNA amidotransferasa	(3.2595)
<b>Mayor (<math>\bar{X} + \sigma</math>) de expansión</b>	Glutamil-tRNA amidotransferasa	(7.9219)
<b>Mayor núm. de expansiones en los genomas</b>	Glutaminil-tRNA sintetasa	(605)
<b>Menor núm. de homólogos</b>	MurM	(26)
<b>Menor <math>\bar{X}</math> del núm. de homólogos</b>	MurM	(0.0209)
<b>Menor <math>\sigma</math> del núm. de homólogos</b>	Proteína ribosomal 50s L19	(0.08)
<b>Menor (<math>\bar{X} + \sigma</math>) de expansión</b>	MurM	(0.164)
<b>Menor núm. de expansiones en los genomas</b>	Proteína ribosomal 50s L19	(2)
	Asp/glu-tRNA amidotransferasa sub B	(2)
	tRNA-pseudouridina sintasa A	(2)
	Ribonucleasa PH	(2)

Como podemos observar en la **Tabla 10**, la familia de tRUEs con el mayor número de expansiones enzimáticas fue la glutaminil-tRNA sintetasa, cuya función en el metabolismo central es la unión del aminoácido glutamina (Gln) con su tRNA<sup>Gln</sup> cognado. El número de homólogos de esta familia de tRUEs en los genomas fue de 1112, con un número 'normal' de 1 homólogo por genoma. Sin embargo, estos valores son relativamente bajos al compararlos con la familia Glutamil-tRNA amidotransferasa, la familia con el mayor número de homólogos detectados (5800)

y mayor número 'normal' de homólogos por genoma (4-5). Este comportamiento de los datos es similar al comparar otros pares de enzimas, lo cual nos podría indicar que no necesariamente existe una relación directa entre el número de homólogos de una familia de enzimas con el número de expansiones enzimáticas. En otras palabras, no podemos asumir que si una familia tiene un alto número de homólogos será directamente proporcional al número de expansiones enzimáticas. Ocurre de manera similar con el otro extremo de los datos, es decir, con las familias de tRUEs con el menor número de homólogos, menor promedio y menor número de expansiones enzimáticas.

Regresando a la tubería bioinformática, EvoMining procedió a comparar las 71 familias de tRUEs con expansiones enzimáticas detectadas contra la base de datos BGCs-DB para determinar posibles eventos de reclutamiento de tRUEs expandidas en rutas de biosíntesis de NPs. A continuación, los resultados.

### **5.2.2 Eventos de reclutamiento de tRUEs en biosíntesis de NPs**

El siguiente conjunto de resultados preliminares del análisis EvoMining, es una tabla que muestra una relación de homología entre familias de tRUEs con expansiones enzimáticas y genes pertenecientes a un BGC de NPs (BLASTp de la base de datos interna FEExp-DB contra BGCs-DB). Por lo tanto, esta tabla representa las familias de tRUEs expandidas que han sido reclutadas en BGCs de NPs. En nuestro análisis, EvoMining encontró que 47 de las 71 familias de tRUEs expandidas contenían enzimas homólogas al menos a un gen dentro de un BGC de la base de datos BGCs-NPs (**Tabla 11**). En esta ocasión, la familia con el mayor número de tRUEs homólogas dentro de BGCs de NPs fue Leucil/fenilalanil-tRNA transferasa.

**Tabla 11. Familias tRUEs expandidas que están reclutadas por BGCs de NPs.**

Familias de tRUEs	EC.núm.	Núm. de BGCs que reclutan homólogos de tRUEs expandidas	Clases de los BGCs que reclutan tRUEs expandidas
Leucil/fenilalanil-tRNA transferasa	2.3.2.6	958	Varios
Asp/Glu-tRNA amidotransferasa sub A	6.3.5.-	479	Varios
Queuina-tRNA ribosiltransferasa	* 2.4.2.29	158	Varios
Amidasa	3.5.1.4	68	Varios
Lisil-tRNA sintetasa	6.1.1.6	53	Varios
Lisil-fosfatidilglicerol sintetasa	2.3.2.3	51	Varios
Lisil-tRNA sintetasa integral de membrana	6.1.1.6	51	Varios
RAK-fosfatidilglicerol sintetasa	* 2.3.2.-	50	Varios
Alanil-diacilglicerol sintetasa	* 2.3.2.-	50	Varios
Glutamil-tRNA amidotransferasa	6.3.5.-	37	Varios
Peptidil-tRNA hidrolasa	3.1.1.29	10	balhimycin, chloroeremomycin, enterobactin, mirubactin, ristocetin, ristomycin_A, vanchrobactin, vancomycin, vulnibactin, thiolutin
Metionil-tRNA sintetasa	2.1.2.9	9	BE-7585A, steffimycin, coelichelin, heterobactin, scabichelin, lipopolysaccharide, oxazolomycin, rhizopodins, albachelin
Alanil-tRNA sintetasa	6.1.1.7	9	lasalocid, cystomanamides, pristinamycin, paenilamicin, zwittermycin_A, kosinostatin
Aminotransferasa dependiente de piridoxal-fosfato (PLP)	* 2.6.1.-	9	calicheamicin, quartromycin, macrotetrolide, SF2575, pyoverdine, ustiloxin_B, lipopolysaccharide, O-antigen
Asparaginil-tRNA sintetasa	* 6.3.5.6	8	vicenistatin, friulimycin, skyllamycin, nikkomycin, streptolydigin, tallysomycin, ustilagic_acid
Triptofanil-tRNA sintetasa	* 6.1.1.2	6	2-deoxystreptamine, gentamicin, sisomicin, dymemicin, indolmycin
Prolil-tRNA sintetasa	6.1.1.15	5	2-amino-4-methoxy-trans-3- butenoic_acid, echinocandin_B, pyoverdine, bactobolin, pneumocandin
Aspartil-tRNA sintetasa	* 6.1.1.12	5	guadinomine, streptolydigin, tallysomycin, thuggacin, ustilagic acid

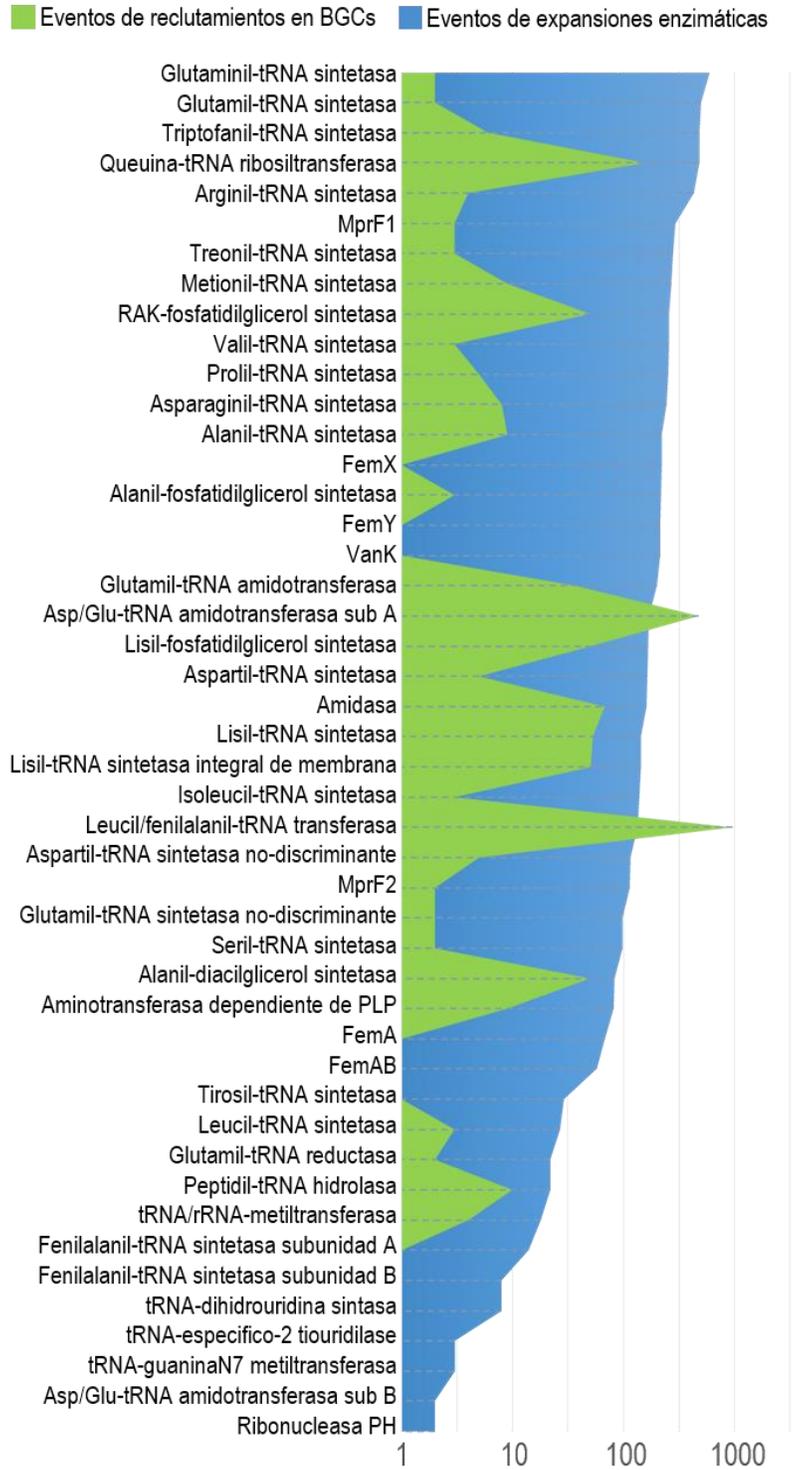
Aspartil-tRNA sintetasa no-discriminante	*	6.1.1.12	5	guadinomine, streptolydigin, tallysomicin, thuggacin, ustilagic_acid
Arginil-tRNA sintetasa		6.1.1.19	4	A40926, arginomycin, dynemicin
tRNA/rRNA-metiltransferasa		2.1.1.-	4	avilamycin_A, heme_D1, pederin
MprF1		2.3.2.-	3	guadinomine, tallysomicin, kalimantacin/batumin
Treonil-tRNA sintetasa	*	6.1.1.3	3	borrelidin, oxazolomicin
Valil-tRNA sintetasa	*	6.1.1.9	3	mupirocin, tubulysin, thiomarinol
Alanil-fosfatidilglicerol sintetasa	*	2.3.2.11	3	guadinomine, tallysomicin, kalimantacin/batumin
Isoleucil-tRNA sintetasa	*	6.1.1.5	3	mupirocin, tubulysin, thiomarinol
Leucil-tRNA sintetasa	*	6.1.1.4	3	mupirocin, tubulysin, thiomarinol
Glutaminil-tRNA sintetasa	*	6.3.5.7	2	aculeximycin, oviedomycin
Glutamil-tRNA sintetasa	*	6.1.1.17	2	aculeximycin, oviedomycin
MprF2		2.3.2.-	2	guadinomine, kalimantacin/batumin
Glutamil-tRNA sintetasa no-discriminante	*	6.1.1.17	2	aculeximycin, oviedomycin
Seril-tRNA sintetasa	*	6.1.1.11	2	diaphorin, valanimycin
Glutamil-tRNA reductasa	*	1.2.1.70	2	heme_D1
FemX	*	2.3.2.16	1	A47934
FemY	*	2.3.2.-	1	A47934
VanK	*	2.3.2.-	1	A47934
FemA	*	2.3.2.17	1	A47934
FemAB	*	2.3.2.-	1	A47934
Tirosil-tRNA sintetasa	*	6.1.1.1	1	tyrocidine
Fenilalanil-tRNA sintetasa subunidad A	*	6.1.1.20	1	streptomycin
Fenilalanil-tRNA sintetasa subunidad B	*	6.1.1.20	1	streptomycin
tRNA-dihidrouridina sintasa		1.3.1.-	1	capsular_polysaccharide
tRNA-especifico-2 tiouridilase		2.8.1.132	1	albonoursin
tRNA-guaninaN7 metiltransferasa		2.1.1.33	1	pactamycin
Asp/Glu-tRNA amidotransferasa sub B	*	6.3.5.-	1	ET-743
Ribonucleasa PH		2.7.7.56	1	bottromycin_A2

\*Familias de tRUEs de las que se construyó su filogenia en el presente trabajo

Sin embargo, cuando observamos estos resultados nos pareció sorprendente que algunas de las familias de tRUEs tuvieran un gran número de eventos de reclutamiento en BGCs, como fue el caso de Leucil/fenilalanil-tRNA transferasa

(958), Asp/Glu-tRNA amidotransferasa sub A (479), Queuina-tRNA ribosiltransferasa (158), Amidasa (68), Lisil-tRNA sintetasa (53), Lisil-fosfatidilglicerol sintetasa (51), Lisil-tRNA sintetasa integral de membrana (51), RAK-fosfatidilglicerol sintetasa (50), Alanil-diacilglicerol sintetasa (50) y Glutamil-tRNA amidotransferasa (37); mientras que las otras 36 familias de tRUEs tenían eventos de reclutamientos dentro de un rango de 1-10. Para comprender este fenómeno, analizamos detalladamente los resultados del BLASTp entre la base de datos interna FEExp-DB (familias de enzimas expandidas) y la BGCs-DB. En estos archivos encontramos que se habían detectado homólogos anotados como tRUEs que contenían dominios adicionales al resto de la familia, e.g. detectamos que un par de homólogos de la familia Leucil/fenilalanil-tRNA transferasa contenían un dominio de adenilación (A). Particularmente este dominio A es una de las firmas conservadas para detectar una de las clases más abundantes de BGCs, que son las NRPSs. Al revisar nuevamente la **Tabla 8**, pudimos observar que hay al menos 381 BGCs (NRPs e híbridos) que pudieron haber alineado con estas enzimas que contienen el dominio A. Por lo tanto, la predicción de este gran número de reclutamientos es un resultado falso-positivo, ya que no representan idóneamente a la familia analizada.

Dejando a un lado estos falsos-positivos, nos enfocamos a analizar el resto de los datos. Específicamente, surgió la duda si existía una correlación entre los eventos de expansiones enzimáticas (detectados en el paso anterior) y los eventos de reclutamiento en BGCs de NPs. Para esto, tomamos los datos de las Tablas 9 y 11 para crear un gráfico de correlación en escala logarítmica ( $\log_{10}$ ) (**Fig. 22**).



