



**UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO**

---

---

**FACULTAD DE QUÍMICA**

**“MINERÍA GENÓMICA DE SINTETASAS DE PÉPTIDOS NO  
RIBOSOMALES EN BACTERIAS ASOCIADAS A PLANTAS”**

*Tesis*

**QUE PARA OBTENER EL TÍTULO DE  
QUÍMICO**

**PRESENTA**

**Reynaldo Villanueva Enríquez**

**ASESORA PRINCIPAL**

**Dra. Corina-Diana Ceapă**

**CDMX**

**AÑO 2023**





Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**JURADO ASIGNADO:**

**PRESIDENTA:** Profesor: GARCIA VELAZQUEZ LIZBETH ESMERALDA

**VOCAL:** Profesor: MARTINEZ CABALLERO CAROL SISETH

**SECRETARIA:** Profesor: CEAPĂ CORINA-DIANA

**1er. SUPLENTE:** Profesor: ORTEGA VAZQUEZ ALBERTO

**2° SUPLENTE:** Profesor: PEDROZA GARCIA JOSE ANTONIO

**SITIO DONDE SE DESARROLLÓ EL TEMA: LABORATORIO DE MICROBIOLOGÍA DEL INSTITUTO DE QUÍMICA**

**ASESOR DEL TEMA:**

**DRA. CEAPĂ CORINA DIANA**



A handwritten signature in black ink, appearing to read 'Ceapă', is written above a horizontal line.

**SUSTENTANTE:**

**REYNALDO VILLANUEVA ENRÍQUEZ**



A handwritten signature in dark red ink, appearing to read 'RV', is written above a horizontal line.

**A mis padres, Consuelo y Gerardo,**  
por darme su amor y apoyo incondicional  
todos estos años.

**A mis hermanos, Eduardo y Gerardo,**  
por ser mis amigos, por acompañarme  
en esta aventura que es la vida y crecer junto a mí.

Sin ustedes este logro no habría sido posible.

¡Los amo!



## **Agradecimientos**

En esta sección quiero permitirme expresar lo agradecido que me encuentro con todas las personas que fueron parte de este proceso y que con ello me permitieron hacer posible este logro. Podría redactar otro libro completo para tratar de escribir las palabras que demuestran lo afortunado que me siento de haber contado con todos, sin embargo, solo se me permite escribir una sección, así que pido disculpas de antemano si es que olvidé a alguien al momento de escribirla.

Antes que a nadie quiero agradecer a mis padres, Gerardo y Consuelo, pues ellos me demostraron la primera prueba de amor que tuve al llegar a este mundo cuando me aceptaron como su hijo. Es el mismo amor que ellos me enseñaron, el que traté de incluir en mis estudios, en mi trabajo y sigo tratando de incluir en mi vida, espero lograr que estén orgullosos de mí.

A mis hermanos, Gerardo y Eduardo, que fueron mis primeros amigos y que son mis compañeros de vida, no veo como no relacionar todo lo que he logrado hasta hoy con cada cosa que aprendimos juntos, no habría elegido a nadie más para que ocupe su lugar.

A mis amigos, Luis, Julio, Alejandro y Tadeo, fueron la familia que elegí y me aceptaron dentro de la suya, me han aconsejado, me han apoyado y han estado en los mejores y en los más difíciles momentos de mi vida, espero seguir contando con su amistad y yo les ofrezco siempre la mía.

A Paola Stefanía, pues fue por tu influencia que tuve el valor y la madurez para tomar una de las decisiones más importantes de mi vida en un momento tan complicado, desde entonces, y hasta el día de hoy, me sigues inspirando y me incentivas a ser la mejor versión de mí.

A mis compañeros de la facultad, Lalo, Kean, David, Shaina, Itzel, Israel y Rodrigo, pues fue gracias a su apoyo y amistad, lo cual es muy difícil de encontrar en nuestro ambiente, que logré aprender tantas cosas nuevas en el mundo de la química y con ello terminar la licenciatura.

A mi tutora, la Dra. Corina, pues gracias a su enseñanza, su asesoría y la presión que puso constructivamente en mí, es que este trabajo se finalizó. Fue mi primera impresión en el mundo de la investigación y es gracias a su ejemplo que trato de construir mi propia imagen como científico.

A mis sinodales, por su interés, su apoyo y su paciencia en la realización de este trabajo, en él se ve reflejado no solo mi esfuerzo sino también el suyo.

A mis profesores, en especial a la Dra. Lizbeth García, fueron sus clases las que despertaron en mí el interés y el gusto por los temas que me llevaron a elegir y a disfrutar este como mi proyecto de tesis.

Adicionalmente quiero agradecer a las instituciones y programas que promovieron de una u otra manera la culminación de este proyecto:

- Investigación realizada gracias al Programa UNAM-PAPIIT, proyecto IA201721 y al CONACYT, proyecto CF-2022/319816, "Exploración de la "materia oscura microbiana" utilizando la genómica", parte de la convocatoria Paradigmas y Controversias de la Ciencia 2022.
- A la beca otorgada para "Licenciatura / Titulación" por el CONACYT, proyecto CF-2022/319816, "Exploración de la "materia oscura microbiana" utilizando la genómica", parte de la convocatoria Paradigmas y Controversias de la Ciencia 2022, en el mes de noviembre 2022.
- A la beca otorgada para "Licenciatura / Titulación", por el Programa UNAM-PAPIIT en el proyecto de IA 201721 en el periodo noviembre-diciembre 2021.
- Al Instituto de Química por el otorgamiento de la beca interna de licenciatura 2022 recibida en el periodo de mayo a octubre de 2022.
- A la Universidad Nacional Autónoma de México por el otorgamiento de la BECA DE TITULACIÓN-EGRESADOS ALTO RENDIMIENTO 2022-2023
- A la Universidad Nacional Autónoma de México (UNAM) y al Instituto de Química (IQ).

## Índice de Contenido

1. Abstract .....	6
2. Resumen.....	8
3. Introducción .....	10
3.1 Productos naturales y su importancia en el desarrollo de la civilización.....	10
3.2. Bacterias Asociadas a Plantas, una fuente de metabolitos especializados .....	13
3.3. Sintetasas de péptidos no ribosomales (SPNR).....	17
3.4. Minería genómica en el descubrimiento de nuevos metabolitos especializados .....	27
4. Justificación .....	35
5. Hipótesis.....	36
6. Objetivos .....	37
6.1. Objetivo general.....	37
6.2. Objetivos particulares .....	37
7. Metodología .....	38
7.1. Recolección de genomas de BAP .....	38
7.2. Selección de un dominio característico de las SPNR.....	39
7.3. Búsqueda por homología con BLAST.....	40
7.4. Minería genómica .....	40
7.5. Construcción de árboles filogenéticos de los genomas resultantes de BAP.....	40
7.6. Construcción de redes de similitud y agrupación de los CGB en FCG por BiG-SCAPE .....	41
7.7. Propuesta de candidato final .....	41
8. Discusión y resultados.....	42
8.1. Recolección de genomas de BAP .....	42
8.2. Selección de un dominio característico de las SPNR.....	42
8.3. Resultados de la búsqueda por homología con BLAST .....	48
8.4. Minería genómica .....	55
8.5. Análisis filogenéticos de los genomas resultantes de BAPs.....	58
8.6. Análisis de redes de similitud .....	64
8.7. Propuesta de un nuevo CGB de una SPNR para la producción de un nuevo compuesto bioactivo.....	72
9. Conclusiones y perspectivas.....	80
8. Referencias .....	82
9. Anexo.....	90

9.1 Tablas del Anexo .....	90
9.2 Figuras del Anexo .....	106
9.3 Secuencia 1.....	107
9.4 Eventos y distinciones en los que se participó con el presente trabajo .....	108

## Índice de Figuras

Figura 1. Esquema de biosíntesis de la surfactina. ....	26
Figura 2. Esquema general de la metodología.....	38
Figura 3. Reacción catalizada por la enterobactina sintasa. ....	44
Figura 4. CGB de MiBIG similares al CGB de la enterobactina .....	45
Figura 6. Especies de plantas de las cuales se aislaron las 214 BAP. ....	50
Figura 7. Regiones de donde se tomaron las muestras de las que se aislaron las 214 BAP. ....	54
Figura 8. Presencia promedio de cada uno de las clases de CGB de metabolitos especializados predichos por AntiSMASH.....	55
Figura 9. Presencia promedio de cada una de las clases de CGB de metabolitos especializados predichos por AntiSMASH por género bacteriano.....	56
Figura 10. Árbol de BAP Gram positivas.....	59
Figura 11. Árbol de BAP Gram negativas. ....	60
Figura 12. FCG resultado del análisis de redes de similitud.....	65
Figura 13. Mapa jerárquico con 13 de las FCG con mayor número de CGB. ....	66
Figura 14. Clan de Familias de Clústeres con un gen homólogo a EntF. ....	70
Figura 15. FCG número 100.....	72
Figura 16. FCG número 13.....	74
Figura 17. CGB predicho por AntiSMASH v6.1.1 de una SPNR identificada en el genoma de Azospirillum brasilense strain Az39.....	75
Figura 18. Composición de módulos y dominios del gen biosintético central del CGB propuesto.....	76
Figura 19. Composición de módulos y dominios del gen biosintético central del CGB propuesto.....	76
Figura 20. Composición de dominios que conforman el gen biosintético central del CGB propuesto predicho por PRISM.....	77
Figura 21. Esquema genético de la SPNR de la brabantamida.....	78
Figura 22. Estructura química del compuesto Brabantamida.....	78

Figura 23. Análisis de BLAST realizado en la base de datos de NCBI usando la secuencia del gen biosintético central del CGB propuesto. ....	79
Figura I. Árbol de bacterias del género Bacillus resultantes de la búsqueda por BLAST. ....	106
Figura II. Árbol de bacterias del género Pseudomonas resultantes de la búsqueda por BLAST. ....	107
Figura III. Constancia de participación en el simposio interno del Instituto de Química 2021. ....	109
Figura IV. Cartel presentado en el Simposio Interno del Instituto de Química 2021 .....	110
Figura V. Cartel presentado.....	111
Figura VI. Constancia de participación en XXXIII Congreso Nacional de Bioquímica. ....	112
Figura VII. Diapositiva presentada en el XXXIII Congreso Nacional de Bioquímica.....	113
Figura VIII. Constancia de participación en XXXIII Congreso Nacional de Bioquímica. ....	114
Figura IX. Constancia de participación en tercer Congreso virtual Latinoamericano de LatinXChem. ....	115
Figura X. Constancia de distinción de uno de los dos mejores trabajos en la categoría de licenciatura en el simposio interno del IQ 2022. ....	115

## Índice de Tablas

Tabla de abreviaturas.....	4
Tabla 1. Características de PN de origen en SPNR descubiertos desde el año 2000 .....	19
Tabla 2. Características del componente F de la enterobactina sintasa.....	43
Tabla 3. Características generales y farmacológicas de las plantas hospederas de BAPs con dominios C.....	51
Tabla I. Genomas de BAP empleados en este estudio. ....	90
Tabla II. Tabla de genomas de BAP con secuencias homólogas al dominio C resultado de un análisis de BLAST. ....	100

Tabla de abreviaturas

Abreviatura	Significado
A	Dominio de adenilación
AG	Ácido Graso
AIA	Ácido Indol-3-Acético
Ala	Alanina
AntiSMASH	Antibiotics and secondary metabolite analysis shell
Asp	Ácido Aspártico
BAP	Bacteria(s) Asociada(s) a Plantas
BiG-SCAPE	Biosynthetic Gene Similarity Clustering And Prospecting Engine
BV-BRC	Bacterial and Viral Bioinformatics Resource Center
C	Dominio de Condensación
CFC	Clan(es) de Familias de Clústeres
CGB	Clúster(es) de Genes Biosintéticos
Cl	Cloruro
CORASON	Core Analysis of Syntenic Orthologues to prioritize Natural product gene clusters
Dab	Modificación C terminal
Dhb	Dehidrobutirina
DKP	2,5-dicetopiperazina(s)
DNA	Ácido desoxirribonucleico
FCG	Familia(s) de Clústeres de Genes
GBC	Gen(es) Biosintético(s) Central(es)
DL	Aminoácido con cualquier configuración
GSC	Genoma Secuenciado Completamente
Gln	Glutamina
HPLC	Cromatografía líquida de alto rendimiento
Glu	Ácido glutámico
Gly	Glicina
IR	Espectroscopía de Infrarrojo
His	Histidina
MPA	Monofosfato de adenosina
MS	Espectroscopía de Masas
NCBI	National Center of Biotechnology Information
Hse	Modificación C terminal
pb	Pares de bases
PK	Policétido
PKS	Policétido Sintetasas
PN	Productos Naturales
PNR	Péptidos No Ribosomales
ppan	Cofactor 4'-fosfopanteteinilo
PPI	Pirofosfato inorgánico

Hy	Modificación C terminal
PsRMP	Péptido sintetizado Ribosomalmente y modificado Postraduccionalmente
PTP	Dominio de la proteína transportadora de péptidos
RAM	Resistencia Antimicrobiana
RMN	Resonancia Magnética Nuclear
Ile	Isoleucina
SPNR	Sintetasas de Péptidos no Ribosomales
TE	Dominio Tioesterasa
THG	Transferencia Horizontal de Genes
Leu	Leucina
UV	Espectroscopía de absorción Ultravioleta
Lys	Lisina
Orn	Ornitina
Pro	Prolina
Ser	Serina
Thr	Treonina
Val	Valina
Unk	Aminoácido desconocido o pudiera ser más de un aminoácido
xile	Modificación C terminal
Abu	Ácido-2-aminobutírico
Tyr	Tirosina
OH	Grupo Hidroxilo
Me	Grupo metilo
TS	Terpeno Sintetasa

## 1. Abstract

Plant-Associated Bacteria (PAB) have successfully adapted to eukaryotic systems through mechanisms such as protection, nutrient acquisition, pollutant degradation, phytohormones production, etc.

The establishment of symbiosis by bacteria are due to the so-called "specialized metabolites" which can be traced at genetic level to a Biosynthetic Gene Cluster (BGC) using genome mining, a technique that combines gene homology comparisons and bioinformatic analysis to explore the biosynthetic potential of microorganisms.

This project is aimed at making use of genome mining to propose Plant Associated Bacteria as one of the most important sources of specialized metabolites, especially Non-Ribosomal Peptides, whose activities are reported from toxins to antibiotic and anticancer compounds.

A database of 388 genomes of 57 bacterial genera was established on the BV-BRC platform (Bacterial and Viral Bioinformatics Resource Center).

The sequence of the central condensation domain of the *enterobactin synthase* was used to perform a homology search using the BLAST tool to identify genomes with domains characteristic of a Non-Ribosomal Peptide Synthetase (NRPS).

The resulting genomes were analyzed by genome mining using AntiSMASH and with that a profile of specialized metabolites by bacterial genus was constructed.

534 NRPS Biosynthetic Gene Clusters (BGC) were analyzed by a similarity network using BiG-SCAPE, which grouped the NRPS BGC into 85 Gene Cluster Families (GCF) and 169 orphan clusters.



The CORASON program was used to perform phylogenetic analysis of the GCF to find evolutive patterns within the genetic and domain composition of the BGC.

A presence-absence summary of the families of clusters and phylogenetic analysis resulted in the establishment of a profile of specialized metabolites in PAB genomes associated with phylogenetic clades.

NRPS were the most abundant metabolite class in the analyzed bacterial genera. It is believed to be due to their great structural and functional diversity, that makes them an adaptation tool in almost any ecosystem and for almost any organism.

For this type of BGC, patterns associated with the evolution of the bacteria were found and mechanisms of vertical and horizontal gene transfer were confirmed.

A BGC for each GCF found by BiG-SCAPE was compared in the BiG-FAM database, resulting in the discovery of 59 novel NRPS with potential bioactivity. Finally, a conserved cluster in the genomes of the bacterial genus *Azospirillum* was proposed for its production and bioactivity tests.

## 2. Resumen

Las bacterias asociadas a plantas (BAP) se han adaptado con éxito a los sistemas eucariontes a través de mecanismos como la protección, adquisición de nutrientes, degradación de contaminantes, producción de fitohormonas, etc.

El establecimiento de la simbiosis por parte de las bacterias se debe a los llamados "metabolitos especializados", que se pueden rastrear a nivel genético hasta los clústeres de genes biosintéticos (CGB) utilizando la minería genómica, una técnica que combina comparaciones por homología de genes y análisis bioinformáticos para explorar el potencial biosintético de los microorganismos.

Este proyecto tiene como objetivo hacer uso de la minería genómica para evaluar a las bacterias asociadas a plantas como una de las fuentes más importantes de metabolitos especializados, en especial a los Péptidos No Ribosomales, cuyas actividades cuentan con reportes que van desde toxinas hasta compuestos antibióticos y anticancerígenos.

Se estableció una base de datos de 388 genomas de 57 géneros bacterianos en la plataforma BV-BRC (Centro de Recursos de Bioinformáticos Bacteriana y Viral, por sus siglas en inglés).

La secuencia del dominio de condensación central de la *enterobactina sintasa* se utilizó para realizar una búsqueda de homología utilizando la herramienta BLAST para identificar genomas con dominios característicos de una Sintetasa de Péptidos No Ribosomales (SPNR).

Los genomas resultantes se analizaron mediante minería genómica utilizando AntiSMASH y con ello se construyó un perfil de metabolitos especializados por género bacteriano.

Se analizaron 534 CGB de NRPS mediante una red de similitud construida usando BiG-SCAPE, que agrupó los CGB de NRPS en 85 Familias de Clústeres de Genes (FCG) y 169 clústeres huérfanos.

El programa CORASON se utilizó para realizar un análisis filogenético del FCG para encontrar patrones evolutivos en la composición genética y de dominios de los CGB.

Un análisis de presencia-ausencia de las familias de clústeres en comparación con análisis filogenéticos dio como resultado el establecimiento de un perfil de metabolitos especializados en genomas de BAP ligado a sus clados filogenéticos.

Las SPNR fueron la clase de metabolitos más abundantes en los géneros bacterianos analizados. Se cree que es debido a su gran diversidad estructural y funcional, que las convierte en una herramienta de adaptación en casi cualquier ecosistema y para casi cualquier organismo.

Para este tipo de CGB se encontraron patrones asociados al desarrollo evolutivo de la bacteria y con ello se confirmaron mecanismos de transferencia vertical y horizontal de genes.

Se comparó un CGB por cada FCG encontrada por BiG-SCAPE en la base de datos de BiG-FAM, lo que resultó en el descubrimiento de 59 SPNR novedosas con potencial de bioactividad. Finalmente, se propuso un clúster conservado en los genomas del género bacteriano *Azospirillum* para su producción y pruebas de bioactividad.

### 3. Introducción

#### 3.1 Productos naturales y su importancia en el desarrollo de la civilización

*"Jamás la sutileza humana ideará una invención más bella, más sencilla o más directa que como lo hace la Naturaleza, porque en sus invenciones nada falta y nada sobra..."*

Leonardo da Vinci

Los productos naturales (PN) hacen alusión a aquellos compuestos funcionales producidos por organismos vivos.

Los PN incluyen a los conocidos anteriormente como "metabolitos secundarios", los cuales no son parte del metabolismo primario pues no cumplen funciones asociadas al crecimiento, desarrollo o reproducción.<sup>1,2</sup> En la literatura más reciente, se maneja con cada vez más frecuencia el término "metabolito especializado" para referirse a estos, tratando de reforzar la importancia de sus funciones biológicas y ecológicas.

El metabolismo primario se encarga de la biosíntesis y degradación de macromoléculas como proteínas, grasas, carbohidratos y ácidos nucleicos. Estos procesos se encuentran distribuidos en el árbol de la vida debido a que son necesarios para la viabilidad autónoma de la vida misma en la mayoría de las condiciones probadas.<sup>2,3</sup>

Los genes asociados a procesos del metabolismo primario se encuentran conservados y su evolución ocurre primordialmente a través de duplicación de genes y neofuncionalización.<sup>4</sup>

Por el contrario, los procesos del metabolismo especializado están frecuentemente siendo diversificados y evolutivamente seleccionados,

tratando de otorgar las herramientas necesarias a los organismos para sobrevivir a un ecosistema específico en constante cambio.<sup>4</sup>

La hipótesis evolutiva de la matriz química dinámica menciona que el surgimiento de metabolitos especializados se da por el reclutamiento de genes codificantes de enzimas de alta promiscuidad (capacidad de interaccionar con más de un sustrato sin ser de alta especificidad) del metabolismo primario y posterior agrupamiento por eventos de expansión genética (duplicación, transferencia horizontal de genes o recombinación).<sup>2</sup>

Un caso particular de la adaptación gracias al metabolismo especializado es el surgimiento de bacterias multirresistentes. La exposición de los microorganismos a compuestos antimicrobianos en cantidades no naturales en sus ecosistemas provocó una presión biológica selectiva sobre los organismos para generar mecanismos de resistencia.

Los genes asociados a estos mecanismos fueron adquiridos por mutaciones y compartidos por Transferencia Horizontal de Genes (THG). Su acumulación originó al surgimiento de bacterias multirresistentes, las cuales son uno de los riesgos a la salud pública más amenazantes al día de hoy.<sup>4,5</sup>

Las bacterias multirresistentes son uno de los mejores ejemplos de la importancia del metabolismo especializado y la forma en cómo puede afectar a otros miembros del ecosistema.

Desde su descubrimiento, los PN han sido aprovechados por el ser humano para diversas aplicaciones biotecnológicas. Los extractos de compuestos de plantas, algas, hongos, esponjas marinas, etc., han sido empleados en la industria farmacéutica, alimenticia, cosmética y agrícola.<sup>1,6,7</sup>

Los primeros registros del aprovechamiento de PN datan del año 2900 A.C., en el papiro Ebers, un documento que describe más de 700 fármacos basados en plantas. En el 2600 A.C., en la civilización de Mesopotamia se documenta en tabletas de barro el uso de aceites de *Cupressus sempervirens* (ciprés común) y especies de *Commiphora* (mirra) para tratar la tos, resfriados e inflamación, remedios usados aún en la actualidad.<sup>1</sup>

Probablemente una de las aplicaciones más importantes de los PN ha sido la que le ha dado la industria farmacéutica a compuestos que probaron tener propiedades terapéuticas. Esto tuvo su auge después de la segunda guerra mundial, donde la producción de cantidades masivas de penicilina fomentó el interés en el descubrimiento de nuevos antibióticos.<sup>8</sup>

El descubrimiento de la estreptomicina, gentamicina y tetraciclina logró financiar programas de descubrimiento de productos naturales.<sup>1,8</sup> Sin embargo a finales de los noventa, las metodologías basadas en el análisis de extractos llevaban a constantes redescubrimientos y las metodologías requerían una síntesis y derivatización complicada, lenta y costosa.<sup>1,8</sup>

Pronto se optó por la síntesis y simulación computacional, proyecciones de alto rendimiento y química combinatorial como nuevos métodos de descubrimiento, basados en probar bioactividades de extractos naturales conformados por una mezcla de PN.<sup>1,8</sup>

El desarrollo de las ciencias ómicas, en especial la metabolómica, complementaron el descubrimiento empleando técnicas de caracterización y purificación de compuestos como la Resonancia Magnética Nuclear (RMN), Espectroscopía de Absorción Ultravioleta (UV), Espectroscopía de Infrarrojo (IR), Espectroscopía de Masas (MS) y Cromatografía líquida de alto rendimiento (HPLC) o combinaciones de las

mismas, estableciendo lo que es conocido ahora como el descubrimiento clásico de productos naturales.<sup>1,9</sup>

Con los esfuerzos anteriores se ha logrado introducir al mercado la mayoría de los fármacos (60% entre 1981 y 2014) de origen en PN, basados en ellos o modificados a partir de los mismos.

Sin embargo, sólo se han explorado un estimado del 10% de las fuentes de PN en el mundo, lo cual deja abierto un gran campo de conocimiento a nuevos compuestos con actividades biológicas de interés.<sup>10,11</sup>

Los PN no sólo suponen una fuente diversa de una gran cantidad de productos funcionales con aplicaciones biotecnológicas que han permitido a la humanidad desarrollarse a lo largo de la historia, sino que también son una fuente de recursos renovables, más económica y amigable con el medio ambiente.<sup>1,12,13</sup>

El problema del método clásico radicó principalmente en la incapacidad de producir todos los posibles PN en condiciones de laboratorio, por lo que actualmente se emplea la técnica de minería genómica, que gracias a los avances en la genómica y técnicas bioinformáticas hace posible determinar todos los metabolitos especializados producidos por una especie a nivel genético.<sup>14</sup>

### **3.2. Bacterias Asociadas a Plantas, una fuente de metabolitos especializados**

Las bacterias son una de las principales fuentes de PN empleados por la industria en la elaboración de nuevos fármacos debido a su capacidad de producir compuestos con actividad antimicrobiana.<sup>15,16</sup>

Tan solo el género bacteriano *Streptomyces* es precursor del 55% de los antibióticos descubiertos entre 1945 y 1978 que actualmente se siguen empleando en el tratamiento de enfermedades infecciosas.<sup>17,18</sup>

El 80% de los antibióticos usados hoy en día provienen de actinobacterias, las cuales se han vuelto el principal blanco para estudios de minería genómica.<sup>19</sup>

Las bacterias han demostrado ser una fuente efectiva de compuestos capaces de inhibir o ralentizar el crecimiento de otros microorganismos, pues en sus ecosistemas se encuentran en una competencia constante por los recursos disponibles, como una zona de guerra donde hay una carrera armamentista y la bacteria que logre desarrollar el mejor antimicrobiano será la vencedora.

La mayoría de los ambientes de los que se extraen estas bacterias son ecosistemas multidiversos como ambientes marinos o muestras de suelo con diferentes clases taxonómicas.<sup>1</sup>

Recientemente han llamado la atención las Bacterias Asociadas a Plantas (BAP) por los mecanismos por los que establecen relaciones simbióticas con sus hospederas, entre los cuales destacan la adquisición de nutrientes del medio, fijación de nitrógeno, degradación de xenobióticos, producción de vitaminas y terpenos y la regulación de fitohormonas como el ácido indol-3-acético (AIA).<sup>19-23</sup>

Las BAP incluso han llegado a ser consideradas como la segunda línea de defensa en contra de patógenos u otras plagas por la producción de compuestos antibióticos, antifúngicos e insecticidas.<sup>15,16,24</sup>

A través de estos procesos, la bacteria promueve el crecimiento de la planta aún en ambientes con temperaturas extremas o baja disponibilidad de agua. A su vez, la planta ofrece a la bacteria un ambiente seguro con alta disponibilidad de nutrientes.<sup>25</sup>

Por varios años se ha planteado la hipótesis de que existen mecanismos de intercambio genético como THG entre las células de plantas con las



bacterias y que la presencia de compuestos propios de la hospedera induce en las bacterias la adquisición orientada de secuencias que sinteticen los mismos compuestos, por lo que muchas BAP podrían compartir propiedades farmacológicas con sus hospederas.<sup>26,27</sup>

El caso más conocido de la hipótesis anterior, es el del fármaco anticancerígeno taxol (paclitaxel), el cual fue aislado por primera vez del árbol *Taxus brevifolia*.<sup>26</sup> Después por medio del descubrimiento clásico de PN se encontró el mismo compuesto producido por la bacteria *Kitasatospora sp.* aislada de *Taxus baccata*.<sup>27</sup>

Registros fósiles han permitido rastrear la adaptación de BAP a sus hospederas desde hace aproximadamente 400 millones de años.<sup>28</sup> Durante este tiempo, los productos funcionales que promueven la simbiosis se han modificado para tener baja o nula citotoxicidad y no provocar daños o interferir con el sistema del hospedero eucarionte.<sup>26</sup>

Las BAP han podido establecer mecanismos de comunicación con las células de la planta por quimiotaxis y *quorum sensing*, un sistema de comunicación célula-célula mediado por autoinductores y respuesta a estímulos externos.<sup>29</sup>

En general podemos encontrar tres tipos de BAP, **las bacterias de la rizósfera** son aquellas que colonizan el suelo alrededor del que crece la planta y las raíces de esta, **las bacterias endófitas** colonizan y se establecen en los tejidos sanos de la planta abarcando regiones intra e intercelulares (endósfera de la planta) y finalmente **las bacterias epífitas** son aquellas que se encuentran en la superficie expuesta (filósfera) de la planta.<sup>30-34</sup>

Los tres tipos de BAP se especializan en diversos mecanismos tratando de adaptación, en beneficio de la simbiosis, por lo que en los tres tipos de

bacterias podemos encontrar una gran variedad de metabolitos especializados.

El estudio de la microbiota de las plantas puede extenderse al estudio específico de las regiones de la misma (tierra, raíces, tallos, hojas, flores e incluso los frutos), lamentablemente en las bases de datos donde no se suele reportar el tejido exacto del que se aisló la bacteria y se limitan a reportar el organismo en el que habitaba.

El estudio de las BAP es relativamente nuevo, si bien se han logrado muchos avances y descubrimientos en el estudio evolutivo de la adaptación de ambos organismos no se ha logrado abarcar casi nada del potencial de estas bacterias por su gran variedad de metabolitos especializados.

Sólo el 1-2% de las 300,000 especies de plantas conocidas han sido muestreadas con esta intención y muchos de estos experimentos se vieron limitados por la incapacidad de crecer microorganismos en condiciones de laboratorio.<sup>24,35</sup>

Recientes avances en microbiología y biología molecular han permitido activar estas células durmientes, haciendo accesible el descubrimiento de nuevas especies de BAP.<sup>35-37</sup>

Los mecanismos por los cuales tuvieron que adaptarse las bacterias a organismos eucariontes, lograron permitir que los compuestos bioactivos que benefician la simbiosis pudieran ser aprovechados por las diferentes industrias.

Los compuestos anticancerígenos y antimicrobianos pudieran ser una respuesta a dos de las amenazas actuales más grandes contra la salud pública, la industria cosmética se vería beneficiada en el uso de PN para desarrollar bloqueadores solares hechos a base de terpenos, la sustitución

de químicos tóxicos por bioplaguicidas naturales permitirían hacer uso del suelo sin degradarlo, aumentando su capacidad de siembra a lo largo del tiempo, además, muchos de los compuestos encontrados en BAP permitirían sembrar alimentos en zonas que se creían infértiles, logrando la independencia alimentaria por parte de varios países y más.

El desarrollo de las ciencias ómicas, la llegada de la minería genómica y los avances en microbiología y la biología molecular permiten hoy en día aprovechar al máximo el potencial de las BAP, lo que las vuelve un excelente punto de partida para demostrar la eficiencia de estas nuevas metodologías.

### **3.3. Sintetasas de péptidos no ribosomales (SPNR)**

Debido al aumento de genomas depositados en las bases de datos públicas, se ha vuelto una tarea muy complicada realizar análisis completos de todo el metabolismo especializado de una sola especie, más aún si se desean comparar diferentes clases taxonómicas.

Es necesario realizar un enfoque más preciso a una clase de metabolito especializado para aprovechar al máximo las herramientas bioinformáticas.

En este trabajo se seleccionaron las Sintetasas de Péptidos No Ribosomales (SPNR), un grupo de enzimas precursoras de los Péptidos No Ribosomales (PNR), el verdadero compuesto bioactivo que cuenta con numerosos reportes en la literatura por su actividad antimicrobiana y pesticida. En la Tabla 1 se muestran algunos de los fármacos descubiertos desde el año 2000 que tienen su origen de CGB de SPNR.<sup>15,16,38</sup>

Los PNR son una clase de metabolito abundante en genomas de BAP con numerosos reportes de bioactividad aplicada por la industria como plaguicidas (Xantolisina), promotores del crecimiento de plantas, captadores de metales pesados y biosurfactantes (sideróforos), actividad

de nucleación de hielo (por partículas que alteran la temperatura de congelamiento del agua), antimicrobianos (Vancomicina), anticancerígenos (Bleomicina), antiinflamatorios (Ciclosporina), toxinas ( $\alpha$ -amanitina), etc.<sup>39-43</sup>

En general, los PNR empleados en la industria farmacéutica pueden clasificarse en lipopéptidos lineales y cíclicos, que incluyen a los péptidos aromáticos, en 2,5-dicetopiperazinas (DKP) (péptidos cíclicos sencillos compuestos por uno o dos aminoácidos que pueden tener modificaciones adicionales) y glicopéptidos como la vancomicina, teicoplanina, balhimicina y la kistamicina.<sup>15,16,44</sup>

La variedad funcional dentro de la misma clase de metabolitos podría estar explicada por la variedad estructural de los mismos.

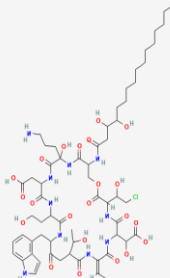
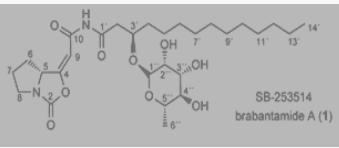
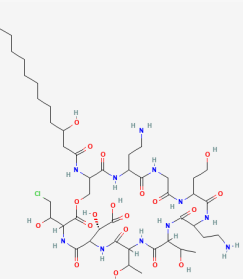
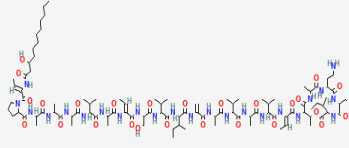
Las SPNR son capaces de añadir al péptido final bloques de construcción no convencionales, como aminoácidos no proteínogénicos,  $\beta$ -aminoácidos, aminoácidos de configuración D, ácidos grasos, azúcares, elementos heterocíclicos,  $\alpha$ -hidroxiácidos, etc. Además, estos residuos pueden sufrir modificaciones al ser reclutados lo que permite obtener varias estructuras a partir de un solo CGB de SPNR.<sup>44,47-49</sup>

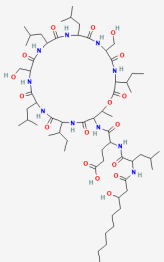
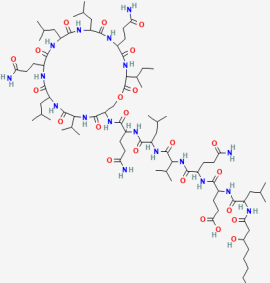
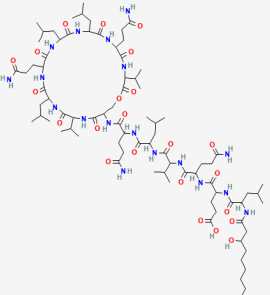
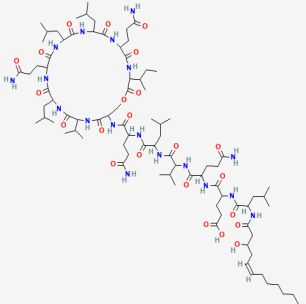
Las SPNR son mega enzimas compuestas por módulos, cada uno de estos se encarga de la adición de un sustrato y la modificación estructural del mismo, en caso de ser necesario.

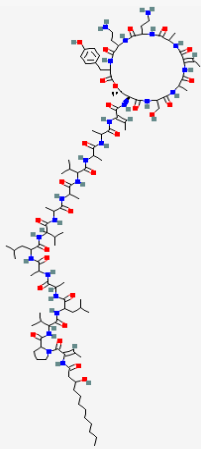
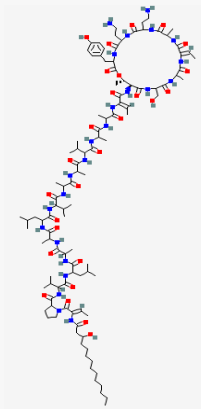
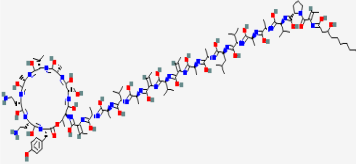
Los PNR resultantes de la biosíntesis se componen de un número de residuos igual al número de módulos que conforman la SPNR precursora.

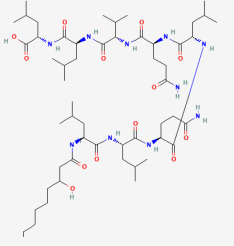
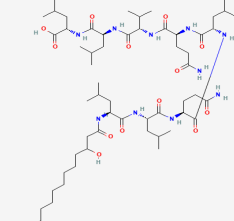
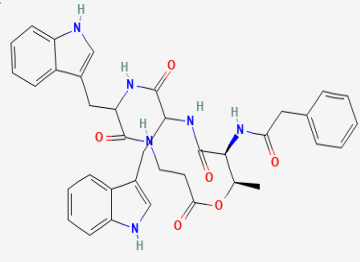
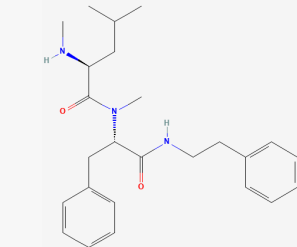
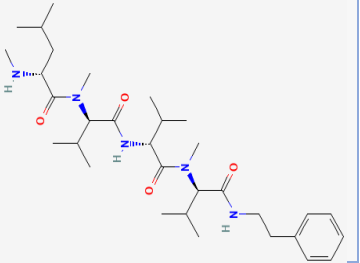
Además, estos residuos son añadidos en el mismo orden en el que son seleccionados por la enzima, por lo que la secuencia de bloques de construcción del PNR coincide con el orden en el que se presentan los módulos en la SPNR.<sup>67,68</sup>

Tabla 1. Características de PN de origen en SPNR descubiertos desde el año 2000<sup>16</sup> (imágenes e información obtenida de PubChem,<sup>38</sup> los asteriscos en la estructura primaria indican AA unidos covalentemente).

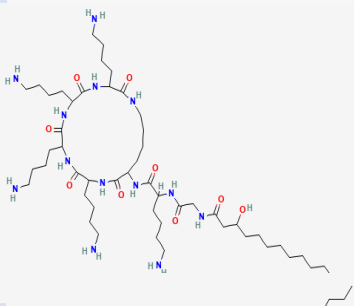
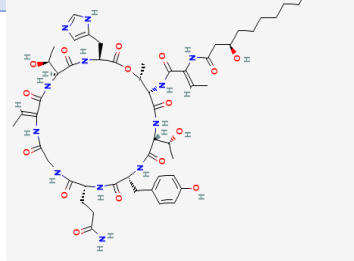
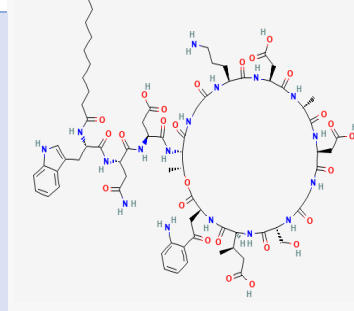
Antimicrobiano	Tipo de SPNR	Número de genes centrales	Número de módulos por GBC	Tamaño del CGB	Género bacteriano de origen	Características	Estructura primaria	Estructura secundaria
Tanamicina	Lipopéptido cíclico	2	5,4	32 kilobases (kb)	<i>Pseudomonas</i>	Antibacterial y antifúngico	Ácido Graso (AG) (desconocido)-Ser*-Orn-Asp-Lys-His-Thr-Dhb-3-Hy-Asp-4-Cl-Thr*	
Brabantamida <sup>45</sup>	Lipopéptido lineal	1	2	12 kb	<i>Pseudomonas</i>	Dipéptido/ Antibacterial y antituberculótico	Derivado estructural de Rhamnosa AG(3-Hy-n-C <sub>14</sub> )-Ser-Pro	
Nunamicina <sup>46</sup>	Lipopéptido cíclico	2	5,9	Isla genómica de 100 kb que contiene también la nunapeptina	<i>Pseudomonas</i>	Antifúngico	AG (3-Hy-n-C <sub>14</sub> AG)-Ser* <sup>-10</sup> Dab-Gly-Hse-Dab-Thr-Thr-3-Hy-Asp-4-Cl-Thr*	
Nunapeptina <sup>46</sup>	Lipopéptido cíclico	3	9,4,9	Isla genómica de 100 kb que contiene también la nunamicina	<i>Pseudomonas</i>	Antifúngico	AG (3Hy-n-C <sub>12</sub> AG)-Dhb-Pro-Ala-Ala-Ala-Val-Ala-Dhb-Ser-Val-Ile-Dhb-Ala-Val-Ala-Val-Dhb-Thr*-Ala-Dab-Ser-Ile*	

Poamida B <sup>47</sup>	Lipopéptido cíclico	3	2,4,4	Dos regiones cromosómicas abarcando 10 y 31 kb	<i>Pseudomonas</i>	Antifúngico	AG (3-Hy-n-C <sub>10</sub> )-Leu-Glu-Thr*-Leu-Leu-Ser-Leu-Leu-Ser-Ile*	
Xantolisina A <sup>48</sup>	Lipopéptido cíclico	3	2,8,4	41 kb	<i>Pseudomonas</i>	2 dominios TE/Antibacterial, antifúngico e insecticida	Unk-DL-Leu-DL-Glu-DL-Gln-DL-Val-DL-Leu-DL-Gln-DL-Ser*-DL-Val-DL-Leu-DL-Gln-DL-Leu-DL-Leu-DL-Gln-DL-xille*	
Xantolisina B <sup>48</sup>	Lipopéptido cíclico	3	2,8,4	41 kb	<i>Pseudomonas</i>	2 dominios TE/Antibacterial, antifúngico e insecticida	Unk-DL-Leu-DL-Glu-DL-Gln-DL-Val-DL-Leu-DL-Gln-DL-Ser*-DL-Val-DL-Leu-DL-Gln-DL-Leu-DL-Leu-DL-Gln-DL-Val*	
Xantolisina C <sup>48</sup>	Lipopéptido cíclico	3	2,8,4	41 kb	<i>Pseudomonas</i>	2 dominios TE/Antibacterial, antifúngico e insecticida	Unk-DL-Leu-DL-Glu-DL-Gln-DL-Val-DL-Leu-DL-Gln-DL-Ser*-DL-Val-DL-Leu-DL-Gln-DL-Leu-DL-Leu-DL-Gln-DL-xille*	

Siringopeptina SP508 A <sup>49</sup>	Lipopéptido cíclico	3	5,5,12	80 kb	<i>Pseudomonas</i>	Antibacterial	Unk-Abu(2,3- dehydro)-DL-Pro- DL-Val-DL-Leu- DL-Ala-DL-Ala-DL- Leu-DL-Val-DL- Ala-DL-Ala-DL- Val-DL-Ala-DL- Ala-Abu(2,3- dehydro)-aThr*- DL-Ser-DL-Ala- Abu(2,3- dehydro)-DL-Ala- DL-Dab-DL-Dab- DL-Tyr*	
Siringopeptina SP508 B <sup>49</sup>	Lipopéptido cíclico	3	5,5,12	80 kb	<i>Pseudomonas</i>	Antibacterial	Unk-Abu(2,3- dehydro)-DL-Pro- DL-Val-DL-Leu- DL-Ala-DL-Ala-DL- Leu-DL-Val-DL- Ala-DL-Ala-DL- Val-DL-Ala-DL- Ala-Abu(2,3- dehydro)-aThr*- DL-Ser-DL-Ala- Abu(2,3- dehydro)-DL-Ala- DL-Dab-DL-Dab- DL-Tyr-*	
Cichopectina	Lipopéptido cíclico	6	4,3,2,4,2, 7	?	<i>Pseudomonas</i>	Antibacterial	AG(desconocido)- Dhb-Pro-Ala-Ala- Ala-Ala-Val-Dhb- Gly-Val-Ile-Gly- Ala-Val-Ala-Val- Dhb-Thr*-Ala- Dab-Ser- Val*/Ile*/Leu*	

Cichofactina A	Lipopéptido lineal	2	3,5	?	<i>Pseudomonas</i>	Antibacterial y antifúngico	Unk-Leu-Leu-Gln-Leu-Gln-Val-Leu-Leu-OH	
Cichofactina B	Lipopéptido lineal	2	3,5	?	<i>Pseudomonas</i>	Antibacterial y antifúngico	Unk-Leu-Leu-Gln-Leu-Gln-Val-Leu-Leu-OH	
Xenematides	Péptidos aromáticos	1	4	5 kb	<i>Xenorhabdus</i>	Módulo de terminación atípico/ Antibacterial y antifúngico/ Sin genes biosintéticos adicionales	Aminoácido (AA) (ácido fenilacético)-Thr*-Trp-b-Ala*	
Xenortides	Péptidos aromáticos	2	1,1	?	<i>Xenorhabdus</i>	Péptidos aromáticos/ Antiprotozoario	AA (fenetilamina o triptamina)-N-Me-Phe-N-Me-Leu/Val	
Rabdopéptidos	Lipopéptido lineal	3	1,1,1	?	<i>Xenorhabdus</i>	Péptidos aromáticos/ Antiprotozoario	AA (fenetilamina o triptamina)-N-Me-Val-N-Me-Val/Val-N-Me-Leu/N-Me-Val-N-Me-Val	



PAX	lipopéptidos cíclicos	3	1,3,3	?	<i>Xenorhabdus</i>	Antibacterial y antifúngico	AG (3-Hy-iso-C <sub>14</sub> )-Gly-Lys/Arg-Lys*-Lys-Lys-Lys-Lys*	
Jagaricina <sup>50</sup>	lipopéptidos cíclicos	4	2,2,3,2	33 kb	<i>Janthinobacterium</i>	Antifungico	AG (3-Hy-n-C <sub>14</sub> AG)-Dhb-Thr*-Dhb-Tyr-His-Gln-Gly-N-Me-Thr*	
Daptomicina <sup>51</sup>	Lipopéptidos cíclicos	3	5,6,2	47 kb	<i>Streptomyces</i>	Antibacterial	AG (Decanoil)-Trp-D-Asn-Asp-Thr*-Gly-Orn-Asp-D-Ala-Asp-Gly-D-Ser-Glu(3R-Me)-Asp(Ph(2-NH <sub>2</sub> ))*	

Si bien, la variedad estructural de los PNR se debe a la capacidad de emplear una gran variedad de residuos para su construcción, así como la producción de múltiples PNR a partir de una sola SPNR, la selección y unión covalente de estos está sujeta a la función conservada de sus dominios funcionales.