**Figura 22. Comportamiento de los eventos de expansión de las tRUEs y sus reclutamientos en BGCs de NPs.** Los valores de los eventos de expansiones enzimáticas (azules) representan el número de genomas con la correspondiente familia de tRUEs expandida, mientras que los valores de los eventos de reclutamiento (verdes) representan los BGCs que contienen homólogos de las tRUEs expandidas. En ambos casos, los valores fueron ajustados a una escala logarítmica ( $\log_{10}$ ).

En esta figura podemos visualizar el comportamiento de los datos de expansión y reclutamiento. Es evidente que no existe una relación directamente proporcional entre los eventos de expansiones enzimáticas (azules) y los eventos de reclutamiento en BGCs (verdes). El número de clases únicas de BGCs que conocemos actualmente es relativamente escaso (1294 en MIBiG) en comparación con lo que seguramente existe en la naturaleza. Es de esperarse que solo unas cuantas familias expandidas puedan ser asociadas a los pocos BGCs que se conocen. Para poder asignar una correlación a todas las familias expandidas, necesitaríamos el repertorio completo de BGCs existentes en la Naturaleza, lo cual es imposible hoy en día. Por lo tanto, entre más enriquezcamos nuestras bases de datos de BGCs con clases nuevas, podríamos detectar un mayor número de reclutamientos y a su vez, nuevos metabolitos.

Este razonamiento conllevó a no perder de vista a las otras 25 familias de tRUEs expandidas a las que no fue posible asignar una relación con la base de datos BGCs-DB (**Tabla 12**). Resulta ilógico aceptar que algunas de las familias con el mayor número de eventos de expansión, e.g. Metionil-tRNA formiltransferasa (345), tRNA-adenosina deaminasa (324), o Cisteinil-tRNA sintetasa (66), no tienen reclutamientos en BGCs. Sin duda, el análisis funcional profundo de estas familias no reclutadas es de suma importancia, ya que podríamos desenmascarar funciones ocultas de las enzimas expandidas. Sin embargo, este análisis queda como una perspectiva del presente proyecto.

**Tabla 12. Familias de tRUEs expandidas que no fueron asociadas a algún BGCs**

<b>Familia de tRUEs</b>	<b>Eventos de expansión</b>
Metionil-tRNA formiltransferasa	345
tRNA-adenosina deaminasa	324
S-adenosilmetionina-tRNA ribosiltransferasa isomerasa	319
L-seril-tRNA <sup>Sec</sup> seleno transferasa	307
Selenocisteinil-tRNA sintetasa	307
Cisteinil-tRNA sintetasa	66
FemB	39
Epr	38
tRNA-modificación GTPasa	38
Lif	35
tRNA-2-metiltio-N6-dimetilaliladenosina sintasa	32
MurN	31
MurM	27
Histidil-tRNA sintetasa	21
tRNA-delta-2-isopentenilpirofosfato transferasa	9
tRNA-dimetilaliltransferasa	9
tRNA-adenina-5(8) N1-metiltransferasa	6
D-tirosil-tRNA deaciclaza	5
Factor de elongación P	5
tRNA-guaninaN1 metiltransferasa	4
Glicil-tRNA sintetasa	3
tRNAIle-lisidina sintasa	3
tRNA-pseudouridina sintasa B	3
Proteína ribosomal 50s L19	2
tRNA-pseudouridina sintasa A	2

## **5.3 Análisis cualitativo de EvoMining enfocado a tRUEs**

### **5.3.1 Reconstrucciones filogenéticas de las familias tRUEs reclutadas en BGCs de NPs**

Los últimos pasos de la tubería bioinformática de EvoMining es la reconstrucción filogenética y anotación funcional de las familias de enzimas en las que se detectaron tanto eventos de expansión enzimática como reclutamientos en BGCs de NPs. Para fines del presente proyecto, elegimos construir la filogenia de 29 de las 46 familias de tRUEs con eventos de reclutamiento en BGCs que habíamos detectado (**Apéndice**).

Las filogenias que reconstruye EvoMining simbolizan la historia evolutiva de cada una de las familias de enzimas, donde cada rama representa un homólogo de las tRUEs que se usaron como referencia. Por lo tanto, el número total de ramas en un árbol es igual al número de homólogos de dicha familia detectados en los genomas más el número de hits rescatados de la base de datos MIBiG. La topología general del tipo de árboles que reconstruye EvoMining se caracteriza por la presencia de un clado visiblemente divergente, el cual representa las expansiones enzimáticas de la familia en cuestión. Idealmente, el clado divergente debería contener todos los posibles homólogos de la familia de tRUEs particular asociados con BGCs de NPs. De acuerdo a la anotación funcional que EvoMining realiza sobre estos árboles (predicciones detectadas por EvoMining en color verde, predicciones detectadas por EvoMining y antiSMASH color cian y los hits homólogos encontrados en la base de datos MIBiG en color azul), esperaríamos ver ramas en color azul, cian y verde.

Una rama en color azul dentro del clado divergente, nos indica que todo ese clado podría estar asociado a biosíntesis de NPs, es decir, que los homólogos que se encuentran en ese clado tienen historias evolutivas parecidas y dado que la rama azul está asociada a BGCs (hit rescatado de MIBiG, homólogo a la familia de tRUEs particular) es probable que los demás también. Las ramas en color cian, predicciones de EvoMining pero que también detecta antiSMASH, funcionan como otro control positivo del clado divergente, ya que representan homólogos de la familia de tRUEs particular que fueron detectados en la anotación previa de los genomas con antiSMASH, lo cual nos indica que este tipo de homólogos sí podrían estar asociados a biosíntesis de NPs. Por lo tanto, si una rama de color verde (predicción de EvoMining) se localiza dentro del clado divergente y está cercana a ramas azules y cian, es altamente probable que ese homólogo sea parte de una ruta biosintética. Las ramas verdes representan las posibles nuevas rutas biosintéticas que detecta EvoMining.

La topología de las 29 filogenias de tRUEs reconstruidas fue muy variable, sin embargo todas contienen el clado divergente característico de las familias de enzimas expandidas con una anotación funcional también diversa. Debido a la magnitud del elevado número de homólogos con que se reconstruyeron algunas filogenias, la cuantificación exacta de las predicciones propias de cada familia fue complicada. Por lo cual, únicamente realizamos una estimación representativa del número de predicciones de BGCs (**Tabla 13**).

**Tabla 13. Características generales de la anotación funcional de los clados divergentes de las reconstrucciones filogenéticas de las tRUEs**

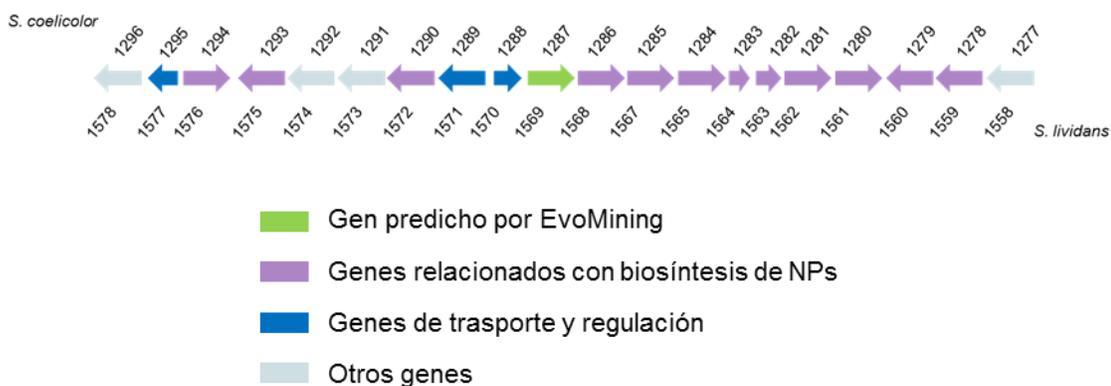
<b>Familias enzimáticas</b>	<b>Estimación del número de ramas marcadas como predicciones de EvoMining</b>	<b>Estimación del número de ramas marcadas como predicciones de antiSMASH</b>	<b>Estimación del número de ramas marcadas como hits con MIBiG</b>
Aminotransferasa dependiente de PLP	++++	++++	+++
Asp/glu-tRNA amidotransferasa sub. B	+	++	+
Asparaginil-tRNA sintetasa	+	++	+++
Aspartil-tRNA sintetasa	+	++	+++
Aspartil-tRNA sintetasa no-discriminante	+	++	+++
Fenilalanil-tRNA sintetasa subunidad A	+	++	+
Fenilalanil-tRNA sintetasa subunidad B	+	+++	+
Glutamil-tRNA sintetasa	+++	+++	++
Glutamil-tRNA sintetasa no-discriminante	+	+	++
Glutaminil-tRNA sintetasa	+++	++++	++
Isoleucil-tRNA sintetasa	++	++	++
Leucil-tRNA sintetasa	+	++	++
Seril-tRNA sintetasa	++++	+++	++
Tirosil-tRNA sintetasa	++++	+++	+
Treonil-tRNA sintetasa	++++	++	++
Triptofanil-tRNA sintetasa	++++	++++	+++
Valil-tRNA sintetasa		+	++
FemA	+	+	+
FemAB	+	+	+
FemX	++++	+++	+
FemY	++++	+++	+
VanK	++++	+++	+
Alanil-diacilglicerol sintetasa	+	++	+++++
Alanil-fosfatidilglicerol sintetasa	+	++	++
RAK-fosfatidilglicerol sintetasa	++	++	+++++
MprF1	++	++	++
MprF2	++	++	++
Glutamil-tRNA reductasa	++++	+++++	+
Queuina-tRNA ribosiltransferasa	++	+++	+++++

Posterior a este análisis semi-cuantitativo, nos dimos a la tarea de analizar el contexto genómico de la mayoría de las predicciones de EvoMining (homólogos divergentes en color verde) de nuestras filogenias. Para esto, EvoMining cuenta con un subprograma integrado en su interfaz gráfica el cual arroja una imagen con el contexto genómico de cualquier homólogo(s) que el usuario seleccione, respetando la escala del tamaño de los genes, su orientación y la estructura general del gen, además de indicar el organismo al que pertenece. Dado que estos contextos genómicos se muestran anotados, el usuario puede determinar si los homólogos seleccionados realmente pertenecen a una ruta biosintética de acuerdo a la anotación de los genes de la vecindad genómica (el usuario requiere ciertos conocimientos previos para realizar estas estimaciones manuales).

Dentro de estos contextos genómicos analizados, logramos detectar diversos BGCs que muy seguramente codifican para nuevas clases de NPs, ya que la anotación de sus genes y su combinación son muy peculiares y no fueron detectados por antiSMASH. Estos resultados conllevan a concluir preliminarmente que las tRUEs son enzimas altamente probables en BGCs, más de lo que previamente imaginamos, y que podrían ser buenas pautas para herramientas de minería genómica convencionales.

Adicional a este análisis *in silico* de las familias de tRUEs, llevamos a cabo la caracterización experimental de un BGCs que contiene una tRUE predicho por EvoMining en un trabajo anterior (Trabajo de tesis de doctorado, Pablo Cruz Morales, 2013). Lo más interesante de haber elegido este clúster para nuestros experimentos es que es un BGC nunca antes visto y que se detectó en dos de las

cepas modelo de laboratorio más estudiadas y minadas: *S. coelicolor* A3 (2) y *S. lividans* 66 (**Fig. 23**). Este BGC se derivó de la reconstrucción de la historia evolutiva de la familia FemXAB en actinobacterias. La tRUE presente en este clúster es la O-seril-isobutilhidroxilamina sintasa, una enzima previamente reportada en el clúster biosintético de la valanimicina, la cual se presume ser la responsable de la estructura base del grupo azoxy de este compuesto (Grag, Ma, Hoyt, & Parry, 2002). Por tal razón, nombramos a este clúster: azoxy-BGC. A continuación, se describe a detalle la caracterización de este BGC.

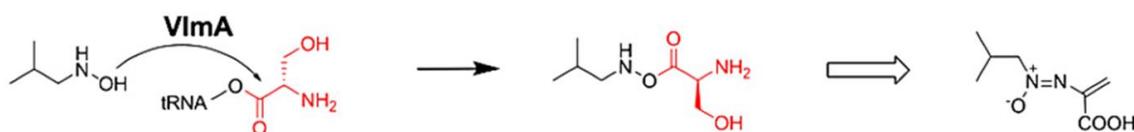


**Figura 23. Estructura del azoxy-BGC detectado en *S. coelicolor* y *S. lividans***

## 5.4 Expresión del azoxy-BGC

### 5.4.1 Propuesta bioinformática

Un análisis tipo EvoMining de la familia FemXAB, reveló un nuevo BGC codificado en los genomas de las cepas modelo *S. lividans* y *S. coelicolor*, dos de los organismos más estudiados y minados por expertos alrededor del mundo. Hasta la fecha, se sabe que cada uno de estos organismos contiene 27 BGCs de NPs detectados con antiSMASH y 63-68 BGCs putativos detectados por ClusterFinder (complemento de antiSMASH). Sin embargo, EvoMining detectó un nuevo clúster biosintético que contiene una enzima de la familia de tRUEs FemXAB. Este hallazgo certifica el poder de predicción que tiene EvoMining, al ser capaz de desenmascarar BGCs ocultos, incluso en genomas de organismos que han sido examinados profundamente. La anotación funcional de los genes codificados en este clúster biosintético (**Tabla 8**) sugiere que el producto final es un compuesto con un grupo funcional azoxy. Esta hipótesis surge de haber identificado la enzima isobutilhidroxilamina sintasa (VImA) dentro del clúster, la cual es considerada como una enzima esencial para la formación del grupo azoxy ( $R - N^+ \equiv N^+ - R'$ ) de la valanimicina (Grag, Qian, Alemany, Moran, & Parry, 2008) (**Fig. 24**). Este grupo funcional es una entidad química rara en la naturaleza, lo cual nos despertó el interés por caracterizar el azoxy-BGC detectado en *S. coelicolor* y *S. lividans*.



**Figura 24. Mecanismo de reacción de VImA correspondiente a la ruta de la valanimicina.** VImA es una enzima dependiente de tRNA que lleva a cabo la reacción clave para formar el esqueleto de la valanimicina.