Los módulos de las SPNR están compuestos por dominios, los más característicos, por ser esenciales para la construcción de un PNR, son el dominio de adenilación (A), el dominio de condensación (C), el dominio de la proteína transportadora de péptidos (PTP) y el dominio tioesterasa (TE).

El dominio A, junto al dominio PTP son los encargados de comenzar con la selección y activación del inicio de la elongación del PNR.

El dominio A se compone de 550 AA conservados. Todos los dominios A conocidos comparten un 30-60% de similitud de secuencia y su función principal es el reconocimiento y la selección del dominio que incorporará el módulo de elongación de la SPNR.

La activación del sustrato se da por una reacción química de dos pasos: Primero la enzima cataliza la formación de un intermediario de aminoacil adenilato del residuo seleccionado, posteriormente, el intermediario oxoéster AA-O-Monofosfato de adenosina (MPA) es transformado a un tioéster por un ataque nucleofílico por el cofactor 4'-fosfopanteteinilo (ppan) libre de tiol del dominio adyacente de PTP.<sup>67,68</sup>

Los dominios A son altamente específicos para cada residuo. Existen herramientas como NRSPredictor que son capaces de identificar la secuencia HMMER (secuencia de nucleótidos o AA estadísticamente conservados por cada dominio funcional identificado) del dominio A y asociarla a un residuo específico. Esta herramienta está incluida en las versiones más recientes de AntiSMASH.<sup>52,53</sup>

El dominio PTP es el más corto de las secuencias esenciales de las SPNR con tan sólo 100 AA, se encuentra río abajo del dominio A y es el responsable del transporte y elongación de los centros catalíticos.

Un brazo móvil del cofactor ppan de la coenzima A está unido covalentemente a este dominio y sirve como una "grúa" de entrega de los bloques de construcción.<sup>67,68</sup> Los residuos seleccionados por el dominio A terminan unidos covalentemente a este brazo móvil del dominio PTP esperando a ser unidos con un residuo de un módulo adyacente por efecto del dominio C.

El dominio C mide 450 AA y puede clasificarse en varios subtipos dependiendo de los de residuos que esté uniendo.<sup>54,55</sup> Este cataliza una reacción de ataque nucleofílico del aceptor de electrones del AA, unido al dominio PTP río abajo del dominio C, con el grupo  $\alpha$ -amino libre del tioéster del péptido o AA unido al dominio PTP río arriba.<sup>67,68</sup>

Se establece una reacción en cadena secuencial entre los módulos de elongación conformados por los tres dominios A, C y PTP hasta llegar a un último módulo que contiene un único dominio TE.

El dominio TE está constituido por 280 AA y cataliza una reacción de un ataque nucleofílico en el intermediario de tioéster unido al último dominio PTP. El nucleófilo empleado en esta reacción depende de si el péptido será lineal, en cuyo caso emplea una molécula de agua para liberar al péptido por una reacción de hidrólisis, o en el caso de que el PNR sea cíclico, se utilizará un nucleófilo interno de la estructura del último intermediario de tioéster unido al dominio PTP para liberar al PNR por ciclación.<sup>67,68</sup>

Existen dominios que no son esenciales para la producción de un PNR que realizan modificaciones sobre los sustratos como el dominio de metilación, el dominio de epimerización que se encarga de cambiar la configuración de L-AA a D-AA, dominios de oxidación, o dominios que se encargan de

realizar liberaciones alternativas como el dominio de reducción, que libera al péptido obteniendo un extremo con un grupo funcional de aldehído o alcohol. Sin embargo, estos no se encuentran con tanta frecuencia como los dominios esenciales.

En la Figura 1 se muestra el esquema de biosíntesis de la surfactina por su respectiva SPNR.

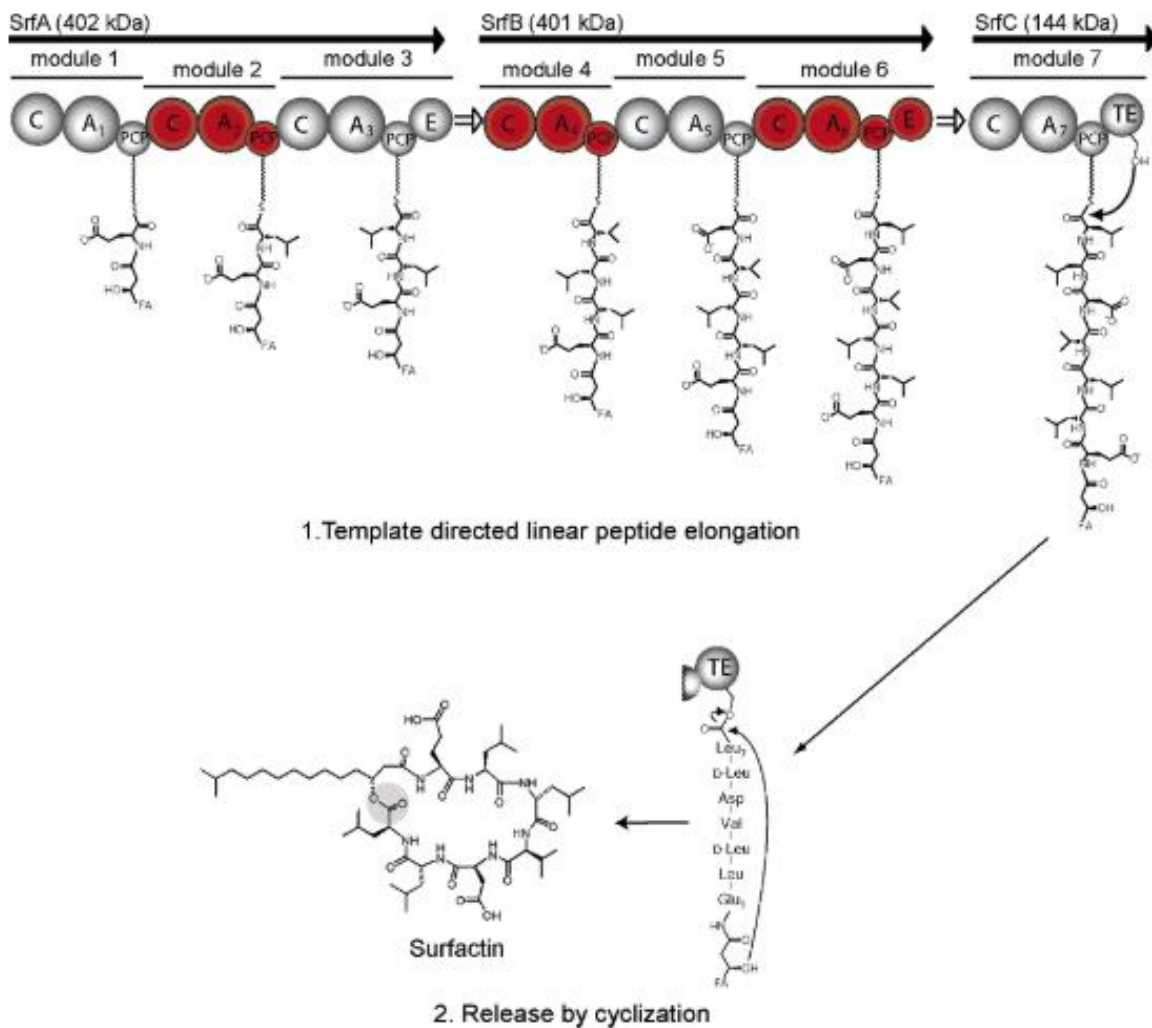


Figura 1. Esquema de biosíntesis de la surfactina. La SPNR está codificada en los tres genes *SrfA*, *SrfB* y *SrfC*, en los que se codifican los módulos 1-3, 4-6 y 7, respectivamente. Consta de 23 dominios encargados de la elongación del péptido y un único dominio de liberación por ciclación (TE). El PNR final está formado por 7 residuos, el mismo número de módulos de la enzima precursora. Tomado de Sieber y Marahiel.<sup>68</sup>

Las SPNR enzimas capaces de producir más de un compuesto a partir de un solo CGB. Al tener una vía de síntesis independiente de los ribosomas, los PNR cuentan con una gran versatilidad estructural, siendo constituidos por elementos proteínogénicos y otros elementos no convencionales, lo cual a su vez les otorga un increíble potencial de bioactividad, que en la literatura se ha reportado que va desde toxinas hasta compuestos anticancerígenos.

En la búsqueda de nuevos compuestos con potencial biotecnológico, las SPNR son sin duda uno de los metabolitos especializados que llaman más la atención. Las herramientas y técnicas más actuales permiten rastrear el origen de estas enzimas a nivel genético y reproducirlas e incluso modificarlas para su producción en condiciones de laboratorio.

### **3.4. Minería genómica en el descubrimiento de nuevos metabolitos especializados**

Gracias a la llegada de la secuenciación de siguiente generación, el tiempo en el que se puede secuenciar un genoma completo ha pasado de años a cuestión de horas, optimizando no sólo el tiempo del proceso sino también los costos.<sup>56</sup>

Debido a la facilidad con la que se pueden obtener los genomas de distintos organismos, bases de datos públicas como las de NCBI (National Center of Biotechnology Information) y BV-BRC (BACTERIAL AND VIRAL BIOINFORMATICS RESOURCE CENTER) han aumentado masivamente la cantidad de secuencias de genomas, genes, proteínas, etc., contenidas en ellas.<sup>57,58</sup>

Además, los programas y softwares para analizar la información genómica han sido constantemente desarrollados y actualizados en los últimos años.

La minería genómica es una metodología desarrollada en torno al aprovechamiento de estos avances en genómica en conjunto con la química combinacional.<sup>14</sup> Hace uso de diferentes herramientas bioinformáticas para realizar comparaciones de secuencias nuevas en los genomas de plantas, hongos y bacterias contra la información depositada en las bases de datos y puesta a la disposición del público.<sup>59</sup>

El fundamento de la técnica es realizar una comparación de secuencias de algo desconocido contra lo ya conocido, con esto se pretende aprovechar la conservación de genes, dominios o módulos de Clústeres de Genes Biosintéticos (CGB), los cuales son los precursores genéticos de los metabolitos especializados.

Los CGB, definidos por Medema et al. como “Un conjunto de dos o más genes agrupados físicamente en un genoma particular que juntos codifican una ruta biosintética para la producción de un metabolito especializado”, pueden clasificarse dependiendo del tipo de metabolito que serán precursores.<sup>60</sup>

Entre las clases más importantes de metabolitos especializados podemos encontrar a los Péptidos sintetizados Ribosomalmente y modificados Postraduccionalmente (PsRMP), los Policétidos (PKs), sintetizados por las Policétido Sintetasas (PKSs), los terpenos provenientes de las terpeno sintetasas (TSs), los sideróforos y los Péptidos no Ribosomales (PNR) que son producidos por las Sintetasas de Péptidos no Ribosomales (SPNR).<sup>61-</sup>

68

A excepción de los PNR, terpenos y PKs, las secuencias de DNA que forman los CGB pueden compararse directamente con genes ya conocidos y depositados en las bases de datos de metabolitos de la misma clase (los terpenos, PNR y PKs deben rastrearse por medio de sus enzimas precursoras, las TSs, SPNR y las PKSs).

Las PKSs comparten muchas características en común con las SPNR, por ejemplo, ambas enzimas tienen una composición modular que coincide con el número de residuos incorporados en el producto final.

Las PKSs sintetizan PKs usando precursores de Acil-CoA y han despertado interés debido a su capacidad de biosintetizar a través de métodos combinatoriales productos con gran versatilidad, difíciles de lograr por otros medios.

Se conocen 3 tipos de PKSs, el tipo I son enzimas multifuncionales organizadas en módulos, que actúan de forma no iterativa, responsables de la catálisis de un ciclo de elongación de la cadena de policétidos, las PKSs tipo II son complejos multienzimáticos que llevan un conjunto único de actividades que actúan iterativamente y las PKSs de tipo III son enzimas homodiméricas de condensación que actúan iterativamente.<sup>63</sup>

Se cree que tanto las SPNR como las PKSs convergen desde un ancestro en común, lo cual tiene sentido debido al modo de evolución de ambas enzimas, a partir de la unión de subclústeres independientes en el genoma.

Los CGB de metabolitos especializados cuentan con regiones ampliamente conservadas entre todos los PN de la misma clase, por ejemplo, los CGB de lantipeptidos clase II (de la clase de los PsRMP) están compuestos por un gen que codifica un péptido precursor y del gen *LanM*, que se traduce en una enzima bifuncional, encargada de realizar modificaciones postraduccionales.

La enzima encargada de modificar el péptido precursor se compone de un dominio con función de deshidratación N-terminal y otro con función de ciclación en el extremo C-terminal.<sup>61</sup>

A pesar de que el péptido precursor puede variar entre un CGB de lantipéptido a otro, otorgando variedad estructural dentro de la misma clase de metabolito, la secuencia del gen *LanM* se encuentra muy conservada debido a que su función no varía, y es por medio este gen que se identifica una secuencia como un CGB de lantipeptido clase II.

Se han desarrollado diferentes softwares automatizados de procesamiento de datos genómicos para realizar estas búsquedas basadas en homología en cuestión de minutos, los cuales pueden utilizarse de forma intuitiva sin ser necesario ser especialista en el tema.

En palabras de Maldonado-Carmona et al. "AntiSMASH se ha convertido en la plataforma estándar de oro actual para la identificación de CGB".<sup>9</sup>

AntiSMASH (antibiotics and secondary metabolite analysis Shell) es un ecosistema de herramientas independientes incorporadas, para la predicción y análisis de 71 diferentes clases de CGB en genomas bacterianos.<sup>69</sup>

La minería genómica es una técnica casi tan revolucionaria en el descubrimiento de productos naturales como lo es la secuenciación de siguiente generación en el campo de la genómica.

Así como aumentaron de manera indiscriminada el número de genomas secuenciados, lo hizo también el número de CGB de metabolitos especializados, incluso los caracterizados por el descubrimiento clásico de productos naturales, los cuales hoy se conocen a detalle sus respectivos CGB y se encuentran depositados en bases de datos como la de AntiSMASH y la de MiBIG (la cual es una base de datos se que recopila información detallada de CGB explicada de forma integral y estandarizada, descrita en la literatura).<sup>60</sup>



Estas bases de datos no solamente han permitido predecir con mayor precisión los CGB en secuencias de DNA nuevas, sino que permite realizar comparaciones acertadas de los metabolitos especializados ya caracterizados, conocidos e incluso comercializados evitando así el redescubrimiento de compuestos químicos, y resolviendo el problema principal del descubrimiento clásico de productos naturales.

Aún con todo el progreso, la técnica de la minería genómica no es perfecta y se enfrenta a predecir CGB de compuestos sin actividad biológica o compuestos con estructuras químicas y genéticas diferentes, y actividades idénticas a la de compuestos ya descubiertos.

Este problema pretende resolverse complementando la información obtenida con estudios bioinformáticos sobre la conservación de productos funcionales en ciertas especies bacterianas, así como su abundancia en distintas cepas y la relación entre sus CGB.

Entre las principales herramientas para estudiar los CGB están BiG-SCAPE (Biosynthetic Gene Similarity Clustering And Prospecting Engine) y CORASON (Core Analysis of Syntenic Orthologues to prioritize Natural product gene clusters), desarrollada en el 2020 por Navarro-Muñoz, Selem-Mojica et al.<sup>70</sup>

BiG-SCAPE es un programa interactivo e intuitivo para construir redes de similitud de secuencias de BGC obtenidos de AntiSMASH o MiBIG.

El algoritmo de BiG-SCAPE consiste en la descomposición de genes que constituyen un CGB en dominios funcionales clasificados acorde a la base de datos de Pfam, posteriormente estos dominios son analizados por su porcentaje de recurrencia, su porcentaje de dominios adyacentes y el porcentaje de identidad de la secuencia de los mismos.<sup>71</sup>

Con la información obtenida es posible primeramente clasificar a los CGB por clase de metabolito especializado y posteriormente son agrupados en Familias de Clústeres de Genes (FCG), en las cuales los integrantes comparten características genéticas que permiten predecir estructuras químicas y funciones similares.

Finalmente se realiza un segundo agrupamiento de propagación por afinidad de CGB para agrupar FCG en Clanes de Familias de Clústeres (CFC).

Los CGB son mostrados como puntos en una representación arquitectónica; aquellos que estén relacionados al nivel de ser considerados dentro de la misma FCG, estarán unidos por líneas.

Por otra parte, el programa CORASON pretende elucidar relaciones filogenéticas por medio del alineamiento de los CGB dentro de una FCG empleando un clúster de referencia.

La representación visual por medio de un árbol filogenético construido por el método de máxima verosimilitud permite, no solamente analizar la relación evolutiva de estos clústeres de genes, sino también comprender el contexto genético de los Genes Biosintéticos Centrales (GBC) y la recurrencia de genes adicionales, ambos procesos fundamentales en el descubrimiento de nuevos PN.

A pesar de que el agrupamiento de CGB en familias es una metodología relativamente novedosa, el aumento de CGB depositados en las bases de datos y la necesidad de comparación contra clústeres de referencia de compuestos ya caracterizados experimentalmente ha fomentado que este tipo de análisis sea cada vez más frecuente, aumentando la necesidad de optimizar estos procesos de comparación.

BiG-FAM<sup>72</sup> es una plataforma donde se agrupan 1,225,071 CGB en FCG, provenientes de la base de datos de MiBIG (caracterizados experimentalmente) y predichos de 209,206 genomas de NCBI. El resultado fue la obtención de 29,955 FCG de 5 clases de metabolitos diferentes (PNR, PKs, Terpenos, PsRMP y otros).

En esta base de datos es posible comparar cualquier CGB obtenido por AntiSMASH contra las casi treinta mil familias de clústeres de BiG-FAM, facilitando el descubrimiento de productos naturales y mapeando la diversidad estructural y de origen taxonómico de los CGB analizados por programas como BiG-SCAPE.

No solamente el análisis y comparación de CGB es suficiente para garantizar el descubrimiento de nuevos PN; análisis bioinformáticos de alineamientos de secuencias múltiples de genomas completos o genes en común, permite identificar y comparar taxonómicamente las distintas especies de las cuales obtenemos un gran número de metabolitos especializados.<sup>73</sup>

Con lo anterior podemos construir una representación visual denominada árbol filogenético que permite realizar análisis evolutivos y encontrar patrones en los organismos que biosintetizan PN empleando distintas herramientas puestas a disposición del dominio público.<sup>74</sup>

El descubrimiento de nuevos compuestos por comparación con estructuras químicas de compuestos ya conocidos es un estudio relativamente nuevo.

Lo que se conoce como metabologenómica es la asociación de un CGB con una estructura química, determinada usualmente por MS. Usando esta técnica es posible comparar y asociar redes de similitud de varios CGB con redes de distintas estructuras químicas, acelerando el descubrimiento de nuevos PN.<sup>75</sup>

Stowell et al., a través de una serie de mutaciones y determinaciones estructurales en una TS de clase II rastrearon los residuos esenciales para la catálisis y control selectivo para la formación del producto final en esta clase de TS.<sup>76</sup>

De la misma manera Pan et al. a través de caracterizaciones cristalográficas de proteínas, encontraron dominios funcionales propios de terpeno ciclasas clase II en ciclasas sesquiterpénicas, cuyas estructuras únicamente habían sido asociadas a dominios de TS clase I.<sup>77</sup>

El aprendizaje automático y el desarrollo de la Inteligencia Artificial (IA) permitieron el descubrimiento de la Halicina, un antibiótico activo contra un gran número de bacterias multirresistentes, alimentando la base de datos de la IA con 6000 estructuras de compuestos empleados en la medicina humana con el fin de que esta aprendiera cómo se ve un antibiótico funcional.<sup>75</sup>

Finalmente, la ingeniería genética en conjunto con análisis bioinformáticos de sinergia de compuestos que al unirse tienen un mayor carácter antibiótico, han permitido la edición de CGB con el fin de crear productos no naturales con mayor potencial biotecnológico.

Para la ingeniería genética se emplean técnicas de edición de DNA como enzimas *CRISPR-Cas9* o la unión de clústeres alejados en genomas, que pareciera que se regulan mutuamente con el fin de tener una producción similar de sus respectivos metabolitos especializados.<sup>75,78</sup>

En conjunto, la técnica de minería genómica complementada con análisis bioinformáticos y metabolómicos no solamente ha facilitado y propiciado exponencialmente el descubrimiento de nuevos PN, sino que permiten comprender las relaciones evolutivas y rastrear el origen de los mismos, lo cual abre las posibilidades de comprender el papel de este tipo de compuestos en la naturaleza, lo cual en un futuro podría ser la clave para

resolver varias de las problemáticas más amenazantes para la civilización humana.

#### **4. Justificación**

Los recientes avances en las ciencias ómicas, la tecnología de secuenciación de siguiente generación, la minería genómica y las mejoras en los campos de la microbiología y biología celular han facilitado el descubrimiento de Productos Naturales enormemente.

Lo anterior llevó a la acumulación y aumento de información genética y de CGB en bases de datos, por lo que, en el descubrimiento de nuevos PN se volvió necesario definir un blanco específico, con gran abundancia de clases y estructuras de metabolitos especializados con actividades que sean aplicables a la resolución de problemas que aquejan a la sociedad.

Las bacterias asociadas a plantas son microorganismos que han llamado la atención recientemente debido a su capacidad de adaptarse a sistemas eucariotas y beneficiarse junto a sus hospederas a través de la producción de una gran variedad de compuestos con funciones que van desde la protección contra agentes infecciosos hasta la promoción del crecimiento y mejora de la salud de las plantas, todo sin provocar ningún daño aparente.

Entre estos compuestos destacan las Sintetasas de Péptidos no Ribosomales, las enzimas precursoras de los Péptidos no Ribosomales, que tienen reportes en la literatura de actividades biológicas que van desde insecticidas hasta anticancerígenas.

Las herramientas bioinformáticas en conjunto con el análisis evolutivo de este tipo de metabolitos abren la posibilidad entender las funciones de estos compuestos en la adaptación de los microorganismos al medio ambiente y con ello la proposición de soluciones a problemas por las que

el ser humano también se ve afectado, como el surgimiento de bacterias multirresistentes e infertilidad o desgaste de la tierra en muchas regiones del mundo.

El uso de productos naturales en diferentes sectores abre la posibilidad de reemplazar los productos usados frecuentemente, que con el paso del tiempo se han vuelto perjudiciales para los ecosistemas y los organismos que viven en ellos.

Los compuestos que son generados de manera natural por microorganismos tienen aplicaciones que van desde plaguicidas, promotores del crecimiento de plantas, fármacos, etc., que, además de demostrar efectividad en sus respectivos ambientes, su uso y producción biotecnológica los vuelve una opción mucho menos contaminante que las alternativas actuales.

La metodología propuesta pretende promover el descubrimiento de nuevos compuestos para aplicaciones biotecnológicas que logren satisfacer necesidades humanas y sean aprovechadas por las diferentes industrias como una alternativa más eficiente, económica y amigable con el medio ambiente.

## **5. Hipótesis**

Por su versatilidad estructural y funcional, las sintetasas de péptidos no ribosomales deberían tener un papel fundamental en la adaptación de las bacterias asociadas a plantas con sus respectivas hospederas a través de mecanismos que promuevan el transcurso del ciclo de vida de la bacteria y hospedero, aún en presencia de amenazas ambientales, como otros microorganismos infecciosos.

Si tienen presencia en géneros bacterianos distintos, las SPNR podrían ser fundamentales en la evolución y supervivencia de los organismos.

## 6. Objetivos

### 6.1. Objetivo general

Identificar la importancia de las SPNR en la adaptación y supervivencia de las bacterias asociadas a plantas y de sus hospederas, tanto funcionalmente como evolutivamente.

### 6.2. Objetivos particulares

- Evaluar a las BAP como una fuente de metabolitos especializados con potencial de bioactividad.
- Determinar la importancia en términos de bioactividad de las SPNR en BAP.
- Demostrar que la minería genómica en conjunto con análisis bioinformáticos, son una técnica efectiva en el descubrimiento de nuevos productos naturales.
- Construir un perfil de metabolitos especializados de genomas de BAP para encontrar patrones evolutivos entre los genomas completos y sus CGB.
- Identificar patrones de presencia de metabolitos especializados en los géneros bacterianos para determinar sus funciones específicas en cada una de las diferentes regiones de la planta donde habitan las diferentes clases taxonómicas de BAP.
- Proponer un CGB basado en sus características evolutivas, de origen y estructurales, para producir por técnicas de biología molecular y posteriormente realizar pruebas de actividad biológica del compuesto final.

## 7. Metodología

El esquema general de la metodología se muestra en la Figura 2. Éste se divide en seis partes, de las cuales, las primeras tres fueron consecutivas y culminaron con el análisis de minería genómica. A partir de este punto se realizaron los pasos subsecuentes sin ningún orden en particular. El análisis del conjunto de las seis secciones permitió alcanzar los objetivos planteados en la sección anterior.

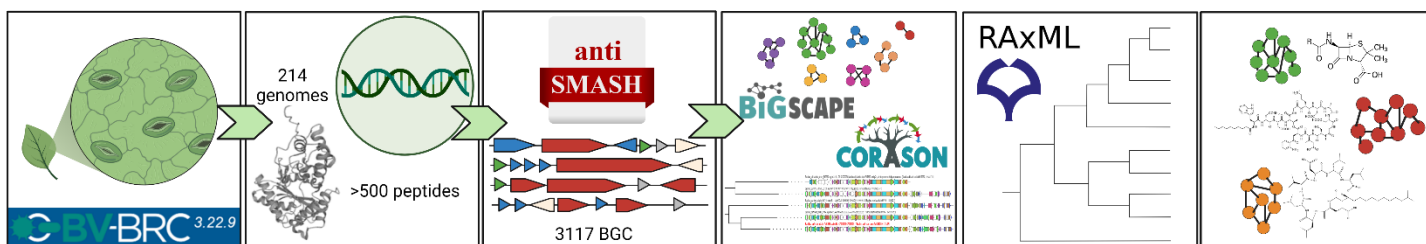


Figura 2. Esquema general de la metodología. En orden de izquierda a derecha: 1) Recolección de genomas de BAP en la plataforma BV-BRC.<sup>58</sup> 2) En la misma plataforma se realizó una búsqueda de secuencias homólogas a un dominio de una SPNR en los genomas recolectados, para ello se usó la herramienta BLAST. 3) Los genomas resultantes de la búsqueda se analizaron por minería genómica empleando AntiSMASH.<sup>69</sup> 4) Los CGB predichos fueron analizados por redes de similitud y árboles filogenéticos construidos por los programas BiG-SCAPE y CORASON respectivamente.<sup>70</sup> 5) Se construyeron árboles filogenéticos usando genomas de las bacterias precursoras de las SPNR. 6) Finalmente se propuso un CGB perteneciente a una FCG obtenida en el inciso 4.

### 7.1. Recolección de genomas de BAP

Para construir una base de datos de genomas de BAP, se empleó la plataforma BV-BRC.<sup>58</sup>

Empleando la barra de búsqueda, con apoyo de los artículos realizados por Ulloa-Muñoz et al., Martínez-Klimova et al., Ali et al., Cordero et al. y Liu et al.<sup>29,79-82</sup> y con los filtros de búsqueda de la misma plataforma se seleccionaron 384 genomas que cumplieran los criterios de inclusión de:

1) Ser bacterias cuyo género o especie haya sido reportado por uno de los artículos antes mencionados como aislado de alguna planta o en su defecto haber sido subido a la plataforma con tal característica.



2) Su genoma sea reportado como de buena calidad dentro de la misma base de datos (BV-BRC) y tuviera un número de contigs igual o menor a cien.

3) De preferencia ser un genoma representativo y/o genoma completo o Genoma Secuenciado Completamente (GSC).

Los genomas seleccionados y sus características se encuentran enlistados en la Tabla I de la información suplementaria.

## **7.2. Selección de un dominio característico de las SPNR**

Se empleó la plataforma Pfam del Instituto Europeo de Bioinformática (EMBL-EBI), la cual se encuentra contenida ahora en InterPro, manteniendo las mismas herramientas que se mencionan en esta metodología.<sup>83</sup>

Se utilizó la herramienta de búsqueda por palabras clave para encontrar una proteína con dominios esenciales de las SPNR, se emplearon los filtros de búsqueda para seleccionar una enzima que tenga el mayor grado de confianza, de preferencia con evidencia a nivel proteína revisada por curadores de UniProt.<sup>84</sup>

La enzima seleccionada fue analizada a nivel de genes y dominios, usando herramientas como BLAST, ActiveSiteFinder y Cluster Pfam Analysis, todas ellas incorporadas en la versión más actual de AntiSMASH. Se determinó la composición estructural y secuencial de dichas secuencias, así como los residuos que serían seleccionados por los dominios A predichos, usando la herramienta NRPSpredictor.

Lo anterior con la intención de seleccionar un dominio característico de las SPNR comparando con las familias de dominios de InterPro.

La importancia de la familia de proteínas seleccionada se confirmó con información de la literatura científica.<sup>44,86</sup>

### **7.3. Búsqueda por homología con BLAST**

Se utilizó la secuencia determinada en el paso anterior para realizar una búsqueda por homología con la herramienta de blastp, incorporada a plataforma BV-BRC, dentro del grupo de genomas de BAP recolectados en la primera parte de la metodología.<sup>87</sup>

Los parámetros de búsqueda fueron modificados con un valor de E de 0.01 para hacer menos restrictiva la búsqueda y obtener una mayor variedad funcional de compuestos con presencia de dominios de condensación.

### **7.4. Minería genómica**

Los genomas con presencia de dominios homólogos al dominio determinado en la sección 7.2, fueron descargados en formato FASTA de la base de datos BV-BRC y subidos a la plataforma de antiSMASH v.6.1.1 para identificar posibles CGB de metabolitos especializados.

Se empleó un rigor de detección “relajado” con el fin de evitar en lo posible los CGB falsos positivos y descubrir secuencias nuevas con suficiente grado de diversidad en comparación con las depositadas en la base de datos de antiSMASH y MiBIG.

Se activaron todas las características adicionales que provee la plataforma.

### **7.5. Construcción de árboles filogenéticos de los genomas resultantes de BAP**

Los árboles filogenéticos fueron construidos en la plataforma BV-BRC, utilizando el servicio “Construcción de árbol filogenético”, el cual emplea el método de árbol de codones, basado en la selección de PGFams de BV-BRC de una sola copia, compartidas entre los genomas.<sup>85</sup>

Estas fueron alineadas usando el programa MAFFT y analizados por el programa RAXML (Randomized Axelerated Maximum Likelihood).<sup>73,74</sup> Los

parámetros empleados para construir los árboles fueron 100 genes compartidos, 0 deleciones y 0 duplicaciones permitidas con el fin de abarcar una mayor variedad de genes y productos proteicos.

#### **7.6. Construcción de redes de similitud y agrupación de los CGB en FCG por BiG-SCAPE**

Se descargaron los archivos .gbk (genbank) de los CGB de SPNR obtenidos del análisis de minería genómica por antiSMASH v6.1.1 de genomas con presencia de dominios homólogos a un dominio C.

Para la construcción de la red de similitud se empleó el programa BIG-SCAPE v1.1.0. Se probaron varios cortes de distancia sin procesar, que iban de 0 a 1, intercalando los valores por distancias de 0.1.

Finalmente se seleccionó 0.3 (igual al valor por defecto del programa), siendo este el valor donde comenzaba a aparecer el escenario más restrictivo, agrupando los CGB en un mayor número de familias.

#### **7.7. Propuesta de candidato final**

Como resultado de los análisis de los resultados de los puntos anteriores de la metodología, se propuso un CGB basado en la conservación de las secuencias con el fin de asegurar la producción de un compuesto funcional, en la estructura química y genética del CGB, y en la filogenia de la especie productora, así como el origen de la misma.

## 8. Discusión y resultados

### 8.1. Recolección de genomas de BAP

En la Tabla I del material suplementario se reportan los 384 genomas empleados en el estudio, así como su respectiva clave de acceso en las bases de datos públicas y sus características genéticas más importantes.

En ella se reúnen genomas de 57 géneros bacterianos encontrados en más de 150 especies de plantas y su rizosfera.

Estos tienen un tamaño que va desde las 1,863,914 pares de bases (pb) (*Pediococcus* sp. EKM201D) a 11,715,108 pb (*Actinomadura barringtoniae* strain GKU 128) y un número de genes que va desde los 1,835 a los 10,935 (mismos genomas, respectivamente). Por lo que, en general, a pesar de ser todas bacterias aisladas de plantas hay suficiente variedad genómica para sustentar los resultados de este estudio.

### 8.2. Selección de un dominio característico de las SPNR

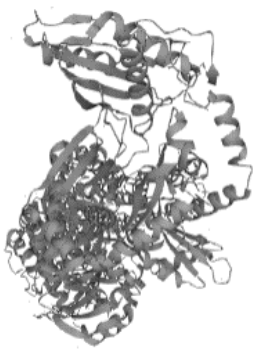
Se emplearon las palabras "Condensation NRPS", para la búsqueda por palabras clave en Pfam.

De los resultados se seleccionó la familia de proteínas del mismo nombre (Pfam: PF00668, InterPro: IPR001242) la cual contiene un aproximado de 111 mil proteínas de las cuales sólo 254 de estas son revisadas por los curadores de UniProt, de éstas se seleccionaron aquellas que alcanzaban el máximo nivel del contenido de la anotación de la proteína en una escala de 5 puntos dada por la base de datos de UniProt.

Fueron 5 las proteínas que obtuvieron el puntaje perfecto que corresponde a una anotación que contenga evidencia a nivel proteína, de ellas se seleccionó el componente F de la *enterobactina sintasa*.

En la Tabla 2 se presentan las características más importantes de dicho componente.

Tabla 2. Características del componente F de la enterobactina sintasa.

Características	Descripción
Péptido	Componente F de la <i>enterobactina sintasa</i>
Número de acceso	P11454 (UNIPROT), BGC0002476 (MiBIG)
Gen	<i>EntF</i>
Organismo de Origen	<i>Escherichia coli</i> (cepa K12)
Tamaño	1293 AA
Función	Involucrada en la producción del sideróforo enterobactina, <i>EntF</i> cataliza la activación de L-serina por medio de una reacción de intercambio PPi dependiente de ATP para formar seriladenilato.
Localización subcelular	Citoplasma
Interacción	Las proteínas <i>EntB</i> , <i>EntD</i> , <i>EntE</i> y <i>EntF</i> son componentes de la enterobactina sintasa
Producto final	Enterobactina
Estructura del componente PDB: 5JA1	
Dominios	<p>Dominio condensación (2-442): encargado de la elongación del péptido.</p> <p>Dominio de adenilación (482-886): Activa el residuo de serina en un éster de aminoacil-AMP.</p> <p>Dominio de PTP (971-1046): Lleva un brazo de fosfopanteteinilo para unir el aminoácido activado y un dominio de tioesterasa (Probable).</p> <p>Dominio tioesterasa (1067-1283): Es tanto una lactona sintetasa ciclotrimerizante como un catalizador de elongación para la formación de enlaces éster entre restos DHB-Ser unidos covalentemente.</p>

El componente F de la *enterobactina sintasa* cataliza la activación de L-serina por medio de una reacción de intercambio PPi dependiente de ATP para formar seriladenilato.

En conjunto con los componentes de un solo dominio B y E, de la misma enzima, se cataliza la reacción que produce la enterobactina (Figura 3)

por la SPNR de seis dominios la cual es finalmente es transportada por el producto del componente D.<sup>86</sup>

La enterobactina es un sideróforo con alta afinidad por el hierro, es una trilactona cíclica de N-(2,3-dihidroxi-benzoil)serina.

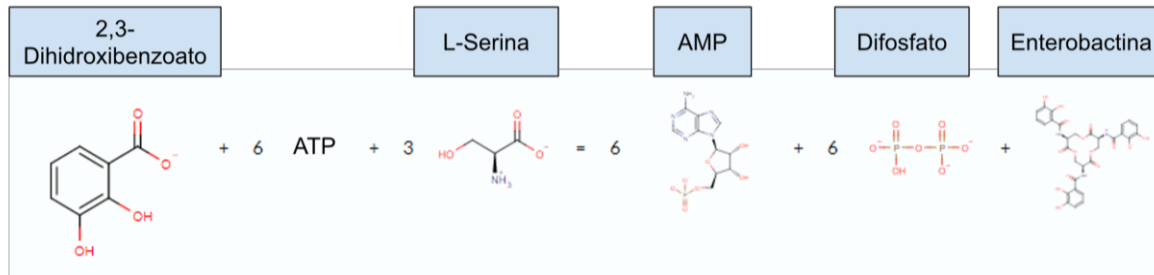


Figura 3. Reacción catalizada por la enterobactina sintasa. El primer dominio de adenilación de la SPNR (codificado en el gen *EntE*) selecciona al ácido carboxílico 2,3-dihidroxi-benzoato, el cual es transportado al dominio PTP del segundo módulo (codificado en el gen *EntB*), mientras tanto el dominio de adenilación del último módulo (codificado en el gen *EntF*) selecciona y modifica la L-serina a seriladenilato para posteriormente ser unidos por el único dominio de condensación de la SPNR, este proceso es llevado a cabo dos veces para formar la enterobactina.

La herramienta de BLAST de clústeres conocidos en la base de datos de MiBIG, muestra un porcentaje de similitud entre diferentes PNR con el clúster analizado que van desde el 30% al 69% (Figura 4), un intervalo lo suficientemente significativo para asegurar que se trata de una enzima SPNR.

Los genes involucrados en cada uno de los CGB resultantes parecieran tener el mismo orden, con excepciones, resultado de la inserción de uno o más genes en sus respectivos clústeres o de la delección o duplicación de algunos de ellos, dando como resultado diferentes productos funcionales de enzimas con composiciones genéticas similares, esto concuerda con el modo en el que evolucionan las SPNR.

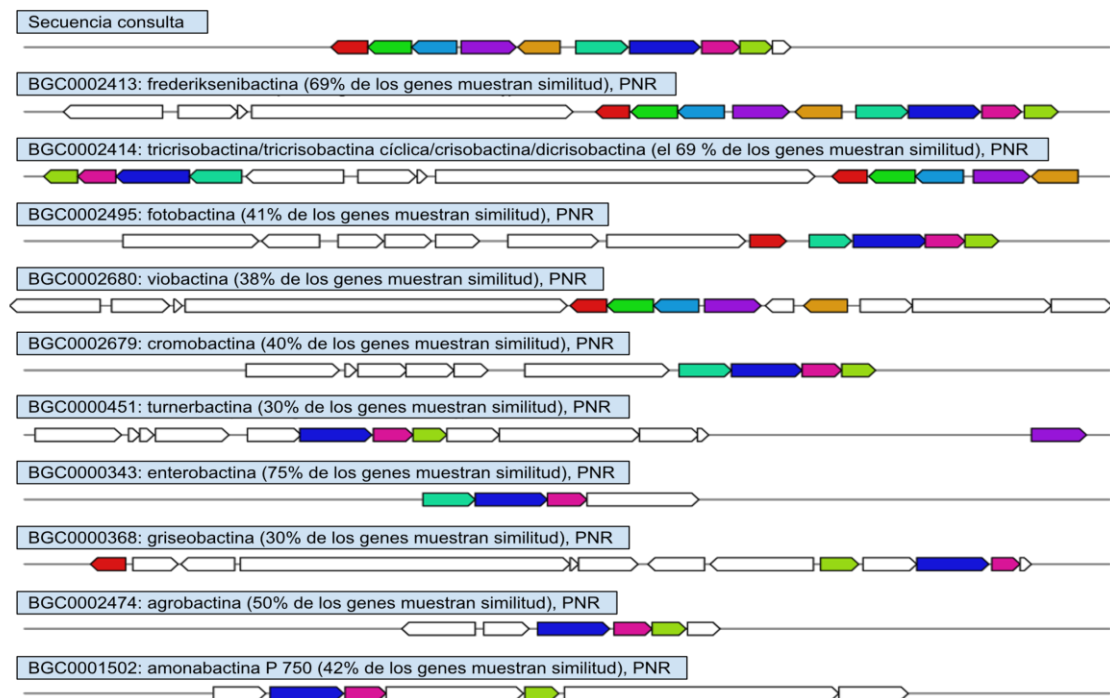


Figura 4. CGB de MiBIG similares al CGB de la enterobactina. El CGB de la parte superior corresponde al grupo de genes que codifican la *enterobactina sintasa*, los genes biosintéticos centrales se encuentran el azul fuerte (*EntE*) y rosa (*EntB*). Hay genes adicionales que realizan modificaciones postraduccionales como los resaltados en color azul celeste (*EntC*) y rosa pastel (*EntA*). El resto son genes asociados al transporte del producto final. Imagen modificada de la entrada de la enterobactina en la base de datos de MiBIG.<sup>19</sup>

La *enterobactina sintasa* se constituye por dos módulos, el primero está codificado en ambos genes *EntE* y *EntB*, y el segundo módulo se encuentra codificado en el gen *EntF*.

Estos módulos a su vez se dividen en seis dominios. El gen *EntE* codifica a un único dominio de adenilación mientras que el gen *EntB* se traduce en un único dominio de PTP. Por último, el componente resultante del gen *EntF* contiene los cuatro tipos de dominios esenciales para una SPNR.

Los dominios codificados en *EntF* se encuentran en el orden C-A-PCP-TE, por esta característica, el componente F de la *enterobactina sintasa* es

más completo y autosuficiente por sí mismo, y por ende *EntF* es un gen más característico de este tipo de enzimas.

Los dominios que conforman el componente F de la *enterobactina sintasa* están contenidos en 8 super familias en la base de datos de InterPro (Figura 5), sin embargo, estos pueden reducirse a un dominio de condensación (C), un dominio de adenilación (A), un dominio de PTP y un dominio de terminación o tioesterasa (TE).

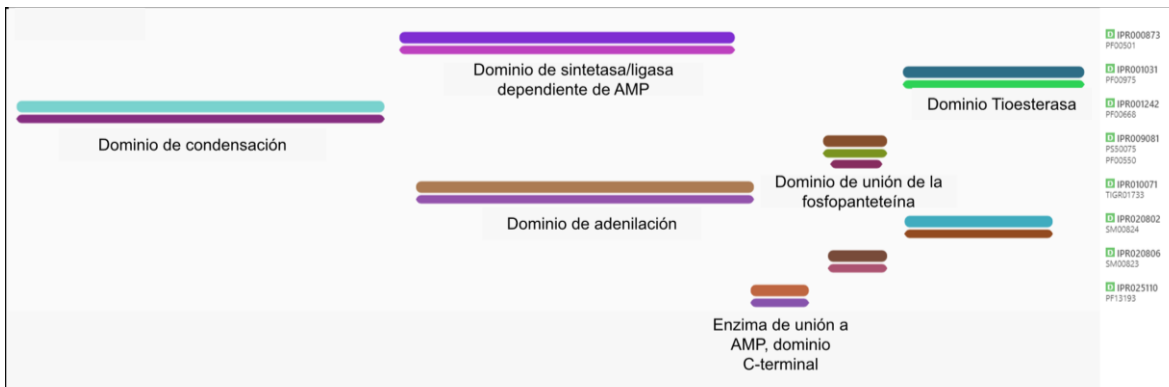


Figura 5. Familias de dominios de InterPro del componente F de la enterobactina sintasa.

Los dominios que conforman el gen monomodular de la *enterobactina sintasa*; se compone de los cuatro dominios característicos de las SPNR, estos a su vez están clasificados dentro de 8 superfamilias de proteínas de InterPro. Imagen tomada y modificada de InterPro.<sup>83</sup>

De acuerdo con el estudio de Ehmann et al.<sup>86</sup>, el dominio de adenilación del gen *EntF* es altamente específico para L-Serina y su uso en una búsqueda por homología delimitaría los resultados a aquellos PNR que contengan este aminoácido.

El dominio TE restringe la búsqueda a péptidos liberados por hidrólisis (péptidos lineales) o por macro ciclación.<sup>67</sup>

La selección del dominio A o TE no habría aportado los suficientes grados de libertad a la búsqueda por BLAST para obtener una gran variedad estructural de PNR.



El dominio de PTP es un dominio que, si bien es esencial para las SPNR, no es exclusivo de las mismas y puede ser encontrado en PKSs o en CGB híbridos con este tipo de metabolitos.

Finalmente, el dominio de condensación seleccionado para la búsqueda por homología tiene la función de unir un residuo no proteico (ácidos carboxílicos de arilo) con un aminoácido (L-Serina).<sup>86</sup>

Se sabe por el estudio de Kaniusaite et al. que los dominios de condensación que se encuentran río arriba de un dominio de adenilación son los responsables de mediar las modificaciones estructurales que pudieran tener los residuos seleccionados por este último, como es el caso de la *enterobactina sintasa*, en donde la L-Serina es modificada a seriladenilato, producto intermediario de la reacción de formación de la enterobactina.<sup>44</sup>

Estas dos características permiten la posibilidad de realizar una búsqueda donde se encuentren secuencias que aprovechen la capacidad de las SPNR de incluir residuos no proteinogénicos así como de realizar modificaciones estructurales específicas y con esto explotar la diversidad estructural que ofrecen los PNR.

En el material anexo, se presenta la secuencia del dominio de condensación del gen *EntF* en formato FASTA, obtenida de la base de datos de InterPro (Secuencia 1).

### 8.3. Resultados de la búsqueda por homología con BLAST

Se empleó la secuencia Dominio\_C\_EntF (Secuencia 1) en la herramienta de blastp de la plataforma BV-BRC dentro del grupo de genomas de BAP recolectados en la primera parte de la metodología (Tabla I).

Los resultados del BLAST fueron 500 secuencias homólogas, pertenecientes a distintos tipos de clústeres.

De los 384 genomas recolectados, 214 (55.73%) mostraron tener la presencia de al menos una secuencia homóloga al dominio C. Si bien este tipo de dominio es característico de las SPNR, no se le puede asociar directamente a la presencia de estas enzimas en los genomas resultantes. Se requiere un análisis de minería genómica para confirmar estas predicciones.

Las características más importantes de los genomas resultantes del BLAST se muestran en la Tabla II.