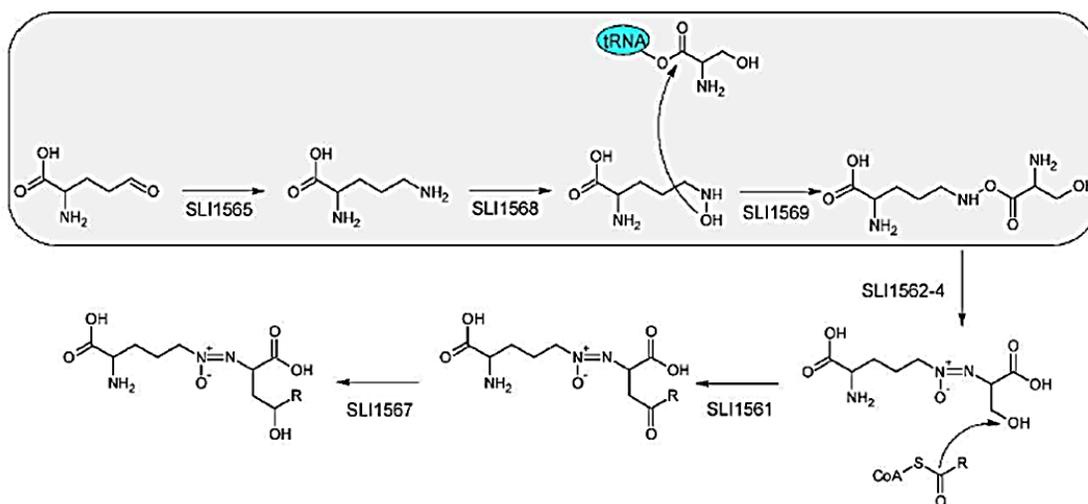
La anotación funcional del azoxy-BGCs de *S. coelicolor* y *S. lividans* se realizó usando dos programas distintos. El primero fue el servidor RAST (Rapid Annotations using Subsystems Technology), uno de los mejores anotadores para genomas bacterianos basado en el sistema SEED (Aziz, y otros, The RAST Server: rapid annotations using subsystems technology, 2008). Y el segundo fue Phyre2, basado en el reconocimiento de proteínas homólogas o análogas en base a su plegamiento (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015). La anotación con ambos programas es bastante similar, con algunas excepciones (**Tabla 14**).

**Tabla 14. Anotación del azoxy-BGC con RAST y Phyre2**

Gen-ID por organismo		Anotación con RAST	Anotación con Phyre2
SCO	SLI		
SCO1277	SLI1558	Proteína hipotética	Colágeno simple tipo 1
SCO1278	SLI1559	Proteína de unión a ATP/GTP	Cinasa
SCO1279	SLI1560	Metiltransferasa dependiente de SAM	Metiltransferasa dependiente de SAM
SCO1280	SLI1561	Diacilglicerol O-acil transferasa	Tirocidina sintetasa 3
SCO1281	SLI1562	Oxidoreductasa dependiente de FAD	Dimetilglicina deshidrogenasa
SCO1282	SLI1563	Oxidoreductasa	Oxidoreductasa
SCO1283	SLI1564	Ciclasa deshidratasa	Ciclasa deshidratasa
SCO1284	SLI1565	Ornitina aminotransferasa	Aminotransferasa
SCO1285	SLI1567	Cetoreductasa	Cetoreductasa // Probable NRPS
SCO1286	SLI1568	Amino N-hidrolasa	Oxidoreductasa
<b>SCO1287</b>	<b>SLI1569</b>	<b>O-seril-isobutilhidroxilamina sintasa</b>	<b>L-alanil-fosfatidilglicerol sintasa</b>
SCO1288	SLI1570	Proteína integral de membrana	Fosfatidilinositol 3-cinasa (Vps34)
SCO1289	SLI1571	Regulador transcripcional	Proteína reguladora de la fam GntR
SCO1290	SLI1572	Alcalina fosfatasa	Alcalina fosfatasa D
SCO1291	SLI1573	Proteína hipotética	Factor de procesividad de la DNApol
SCO1292	SLI1574	Proteína hipotética	Proteína hipotética
SCO1293	SLI1575	Acetiltransferasa (GNAT)	Lisina acetil-transferasa
SCO1294	SLI1576	Cistationa gama-sintasa	O-succinilhomoserina sulfidrilasa
SCO1295	SLI1577	Regulador transcripcional fam AsnC	Proteína reguladora - responde a Leu
SCO1296	SLI1578	Proteína hipotética	Lipoproteína

\* SCO = *S. coelicolor* A3 (2); SLI = *S. lividans* 66

De acuerdo a esta anotación y al antecedente que teníamos de la ruta biosintética de la valanimicina, surgió una propuesta de la posible estructura del metabolito producido por el azoxy-BGC. Esta predicción involucra la acción de los genes Sli1561-Sli1569 los cuales podrían estar produciendo la molécula que se muestra en la **figura 25**. La naturaleza de esta molécula es interesante debido al grupo funcional azoxy que contiene, lo cual inspiró el diseño experimental para elucidar la molécula producida por este clúster biosintético detectado por EvoMining.



**Figura 25. Predicción de la posible estructura del producto final del azoxy-BGC**

#### 5.4.2. Caracterización experimental del azoxy-BGC

La estrategia general de experimentación dirigida para la expresión del azoxy-BGC consistió en la construcción de mutantes del clúster, un análisis comparativo de fenotipo entre cepas mutantes y silvestres bajo diferentes condiciones de crecimiento, análisis cromatográficos (HPLC) y un análisis de expresión de genes (RT-PCR).

##### *Construcción de mutantes*

La mayoría de las bacterias no son transformables con DNA linear debido a la presencia de exonucleasas *recBCD* intracelulares que degradan el DNA linear. Existen varios métodos que han logrado la recombinación homóloga de *E.coli* sobre fragmentos de unos cientos pares de bases usando el sistema de recombinación del bacteriófago  $\lambda$  (sistema  $\lambda$ -Red) (Poteete, 2001; Court, Sawitzke, & Thomason, 2002). Sin embargo, cuando se trata de manipular todo un clúster biosintético (10 – 30kb) de *Streptomyces*, no es tan simple. Afortunadamente, Gust y colaboradores crearon una versión adaptada del sistema  $\lambda$ -Red para manipular exitosamente BGCs de *Streptomyces* (Gust, Kieser, & Chater, 2002). Esta estrategia consiste en reemplazar DNA de un inserto de un cósmido de *S. coelicolor* secuenciado (los cuales pueden consultarse de <http://strepdb.streptomyces.org.uk/>) por un marcador de selección (gen de resistencia a antibióticos) que ha sido generado por PCR con oligonucleótidos que contienen brazos homólogos de 39 nucleótidos. El casete de interrupción incluye el sitio *oriT*, el cual permite la conjugación intergénica mediada por RP4, para insertar el DNA de cósmido de *E. coli* a *S. coelicolor*. Para evadir la potente

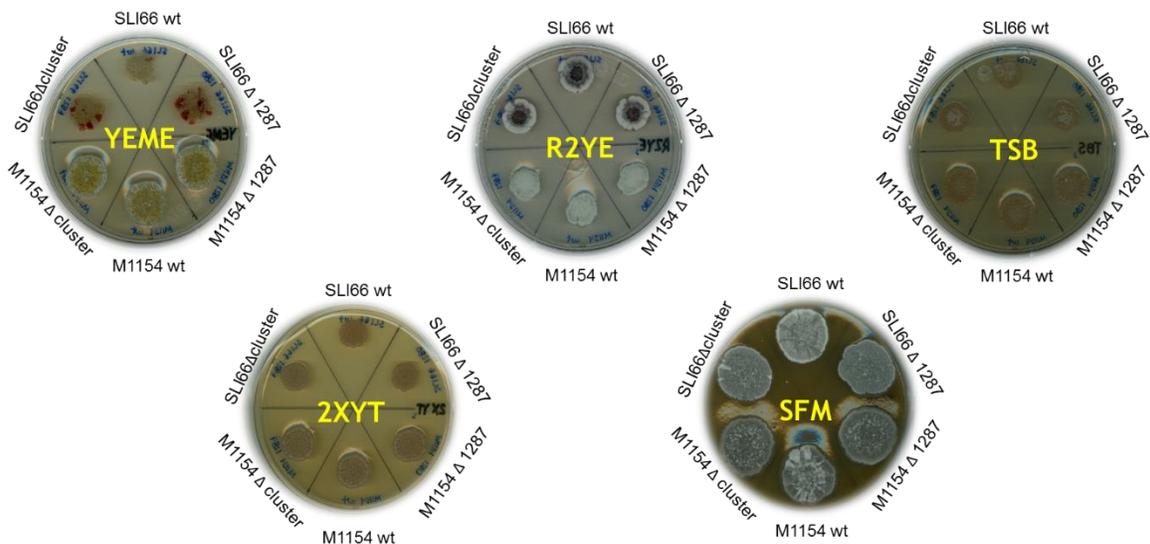
restricción por metilación de *Streptomyces*, se emplea *E. coli* ET12567, una cepa con metilación deficiente. Los vectores que contienen *oriT*, son móviles *in trans* en *E. coli* por el plásmido pUZ8002. Los marcadores de selección que se usan en los casetes deben funcionar tanto para *E. coli* como para *Streptomyces*, en nuestro caso usamos kanamicina y apramicina.

Usando este sistema construimos dos mutantes del azoxy-BGC distintas en *S. lividans* 66 y *S. coelicolor* M1154. La primera de *S. lividans* 66 fue sobre la enzima isobutilhidroxilamina sintasa (*Sli66Δ1569*) y en la segunda, mutamos del gen *Sli1561* al *Sli1570* del clúster (*Sli66Δcluster*). En el caso de *S. coelicolor*, trabajamos con la cepa M1154 (*ScoM1154*), una mutante derivada de la cepa M145 que contiene deleciones en los clústeres biosintéticos de la actinorrodina (*act*), prodiginina (*red*), CPK (*cpk*) y CDA (*cda*) para reducir el potencial competitivo del suministro de carbono y nitrógeno, así como mutaciones puntuales en los genes *rpoB* (*[C128T]*) y *rpsL* (*[A262G]*) para incrementar pleiotrópicamente la producción de NPs (Gomez-Escribano & Bibb, 2011). En este fondo genético, mutamos su enzima O-seril-isobutilhidroxilamina sintasa (*M1154Δ1287*) y del gen *Sco1280* al *Sco1288* (*M1154Δcluster*). Una vez que las mutantes fueron confirmadas por PCR, se almacenó un stock de esporas para posteriores experimentos.

### ***Análisis comparativo entre el fenotipo de mutantes vs silvestres***

Los streptomycetes son organismos con una capacidad de adaptación sorprendente, equipados con la maquinaria metabólica necesaria para aprovechar diferentes fuentes de carbono y nitrógeno. De acuerdo al tipo de sustrato que

adquieren del medio en que crecen, estas bacterias activan diferentes rutas metabólicas tanto para llevar a cabo sus funciones vitales como para la producción de NPs. Por tal razón, realizamos un análisis comparativo de fenotipos entre nuestras cepas silvestres y mutantes de Sli66 y ScoM1154, en medios de cultivo ricos en nutrientes (YEME, R2YE, TSB, 2XYT, SFM) y medios con composición mínima (MM, SMM, SMMS, NMMP). En estos últimos, añadimos la variable del fósforo (un medio de cultivo sin fósforo (0%) y otro con fósforo  $\text{KH}_2\text{PO}_4$ , 0.5%), un mineral que actúa como inductor de metabolismo especializado. Lo que esperábamos ver en estas comparaciones era un cambio morfológico, e.g. pigmentación, esporulación, etcétera, entre cepas silvestres y mutantes que pudiera estar asociado a su fisiología afectada por la presencia/ausencia del azoxy-BGC.

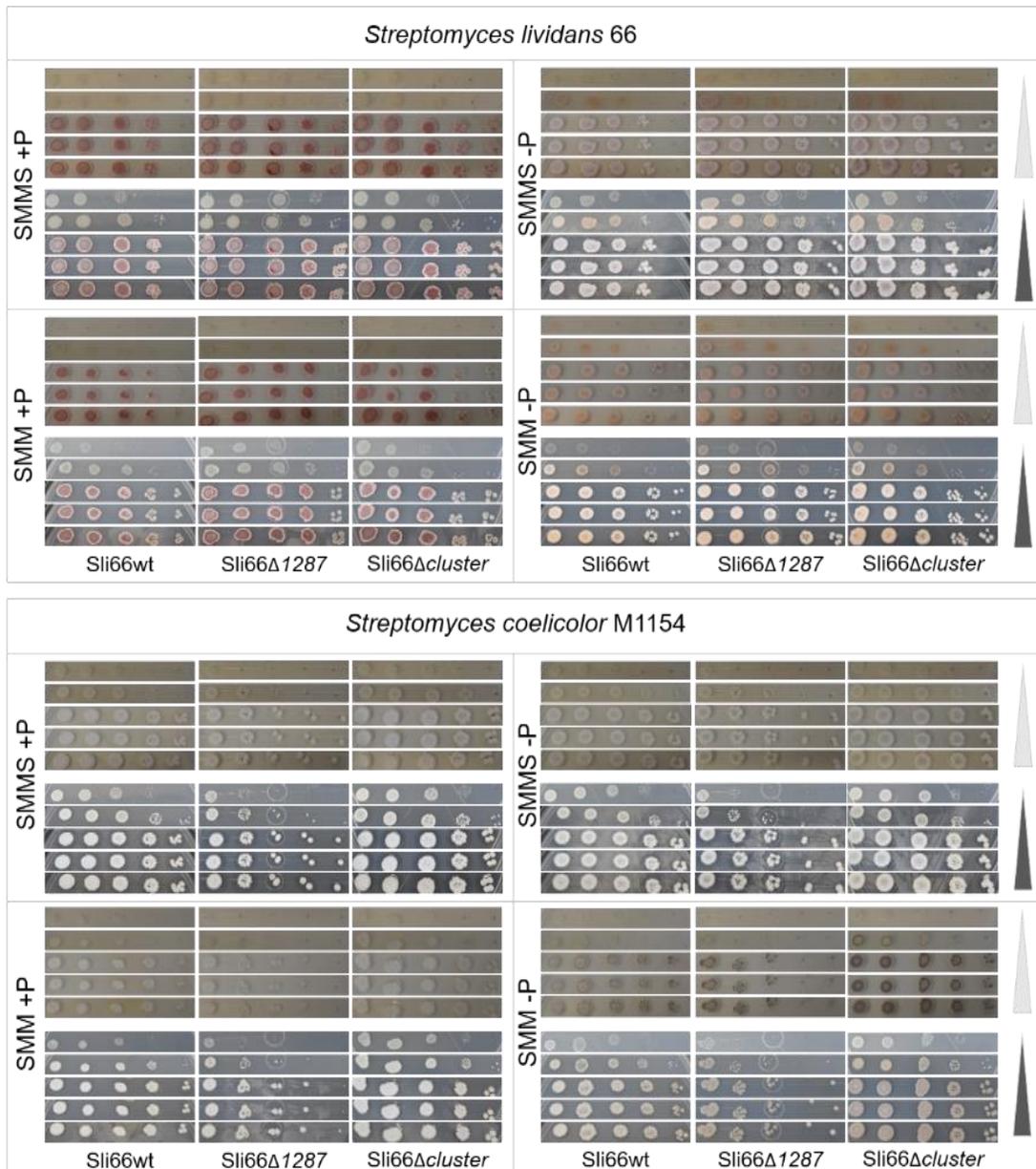


**Figura 26. Diferencias fenotípicas entre mutantes y silvestres creciendo en medios ricos en nutrientes.** En estas placas se inoculó el mismo número de esporas por gota (~100,000). El tiempo de crecimiento al momento de las fotografías fue de 4 días, creciendo a una temperatura de 30°C. YEME= Yeast Extract-Malt Extract; R2YE= R2 Yeast Extract; TSB= Tryptic Soy Broth (más agar); 2XYT= Yeast Extract Tryptone; SFM= Soya Flour Manitol.

En el caso de los experimentos de fenotipo en placas de medios ricos en nutrientes, no observamos grandes cambios morfológicos entre cepas silvestres y mutantes, es decir, entre Sli66wt (silvestre) vs Sli66 $\Delta$ 1569 y Sli66 $\Delta$ cluster (mutantes) (**mitad superior de las placas, Fig. 26**) ni entre M1154wt (silvestre) vs M1154 $\Delta$ 1287 y M1154 $\Delta$ cluster (mutantes) (**mitad inferior de las placas, Fig. 26**). Sin embargo, comparando entre placas de distintos medios, si se pueden apreciar diferencias en el tiempo de esporulación, pigmentación y modo de dispersión de las esporas. Estas diferencias se deben principalmente al tipo de fuentes de carbono y nitrógeno, así como minerales presentes en cada tipo de medio. Generalmente, estos medios se utilizan para propagación de esporas del género *Streptomyces*. Otra diferencia que podemos observar es que algunas cepas de Sli66, tanto silvestres como mutantes, tienen pigmentos rojos que no están presentes en las cepas de ScoM1154. Esto se debe a que, como mencionamos anteriormente, ScoM1154 wt es una cepa a la que le fueron mutados los BGCs de actinorrodina (act), prodiginina (red), CPK (cpk) y CDA (cda), lo cual está afectando la producción de este pigmento desde un principio y no por efecto de la mutación en el azoxy-BGC.

De todos los medios probados, donde se observaron mayores diferencias fenotípicas entre silvestres y mutantes, fueron los medios SMM y SMMS sin fósforo. Ambos medios se caracterizan por tener una composición mínima de nutrientes, pero suficiente para que crezcan las bacterias. Cuando los streptomycetes crecen en este tipo de medios, se ven obligados a activar de manera temprana su metabolismo especializado en respuesta al estrés

ocasionado por la deficiencia de nutrientes, lo cual convierte a estos medios en predilectos para la producción de NPs. La diferencia fenotípica más clara que observamos en este estudio fueron cambios en la pigmentación y diferencias en el tiempo de esporulación comparando cepas silvestres y mutantes (**Fig. 27**).



**Figura 27. Diferencias fenotípicas entre mutantes y silvestres creciendo en medios mínimos.** Esta figura muestra el monitoreo de 5 días de crecimiento de las cepas analizadas (de arriba hacia abajo, indicado por los triángulos de la derecha). Las fotografías fueron tomadas tanto en fondo claro como oscuro (parte superior e inferior de cada panel, respectivamente). Se inocularon 100mil, 10mil, mil, cien y diez esporas por línea (de izq. a der.) Las diferencias más visibles entre silvestres y mutantes son la pigmentación y la esporulación.

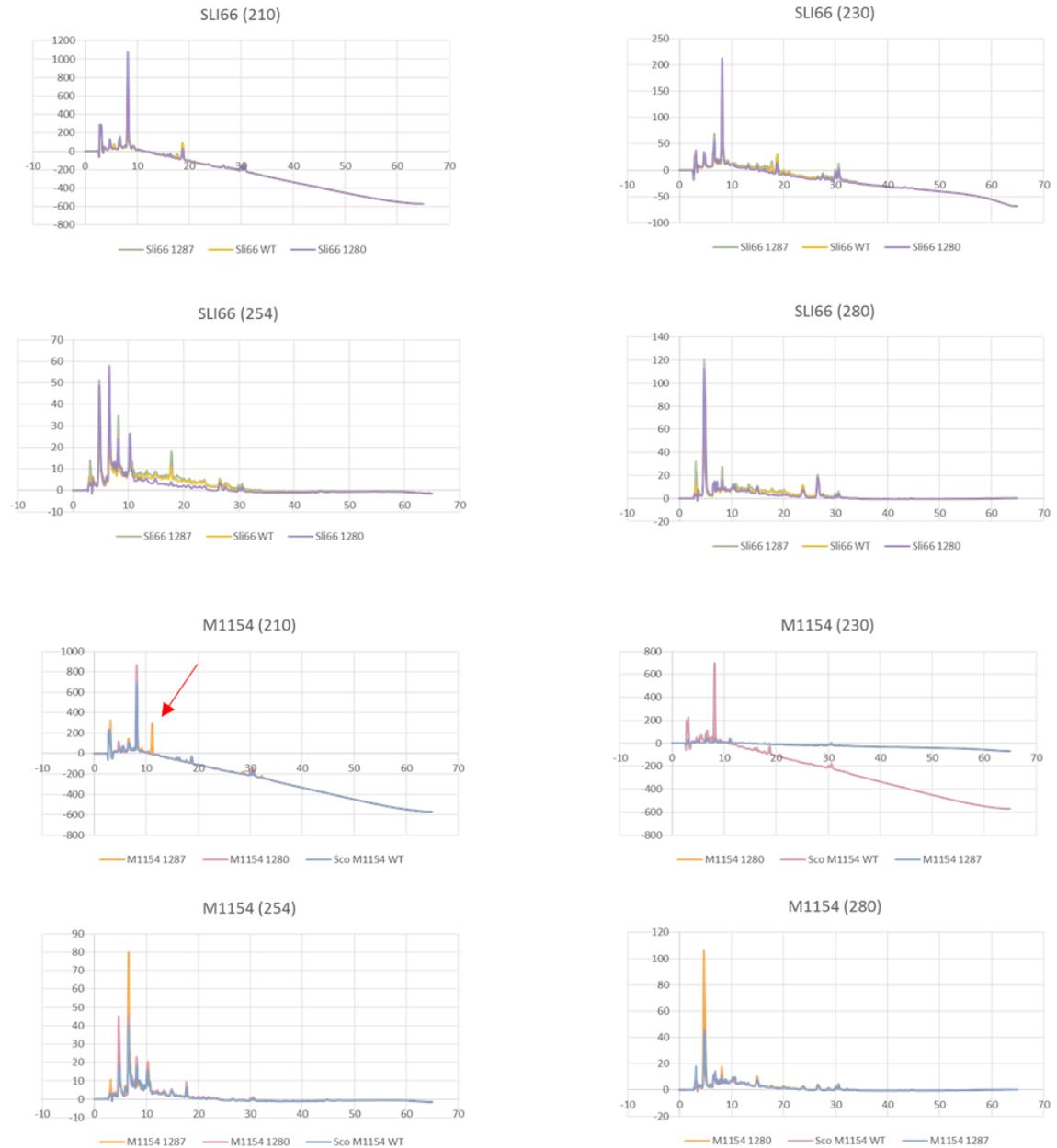
En el caso de Sli66, podemos observar que la pigmentación en color rojo aparece antes en las cepas mutantes que en las silvestres. Este efecto puede deberse a que los genes mutados podrían estar involucrados en una red de regulación que se ve afectada por la interrupción del azoxy-BGC. Por otro lado, en las cepas ScoM1154 no observamos este cambio de pigmentación, dado que este fondo genético contiene las mutaciones (*act*, *red*, *cpk* y *cda*) que impiden la producción de los pigmentos azul y rojo. En este caso, observamos que la esporulación ocurre más temprano en las cepas mutantes que las silvestres. La esporulación es la última etapa del ciclo celular de los streptomycetes, lo cual indica la culminación de su desarrollo (Barka, y otros, 2015). Esto quiere decir que, al haberse agotado los nutrientes disponibles a su alrededor, el microorganismo asegura su descendencia produciendo esporas. Una etapa antes, justo en el momento en que los nutrientes se vuelven escasos, debió haber activado su metabolismo especializado para el aprovechamiento máximo de sustratos del medio (Barka, y otros, 2015).

En base a estos resultados, determinamos que los medios de cultivo SMM, SMM – P (sin fósforo), SMMS y SMMS –P (sin fósforo) eran los indicados para realizar los próximos experimentos de expresión del azoxy-BGC. Hicimos fermentaciones de las 6 cepas (mutantes y silvestres) en los medios seleccionados producir biomasa y extraer los metabolitos derivados de esta fermentación. El cultivo producto de la fermentación (medio sin células), se utilizó para los análisis cromatográficos por HPLC para tratar de identificar el metabolito asociado al azoxy-BGC. Mientras que

la biomasa obtenida fue usada para aislar RNA total, el cual se usó para un análisis de expresión por medio de RT-PCR.

### ***Análisis de HPLC de los extractos de fermentación***

Debido a que no conocemos la estructura exacta de la molécula que tratamos de identificar, no existía un método de separación cromatográfica dirigido para nuestra molécula. Por lo tanto, nuestros experimentos de HPLC fueron necesariamente exploratorios. Basándonos en la predicción de la molécula guiada por la anotación de los genes del clúster (Sección 5.4.1), se planteó que la mejor estrategia para separar el metabolito azoxy del extracto de fermentación era usar el método de fase reversa con una columna C-18 (ZORBAX®). C-18 es una columna de sílica ideal para separar metabolitos de naturaleza peptídica. Además, usamos un cromatógrafo de HPLC Agilent con arreglo de diodos que nos permitió obtener lecturas a 4 longitudes de onda distintas 210, 230, 254 y 280 nm. Analizamos alrededor de 72 extractos de fermentaciones: 6 cepas (Sli66wt, ScoM1154wt y mutantes), 3 tiempos (24, 48 y 72 h) y 4 medios (SMM y SMMS, con y sin fosfatos). En estos cromatogramas esperábamos ver picos diferenciales entre cepas silvestres (pico alto) y mutantes (pico nulo o bajo), los cuales serían recolectados y posteriormente analizados mediante cromatografía de masas. Sin embargo, de todas nuestras muestras analizadas, solo observamos un pico diferencial en la muestra de la fermentación de ScoM1154 en medio SMMS sin fósforo a una longitud de onda de 210nm (**Fig. 28**). Antes de enviar esta fracción a cromatografía de masas, realizamos el análisis de RT-PCR de algunos genes del azoxy-BGCs para saber si el clúster había sido inducido.



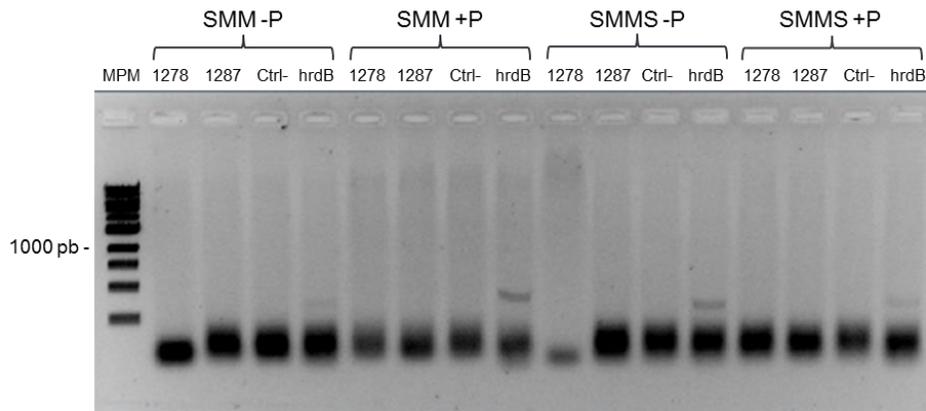
**Figura 28. Cromatogramas de HPLC con picos diferenciales entre cepas mutantes y silvestres.** Este conjunto de cromatogramas representa las muestras de Sli66 y ScoM1154 crecidas en medio SMMS -P a las 72h de fermentación. Únicamente pudimos identificar un pico diferencial (marcado con la flecha roja) entre todos los cromatogramas, el cual estaba presente en la muestra de la cepa silvestre, pero ausente en las mutantes.

### *Análisis de RT-PCR del azoxy-BGC*

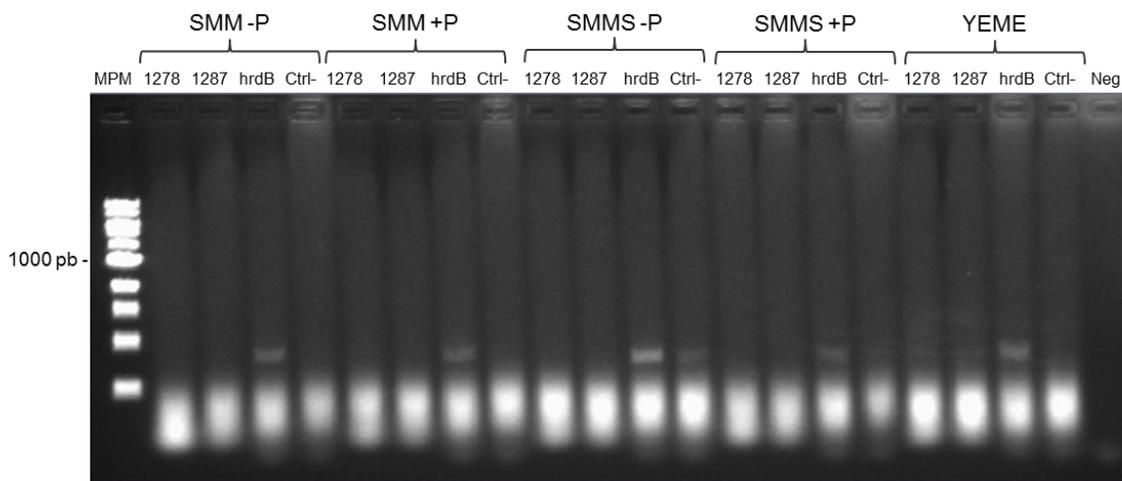
De las 72 fermentaciones que teníamos, seleccionamos sólo un conjunto de muestras que correspondían al tiempo 72 h, de las cepas silvestres crecidas en los 4 medios. Extrajimos RNA de dichas muestras (protocolo que tuvimos que estandarizar para garantizar la calidad del RNA para el RT-PCR) y se probó la expresión de 4 genes: Sco1278 y Sco1289 (genes fuera del clúster), Sco1286 y Sco1287 (genes dentro del clúster). De acuerdo a la condición de crecimiento de cada cepa (medio de cultivo) y su fondo genético (ScoM1154 o Sli66), esperábamos ver distintos patrones de expresión de los genes seleccionados. Por un lado, si el clúster estaba apagado en cierta condición, esperábamos ver que los genes fuera del clúster mantuvieran su expresión. Por otro lado, si el clúster estaba activo, esperábamos ver los cuatro genes expresados. Otro caso posible era que ninguno de los 4 genes estuviera expresado porque quizás los límites del clúster que habíamos identificado eran incorrectos, y por lo tanto, los genes que consideramos como ajenos al clúster tampoco se expresarían si el clúster estaba apagado.