Hubo mayor representación de algunos géneros bacterianos en la recolección de genomas de BAP, por ejemplo, para el género *Bacillus* y *Pseudomonas* hay un total de 63 y 44 genomas respectivamente mientras que para géneros como *Aeribacillus*, *Erwinia* y *Ralstonia* sólo hubo un genoma por género que cumplió con los criterios mencionados en la primera parte de la metodología. Por esa razón, se tratarán los siguientes resultados en valores absolutos (divididos entre el total de bacterias por género recolectadas en la tabla I).

Sin embargo, estos resultados no son probabilísticamente acertados debido a que no es posible realizar un análisis estadístico representativo ya que los genomas recolectados fueron específicamente seleccionados con base en los criterios mencionados anteriormente, perdiendo el factor de aleatoriedad necesario y a que en algunos casos los datos son insuficientes para tal tipo de estudio.

Aun así, como se mencionó en la sección 8.1, los genomas recolectados cuentan con suficiente variedad genómica (Tabla I) para sustentar los resultados obtenidos por este estudio, asociándose con mayor grado de confianza a sus características genéticas que a la filogenia como tal, que para tal caso, también es posible relacionar estas características genéticas a un género bacteriano específico.

Los cinco géneros bacterianos con menor número de representantes con presencia de dominios de condensación fueron *Micromonospora* (5.6%), *Rahnella* (9.1%), *Rouxiella* (25.0%), *Acinetobacter* (28.6%) y *Curtobacterium* (28.6%).

Por otra parte, los géneros bacterianos con mayor número de representantes, sin contar aquellos que tienen menos de 5 miembros por grupo de género, fueron *Nocardia* (100%), *Enterobacter* (100%), *Bacillus* (96.8%), *Serratia* y *Burkholderia* (84.2% cada uno), y a pesar de tener menos de 5 representantes, todas las bacterias analizadas de los géneros *Actinomadura*, *Rhodococcus*, *Stenotrophomonas* y *Xanthomonas* tuvieron la presencia de al menos un dominio de condensación.

Las 214 bacterias resultantes del análisis por BLAST fueron aisladas de una gran variedad de plantas y su rizósfera (Figura 6), entre ellas predominan 22 especies de las que se aisló más de una bacteria que contuviera al menos un dominio de condensación, siendo que las bacterias no necesariamente provienen de la misma muestra de planta.

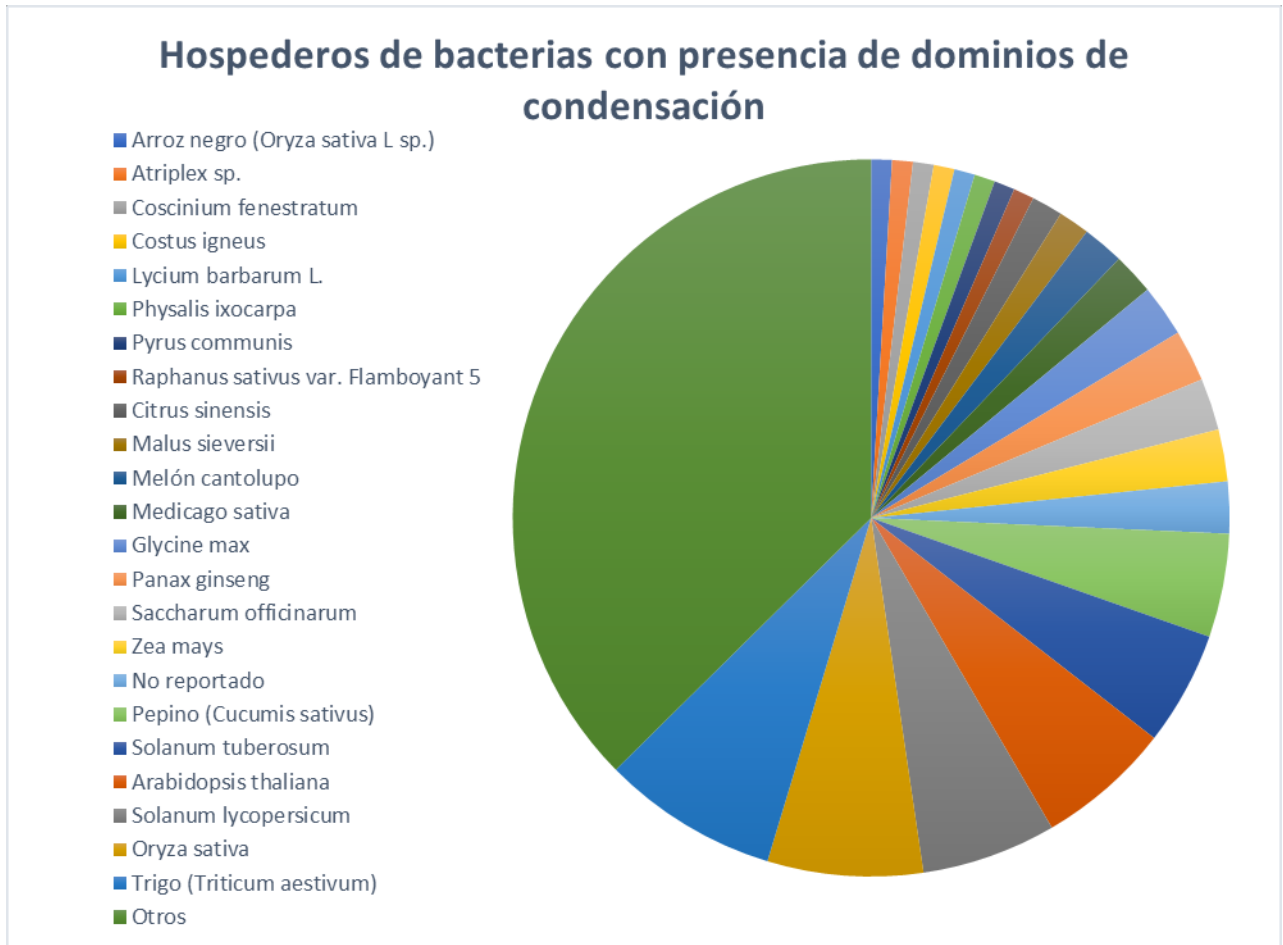


Figura 6. Especies de plantas de las cuales se aislaron las 214 BAP. En el gráfico se muestran las hospederas de las BAP que tienen en sus genomas secuencias homólogas con el dominio de condensación del gen *EntF* así como la recurrencia de BAP con esta característica en cada una de ellas.

Los detalles de estas 22 especies se encuentran contenidos en la Tabla 3. Esta información es relevante debido a la capacidad de algunas BAP de producir compuestos similares o idénticos a los de sus hospederas.

Las especies de plantas con mayor cantidad de bacterias con dominios C fueron *Arabidopsis thaliana*, *Triticum aestivum*, *Oryza sativa*, *Zea Mays*, *Solanum tuberosum*, *Solanum lycopersicum*, *Cucumis sativus*, *Saccharum officinarum* y *Panax ginseng*.

La planta *Arabidopsis thaliana* es una de las plantas más utilizadas en genética y biología molecular debido a sus características de tener un tamaño de genoma pequeño, contenido bajo en regiones de DNA repetitivas dispersas, gran número de marcadores genéticos y su corto ciclo de vida.<sup>88</sup>

Tabla 3. Características generales y farmacológicas de las plantas hospederas de BAPs con dominios C

Hospedero	Bacterias con dominio C aisladas del hospedero	Características del hospedero	Propiedades farmaceuticas
Arroz negro ( <i>Oryza sativa</i> L sp.)	2	La coloración negra es debido al pigmento antocianina, un antioxidante. <sup>92</sup>	Se supone que este arroz aumenta la longevidad. <sup>93</sup>
<i>Atriplex</i> sp.	2	Usado para alimentar animales	Laxante, antioxidante, tratamiento para enfermedad hepática, ictericia. <sup>94,95</sup>
<i>Coscinium fenestratum</i>	2	Considerada como una planta medicinal en peligro crítico, pertenece a la familia Menispermaceae.	Utilizado para el tratamiento de la diabetes mellitus, propiedades antimicrobianas, antidiabéticas, antiinflamatorias, antioxidantes. <sup>96</sup>
<i>Costus igneus</i>	2	Conocida como la "planta insulina".	Tratamiento para diabetes mellitus. <sup>97</sup>
<i>Lycium barbarum</i> L.	2	Contienen diversas variedades de nutrientes y elementos bioactivos.	Antioxidante, inmunomodulador, antiinflamatorio, antiaterogénico, antifibrótico, antimutagénico, antitumor, antirradiación, antifatiga, antienvjecimiento, antiestrés, antilevadura, cardioprotector y cicatrizante. <sup>98</sup>
<i>Physalis ixocarpa</i>	2	Nativo de México y centro américa, crece en regiones tropicales y sub tropicales.	La planta y la fruta del tomatillo contiene fisalinas, que tienen efectos antimicrobianos y antinarcóticos. <sup>99</sup>
<i>Pyrus communis</i>	2	Peral europeo o peral común, es una especie de árbol caducifolio de la familia de las rosáceas.	N/A
<i>Raphanus sativus</i> var. Flamboyant 5	2	Los rábanos se cultivan y consumen en todo el mundo, siendo mayormente consumido crudo.	N/A
<i>Citrus sinensis</i>	3	Planta que da el fruto de la naranja. <sup>6,7</sup>	Su extracto tiene propiedades antitiroideas, hipoglucemiantes y estimulantes de la insulina además de efectos antibacteriales para tratamientos bucales. <sup>95</sup>
<i>Malus sieversii</i>	3	Es una especie de manzana silvestre nativa de las montañas de Asia Central.	N/A

Melón cantalupo	4	Melón variedad de la especie Cucumis melo de la familia Cucurbitaceae.	N/A
<i>Medicago sativa</i>	4	Sus hojas o semillas se venden como hierba en polvo, cápsulas y tabletas para suplementos nutricionales en tiendas naturistas.	Diurético, anticancerígeno, antirreumático, cardiotónico, depurativo, lactagogo, emenagogo, antiescorbútico y en el tratamiento de forúnculos. <sup>99</sup>
<i>Glycine max</i>	5	llamada soya o soja, es una especie de la familia Fabaceae, o de las leguminosas. Se cultiva por sus semillas, de contenido medio en aceite y alto de proteína.	N/A
<i>Panax ginseng</i>	5	El nombre <i>panax</i> significa “cura todo”, nativo de Corea, China y Japón.	antioxidante, antiinflamatorio, vasorelajación, antialérgico, antidiabético y anticancerígeno. <sup>100</sup>
<i>Saccharum officinarum</i>	5	Caña de azúcar, es uno de los cultivos más abundantes en el mundo, posee una variedad de bioactividades debido a sus innumerables constituyentes.	Acantioxidante, anticancerígeno, antifibrótico y antitrombótico. <sup>103</sup>
<i>Zea mays</i>	5	Actualmente es el cereal con el mayor volumen de producción a nivel mundial seguido por el trigo y el arroz.	N/A
Pepino ( <i>Cucumis sativus</i> )	10	Pertenece a la familia Cucurbitaceae, se consume de diferentes maneras como verdura y ensaladas.	actividad antibacteriana, antifúngica, citotóxica, antiácida y carminativa, contra la colitis ulcerosa, hepatoprotectora, hipoglucémica e hipolipidémica. <sup>101</sup>
<i>Solanum tuberosum</i>	11	La papa pertenece a la familia de las solanáceas, es actualmente el cuarto cultivo alimentario básico más importante del mundo.	Actividad antioxidante, anticancerígena, antialérgica, antibacteriana, antiinflamatoria, antiobesidad, antiulcerosa. <sup>90</sup>
<i>Arabidopsis thaliana</i>	13	Su corto ciclo de vida, tamaño pequeño de genoma y peptidos la hacen óptima para experimentos genéticos. <sup>88</sup>	N/A
<i>Solanum lycopersicum</i>	13	Subproductos del procesamiento del tomate, en particular las semillas, son reservorios de macromoléculas que promueven la salud, como las proteínas, carotenoides, polisacáridos, fitoquímicos y vitaminas.	Semillas: Tratamiento de dolencias cardiovasculares, trastornos neurológicos, y agentes antioxidantes, anticancerígenos y antimicrobianos. <sup>104</sup>
<i>Oryza sativa</i>	15	El arroz es el principal alimento básico para más de la mitad de la población mundial.	Remedio contra inflamaciones, dolencias gastrointestinales, hipercolesterolemia, diabetes y enfermedades de la piel. <sup>89</sup>
Trigo ( <i>Triticum aestivum</i> )	17	Uno de los cereales más consumidos del mundo. Excelente fuente de energía, que contiene una alta cantidad de carbohidratos, proteínas y fibra. <sup>91</sup>	N/A

*Triticum aestivum* (trigo), *Oryza sativa* (arroz), *Zea Mays* (maíz) y *Solanum tuberosum* (papa) son los cuatro cultivos más abundantes y consumidos en todo el mundo.<sup>89-91</sup> Casi no hay restricciones geográficas para estos cultivos y el estudio de su microbiota se ha facilitado a nivel global. Estas especies son ricas en almidón y polisacáridos y en menor proporción de compuestos aromáticos.

*Solanum lycopersicum* (jitomate) y *Cucumis sativus* (pepino) también son producidos a gran escala. Ambas especies se componen de raíces, tallo, hojas, flores y en especial de frutos destinados para el consumo humano. Ambos tienen propiedades antimicrobianas.<sup>101-104</sup> Los componentes químicos que abundan en la planta del jitomate son carotenoides, pectina, vitaminas y al igual que la planta del pepino abundan fitoquímicos como los flavonoides.<sup>101-104</sup>

*Saccharum officinarum* (caña de azúcar) es producida a gran escala a nivel global, constan de tallos largos y gruesos con vainas en los extremos, sus constituyentes químicos son en su mayoría compuestos aromáticos, policosanoles, aldehídos de cadena larga, triterpenoides, esteroides y ácidos grasos.<sup>103</sup> Al igual que el *Panax ginseng* (Ginseng), que sólo es cultivado en algunas regiones de Asia, ambos tienen actividad antioxidante y anticancerígena.<sup>100,101</sup> Los compuestos activos de este último son denominados ginsenósidos y en general son esteroides hidrofóbicos de cuatro anillos con azúcares.<sup>100</sup>

Las bacterias resultantes del BLAST fueron aisladas de muestras de regiones distribuidas en América, Asia, África, Europa y Oceanía (Figura 7).

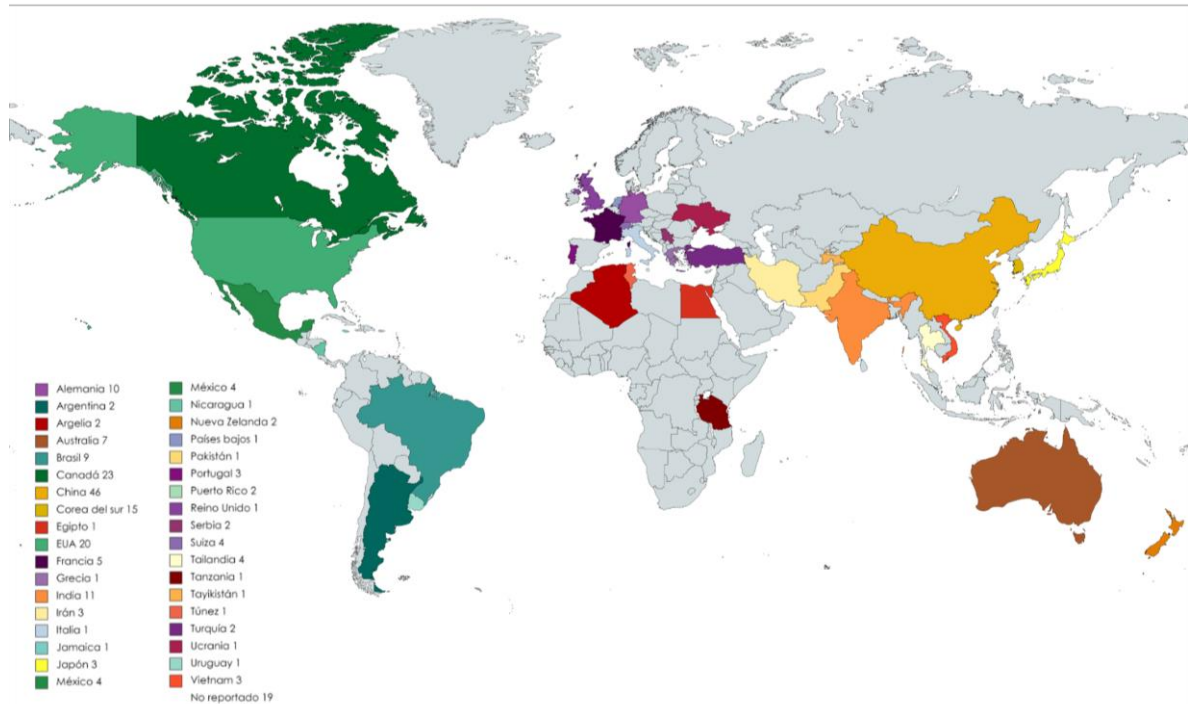


Figura 7. Regiones de donde se tomaron las muestras de las que se aislaron las 214 BAP. En la imagen se muestran resaltadas en color las regiones de las que se aislaron las BAP que tienen en sus genomas secuencias homólogas con el dominio de condensación del gen *EntF*.

Las estructuras biológicas, componentes químicos y distribución de las plantas son muy variadas entre las hospederas.

La presencia de péptidos con dominio C, por tanto, no se ve relacionada con factores ambientales y muy probablemente su propósito no esté asociado a la colonización específica.

Sin embargo, su abundancia sugiere que la bioactividad de estas enzimas tiene una función importante en la simbiosis entre bacteria y planta, tan sólo en los genomas de la Tabla II del material anexo, hubo presencia de CGB de SPNR en 207 de los 214 genomas, y en los 7 restantes hubo CGB híbridos con dominios asociados a SPNR.



#### 8.4. Minería genómica

En total se predijeron 3117 clústeres que se clasificaron en siete grupos de CGB: SPNR, PKS, PsRMP, terpenos, sideróforos, híbridos y otros, se realizó el análisis de la abundancia de estos grupos (Figura 8).

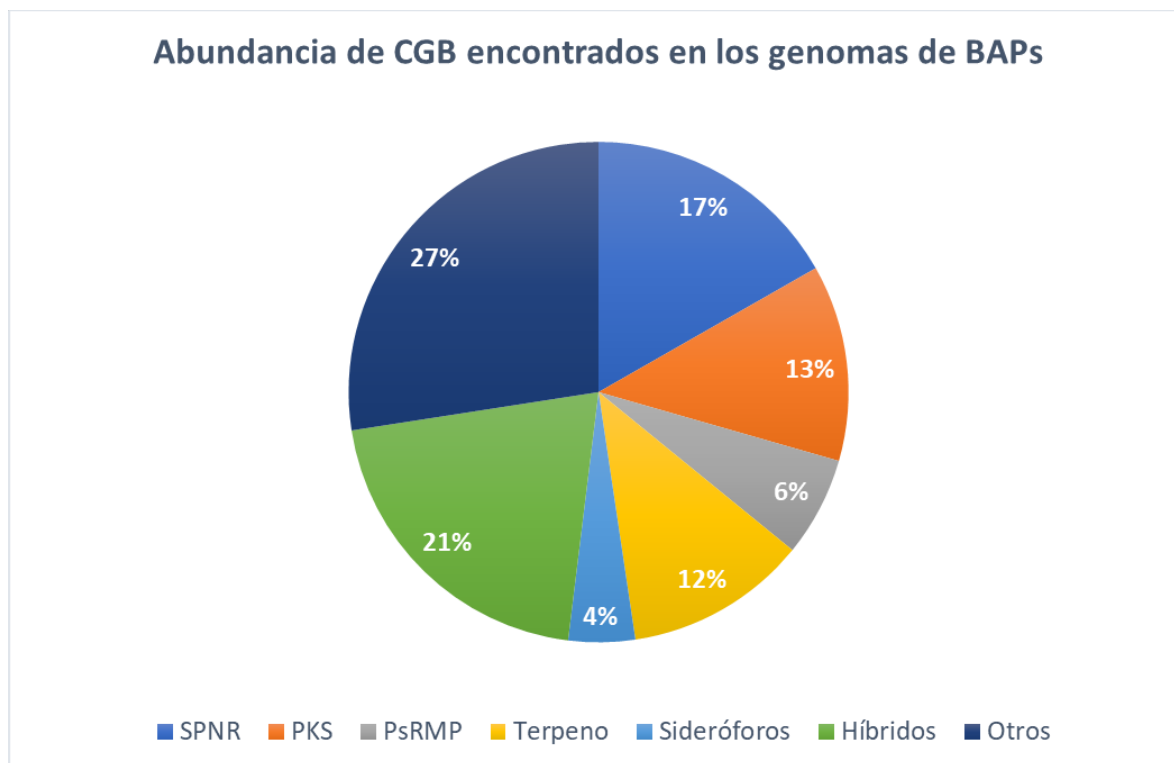


Figura 8. Presencia promedio de cada uno de las clases de CGB de metabolitos especializados predichos por AntiSMASH. En el gráfico se muestra la abundancia de cada clase de metabolito especializado encontrados en los 214 genomas con presencia de dominio C.

En los genomas analizados abundan los híbridos que combinan dos o más clases de metabolitos, de los cuales, la mayoría (70.66% de los CGB híbridos resultantes) contenían genes con dominios putativos de las SPNR.

Esto podría explicarse con el modo de evolución de estas enzimas, que consta en la expansión o duplicación de regiones génicas dentro del mismo clúster para ser alteradas por mutaciones posteriormente.<sup>2</sup>

Después de este proceso es muy poco probable que el producto del CGB adquiera una funcionalidad importante desde el inicio, sin embargo, el deshacerse de esta región se vuelve un proceso complicado y de gran consumo energético para la célula debido a que suelen ser regiones de DNA muy grandes y de alto contenido de guanina y citosina, por lo que se opta por realizar más modificaciones a la secuencia hasta que esta adquiera funcionalidad, como la inserción de genes de otra clase de metabolito, generando así híbridos.<sup>105</sup>

Un ejemplo de esto podría verse en la Figura 4, en donde es claro que la ausencia de genes que rodean a los genes parecidos a *EntE* y *EntB* (azul y rosa fuerte) son consecuencia de los procesos antes mencionados, sin embargo, aún se requiere evidencia experimental para sustentar esta teoría.