Los resultados de este análisis muestran que en ninguna de las condiciones de crecimiento probadas se expresan los genes dentro o fuera del clúster (**Fig. 29**). Todo indica que bajo estas condiciones de crecimiento que probamos, el azoxy-BGC no se está induciendo a pesar de los cambios observados en su morfología entre mutantes y silvestres, y de haber observado un pico diferencial en los análisis de HPLC. Este pico podría ser algún otro metabolito ajeno al azoxy-BGC.

*Streptomyces coelicolor* M1154



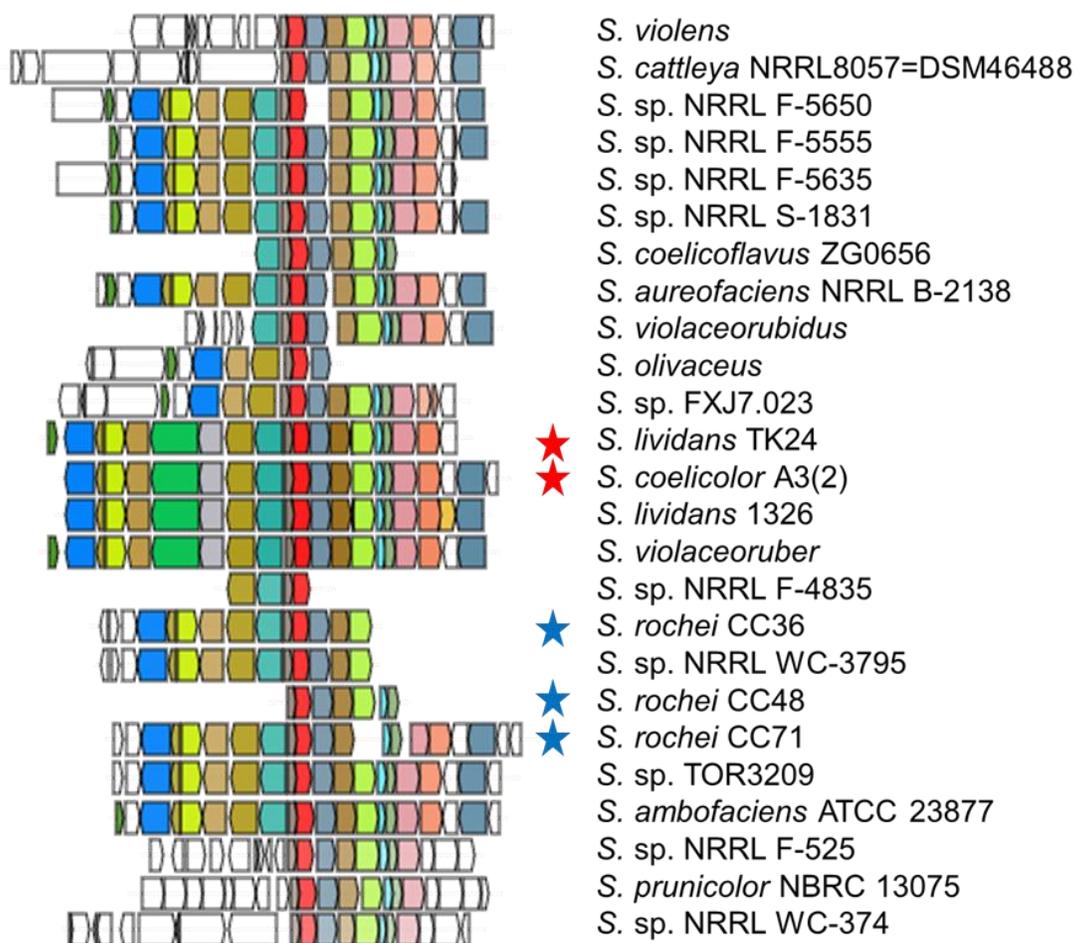
*Streptomyces lividans* 66



**Figura 29. RT-PCR de los genes *Sco1278* y *Sco1287* de las cepas silvestres *Sli66* y *ScoM1154* creciendo en diferentes medios de cultivo.** Las bandas que se ven en los geles corresponden al gen *hrdB*, un gen constitutivo usado como control positivo de la reacción de RT-PCR. El control negativo certifica que no hay presencia de DNA genómico en las muestras de RNA, ya que en estas reacciones, se suprimió la reacción de cDNA. Únicamente se muestran estas dos fotos representativas del análisis de RT-PCR, el resto de las reacciones se veían similares.

Tratando de entender por qué no se expresaban ni los genes fuera del clúster, nos dimos a la tarea de volver a analizar los límites del azoxy-BGC. Para esto empleamos CORASON (CORe Analysis of Syntenic Orthologs to prioritize Natural Product-BGCs), una herramienta visual que detecta los genes core de un clúster

conservados en un conjunto de genomas dado (en nuestro caso actinobacterias), ordenándolos filogenéticamente en base al grado de conservación del clúster (Salem-Mojica, N, *Unpublished*). El resultado de este análisis reveló que el azoxy-BGC estaba conservado en otros *Streptomyces*, incluyendo a tres cepas ambientales del laboratorio identificadas como *S. rochei* 36, 71 y 48, y que los límites del clúster que consideramos para realizar nuestros experimentos estaban errados (**Fig. 30**).



**Figura 30. Análisis de conservación del azoxy-BGC en actinobacterias usando CORASON.** Las estrellas rojas resaltan el azoxy-BGC de *S. lividans* y *S. coelicolor*, Las estrellas azules los clústeres de las cepas ambientales *S. rochei* CC36, CC48 y CC71.

Haber encontrado el azoxy-BGC conservado en nuestras cepas ambientales *S. rochei*, resultó muy interesante ya que recientemente se reportó el aislamiento de un compuesto denominado azoxyalqueno producido por el organismo *S. rochei* 7434AN4 (Kunitake, Hiramatsu, Kinashi, & Arakawa, 2015). En este artículo, únicamente se reporta la estrategia que permitió la producción del metabolito así como su estructura química y las reacciones probables que dan lugar al compuesto (**Fig. 31**), más no se reportan los genes involucrados en la biosíntesis del mismo. Por lo que pudimos observar en la serie de reacciones propuestas por los autores, encontramos cierta relación con las funciones de los genes de nuestro azoxy-BGC.

Como perspectivas a esta parte del proyecto, queda reproducir la estrategia experimental que siguieron los autores del artículo del azoxyalqueno (mutaciones dirigidas, TLC, HPLC -usando una columna COSMOSIL-, ESI-CID-MS/MS y NMR).

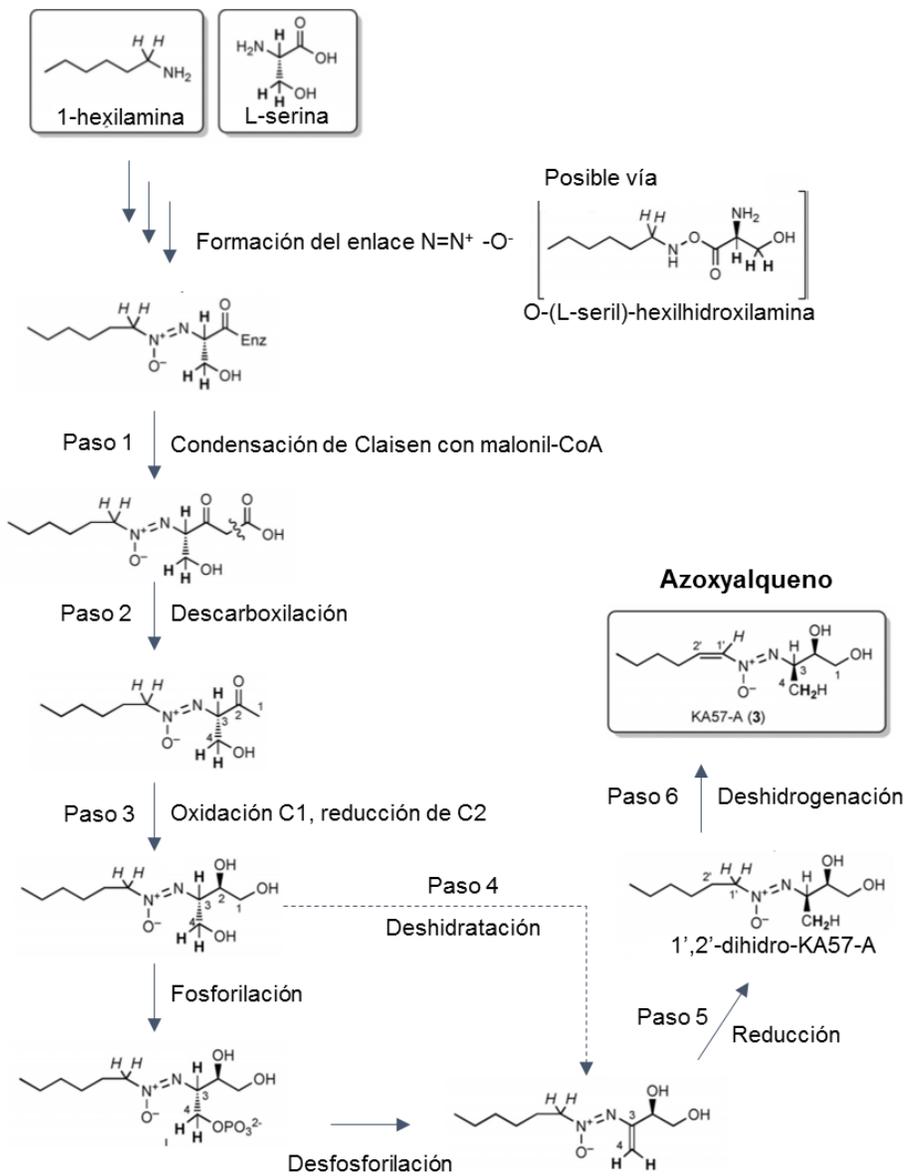


Figura 31. Propuesta de reacciones que conllevan a la producción de azoxyalqueno en *S. rochei* 7434AN4. Modificado de (Kunitake, Hiramatsu, Kinashi, & Arakawa, 2015)

## Capítulo VI. Conclusiones y perspectivas

### 6.1. Conclusiones

A pesar de sus grandes ventajas, la minería genómica convencional de NPs tiene enormes limitaciones para la predicción de nuevas clases de BGCs. Encontrar metabolitos con química nueva es de gran importancia hoy en día, debido al creciente aumento de patógenos resistentes a antibióticos, lo cual se ha convertido en una amenaza para el sistema de salud global. La implementación de nuevas estrategias de minería genómica es imprescindible en el campo de la investigación en NPs como fuentes potenciales para la producción de fármacos eficientes.

El presente proyecto de tesis sirvió para demostrar dos cosas importantes. En primer lugar, que es posible encontrar nuevas rutas de biosíntesis de NPs mediante análisis evolutivos de familias pertenecientes al metabolismo central usando la herramienta EvoMining. Y en segundo lugar, que particularmente las enzimas que utilizan tRNA (tRUEs) –enzimas que habían sido asociadas únicamente con rutas centrales como la síntesis de proteínas, peptidoglicanos y porfirinas, aminoacilación de lípidos y degradación de proteínas- también pueden ser elementos importantes dentro de rutas de biosíntesis de NPs y conducir al descubrimiento de nuevas clases de BGCs.

En este análisis tipo EvoMining enfocado a tRUEs, descubrimos también que las *Actinobacterias* tienen grandes tesoros ocultos en forma de BGCs en sus genomas. A pesar de que algunas bacterias de este filo han sido exploradas

ampliamente por expertos alrededor del mundo, como es el caso de *S. coelicolor* y *S. lividans*, aún es posible encontrar nuevos BGCs (azoxy-BGC) usando estrategias como EvoMining.

EvoMining es una herramienta versátil que puede adaptarse al estudio de un sinnúmero de combinaciones de bases de datos (rutas centrales, genomas, BGCs), lo cual la posiciona como una de las estrategias con un gran potencial para la predicción de nuevos BGCs de NPs. El creciente aumento de genomas disponibles, rutas biosintéticas de NPs caracterizadas y mejores herramientas de anotación, las posibilidades son infinitas. Basta con construir bases de datos (CE-DB, G-DB y BGCs-DB) de buena calidad y EvoMining hará el resto. Teniendo todas esas predicciones, es mucho más sencillo priorizar y diseñar experimentos para aislar nuevos metabolitos.

Los clústeres biosintéticos crípticos predichos por EvoMining son realmente interesantes por el tipo de enzimas que los conforman, la estructura peculiar que pueden tener y la posible novedad química estructural de su molécula producto. Sin embargo, plantear la estrategia para expresar BGCs crípticos es uno de los mayores retos, ya que las posibilidades de encontrar las condiciones de expresión ideales son realmente bajas en condiciones de laboratorio. El organismo productor contiene estos clústeres biosintéticos codificados en su genoma por razones muy específicas y responden a estímulos muy específicos. Hoy en día, se conocen relativamente pocas maneras de inducir BGCs crípticos. Nosotros probamos con la construcción de mutantes del azoxy-clúster, crecimiento de las cepas en diferentes fuentes nutritivas, incluso variando la composición de fósforo –uno de

los pocos inductores del metabolismo especializado conocido en bacterias-. Sin embargo, el azoxy-clúster parece estar respondiendo a otro tipo de estímulos que no fuimos capaces de identificar en el presente proyecto de tesis.

## 6.2. Perspectivas

Durante la realización del presente proyecto, surgieron diversas perspectivas al momento de analizar nuestros resultados, entre ellas:

- a. Optimizar los parámetros de la identificación de homólogos para evitar en la medida de lo posible los falsos-positivos.
- b. No descartar las familias de enzimas expandidas que no fue posible asociar a genes de BGCs conocidos y curados (Tabla 12). Es muy probable que estemos perdiendo valiosa información.
- c. Extender el análisis tipo EvoMining centrado a las familias de tRUEs hacia otros filos bacterianos, *e.g. Cyanobacteria* o *Proteobacteria*, u otros dominios *e.g. Archaea*.
- d. Examinar detalladamente las predicciones encontradas con EvoMining y priorizar un par de casos para su validación experimental.
- e. Reproducir los experimentos que se llevaron a cabo para la expresión del compuesto azoxyalqueno. Es probable que nuestro azoxy-BGCs sea responsable de una molécula del mismo tipo.
- f. Mejorar el modo de ejecución y visualización de resultados de EvoMining, de manera que cualquier persona en cualquier lugar del mundo sea capaz de aplicar esta tecnología a sus datos.

## Capítulo VII. Referencias y Apéndice

### Referencias

- Altschul, S., Gish, W., Miller, W., Mayers, E., & Lipman, D. (1990). Basic local alignment search tool. *J Mol Biol*, 215(3), 403-10.
- Ansari, M., Yadav, G., Gokhale, R., & Mohanty, D. (2004). NRPS/PKS: a knowledge-based resource of analysis of NRPS/PKS megasynthases. *Nucleic Acids Res*, 32, W405-13.
- Arendt, W., Hebecker, S., Jäger, S., Nimtz, M., & Moser, J. (2012). Resistance phenotypes mediated by aminoacyl-phosphatidylglycerol synthetases. *J Bacteriol*, 194(6), 1401-16.
- Asit, K. (2016). Multi-drug resistance genes in bacteria and 21st century problems associated with antibiotic therapy. *Biotechnol Ind J*, 12(7), 113.
- Aziz, R., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., . . . Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, 9, 75.
- Aziz, R., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., . . . Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, 9, 75.
- Baba, T., & Schneewind, O. (1998). Instruments of microbial warfare: bacteriocin synthesis, toxicity and immunity. *Trends Microbiol*, 6, 66-71.

- Bachmair, A., Finley, D., & Varshavsky, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. *Science*, 234(4773), 179-86.
- Barka, E., Vasta, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H., . . . van Wezel, G. (2015). Taxonomy, physiology and natural products of Actinobacteria. *Microbiol Mol Biol Rev*, 80(1), 1-43.
- Barona-Gómez, F., Lautru, S., Francou, F., Leblond, P., Pernodet, J., & Challis, G. (2006). Multiple biosynthetic and uptake systems mediate siderophore-dependent iron acquisition in *Streptomyces coelicolor* A3(2) and *Streptomyces ambofaciens* ATCC 23877. *Microbiology*, 152(Pt 11), 3355-66.
- Battesti, A., & Bouveret, E. (2006). Acyl carrier protein/SpoT interaction, the switch linking SpoT-dependent stress response to fatty acid metabolism. *Mol Microbiol*, 62, 1048-63.
- Belin, P., Mireille, M., Lautru, S., Seguin, J., Pernodet, J.-L., & Gondry, M. (2012). The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways. *Nat Prod Rep*, 29, 961-979.
- Bentley, S., Chater, K., Cerdeño-Tárraga, A., Challis, G., Thomson, N., James, K., . . . Hopwood, D. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 417(6885), 141-7.
- Caetano-Anollés, G., Yafremava, L., Gee, H., Caetano-Anollés, D., Kim, H., & Mittenthal, J. (2009). The origin and evolution of modern metabolism. *Int J Biochem Cell Biol*, 41(2), 285-97.

- CDC. (17 de Julio de 2014). *Centers for Disease Control and Prevention*. Recuperado el 2016 de Septiembre de 2016, de <http://www.cdc.gov/drugresistance/threat-report-2013/pdf/ar-threats-2013-508.pdf>
- Cerdeño, A., Bibb, M., & Challis, G. (2001). Analysis of the prodiginine biosynthesis gene cluster of *Streptomyces coelicolor* A3(2): new mechanisms for chain initiation and termination in modular multienzymes. *Chem Biol*, 8(8), 817-29.
- Chakraborty, R., & Bibb, M. (1997). The ppGpp synthetase gene (*relA*) of *Streptomyces coelicolor* A3(2) plays a conditional role in antibiotic production and morphological differentiation. *J Bacteriol*, 179, 5854-61.
- Challis, G., & Naismith, J. (2004). Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol*, 14(6), 748-56.
- Cimermancic, P., Medema, M., Claesen, J., Kurita, K., Wieland Brown, L., Mavrommatis, K., . . . Fischbach, M. (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*, 158(2), 412-21.
- Cortes, J., Haydock, S., Roberts, G., Bevitt, D., & Leadlay, P. (1990). An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature*, 348(6297), 176-8.
- Cotter, P., Hill, C., & Ross, R. (2005). Bacteriocins: developing innate immunity for food. *Nat Rev Microbiol*, 3, 777-88.

- Court, D., Sawitzke, J., & Thomason, L. (2002). Genetic engineering using homologous recombination. *Annu Rev Genet*, 51(6), 361-88.
- Cragg , G., & Newman, D. (2005). Biodiversity: A continuing source of novel drug leads. *Pure Appl Chem*, 77, 7-24.
- Cruz-Morales, P., Kopp, J., Martínez-Guerrero, C., Yáñez-Guerra, L., Selem-Mojica, N., Ramos-Aboites, H., . . . Barona-Gómez, F. (2016). Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model Streptomyces. *Genome Biol Evol*, 8(6), 1906-16.
- Davies, J. (1990). What are antibiotics? Archaic functions for modern activities. *Mol Microbiol*, 4, 1227-1232.
- de Jong, A., van Hijum, S., Bijlsma, J., Kok, J., & Kuipers, O. (2010). BAGEL2: Mining for bacteriocins in genomic data. *Nucleic Acids Res*, 38, W647-51.
- Dias, D., Urban, S., & Roessner, U. (2012). A historical overview of natural products in drug discovery. *Metabolites*, 2(2), 303-336.
- Donadio, S., Staver, M., McAlpine, J., Swanson, S., & Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. *Science*, 252(5006), 675-9.
- Fernández-Martínez, L., Gomez-Escribano, J., & Bibb, M. (2015). A relA-dependent regulatory cascade for auto-induction of microbisporicin production in *Microbispora corallina*. *Mol Microbiol*, 97(3), 502-14.

- Finn, R., Clements, J., & Eddy, S. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*, W29-37.
- Francklyn, C., & Minajigi, A. (2010). tRNA as an active chemical scaffold for diverse chemical transformations. *FEBS Lett*, 584(2), 366-75.
- Fujishima, K., & Kanai, A. (2014). tRNA gene diversity in the three domains of life. *Front Genet*, 5, 142.
- Gerlt, J., & Babbitt, P. (2001). Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem*, 70, 209-46.
- Giegé, R., Jühling, F., Pütz, J., Stadler, P., Sauter, C., & Florentz, C. (2012). Structure of transfer RNAs: similarity and variability. *Wiley Interdiscip Rev RNA*, 3(1), 37-61.
- Goh, E., Yim, G., Tsui, W., McClure, J., Surette, M., & Davies, J. (2002). Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proc Natl Acad Sci U S A*, 99(26), 17025-30.
- Gomez-Escribano, J., & Bibb, J. (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol*, 4(2), 207-15.
- Gondry, M., Sauquet, L., Belin, P., Thai, R., Amouroux, R., Tellier, C., . . . Pernodet, J. (2009). Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. *Nat Chem Biol*, 5(6), 414-20.

- Grag, R., Ma, Y., Hoyt, J., & Parry, R. (2002). Molecular characterization and analysis of the biosynthetic gene cluster for the azoxy antibiotic valanimycin. *Mol Microbiol*, 46(2), 505-17.
- Grag, R., Qian, X., Alemany, L., Moran, S., & Parry, R. (2008). Investigations of valanimycin biosynthesis: elucidation of the role of seryl-tRNA. *Proc Natl Acad Sci U S A*, 105(18), 6543-7.
- Grasemann, H., Stehling, F., Brunar, H., Widmann, R., Laliberte, T., & et.al. (2007). Inhalation of Moli1901 in patients with cystic fibrosis. *Chest*, 131, 1461-66.
- Gust, B., Challis, G., Fowler, K., Kieser, T., & Chater, K. (2003). PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A*, 100(4), 1541-6.
- Gust, B., Chandra, G., Jakimowicz, D., Yuding, T., Bruton, C., & Chater, K. (2004). Lambda red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Adv Appl Microbiol*, 54, 107-28.
- Gust, B., Kieser, T., & Chater, K. (2002). *REDIRECT technology: PCR-targeting system in Streptomyces coelicolor*. Colney, Norwich, United Kingdom: John Innes Center, Norwich Research Park.
- Haseltine, W., & Block, R. (1973). Synthesis of guanosine tetra- and pentaphosphate requires the presence of a codon-specific, uncharged transfer ribonucleic acid in the acceptor site of ribosomes. *Proc Natl Acad Sci U S A*, 70, 2564-68.

- Herrmann, J., Lukezic, T., Kling, A., Baumann, S., Hüttel, S., Petkovic, H., & Müller, R. (2016). Strategies for the discovery and development of new antibiotics for natural products: three case studies. *Curr Top Microbiol Immunol*, 398, 339-63.
- Hertweck, C. (2009). The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl*, 48(26), 4688-716.
- Hesketh, A., Chen, W., Ryding, J., Chang, S., & Bibb, M. (2007). The global role of ppGpp synthesis in morphological differentiation and antibiotic production in *Streptomyces coelicolor* A3(2). *Genome Biol*, 8, R161.
- Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F., . . . Micklefield, J. (2002). Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem Biol*, 9(11), 1175-87.
- Holley, R., Apgar, J., Everett, G., Madison, J., Marquisee, M., Merrill, S., . . . Zamir, A. (1965). Structure of a ribonucleic acid. *Science*, 147(3664), 1462-5.
- Hurto, R. (2011). Unexpected functions of tRNA and tRNA processing enzymes. *Adv Exp Med Biol*, 722, 137-55.
- Jahn, D., Verkamp, E., & Söll, D. (1992). Glutamyl-transfer RNA: a precursor of heme and chlorophyll biosynthesis. *Trends Biochem Sci*, 17(6), 215-8.
- Jenke-Kodama, H., Müller, R., & Dittmann, E. (2008). Evolutionary mechanisms underlying secondary metabolite diversity. In F. Petersen, & R. Amstutz,

*Natural compounds as drugs, Volume I* (págs. 121-40). Berlin, Alemania: Prog. Drug. Res.

Jenke-Kodama, H., Sandmann, A., Müller, R., & Dittmann, E. (2005). Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol*, 22(10), 2027-39.

Katz, L., & Baltz, R. (2016). Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.*, [Epub ahead of print].

Kelley, L., Mezulis, S., Yates, C., Wass, M., & Sternberg, M. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*, 10(6), 845-58.

Khalidi, N., Seifuddin, F., Turner, G., Haft, D., Nierman, W., Wolfe, K., . . . }. (2010). SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet Biol*, 47(9), 736-41.

Khosla, C., & Keasling, D. (2003). Metabolic engineering for drug discovery and development. *Nat Rev Drug Discov*, 2(12), 1019-25.

Kim, S., Quigley, G., Suddath, F., McPherson, A., Sneden, D., Kim, J., . . . Rich, A. (1973). Three-dimensional structure of yeast phenylalanine transfer tRNA: folding of the polynucleotide chain. *Science*, 147, 285-88.

Knerr, P., & van der Donk, W. (2012). Discovery, biosynthesis, and engineering of lantipeptides. *Annu Rev Biochem*, 81, 479-505.

- Kunitake, H., Hiramatsu, T., Kinashi, H., & Arakawa, K. (2015). Isolation and biosynthesis of an azoxyalkene compound produced by a multiple gene disruptant of *Streptomyces rochei*. *Chembiochem*, 16(15), 2237-43.
- Li, M., Ung, P., Zajkowski, J., Garneau-Tsodikova, S., & Sherman, D. (2009). Automated genome mining for natural products. *BMC Bioinformatics*, 10, 185.
- López, D., Fischbach, M., Chu, F., Losick, R., & Kolter, R. (2009). Structurally diverse natural products that cause potassium leakage trigger multicellularity in *Bacillus subtilis*. *PNAS*, 106(1), 280-285.
- Lubelski, J., Rink, R., Khusainov, R., Moll, G., & Kuipers, O. (2008). Biosynthesis, immunity, regulation, mode of action and engineering of the model lantibiotic nisin. *Cell Mol Life Sci*, 65, 455-76.
- Malik, V. (1980). Microbial secondary metabolism. *Trends Biochem Sci*, 5, 68-72.
- Malpartida, F., & Hopwood, D. (1984). Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. *Nature*, 309(5967), 462-4.
- McIntosh, J., Donia, M., & Schmit, E. (2009). Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. *Nat Prod Rep*, 26, 537-59.
- Medema, M., & Fischbach, M. (2015). Computational approaches to natural product discovery. *Nat Chem Biol*, 11(9), 639-48.

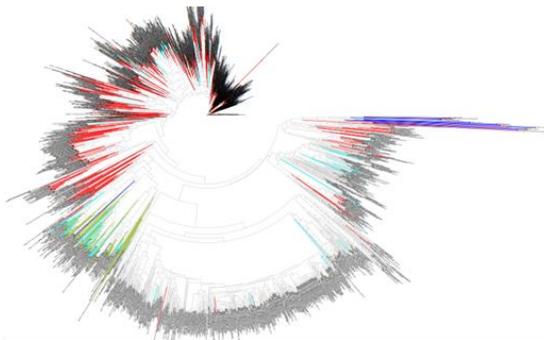
- Medema, M., Blin, K., Cimermanovic, P., de Jager, V., Zakrzewski, P., Fischbach, M., . . . Breitling, R. (2011). antiSMASH: Rapid Identification, Annotation and Analysis of Secondary Metabolite Biosynthesis Gene Clusters in bacterial and fungal genome sequences. *Nucleic Acids Res*, *39*, W339-46.
- Medema, M., Cimermanovic, P., Sali, A., Takano, E., & Fischbach, M. (2014). A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. *PLoS Comput Biol*, *10*(12), e1004016.
- Medema, M., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J., Blin, K., . . . Gomez-Escrib. (2015). Minimum information about a biosynthetic gene cluster. *Nat Chem Biol*, *11*(9), 625-31.
- Nett, M., Ikeda, H., & Moore, B. (2009). Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep*, *26*(11), 1362-84.
- Nolan, E., & Walsh, C. (2009). How nature morphs peptide scaffolds into antibiotics. *ChemBioChem*, *10*(5), 34-53.
- Ortega, M., Hao, Y., Zhang, Q., Walker, M., van der Donk, W., & Nair, S. (2015). Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature*, *517*(7535), 509-12.
- Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G., Papanikolaou, N., Kotoulas, G., . . . Iliopoulos, I. (2015). Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinform Biol Insights*, *9*, 75-88.