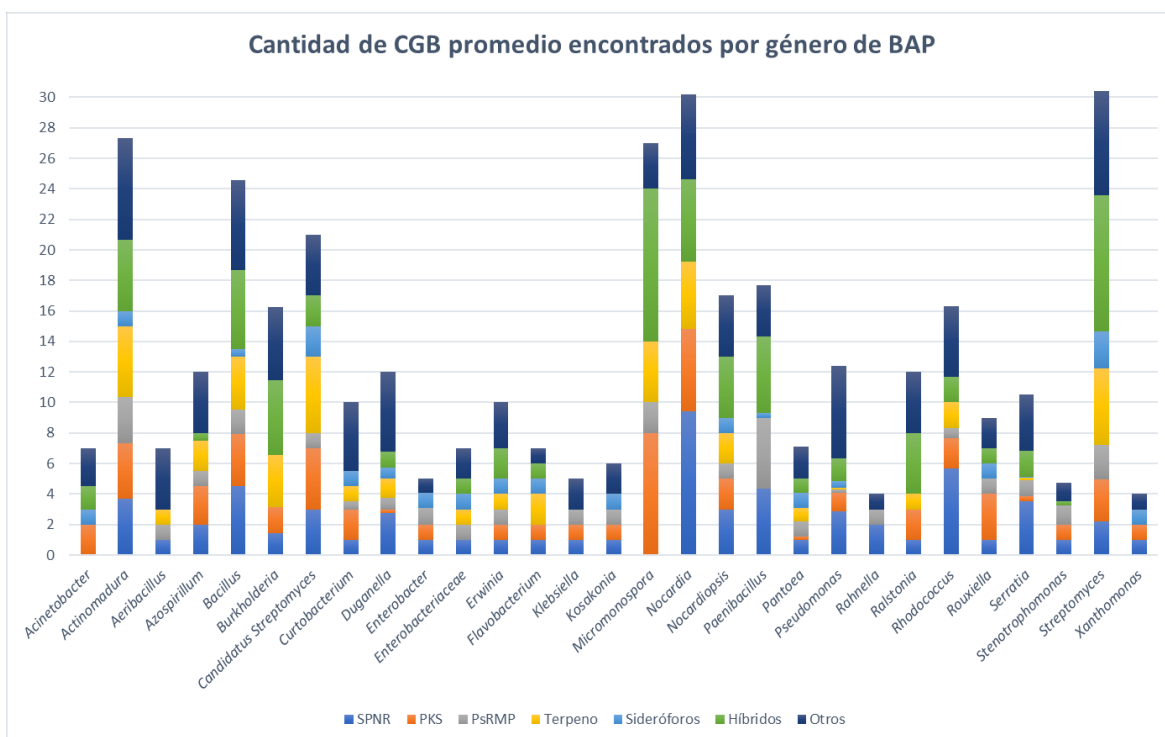


Figura 9. Presencia promedio de cada una de las clases de CGB de metabolitos especializados predichos por AntiSMASH por género bacteriano. En el gráfico se muestra la abundancia de cada clase de metabolito especializado por género bacteriano, encontrados en los 214 genomas con presencia de dominio C.

El siguiente grupo más grande de CGB predichos en los genomas de BAP con presencia de dominios C corresponde a las SPNR (17%).

Se realizó un perfil de metabolitos especializados promedio por género bacteriano (Figura 9) y las SPNR fueron el tipo de CGB más predicho en las especies de *Bacillus*, *Duganella*, *Nocardia*, *Nocardiopsis*, *Pseudomonas*, *Rahnella*, *Rhodococcus* y *Serratia*, además estos CGB estuvieron presentes en todos los genomas analizados.

Para *Acinetobacter* y *Micromonospora* se presentaron en forma de híbridos.

Para el resto de los géneros analizados predominaron los terpenos, PKSs y, para el caso particular de *Paenibacillus*, los PsRMP.

Los terpenos son compuestos comunes en las bacterias asociadas a plantas, en especial en los epífitos (bacterias que habitan la superficie de las plantas) que producen una gran variedad de carotenoides (tetraterpenos) debido a que son capaces de captar y procesar la luz UV de los rayos del sol.<sup>65,106</sup>

Además, se cree que la pigmentación, función de los terpenos y arilpolienos (PKs), es un factor que favorece la colonización de la filosfera (superficie de la planta) por parte de las bacterias.<sup>33,107,108</sup>

Estos procesos son beneficiosos tanto para la bacteria como para la planta, siendo las PKSs y los terpenos importantes para el establecimiento de la simbiosis entre ambos organismos.

La función de los SPNR, como se había mencionado anteriormente, podría estar relacionada de igual forma con mecanismos que promuevan la formación del holobionte.

Entre estos mecanismos se conocen la fijación de nitrógeno, solubilización de fosfatos y metales como hierro y zinc, producción de fitohormonas

como el ácido indol-3-acético (AIA), degradación de compuestos tóxicos como especies reactivas de oxígeno, facilitación de adquisición de nutrientes por transportadores de membrana o sideróforos y la producción de compuestos de carácter antimicrobiano para la producción contra fitopatógenos.<sup>23,26,66,80,86,109-111</sup>

Si bien hay reportes de actividad de captación y solubilización de metales por parte de las SPNR, su bioactividad se relaciona más comúnmente con la defensa contra fitopatógenos.<sup>27,111,112</sup> Esta propiedad de inhibición puede aprovecharse para atacar patógenos que provoquen enfermedades humanas, tal como se muestra en la Tabla 1.

#### **8.5. Análisis filogenéticos de los genomas resultantes de BAPs**

De los 214 genomas resultantes, los géneros de *Bacillus* (Gram positivo) y *Pseudomonas* (Gram negativo) fueron los que tuvieron una mayor representación (61 y 34 cepas, respectivamente), por lo que se construyó un árbol filogenético para ambas especies con el fin de seleccionar un determinado número de cepas representativas.

Las figuras de los árboles se presentan en el material anexo; Figura I para el género *Bacillus* y Figura II para *Pseudomonas*.

Los géneros representativos para cada género, por clado filogenético fueron: *Bacillus safensis* strain PgKB20, *Bacillus altitudinis* strain Ba1449, *Bacillus cereus* strain 21155, *Bacillus licheniformis* strain ME35, *Bacillus subtilis* strain UD1022, *Bacillus velezensis* strain MEJ05, *Pseudomonas viridiflava* strain KF485, *Pseudomonas putida* strain PC2, *Pseudomonas denitrificans* (nomen *rejiciendum*) strain BG1, *Pseudomonas* sp. Q1, *Pseudomonas chlororaphis* subsp. *aureofaciens* strain ChPhzTR18, *Pseudomonas putida* strain PgBE89 y *Pseudomonas brassicacearum* strain TM1A3.

Con estas selecciones se construyeron dos árboles filogenéticos más, uno para las bacterias Gram positivas y otro para las Gram negativas resultantes del BLAST de la sección anterior, los resultados se muestran en las Figuras 10 y 11 respectivamente.

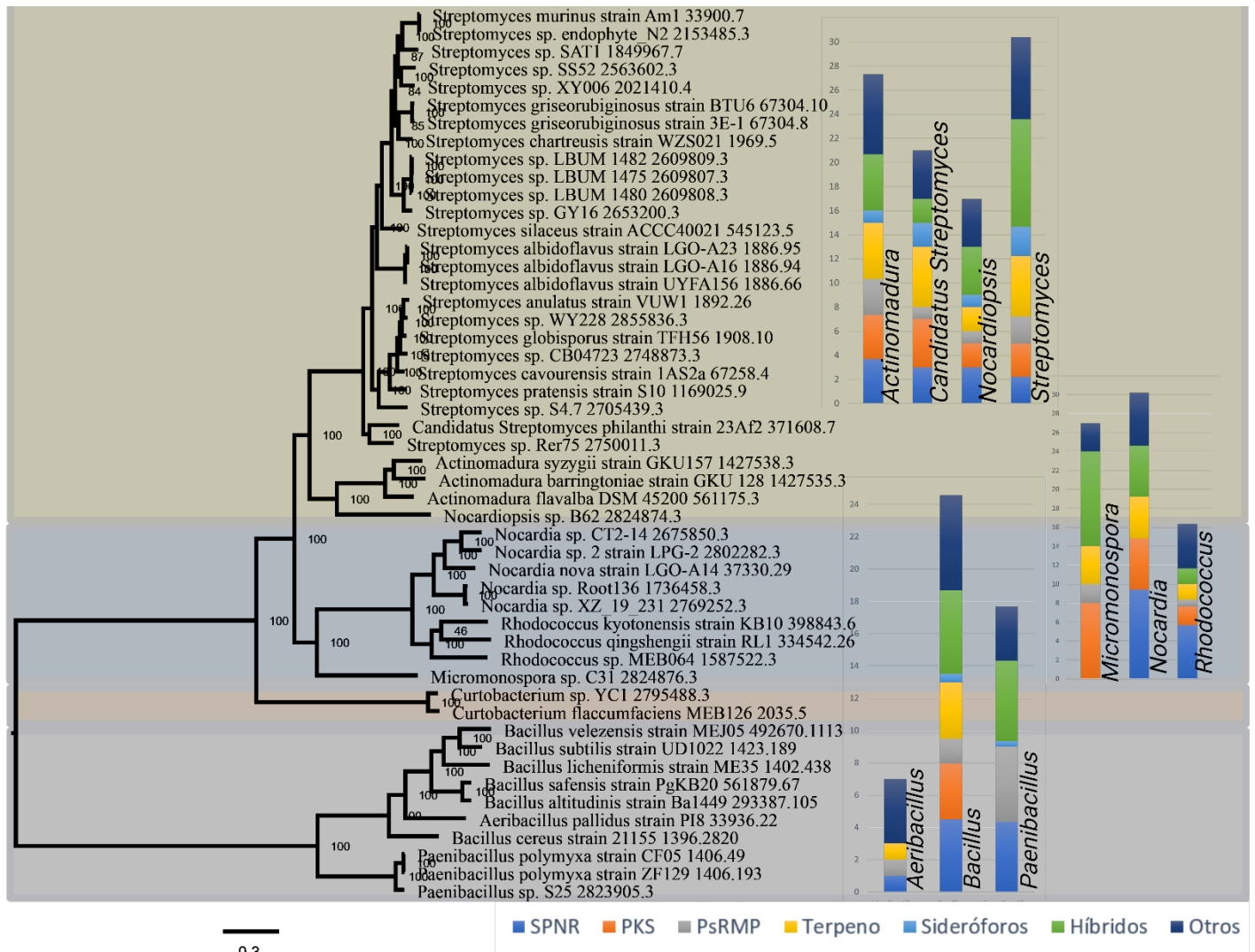


Figura 10. Árbol de BAP Gram positivas. Genomas resultantes del BLAST con un dominio de condensación perteneciente a una SPNR. Alineamiento realizado por MAFFT y analizado por RAXML en la plataforma BV-BRC usando el servicio "Construcción de árbol filogenético". En color azul se encuentran marcados los clados filogenéticos cuyo metabolito especializado más abundante fueron las SPNR y las PKS, en amarillo los terpenos, en naranja los PsRMP y en gris cualquier otro tipo de metabolito.

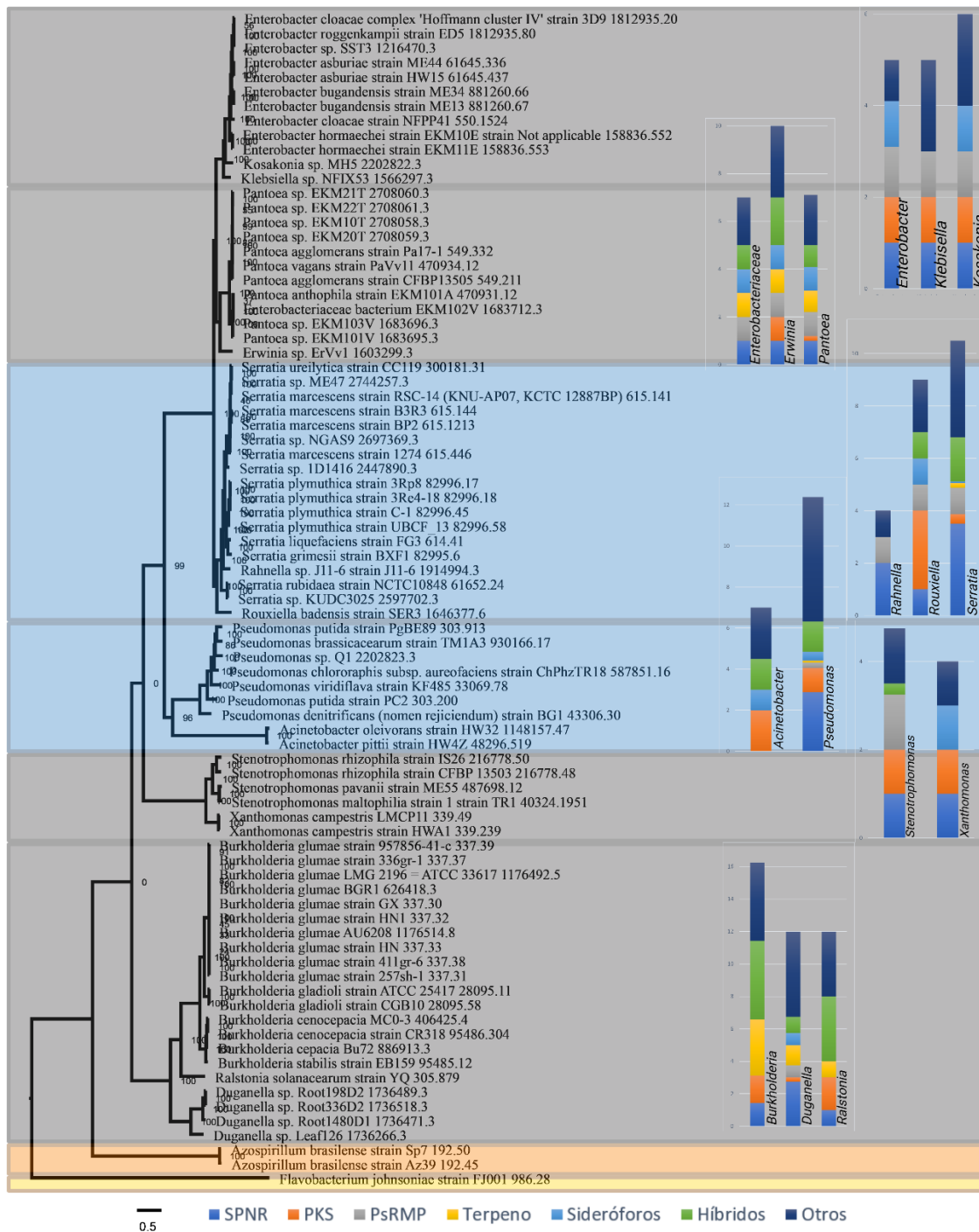


Figura 11. Árbol de BAP Gram negativas. Genomas resultantes del BLAST con un dominio de condensación perteneciente a una SPNR. Alineamiento realizado por MAFFT y analizado por RAxML en la plataforma BV-BRC usando el servicio "Construcción de árbol filogenético". En color azul se encuentran marcados los clados filogenéticos cuyo metabolito especializado más abundante fueron las SPNR y las PKS, en amarillo los terpenos, en naranja los PsRMP y en gris cualquier otro tipo de metabolito.

El perfil de metabolitos especializados parece estar fuertemente ligado a la filogenia de las bacterias productoras.

En la Figura 10 se observa un clado filogenético que contiene a los géneros de *Streptomyces*, *Actinomadura* y *Nocardiopsis*, en general estos tres géneros contienen en su mayoría CGB asociados a terpenos, además los tres géneros tienen presencia de todas las clases de metabolitos analizados en este estudio, en abundancias similares.

El siguiente clado contiene a los géneros *Nocardia*, *Rhodococcus*, *Micromonospora*, los cuales producen mayoritariamente SPNR y PKSs.

A pesar de que se encuentran más metabolitos de cada clase en los genomas de *Nocardia*, parecen tener una proporción y presencia similar con los encontrados en el género *Rhodococcus*.

Por otra parte *Micromonospora* difiere principalmente por su carencia de CGB de SPNR sin embargo este hecho parece compensarse con la abundancia de clústeres híbridos, de los cuales siete de los diez CGB híbridos encontrados en su genoma contienen dominios asociados a la clase de las SPNR.

En el clado filogenético en el que se encuentran los genomas de las bacterias *Bacillus*, *Aeribacillus* y *Paenibacillus* no parece haber similitudes significativas, sin embargo, es un resultado obtenido de comparar 61 cepas de *Bacillus* contra 1 y 3 cepas de las dos últimas mencionadas respectivamente, para hacer afirmaciones acerca de las similitudes en el perfil de metabolitos especializados de estas cepas se requiere un análisis con más datos.

En la Figura 11 se observa el mismo patrón para las bacterias Gram negativas; el clado con los géneros *Enterobacter*, *Kosakonia* y *Klebsiella* son similares en tanto a la clase de metabolitos producidos como a su

abundancia, con excepción de los sideróforos que no se encuentran en los genomas de *Klebsiella*, género que también se encuentra más distanciado dentro del clado filogenético analizado.

El perfil de metabolitos especializados de *Enterobacteriaceae* y *Pantoea* difiere únicamente en la presencia de unos cuantos CGB de PKSs en los genomas de este último, además, el análisis de árbol filogenético los ubica más cercanos entre sí que con el género *Erwinia*, el cual se encuentra en el mismo clado y en cuyos genomas se predijeron metabolitos similares en abundancias, siendo las diferencias más notorias una mayor cantidad de CGB híbridos y de PKSs.

Los géneros de *Stenotrophomonas* y *Xanthomonas* contienen en sus genomas la misma cantidad de CGB de SPNR y PKSs, los cuales constituyen alrededor de la mitad del promedio de los CGB predichos para estos géneros.

El caso en el que más diferencias se encontraron fue el clado constituido por las bacterias *Burkholderia*, *Ralstonia* y *Duganella* donde la predominancia de CGB fue de la clase híbrida o de otro tipo de metabolitos que no entraron en la clasificación de los analizados en este estudio.

Los híbridos no solamente son los más abundantes en los genomas analizados, sino que también se sabe que tienen un gran potencial de bioactividad.<sup>113-114</sup>

Es importante llamar la atención a que en la mayoría de los clados filogenéticos donde abundan las SPNR y PKSs, también se encuentran en gran número híbridos de ambos tipos de metabolitos.

Esto se debe a las grandes similitudes que comparten ambas enzimas, por ejemplo, es posible relacionar el número de residuos que conformarán en producto final (PNR o PK) con el número de módulos que componen la



enzima que los sintetiza, de igual forma esta organización modular está relacionada con la elongación del péptido final.<sup>68,109,115</sup>

Además, se sabe por análisis bioinformáticos que ambas enzimas evolucionan por medio de fusiones entre sub clústeres que funcionan de forma independiente y que sufren modificaciones por mutaciones posteriores o intercambios de información genética por eventos de recombinación robusta.<sup>2,115</sup>

Las diferencias principales entre este tipo de enzimas van en la estructura cuaternaria de las mismas, que para el caso de los PKSs suelen adquirir la forma de homodímeros mientras que las SPNR son monómeros de manera natural, además el mecanismo de elongación de los PKSs está basado en reacciones de tipo condensación de Claisen, mientras que las SPNR establecen la formación de enlaces peptídicos por medio de sus dominios de condensación.<sup>68,116</sup>

El estudio de Theobald et al. demostró una relación evolutiva entre los híbridos de PKSs y SPNR en el género fúngico *Aspergilli*, estos comparten ancestros de CGB de PKSs en común, también se indica que la versatilidad genética de las SPNR es mucho mayor y pueden divergir entre ellas con mayor facilidad.<sup>117</sup>

En el mismo estudio se hipotetiza la divergencia de ambas clases de enzimas de un metabolito especializado que sea un ancestro en común de ambas, lo cual explicaría la concurrencia con la que se presentan estas dos enzimas en los genomas bacterianos.

## 8.6. Análisis de redes de similitud

Las FCG resultantes del análisis se muestran en la Figura 12. El resultado de este análisis fue la agrupación de los CGB en 85 FCG y 169 clústeres huérfanos (que no se agruparon en ninguna familia).

No hubo copias de los CGB en el mismo genoma de ninguna de las bacterias analizadas, esto probablemente por la dificultad de la célula de replicar regiones tan grandes de DNA, lo cual requeriría un gran consumo energético invertido en una enzima que si bien pudiera ofrecer mejores capacidades de adaptación, no es parte del metabolismo central de la célula.

Las duplicaciones que se realizan están más bien asociadas a fragmentaciones del CGB o de genes de este y posteriores fusiones con otros fragmentos o clústeres completos de SPNR como forma de su mecanismo de evolución.<sup>2,115</sup>

En la Figura 13 se muestra un mapa jerárquico de las 20 FCG más grandes, como era de esperarse los géneros bacterianos con mayor representación; *Bacillus* (61), *Burkholderia* (16), *Enterobacter* (10), *Pantoea* (10), *Pseudomonas* (34), *Serratia* (16) y *Streptomyces* (25), fueron también los que aportaron un mayor número de CGB a estas FCG.

Dentro de las 50 familias más numerosas se encuentran también CGB de los géneros de *Azospirillum* (2) con dos familias de dos CGB cada una, *Nocardia* (5) con dos familias de 3 y 2 CGB y *Paenibacillus* (3) con dos familias de tres CGB cada una.

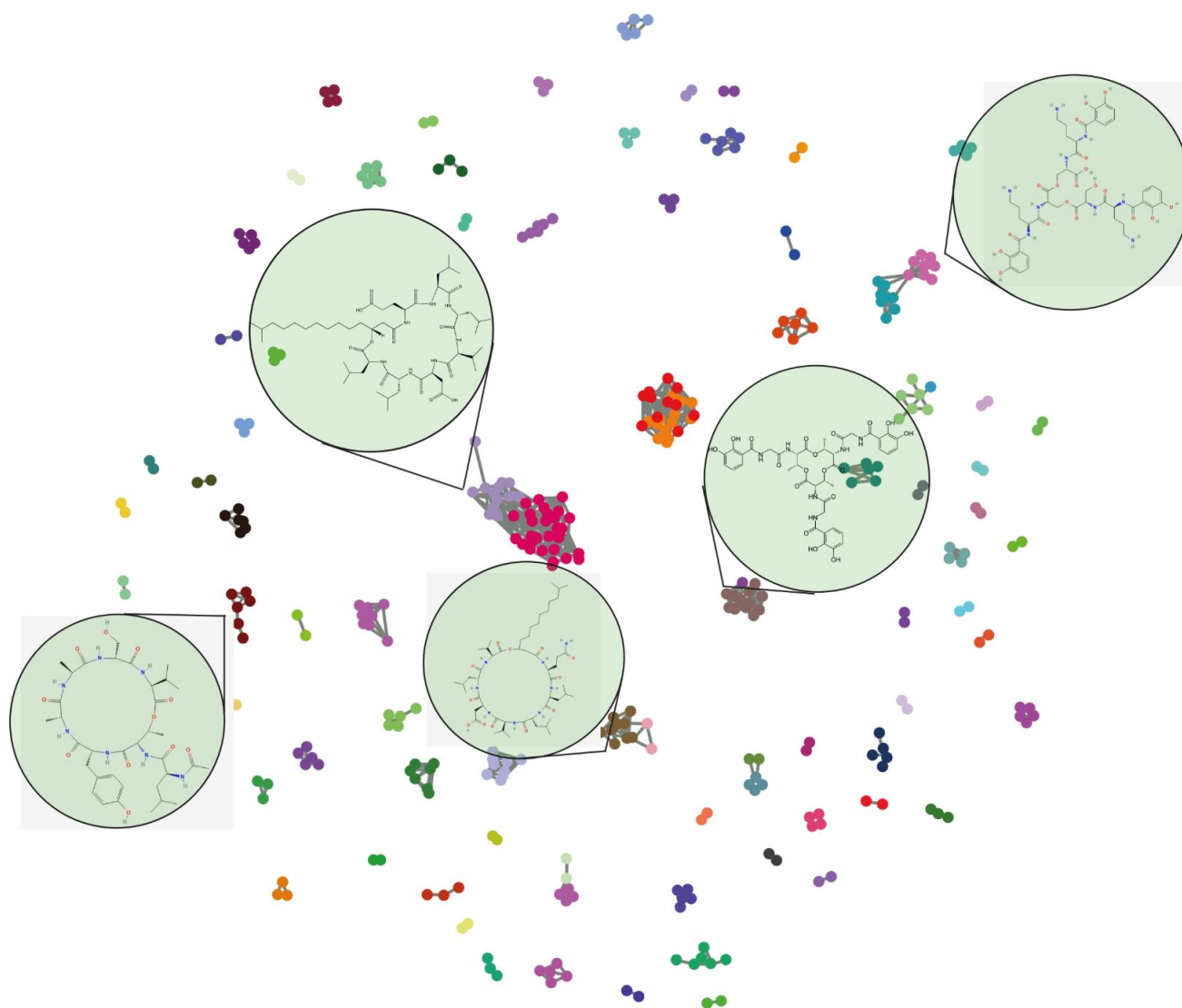


Figura 12. FCG resultado del análisis de redes de similitud. Se muestran las 85 FCG que agrupan 365 CGB de SPNR predichos por el programa BIG-SCAPE utilizando el valor por defecto de corte de distancia sin procesar. Se resaltan en la figura las estructuras de los compuestos con mayor porcentaje de identidad de los CGB representantes por familia, de izquierda a derecha; Rizomida, Surfactina, Liquenisina, Bacillibactina y Turnerbactina

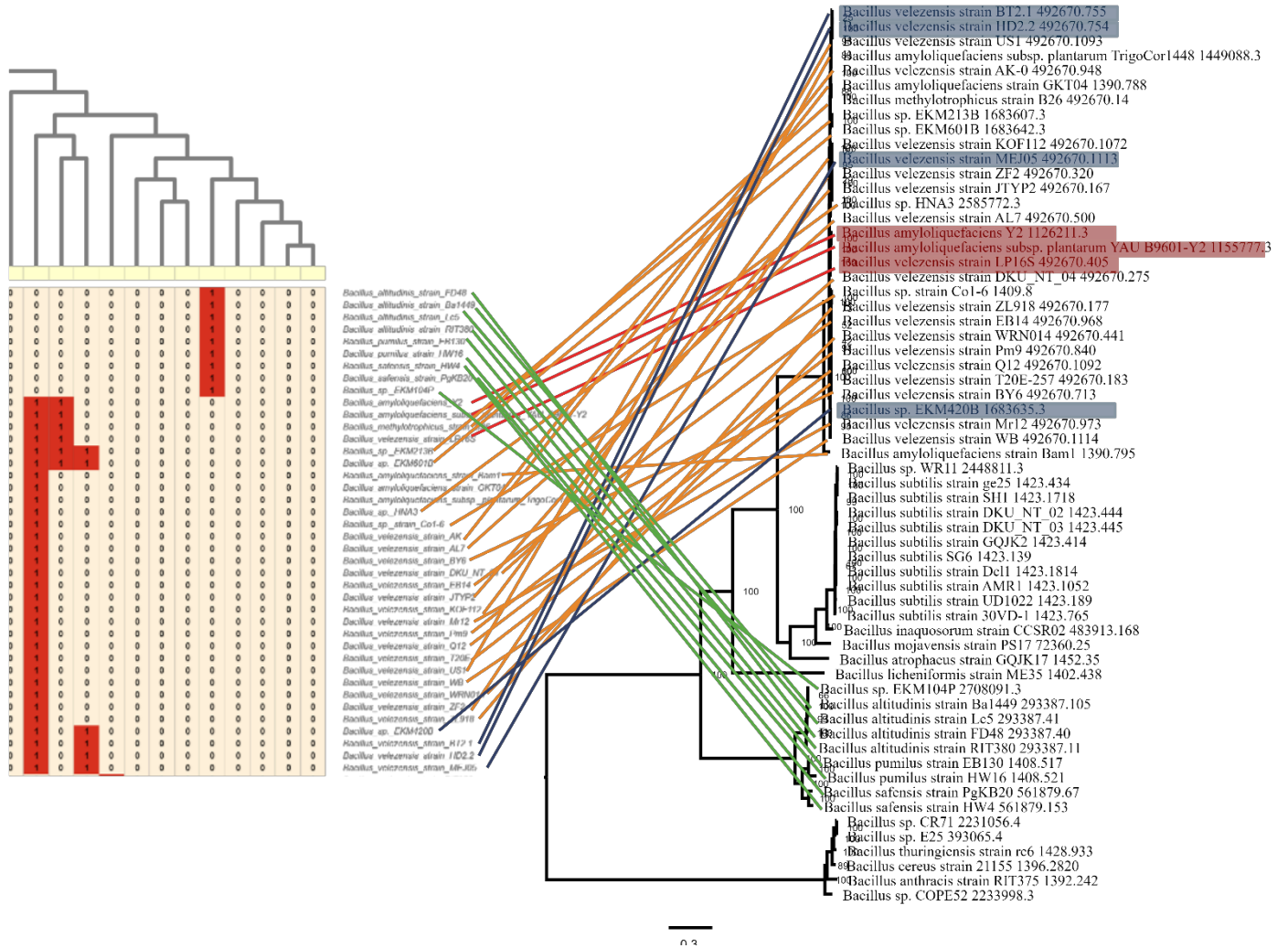


Figura 13. Mapa jerárquico con 13 de las FCG con mayor número de CGB. A la izquierda se muestra un perfil de presencia/ausencia de FCG, se resaltan en orden de izquierda a derecha las FCG 100, 29, 93 y 58. A la derecha se muestran clados con varias cepas de *Bacillus*, Las líneas naranjas asocian la presencia de todos los CGBs de la FCG 100 con todos los miembros del clado más grande del árbol, lo mismo para las líneas verdes con la FCG 58. En rojo y azul se resaltan algunas cepas de *B. amyloliquefaciens* y *B. velezensis*, puede notarse que, aun habiendo más cepas de la misma especie, no todas comparten la presencia de CGB de la FCG 93 (cepas y líneas resaltadas en color azul) y la FCG 29 (cepas y líneas resaltadas en color rojo)

Si bien varias actinobacterias como *Streptomyces* son consideradas como una fuente valiosa de metabolitos especializados, los géneros de

*Pseudomonas*, *Burkholderia*, *Serratia*, *Bacillus* y *Paenibacillus* también producen la mayoría de los péptidos antimicrobianos empleados en la medicina actual que han sido desarrollados desde el año 2000.<sup>15,16,118</sup>

Se sabe que algunas bacterias como las del género *Nocardia* son fuentes importantes de CGB de metabolitos especializados.<sup>54,119</sup>

El estudio realizado por Helfrich et al. del microbioma completo de la planta *Arabidopsis thaliana* probó que los miembros de la familia *Nocardiaceae*, como *Nocardia* y *Rhodococcus* son las únicas actinobacterias con gran abundancia de SPNR en sus genomas.<sup>108</sup> Sin embargo en el mismo estudio se reporta que a pesar de la gran abundancia de metabolitos producidos, estas bacterias no exhibieron actividades de inhibición de crecimiento de otras cepas bacterianas de la misma planta, al menos en condiciones de laboratorio.

Hay una clara predominancia de SPNR en algunos de los genomas analizados en este y otros estudios.

Por el gran tamaño de los genes que las sintetizan pudiera hacerse la hipótesis de que estas se encuentran en los genomas de mayor tamaño, capaces de contener un mayor número de genes y conservar porcentajes grandes de Guanina y Citosina.<sup>105</sup>

Sin embargo, lo anterior se descarta, pues observando la Tabla I de los genomas de BAP analizados; géneros con genomas de tamaño similar al de *Streptomyces*, *Nocardia* y *Burkholderia* como lo son los de *Mesorhizobium*, *Streptacidiphilus* y *Dyadobacter* no se encontró presencia de dominios de condensación asociados a SPNR en el análisis de BLAST por lo que se descarta esta primera hipótesis. Otra hipótesis es que este tipo de enzimas pudieran ser esenciales en las condiciones ambientales en las que se adaptan las bacterias antes mencionadas.

El estudio de Han et al. ubica a las actinobacterias y proteobacterias (*Burkholderia*, *Ralstonia*, *Klebsiella*, *Pseudomonas*, etc.) como las bacterias más abundantes en el suelo y rizósfera de las plantas, sin embargo, de nuevo aparecen excepciones de genomas analizados por BLAST de la Tabla I que no contuvieron dominios de condensación de SPNR como las proteobacterias *Achromobacter*, *Agrobacterium* y *Acinetobacter*.<sup>120,121</sup>

Por otra parte, en la Figura 7 y la Tabla II se pueden observar distintos patrones de distribución geográfica de aislamiento de las bacterias analizadas.

Tan solo las 5 cepas de *Nocardia* analizadas en este estudio provienen de Brasil, China, Alemania y Tailandia. Sumado a esto, un estudio realizado por Männle et al. demostró que los clústeres de metabolitos especializados del género *Nocardia* no tienen una correlación con el ecosistema del que se aisló la cepa. Por lo que se descarta la segunda hipótesis.<sup>121</sup>

A pesar de no contar con los datos suficientes para determinar un patrón de abundancia de SPNR en los genomas de bacterias, es importante destacar el hecho de que de los 169 clústeres huérfanos obtenidos del análisis de BIG-SCAPE, 30 (18%) son de los genomas de *Nocardia*.

Estos no están conservados entre la especie, sino que las bacterias de este género tienden a adquirir nuevos CGB de SPNR con variaciones significativas.

Además, 14 de ellos no tuvieron una correlación importante en ninguna de las familias de clústeres de la base de datos de BiG-FAM, lo que sugiere novedad en estos metabolitos.

Por todo esto, no se descarta esta información como importante y con más investigación pudiera abrir una nueva línea de investigación que ayude a comprender más sobre la presencia de clusters huérfanos dentro

de este género bacteriano, como si la transferencia vertical y horizontal de genes se viera impedida por algún factor específico.

Por último, el género *Azospirillum* ha sido estudiado desde el punto de vista agronómico y se sabe ampliamente desde el descubrimiento clásico de productos naturales que es capaz de producir compuestos como fenilpropanoides, alcaloides y fitohormonas que promueven el crecimiento de las plantas, sin embargo, sus genomas han sido poco explorados desde la minería genómica.<sup>122,123</sup>

Todas las familias estuvieron conformadas por CGB provenientes del mismo género bacteriano a excepción de la familia 220 (Figura 14 A) que estuvo conformada de 10 clústeres de genes provenientes del género *Pantoea* y uno solo del genoma de *Enterobacteriaceae bacterium* EKM102V.

Análisis filogenéticos de la Figura 11 muestran que esta bacteria se encuentra más relacionada con el género de *Pantoea* (familia *Erwiniaceae*) que con *Klebsiella* o *Enterobacter* (miembros de la familia *Enterobacteriaceae*), por lo que puede que el genoma se encuentre mal anotado.

La familia 223 (Figura 14 B) consta de dos clústeres provenientes de los genomas de *Klebsiella* sp. NFIX53 y *Kosakonia* sp. MH5, un análisis de BLAST en la base de datos de NCBI y UNIPROT revelaron que el CGB contiene homólogos del gen *EntF* de la *enterobactina sintasa*, cuya función ha sido detallada en la primera parte de este estudio.

Esta FCG se encuentra relacionada en un clan con las FCG 219 (Figura 14 C), 397 (Figura 14 D) y 220 (Figura 14 A) sin embargo los CGB no fueron agrupados en la misma familia pues la distancia de corte de valores sin procesar fue superior a 0.30.





Por otra parte, es observable en la Figura I y Figura 15 que todos los miembros del clado más grande del género de *Bacillus* (*B. amyloliquefaciens*, *B. methylotrophicus*, *B. velezensis* y las cepas *B. sp.* EKM104P, *B. sp.* EKM213B, *B. sp.* EKM601B, *B. sp.* HNA3, *B. sp.* strain Co1-6 y *B. sp.* EKM420B) también tienen en común la familia 100 del análisis de redes de similitud por BiG-SCAPE, lo que hace inferir que este CGB fue adquirido por todas las especies al ser conservado desde un ancestro en común, confirmando la transferencia vertical de genes.

Por el estudio de Belknap et al. se sabe que miembros de la misma especie, pero de diferente cepa pueden contener en sus genomas una significativa variedad y abundancia de clústeres de metabolitos especializados, este estudio corrobora este planteamiento, específicamente para CGB de SPNR.<sup>118</sup>

Por ejemplo, en los genomas de *B. amyloliquefaciens* Y2, *B. amyloliquefaciens subsp. plantarum* YAU B9601-Y2 y *B. velezensis* strain LP16S y *B. velezensis strain* H2.2, *B. velezensis strain* BT2.1, *B. velezensis strain* MEJ05 se encuentran clústeres de las familias 29, 93, 159 que no se encuentran presentes en el resto de los genomas de la misma especie.

Este fenómeno trata de explicarse con la "Hipótesis de la Reina Negra"<sup>125</sup> la cual propone a estos metabolitos no compartidos dentro de la misma especie como un recurso colectivo, puesto a disposición del resto de las cepas, las cuales pueden enfocarse en conservar solamente los genes esenciales para sobrevivir o aprovechar la bioactividad del metabolito en cuestión, haciendo uso de las ventajas que este pueda proveer sin sufrir las consecuencias de no tenerlo.



Figura 15. FCG número 100. Árbol realizado por el programa CORASON, incluido en la versión 1.1.0 de BIG-SCAPE. Las flechas separadas representan los genes que forman parte del CGB y de sus alrededores mientras que las regiones de diferentes colores representan los dominios funcionales de los mismos.

### 8.7. Propuesta de un nuevo CGB de una SPNR para la producción de un nuevo compuesto bioactivo.

Se seleccionó un CGB por cada una de las 85 FCG y los 169 clústeres huérfanos para ser comparados con las familias de clústeres de la base de datos de BiG-FAM, esta reúne un total de 1,225,071 BGC predichos en genomas de las bases de datos públicas y BGC caracterizados de MiBIG, de los cuales 426,265 (35%) pertenecen a la clase de SPNR y se agrupan en 15,673 familias.<sup>72</sup>

55 de las FCG descubiertas en este estudio mostraron tener una secuencia completa que al ser comparada con la base de datos de BiG-FAM y MiBIG no mostraron coincidencias significativas, lo cual sugiere novedad.

Clústeres que sí fueron agrupados con un nivel significativo con una FCG de la base de datos de BiG-FAM no necesariamente son CGB no novedosos, puesto que muchas de estas familias no tienen evidencia de un producto funcional a nivel experimental y solamente se ha predicho la clase de metabolito especializado a la que pertenecería el producto del CGB, sin embargo, para asegurar el evitar el redescubrimiento estos clústeres fueron descartados de la selección final.

De las 55 FCG se seleccionaron aquellas que se encontraban más conservadas, es decir, aquellas que contenían más de un CGB agrupado en ellas por el programa BiG-SCAPE, esto con el fin de asegurar que el producto final producido por el CGB sea un producto funcional y no una enzima en proceso de evolución y de adquisición de funcionalidad o en proceso de ser silenciado.

Los CGB huérfanos no necesariamente codifican para productos no funcionales, de hecho, estos han llamado recientemente la atención debido a que no se ha mostrado su potencial como fuente de PN, sin embargo, debido al elevado número de CGB huérfanos es necesario realizar análisis que los correlacionen con moléculas pequeñas con actividad demostrada.<sup>75</sup>

Fueron 12 los CGB que se agruparon con al menos un clúster más en una FCG de BiG-SCAPE y además no mostraron una relación significativa con alguna familia en la base de datos de BiG-FAM.

De estos se seleccionó el CGB predicho por AntiSMASH en el genoma de *Azospirillum brasilense* strain Az39, que agrupó junto con un CGB predicho en un genoma del mismo género bacteriano en la FCG 13 (Figura 16).

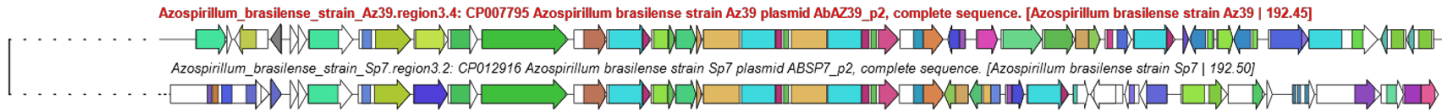


Figura 16. FCG número 13. Árbol realizado por el programa CORASON, incluido en la versión 1.1.0 de BIG-SCAPE. Las flechas separadas representan los genes que forman parte del CGB y de sus alrededores mientras que las regiones de diferentes colores representan los dominios funcionales de los mismos.

Las BAP que contienen los CGB de la FCG 19 fueron aislados de las plantas *Triticum aestivum* y *Digitaria eriantha* en el sur de América, ninguna de las hospederas cuenta con reportes de actividad medicinal relevante sin embargo ambas son cultivadas en climas desde cálidos hasta semiáridos y tienen propiedades altamente nutricionales,<sup>91,126</sup> siendo el trigo uno de los alimentos más cultivados a nivel mundial.

Ambas plantas han logrado adaptarse a ecosistemas con baja disponibilidad de agua y temperaturas altas, manteniendo no sólo cantidades de nutrientes suficientes para sí mismas y su microbiota, sino que también almacenan cantidades grandes de energía que sirve como fuente de alimento principal para ganado y humanos.<sup>126</sup>

Esto vuelve a ambas especies hospederas interesante en términos de mecanismos de adaptación a su ecosistema, ya sean propios o de su microbiota, tomando en cuenta que el género *Azospirillum* ha sido estudiado desde el punto de vista agronómico y se sabe que es capaz de producir compuestos que promueven el crecimiento de las plantas.<sup>122</sup>

El CGB del genoma *Azospirillum brasilense* strain Az39 perteneciente de la FCG 13 predicho por AntiSMASH se muestra en la Figura 17.

Este consta de un único GBC de una SPNR, este gen está compuesto por dos módulos divididos a su vez en siete dominios funcionales.

El programa NRPSpredictor integrado en la versión 6.1.1 de AntiSMASH identifica el segundo dominio A de la enzima como el encargado de

incorporar un residuo de Serina y el primer dominio A como un dominio inactivo (Figura 18), sin embargo, para el momento en el que se escribe esta sección se encuentra disponible una prueba beta de la versión 7.0 de antiSMASH, que al realizar el análisis de minería genómica al mismo genoma identifica el primer dominio A como el encargado de reclutar al aminoácido no proteinogénico de la Ornitina y el segundo como un dominio inactivo (Figura 19).

Si bien ambas versiones pueden estar parcialmente en lo correcto, el software NRPSpredictor no ha tenido actualizaciones nuevas desde el 2011 por lo que el resultado podría ser mejor.<sup>52</sup>

En este tiempo se ha demostrado que la selectividad de un bloque de construcción no se asocia únicamente a un dominio de adenilación sino que el dominio de condensación puede contribuir significativamente en la selección y modificación del residuo, sobre todo si este no es un elemento proteinogénico, como los ácidos grasos, los cuales suelen ser incorporados en SPNR con un dominio de condensación inicial, como es el caso del clúster analizado.<sup>44, 54</sup>

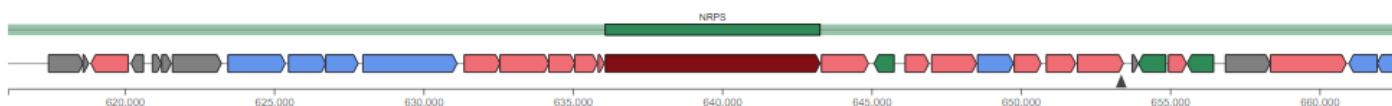


Figura 17. CGB predicho por AntiSMASH v6.1.1 de una SPNR identificada en el genoma de *Azospirillum brasilense* strain Az39. El CGB presentado se agrupó en la FCG 13 por BiG-SCAPE. En rojo se resaltan los genes biosintéticos centrales, en rosa los genes biosintéticos adicionales, en verde los reguladores y en azul los transportadores.

La selectividad modular predicha por AntiSMASH fue corroborada con la herramienta de minería genómica PRISM,<sup>127</sup> el esquema resultante se muestra en la figura 20. En este caso PRISM predijo ambos módulos como módulos activos, cuyos dominios de adenilación seleccionan los residuos de la Ornitina (Módulo 1) y la Serina (Módulo 2), sin embargo, junto con

estos AA predijo un bajo nivel de especificidad para otros residuos como hidroxiacetilornitina, Lys, Arg, Asn, Phe, Gln, Trp para el módulo 1 y Ser, Ala, Gln, Orn, Gly, Arg, Phe, Trp para el módulo 2.

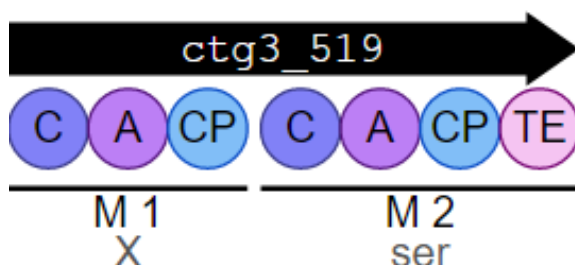


Figura 18. Composición de módulos y dominios del gen biosintético central del CGB propuesto. M1: Módulo 1, M2: Módulo 2, C: Dominio C, A: Dominio A, CP: Dominio PTP, TE: Dominio TE. Análisis modular realizado por la herramienta NRSPredictor2<sup>52</sup> incorporada en la versión 6.0 de AntiSMASH.

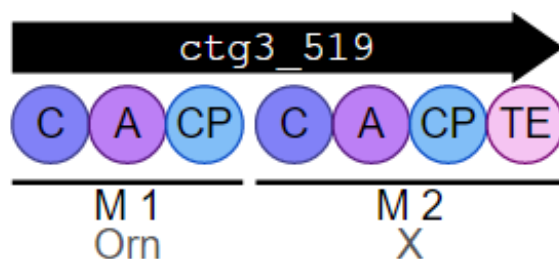


Figura 19. Composición de módulos y dominios del gen biosintético central del CGB propuesto. M1: Módulo 1, M2: Módulo 2, C: Dominio C, A: Dominio A, CP: Dominio PTP, TE: Dominio TE. Análisis modular realizado por la herramienta NRSPredictor2<sup>52</sup> incorporada en la prueba beta de la versión 7.0 de AntiSMASH.



Figura 20. Composición de dominios que conforman el gen biosintético central del CGB propuesto predicho por PRISM. Análisis de dominios realizado por las herramientas incorporadas en la versión más actual de PRISM. C\*: Dominio de condensación inicial. T: Dominio PTP, C: Dominio C, TE: Dominio de TE.<sup>127</sup>

En la tabla 1 se muestran los compuestos Brabantamida y Xenortides, dos péptidos no ribosomales sintetizados por SPNR dimodulares, al igual que la SPNR propuesta, que cuentan con actividad antibacteriana y antiprotozoaria.<sup>46</sup>

En la Figura 21 se muestra el esquema genético del CGB de la Brabantamida, al comparar el CGB propuesto y su estructura de las Figuras 18 y 19, encontramos entre las características en común la selección del residuo de L-Serina por uno de los módulos de sus respectivas enzimas así como la presencia de un dominio de condensación inicial responsable por la adición de un elemento no proteinogénico, lo más probable es que un ácido graso, tal es el caso para la Brabantamida, cuya estructura se presenta en la Figura 22.