- Partida-Martínez, L., & Hertweck, C. (2005). Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature*, *437*, 884-88.
- Poteete, A. (2001). What makes the bacteriophage lambda Red system useful for genetic engineering: Molecular mechanism and biological function. *FEMS Microbiol Lett*, *201*(1), 9-14.
- Rigali, S., Nothhaft, H., Noens, E., Schlicht, M., Colson, S., Muller, M., . . . van Wezel, G. (2006). The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. *Mol Microbiol*, *61*, 1237-51.
- Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A., Hopwood, D., & van Wezel, G. (2008). Feast or famine: the global regulator DasR links nutrient stress to antibiotic production in *Streptomyces*. *EMBO Rep*, *9*, 670-75.
- Rother, M., & Krzycki, J. (2010). Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea. *Archaea*, pii: 453642.
- Rudd, B., & Hopwood, D. (1979). Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). *J Gen Microbiol*, *114*(1), 35-43.
- Sauerwald, A., Zhu, W., Major, T., Roy, H., Palioura, S., Jahn, D., . . . Söll, D. (2005). RNA-dependent cysteine biosynthesis in archaea. *Science*, *307*(5717), 1969-72.

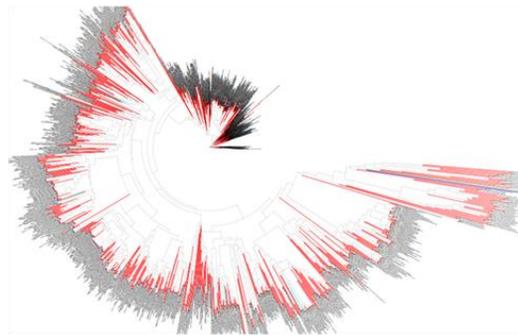
- Schepred, J., & Ibba, M. (2013). Direction of aminoacylated transfer RNAs into antibiotic synthesis and peptidoglycan-mediated antibiotic resistance. *FEMS Lett*, 587(18), 2895-904.
- Sedgwick, T., & Dawson, M. (2009). Novacta targets *C. difficile* the natural way. *MedNows*, 6, 8-9.
- Starcevic, A., Zucko, J., Simunkovic, J., Long, P., Cullum, J., & Hranueli, D. (2008). ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Res*, 36(21), 6882-92.
- Takeda, I., Umemura, M., Koike, H., Asai, K., & Machida, M. (2014). Motif-independent prediction of a secondary metabolism gene cluster using comparative genomics: application to sequenced genomes of *Aspergillus* and ten other filamentous fungal species. *DNA Res*, 21(4), 447-57.
- Tunan, J., Weber, J., Staver, M., Leung, J., Donadio, S., & Katz, L. (1990). Cloning of genes involved in erythromycin biosynthesis from *Saccharopolyspora erythraea* using a novel actinomycete-*Escherichia coli* cosmid. *Gene*, 90(1), 21-29.
- van der Lee, T., & Medema, M. (2016). Computational strategies for genome-based natural product discovery and engineering fungi. *Fungal Genet Biol*, 89, 29-36.
- Vining, L. (1992). Secondary metabolism, inventive evolution and biochemical diversity - a review. *Gene*, 115(1-2), 135-40.

- Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H., Bruccoleri, R., . . . Medema, M. (2015). antiSMASH 3.0 - a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res*, *43*(W1), W273-43.
- Weber, T., Rausch, C., López, P., Hoof, I., Gaykova, V., Huson, D., & Wohleben, W. (2009). CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite gene clusters. *J Biotechnol*, *140*(1-2), 13-17.
- Yim, G., Wang, H., & Davies, J. (2007). Antibiotics as signalling molecules. *Philos Trans R Soc Lond B Biol Sci*, *362*(1483), 1195-200.
- Zähner, H., & Anke, T. (1983). Evolution of secondary pathways. En J. Bennet, & E. Ciegler, *Differentiation and secondary metabolism in fungi* (págs. 153-171). New York: Marcel Dekker.
- Zhang, W., Ostash, B., & Walsh, C. (2010). Identification of the biosynthetic gene cluster for the pacidamycin group of peptidyl nucleoside antibiotics. *Proc Natl Acad Sci U S A*, *107*(39), 16828-33.
- Ziemert, N., Podell, S., Penn, K., Badger, J., Allen, E., & Jensen, P. (2012). The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE*, *7*(3), e34064.

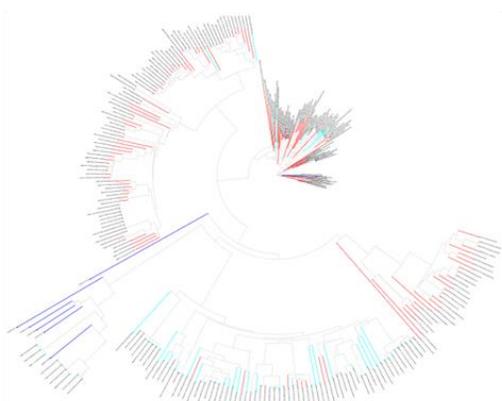
**Apéndice. Reconstrucciones filogenéticas de tRUEs expandidas y reclutadas por BGCs de NPs.**



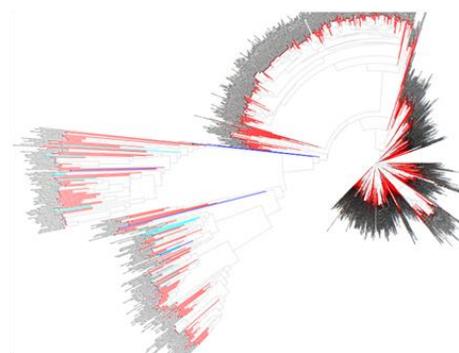
Aminotransferasa dependiente de PLP



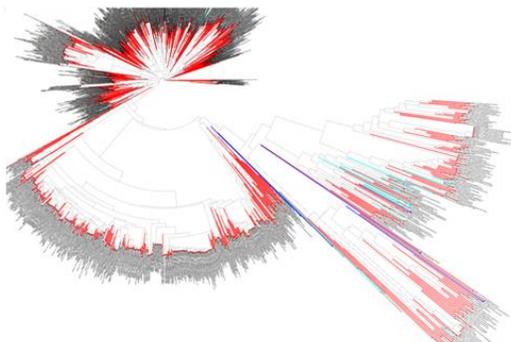
Asp/glu-tRNA amidotransferasa sub. B



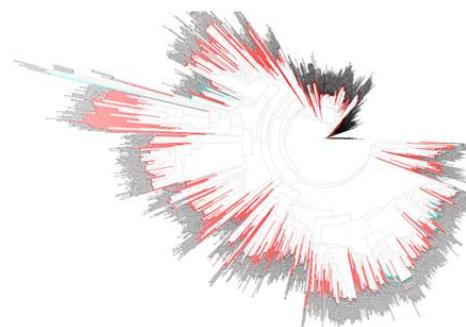
Asparaginil-tRNA sintetasa



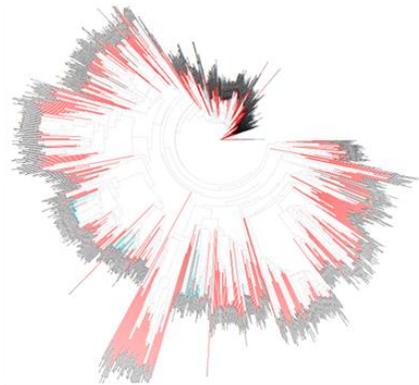
Aspartil-tRNA sintetasa



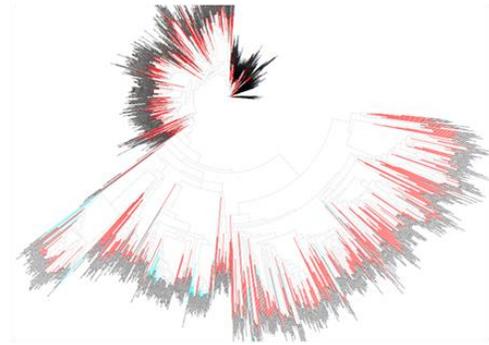
Aspartil-tRNA sintetasa no-discriminante



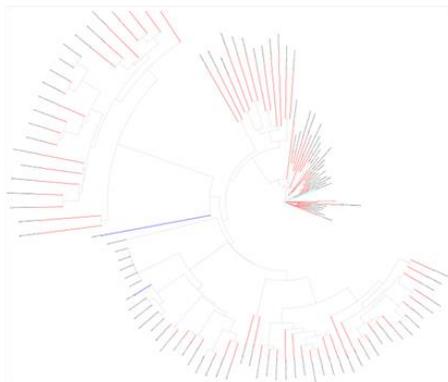
Fenilalanil-tRNA sintetasa sub. A



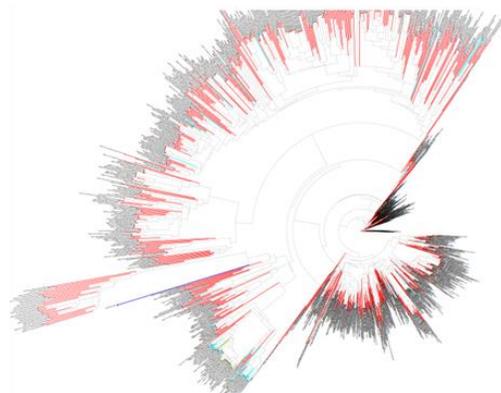
Fenilalanil-tRNA sintetasa sub. B



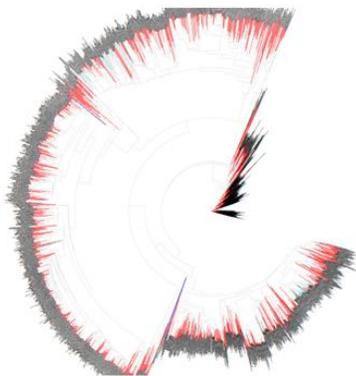
Glutamil-tRNA sintetasa



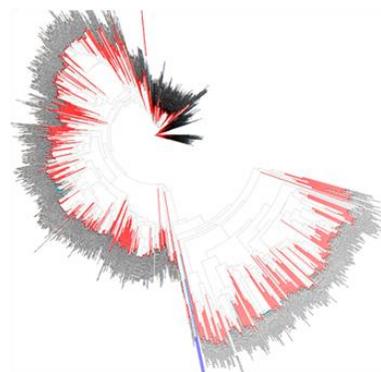
Glutamil-tRNA sintetasa no-discriminante



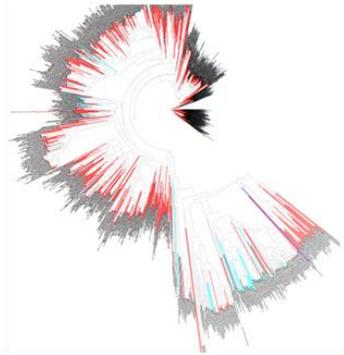
Glutaminil-tRNA sintetasa



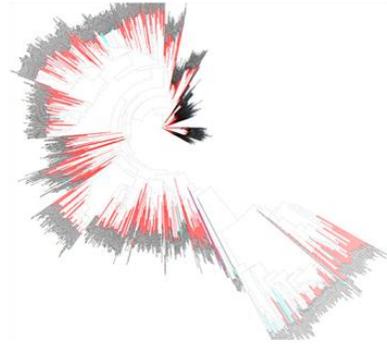
Isoleucil-tRNA sintetasa



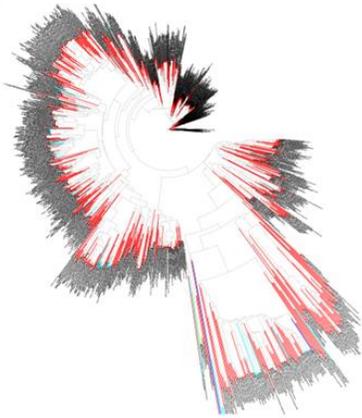
Leucil-tRNA sintetasa



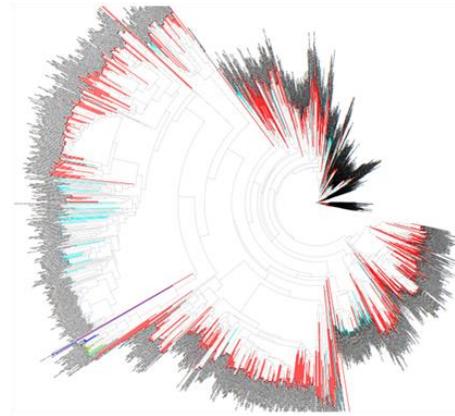
Seril-tRNA sintetasa



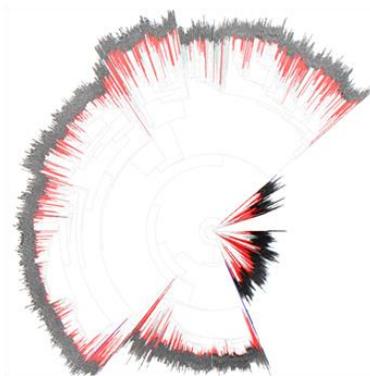
Tirosil-tRNA sintetasa



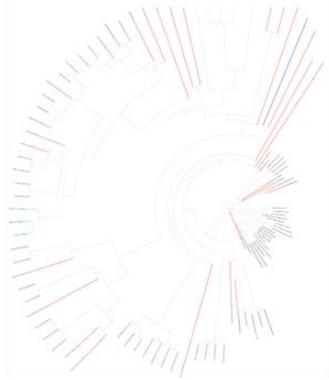
Treonil-tRNA sintetasa



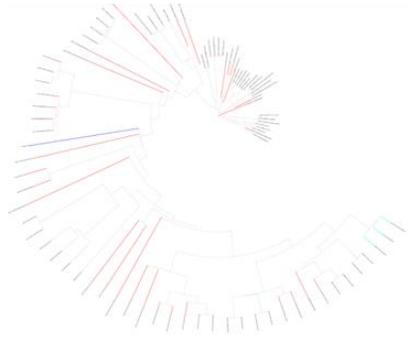
Triptofanil-tRNA sintetasa



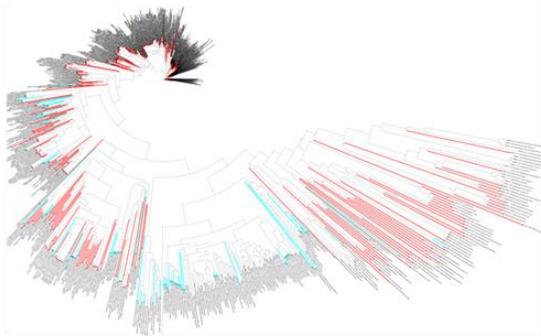
Valil-tRNA sintetasa



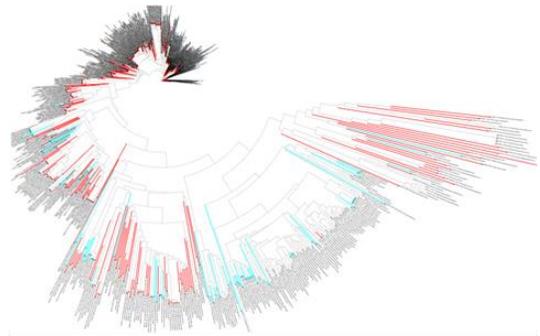
FemA



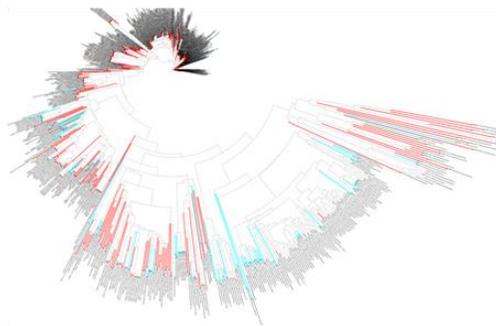
FemAB



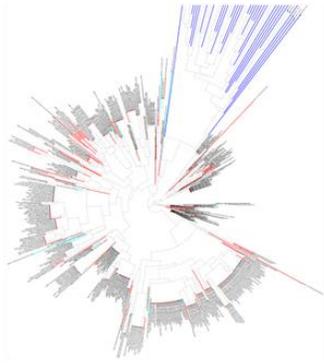
FemX



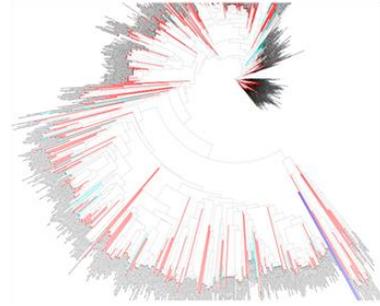
FemY



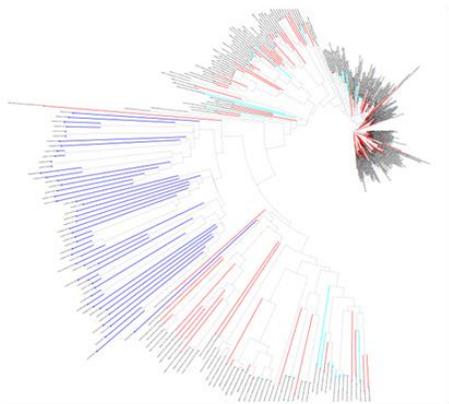
VanK



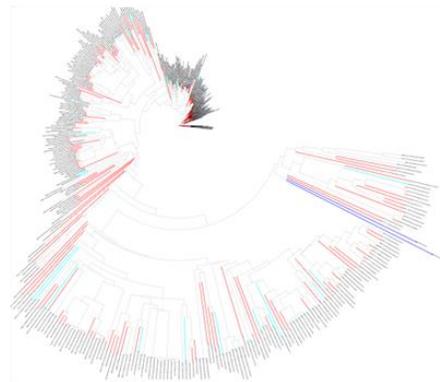
Alanil-diacilglicerol sintetasa



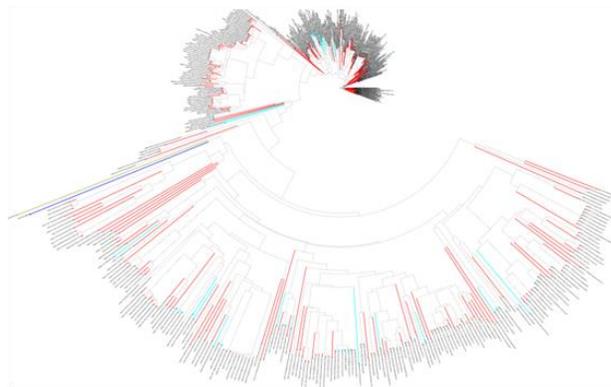
Alanil-fosfatidilglicerol sintetasa



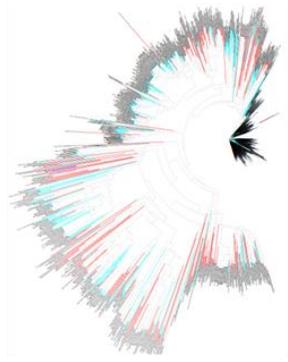
RAK-fosfatidilglicerol sintetasa



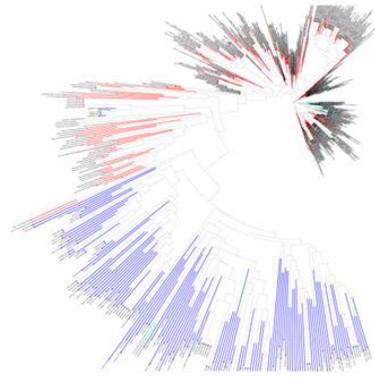
MprF1



MprF2



Glutamyl-tRNA reductasa



Queuina-tRNA ribosiltransferasa

## ***Currículum vitae***

### ***Datos personales***

Paulina Mayell Mejía Ponce - 27 años - paumayell@gmail.com

### ***Educación***

Ingeniería Bioquímica con especialidad en Biotecnología

Instituto Tecnológico de Celaya

Obtención de grado: Diciembre de 2013

Maestría en Ciencias en Biología Integrativa

Centro de Investigación y Estudios Avanzados, Campus Irapuato.

Obtención de grado: Mayo de 2017

### ***Experiencia profesional adquirida durante los estudios de licenciatura***

Estancia de investigación bajo la dirección de la Dra. María Concepción Lora Vilchis en el Laboratorio de Biotecnología con Microalgas, CIBNOR - La Paz, BCS  
Junio 2011 – agosto 2011

Estancia de investigación bajo la dirección del Dr. Alexander de Luna Fors en el Laboratorio de Biología de Sistemas, UGA-CINVESTAV, Irapuato, Gto.  
Junio 2012 – agosto 2012

Residencias y tesis profesional de licenciatura bajo la dirección del Dr. Gerardo Acosta García en el Laboratorio de Biología Molecular del Instituto Tecnológico de Celaya, Celaya, Gto. Septiembre 2012 – Diciembre 2013

Participación con presentación oral en el Congreso Delfín, agosto 2011.

Curso “Control de calidad del resultado de una medición” (UMSNH, febrero 2011)

Taller “Citogenética” (Universidad Autónoma Metropolitana, septiembre 2012)

Taller “Calidad semillas” (Instituto Tecnológico de Celaya, mayo 2013)

Curso “Herramientas Bioinformáticas” (Global Agronomics, septiembre 2013)

### ***Experiencia profesional adquirida durante los estudios de maestría***

Tesis profesional de maestría bajo la dirección del Dr. Francisco Barona Gómez en el Laboratorio de Evolución de la Diversidad Metabólica, UGA-CINVESTAV, Irapuato, Gto. Septiembre 2014 – mayo de 2017 (Presente trabajo de tesis)

Estancia de investigación bajo la dirección del Dr. Kapil Tahlan en el Departamento de Biología, Memorial University of Newfoundland, NL, Canadá. Diciembre 2016 (Presente trabajo de tesis)

Participación con presentación oral del presente trabajo de tesis en el 40° Congreso Nacional de Microbiología, abril 2017.

Participación con presentación oral del presente trabajo de tesis en el 50° Congreso Mexicano de Química, octubre 2015.

Curso “NOM-005, Manejo de sustancias y residuos peligrosos” (abril 2017)

Taller “Analysis of gene families identifies evolutionary processes associated with speciation” (CONABIO – marzo 2017)

Curso “Biosafety” (Memorial University of Newfoundland – diciembre 2016)

Taller “Ensamblaje de genomas” (Langebio-CINVESTAV – julio 2016)

Curso “Escritura de artículos científicos” (CINVESTAV – junio 2016)

### ***Publicaciones derivadas durante los estudios de maestría***

Derivado del presente trabajo de tesis:

Paulina M. Mejía-Ponce, César Aguilar, Nelly Selem-Mojica, Hilda E. Ramos-Aboites, Pablo Cruz-Morales and Francisco Barona-Gómez. An exploration of tRNA-utilizing enzymes using EvoMining reveals novel BGCs in *Actinobacteria*. (*En preparación*).

En colaboración con otros proyectos:

Cruz Morales, Pablo; Ramos-Aboites, Hilda; Licona-Cassani, Cuauhtémoc; Selem-Mójica, Nelly; **Mejía-Ponce, Paulina M**; Souza, Valeria; Barona-Gómez, Francisco. Actinobacteria phylogenomics, selective isolation from an iron oligotrophic environment and siderophore functional characterization, unveil novel desferrioxamine molecular ecological and functional traits. *FEMS Microbiol Ecol* (*En revisión con los editores*).

## **Declaración de independencia**

Por este medio declaro que yo he escrito y preparado este trabajo de tesis de forma independiente y sin ayuda externa. Especialmente declaro que he citado de forma correcta y explícita a los autores y trabajos en los que esta tesis se apoya, así como las contribuciones de las personas que coadyuvaron a su desarrollo.

Irapuato, Guanajuato a 26 de mayo de 2017

Paulina Mayell Mejía Ponce

## Índice de figuras

Figura 1. Línea del tiempo de las principales clases de NPs descubiertas durante la era dorada de los antibióticos. ....	8
Figura 2. . Cronología de los eventos clave de resistencia a antibióticos. ....	10
Figura 3. Decaimiento de la producción y aprobación de nuevos fármacos. ....	11
Figura 4. Morfología de algunos ejemplos de actinobacterias. ....	13
Figura 5. Representación del ciclo de vida del género <i>Streptomyces</i> . ....	15
Figura 6. Línea del tiempo que muestra el origen de la minería genómica. ....	23
Figura 7. Estructura general de un clúster de genes biosintéticos (BGC) de NPs en bacterias. ....	24
Figura 8. Estructuras químicas de tres ejemplos de policétidos derivados de microorganismos. ....	26
Figura 9. Estructura y organización general de las PKSs, ejemplo de la eritromicina. ....	28
Figura 10. Estructuras químicas de tres ejemplos de péptidos no-ribosomales. ....	30
Figura 11. Estructura y organización general de las NRPs, ejemplo de la isopenicilina. .	31
Figura 12. Estructuras químicas de tres ejemplos de terpenos. ....	32
Figura 13. Estructura química característica de los lantipéptidos. ....	34
Figura 14. Esquema general de la biosíntesis de los RIPPs. ....	35
Figura 15. Clasificación de lantipéptidos de acuerdo al tipo de enzimas de modificación post-traducciona que conforman el BGC. ....	36
Figura 16. Estructura secundaria (modelo 'trébol') y terciaria (modelo 'L') del tRNA <sup>Ala</sup> ...	46
Figura 17. Directrices de los aminoacil-tRNAs dentro de la célula. ....	48
Figura 18. Tubería bioinformática de EvoMining. ....	56
Figura 19. Sistema de transformación $\lambda$ -Red (REDIRECT-Technology). ....	65
Figura 20. Estrategia de siembra en placa para el análisis de fenotipo. ....	67

Figura 21. Muestra del mapa de calor representa las familias de tRUEs con expansiones enzimáticas en <i>Actinobacteria</i> .	86
Figura 22. Comportamiento de los eventos de expansión de las tRUEs y sus reclutamientos en BGCs de NPs.	95
Figura 23. Reconstrucción filogenética de la familia FemXAB y estructura del azoxy-BGC detectado en <i>S. coelicolor</i> y <i>S. lividans</i> .	102
Figura 24. Mecanismo de reacción de VImA correspondiente a la ruta de la valanimicina	103
Figura 25. Predicción de la posible estructura del producto final del azoxy-BGC	105
Figura 26. Diferencias fenotípicas entre mutantes y silvestres creciendo en medios ricos en nutrientes.	108
Figura 27. Diferencias fenotípicas entre mutantes y silvestres creciendo en medios mínimos	110
Figura 28. Cromatogramas de HPLC con picos diferenciales entre cepas mutantes y silvestres.	113
Figura 29. RT-PCR de los genes Sco1278 y Sco1287 de las cepas silvestres Sli66 y ScoM1154 creciendo en diferentes medios de cultivo.	115
Figura 30. Análisis de conservación del azoxy-BGC en actinobacterias usando CORASON.	116
Figura 31. Propuesta de reacciones que conllevan a la producción de azoxyalqueno en <i>S. rochei</i> 7434AN4.	118

## Índice de tablas

Tabla 1. Clasificación de los principales productos naturales .....	4
Tabla 2. Descripción general de los tipos de PKSs .....	29
Tabla 3. Bases de datos iniciales necesarias para ejecutar EvoMining .....	51
Tabla 4. Oligonucléotidos .....	63
Tabla 5. Bases de datos iniciales del análisis EvoMining enfocado a tRUEs en <i>Actinobacteria</i> .....	76
Tabla 6. Clasificación por familias de las tRUEs y ruta metabólica central al que pertenecen .....	78
Tabla 7. . Composición de la G-DB .....	82
Tabla 8. Clases de BGCs del repositorio MIBiG.....	84
Tabla 9. Número de homólogos y familias de tRUEs expandidas detectadas por EvoMining.....	88
Tabla 10. Comparación de las tRUEs con el mayor y menor número de homólogos, promedio $X$ , desviación estándar $\sigma$ , valor estadístico $(X + \sigma)$ y número de genomas con expansiones enzimáticas.....	90
Tabla 11. Familias tRUEs expandidas que están reclutadas por BGCs de NPs...	92
Tabla 12. Familias de tRUEs expandidas que no fueron asociadas a algún BGCs .....	97
Tabla 13. Características generales de la anotación funcional de los clados divergentes de las reconstrucciones filogenéticas de las tRUEs.....	100
Tabla 14. Anotación del azoxy-BGC con RAST y Phyre2 .....	104