Por otra parte, se realizó un análisis de homología con BLAST en la base de datos de NCBI empleando la secuencia del gen biosintético central del CGB propuesto, los resultados se muestran en la Figura 23.

En la Figura 23 es observable que el gen biosintético central se encuentra más conservado de lo predicho en el análisis de redes de similitud por BiG-SCAPE al ser encontrado en 10 cepas más del mismo género con un porcentaje de identidad superior al 90%, además, en varias ocasiones este CGB se encuentra repetido en el cromosoma y los plásmidos de las

cepas resultantes, lo cual es un fuerte indicio de su importancia en los genomas analizados.

Las bacterias suelen compartir secuencias genéticas por eventos de THG cuando estas secuencias se encuentran codificadas en sus plásmidos, sólo las secuencias más importantes son conservadas por la bacteria receptora de la información genética, lo cual pudiera ser el caso ya que el CGB se encuentra en plásmidos de más de una especie de *Azospirillum*.

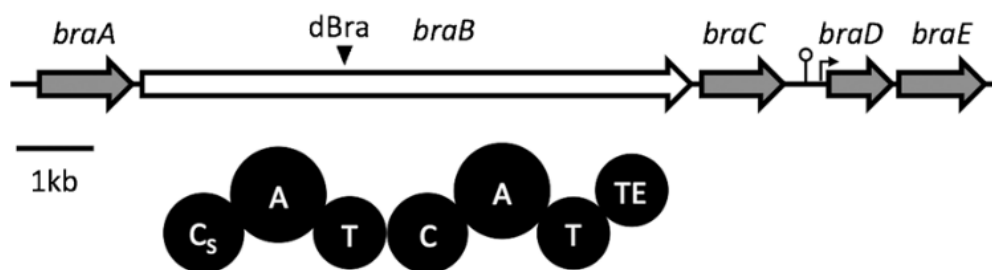


Figura 21. Esquema genético de la SPNR de la brabantamida. El módulo 1 (primeros dominios de izquierda a derecha C-A-T) se encarga de la incorporación del ácido graso así como del residuo Serina, mientras que el módulo 2 (dominios C-A-T-TE subsecuentes) se encargan de la adición de la Prolina y liberación del péptido. Imagen tomada de Schmidt et al.<sup>45</sup>

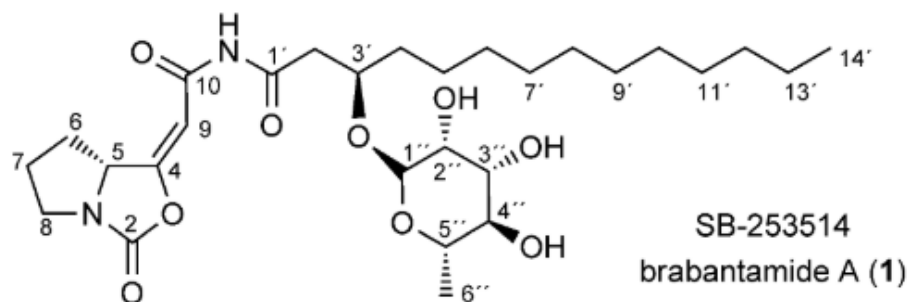


Figura 22. Estructura química del compuesto Brabantamida. Imagen tomada de Schmidt et al.<sup>45</sup>

Por otro lado, un mecanismo para conservar la información genética importante por parte de las bacterias es llevando varias copias del CGB, estas pueden estar distribuidas entre el cromosoma y los plásmidos lo



cual también es el caso en algunos de los genomas resultantes del análisis de BLAST.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<a href="#">Azospirillum brasilense strain Az39 plasmid AbAZ39_p2_complete sequence</a>	<a href="#">Azospirillum argentinense</a>	13280	13280	100%	0.0	100.00%	933960	<a href="#">CP007795.1</a>
<a href="#">Azospirillum argentinense strain MTCC4036 plasmid p2_complete sequence</a>	<a href="#">Azospirillum argentinense</a>	11557	11557	100%	0.0	95.70%	1167812	<a href="#">CP032332.1</a>
<a href="#">Azospirillum argentinense strain MTCC4035 plasmid p2_complete sequence</a>	<a href="#">Azospirillum argentinense</a>	11409	11409	100%	0.0	95.28%	951667	<a href="#">CP032323.1</a>
<a href="#">Azospirillum brasilense strain 2020WEIHUA_K plasmid unnamed2_complete sequence</a>	<a href="#">Azospirillum brasilense</a>	10567	10567	99%	0.0	93.44%	863243	<a href="#">CP059451.1</a>
<a href="#">Azospirillum brasilense strain MTCC4039 plasmid p1_complete sequence</a>	<a href="#">Azospirillum brasilense</a>	10388	10388	99%	0.0	92.97%	1845095	<a href="#">CP032346.1</a>
<a href="#">Azospirillum baldaniorum strain Sp245 plasmid Sp245_p03</a>	<a href="#">Azospirillum baldaniorum</a>	10364	10364	99%	0.0	92.89%	810374	<a href="#">CP022255.1</a>
<a href="#">Azospirillum brasilense Sp245 plasmid AZOBR_p3 complete genome</a>	<a href="#">Azospirillum baldaniorum</a>	10346	10346	99%	0.0	92.84%	778798	<a href="#">HE577330.1</a>
<a href="#">Azospirillum brasilense strain Cd plasmid p3_complete sequence</a>	<a href="#">Azospirillum brasilense</a>	10344	10344	99%	0.0	92.85%	701851	<a href="#">CP033321.1</a>
<a href="#">Azospirillum brasilense strain Cd chromosome_3_complete sequence</a>	<a href="#">Azospirillum brasilense</a>	10344	10344	99%	0.0	92.85%	866433	<a href="#">CP033320.1</a>
<a href="#">Azospirillum brasilense strain Sp 7 plasmid p3_complete sequence</a>	<a href="#">Azospirillum brasilense</a>	10344	10344	99%	0.0	92.85%	700012	<a href="#">CP033315.1</a>
<a href="#">Azospirillum brasilense strain Sp 7 chromosome_3_complete sequence</a>	<a href="#">Azospirillum brasilense</a>	10344	10344	99%	0.0	92.85%	851076	<a href="#">CP033314.1</a>
<a href="#">Azospirillum brasilense strain MTCC4038 plasmid p2_complete sequence</a>	<a href="#">Azospirillum brasilense</a>	10344	10344	99%	0.0	92.85%	824946	<a href="#">CP032341.1</a>
<a href="#">Azospirillum brasilense strain Sp 7 plasmid ABSP7_p2_complete sequence</a>	<a href="#">Azospirillum brasilense</a>	10344	10344	99%	0.0	92.83%	819020	<a href="#">CP012916.1</a>
<a href="#">Azospirillum sp. TSH58 plasmid TSH58_p02_complete sequence</a>	<a href="#">Azospirillum sp. TSH58</a>	9878	9878	100%	0.0	91.50%	1910929	<a href="#">CP022367.1</a>
<a href="#">Blastochloris viridis DNA_complete genome_strain: DSM 133</a>	<a href="#">Blastochloris viridis</a>	1714	1714	79%	0.0	72.85%	3726227	<a href="#">AP014854.2</a>
<a href="#">Blastochloris viridis genome assembly Blastochloris viridis genome chromosome -1</a>	<a href="#">Blastochloris viridis</a>	1714	1714	79%	0.0	72.85%	3723225	<a href="#">LN907867.1</a>
<a href="#">Blastochloris viridis strain ATCC 19567 chromosome_complete genome</a>	<a href="#">Blastochloris viridis</a>	1714	1714	79%	0.0	72.85%	3724841	<a href="#">CP012946.1</a>
<a href="#">Xanthobacter sp. YC-JY1 chromosome_complete genome</a>	<a href="#">Xanthobacter sp. YC-JY1</a>	1472	1472	78%	0.0	72.23%	5392410	<a href="#">CP032674.1</a>
<a href="#">Xanthobacter dioxanivorans strain YN2 chromosome_complete genome</a>	<a href="#">Xanthobacter dioxanivorans</a>	1301	2044	76%	0.0	74.46%	6277310	<a href="#">CP063362.1</a>

Figura 23. Análisis de BLAST realizado en la base de datos de NCBI usando la secuencia del gen biosintético central del CGB propuesto.

Tomando en cuenta las características que hacen parecer una secuencia importante al CGB predicho, la promoción de crecimiento conocida del género *Azospirillum*, las capacidades de crecer en ambientes con baja disponibilidad de agua y temperaturas altas de los hospederos de los genomas que contienen dicho CGB así como las propiedades genéticas que comparte el CGB propuesto con el del antibiótico Brabantamida, hacen que este clúster sea un candidato interesante para su producción por medio de técnicas de biología celular para probar su actividad en el laboratorio.

## 9. Conclusiones y perspectivas

En este estudio se demostró que muchas de las BAP analizadas producen una gran variedad de metabolitos especializados con potenciales aplicaciones biotecnológicas.

Los metabolitos analizados en este estudio no sólo suponen una herramienta por la cual estas bacterias se adaptan a organismos eucariontes, sino que su aprovechamiento abre la posibilidad de obtener nuevos compuestos con características funcionales y estructurales únicas que no perjudiquen a las personas o el medio ambiente para atender necesidades humanas.

Se realizó *in silico* un perfil de metabolitos especializados presentes en genomas de bacterias asociadas a plantas filogenéticamente variadas.

Se logró asociar este perfil a la evolución de las bacterias por medio de la construcción de árboles filogenéticos, demostrando que los cambios evolutivos del metabolismo especializado se encuentran relacionados y/o acoplados a la evolución del metabolismo primario.

Las bacterias relacionadas filogenéticamente presentan también patrones de similitud en la abundancia y clase de compuestos producidos por los CGB codificados en sus genomas.

Entre los metabolitos más destacados se encontraron las SPNR por demostrar tener no sólo una gran abundancia en los genomas de BAP sino que la variedad estructural y funcional que mostraron en el análisis de redes de similitud las vuelve un candidato prometedor para enfocar la búsqueda de nuevos PN por minería genómica.

Esto abre la posibilidad de explotar no solamente los CGB más conservados dentro de las especies, también una gran cantidad de CGB huérfanos, de los cuales aún no se ha logrado comprender su papel en

los ecosistemas que habitan estas bacterias y pudieran tener un importante potencial biotecnológico.

Las predicciones de minería genómica se vieron sumamente facilitadas y complementadas con los análisis bioinformáticos y de filogenia.

Haciendo uso de herramientas como AntiSMASH y PRISM es posible predecir en cuestión de minutos los potenciales CGB que codifican para alguna clase de metabolito, comparando con las bases de datos públicas, y asegurando que estos sean PN novedosos, superando el principal obstáculo del descubrimiento de PN por el método clásico.

Este tipo de análisis complementado con técnicas de metabolómica y comparación con redes de familias moleculares es actualmente la única manera de descubrir productos naturales con actividad biológica importante, por lo que se propone complementar esta metodología con dichos estudios.

Finalmente se propuso un CGB de *Azospirillum brasilense* strain Az39, por sus características de origen; de hospederos que son capaces de adaptarse a medios con baja disponibilidad de agua y altas temperaturas manteniendo un almacenamiento de energía muy grande.

También se tomó en cuenta la composición genética del CGB que comparado con el clúster precursor del antibiótico Brabantamida, hace que la secuencia propuesta resulte de interés para su producción en el laboratorio.

Se espera que la SPNR propuesta sintetice un producto funcional ya que se encuentra conservado en varios genomas del género *Azospirillum*, el cual cuenta con pocos reportes en la literatura de estudios de minería genómica y de la producción de metabolitos especializados con actividades biológicas importantes.

## 8. Referencias

1. Dias, D. A., Urban, S., & Roessner, U. (2012). A Historical overview of natural products in drug discovery. *Metabolites*, 2(2), 303–336. <https://doi.org/10.3390/metabo2020303>
2. Chevrette, M. G., Gutiérrez-García, K., Selem-Mojica, N., Aguilar-Martínez, C., Yañez-Olvera, A., Ramos-Aboites, H. E., Hoskisson, P. A., & Barona-Gómez, F. (2020). Evolutionary dynamics of natural product biosynthesis in bacteria. *Natural Product Reports*, 37(4), 566–599. <https://doi.org/10.1039/c9np00048h>
3. Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hermsdorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C., & Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology*, 1(5), 1–6. <https://doi.org/10.1038/nmicrobiol.2016.48>
4. Thiemann, S., Smit, N., & Strowig, T. (2016). Antibiotic Resistance: Problems and New Opportunities. In *How to Overcome the Antibiotic Crisis Facts, Challenges, Technologies and Future Perspectives* (Vol. 398).
5. Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D. L., Pulcini, C., Kahlmeter, G., Kluytmans, J., Carmeli, Y., Ouellette, M., Outtersson, K., Patel, J., Cavalieri, M., Cox, E. M., Houchens, C. R., Grayson, M. L., Hansen, P., Singh, N., ... Zorzet, A. (2018). Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases*, 18(3), 318–327. [https://doi.org/10.1016/S1473-3099\(17\)30753-3](https://doi.org/10.1016/S1473-3099(17)30753-3)
6. Karoui, I. J., Wannas, W. A., & Marzouk, B. (2010). Refined corn oil aromatization by *Citrus aurantium* peel essential oil. *Industrial Crops and Products*, 32(3), 202–207. <https://doi.org/10.1016/j.indcrop.2010.04.020>
7. Mannucci, C., Calapai, F., Cardia, L., Inferrera, G., D’Arena, G., Di Pietro, M., Navarra, M., Gangemi, S., Ventura Spagnolo, E., & Calapai, G. (2018). Clinical pharmacology of *citrus aurantium* and *citrus sinensis* for the treatment of anxiety. *Evidence-Based Complementary and Alternative Medicine*, 2018. <https://doi.org/10.1155/2018/3624094>
8. Baker, D. D., Chu, M., Oza, U., & Rajgarhia, V. (2007). The value of natural products to future pharmaceutical discovery. *Natural Product Reports*, 24(6), 1225–1244. <https://doi.org/10.1039/b602241n>
9. Maldonado-Carmona, N., Vázquez-Hernández, M., Patiño Chávez, O. J., Rodríguez-Luna, S. D., Jiménez Rodríguez, O., Sanchez, S., & Ceapă, C. D. (2019). Impact of ~omics in the detection and validation of potential anti-infective drugs. *Current Opinion in Pharmacology*, 48, 1–7. <https://doi.org/10.1016/j.coph.2019.02.008>
10. Newman, D. J., & Cragg, G. M. (2016). Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products*, 79(3), 629–661. <https://doi.org/10.1021/acs.jnatprod.5b01055>
11. Cragg, G. M., & Newman, D. J. (2005). Biodiversity: A continuing source of novel drug leads. *Pure and Applied Chemistry*, 77(1), 7–24. <https://doi.org/10.1351/pac200577010007>
12. Gallezot, P. (2007). Catalytic routes from renewables to fine chemicals. *Catalysis Today*, 121(1–2), 76–91. <https://doi.org/10.1016/j.cattod.2006.11.019>
13. Kasana, R. C., & Pandey, C. B. (2018). Exiguobacterium: an overview of a versatile genus with potential in industry and agriculture. *Critical Reviews in Biotechnology*, 38(1), 141–156. <https://doi.org/10.1080/07388551.2017.1312273>
14. Ojima, I. (2008). Miniperspectives : Natural Products in Drug Discovery. *Journal of Medicine Chemistry*, 51(9), 8–9.
15. Zhao, P., Xue, Y., Gao, W., Li, J., Zu, X., Fu, D., Bai, X., Zuo, Y., Hu, Z., & Zhang, F. (2018). Bacillaceae-derived peptide antibiotics since 2000. *Peptides*, 101(December 2017), 10–16. <https://doi.org/10.1016/j.peptides.2017.12.018>
16. Xue, Y., Wang, M., Zhao, P., Quan, C., Li, X., Wang, L., Gao, W., Li, J., Zu, X., Fu, D., Feng, S., & Li, P. (2018). Gram-negative bacilli-derived peptide antibiotics developed since 2000. In *Biotechnology Letters* (Vol. 40, Issues 9–10, pp. 1271–1287). <https://doi.org/10.1007/s10529-018-2589-1>
17. Hutchings, M., Truman, A., & Wilkinson, B. (2019). Antibiotics: past, present and future. *Current Opinion in Microbiology*, 51, 72–80. <https://doi.org/10.1016/j.mib.2019.10.008>

18. Wang, P., Liu, Z., & Huang, Y. (2018). Complete genome sequence of soil actinobacteria *Streptomyces cavourensis* TJ430. *Journal of Basic Microbiology*, 58(12), 1083–1090. <https://doi.org/10.1002/jobm.201800181>
19. Chakraborty, A., Bhakat, K., & Islam, E. (2017). Arsenic Contamination in Agricultural Soil Reduces Metabolic Activity of Total and Free-Living Nitrogen-Fixing Bacteria as Revealed by Real-Time qPCR. *Soil and Sediment Contamination*, 26(7–8), 736–748. <https://doi.org/10.1080/15320383.2017.1413637>
20. Chakraborty, A., Aziz Chowdhury, A., Bhakat, K., & Islam, E. (2019). Elevated level of arsenic negatively influences *nifH* gene expression of isolated soil bacteria in culture condition as well as soil system. In *Environmental Geochemistry and Health* (Vol. 41, Issue 5, pp. 1953–1966). <https://doi.org/10.1007/s10653-019-00261-2>
21. Zheng, X., Aborisde, M. A., Wang, H., He, P., Lu, S., Cui, N., Wang, S., Zhang, H., Ding, H., & Liu, K. (2020). Effect of lignin and plant growth-promoting bacteria (*Staphylococcus pasteurii*) on microbe-plant Co-remediation: A PAHs-DDTs Co-contaminated agricultural greenhouse study. *Chemosphere*, 256. <https://doi.org/10.1016/j.chemosphere.2020.127079>
22. Ahmad, F., Ahmad, I., & Khan, M. S. (2008). Screening of free-living rhizospheric bacteria for their multiple plant growth promoting activities. *Microbiological Research*, 163(2), 173–181. <https://doi.org/10.1016/j.micres.2006.04.001>
23. Kandel, S. L., Firrincieli, A., Joubert, P. M., Okubara, P. A., Leston, N. D., McGeorge, K. M., Mugnozza, G. S., Harfouche, A., Kim, S. H., & Doty, S. L. (2017). An in vitro study of bio-control and plant growth promotion potential of salicaceae endophytes. *Frontiers in Microbiology*, 8(MAR). <https://doi.org/10.3389/fmicb.2017.00386>
24. Strobel, G. (2018). The emergence of endophytic microbes and their biological promise. *Journal of Fungi*, 4(2). <https://doi.org/10.3390/jof4020057>
25. Ahsan, N., & Shimizu, M. (2021). Lysinibacillus species: Their potential as effective bioremediation, biostimulant, and biocontrol agents. In *Reviews in Agricultural Science* (Vol. 9, pp. 103–116). [https://doi.org/10.7831/ras.9.0\\_103](https://doi.org/10.7831/ras.9.0_103)
26. Alvin, A., Miller, K. I., & Neilan, B. A. (2014). Exploring the potential of endophytes from medicinal plants as sources of antimycobacterial compounds. *Microbiological Research*, 169(7–8), 483–495. <https://doi.org/10.1016/j.micres.2013.12.009>
27. Janso, J. E., & Carter, G. T. (2010). Biosynthetic potential of phylogenetically unique endophytic actinomycetes from tropical plants. *Applied and Environmental Microbiology*, 76(13), 4377–4386. <https://doi.org/10.1128/AEM.02959-09>
28. Rodriguez, R., & Redman, R. (2008). More than 400 million years of evolution and some plants still can't make it on their own: Plant stress tolerance via fungal symbiosis. *Journal of Experimental Botany*, 59(5), 1109–1114. <https://doi.org/10.1093/jxb/erm342>
29. Martinez-Klimova, E., Rodríguez-Peña, K., & Sánchez, S. (2017). Endophytes as sources of antibiotics. *Biochemical Pharmacology*, 134, 1–17. <https://doi.org/10.1016/j.bcp.2016.10.010>
30. Alawiye, T. T., & Babalola, O. O. (2021). Metagenomic insight into the community structure and functional genes in the sunflower rhizosphere microbiome. *Agriculture (Switzerland)*, 11(2), 1–21. <https://doi.org/10.3390/agriculture11020167>
31. Wu, L., Wang, J., Wu, H., Chen, J., Xiao, Z., Qin, X., Zhang, Z., & Lin, W. (2018). Comparative metagenomic analysis of rhizosphere microbial community composition and functional potentials under rehmanna glutinosa consecutive monoculture. *International Journal of Molecular Sciences*, 19(8), 1–17. <https://doi.org/10.3390/ijms19082394>
32. McCully, M. E. (2001). Niches for bacterial endophytes in crop plants: A plant biologist's view. *Australian Journal of Plant Physiology*, 28(9), 983–990. <https://doi.org/10.1071/pp01101>
33. Lindow, S. E., & Brandl, M. T. (2003). Microbiology of the phyllosphere. *Applied and Environmental Microbiology*, 69(4), 1875–1883. <https://doi.org/10.1128/AEM.69.4.1875-1883.2003>
34. Lindow, S. E., & Leveau, J. H. J. (2002). Phyllosphere microbiology. *Current Opinion in Biotechnology*, 13(3), 238–243. [https://doi.org/10.1016/S0958-1669\(02\)00313-0](https://doi.org/10.1016/S0958-1669(02)00313-0)
35. Overmann, O., Abt, B., & Sikorski, J. (2017). Present and Future of Culturing Bacteria Changes may still occur before final publication. *Annual Review of Microbiology*, 71(July), 711–730. <https://doi.org/10.1146/annurev-micro-090816-093449>

36. Espina, L. (2020). An approach to increase the success rate of cultivation of soil bacteria based on fluorescence-activated cell sorting. In PLoS ONE (Vol. 15, Issue 8 August). <https://doi.org/10.1371/journal.pone.0237748>
37. Hamza, T. A., & Woldesenbet, F. (2017). Optimization of Culture Growth Parameters for Production of Protease from Bacteria , Isolated from Soil Optimization of Culture Growth Parameters for Production of Protease from Bacteria , Isolated from Soil. 3(September), 1–10.
38. Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., & Bryant, S. H. (2009). PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Research*, 37(SUPPL. 2), 623–633. <https://doi.org/10.1093/nar/gkp456>
39. Pattnaik, S., Mohapatra, B., & Gupta, A. (2021). Plant Growth-Promoting Microbe Mediated Uptake of Essential Nutrients (Fe, P, K) for Crop Stress Management: Microbe–Soil–Plant Continuum. *Frontiers in Agronomy*, 3(August), 1–20. <https://doi.org/10.3389/fagro.2021.689972>
40. Faylor, K. C., Liu, H., Llontop, M. E. M., LeBlanc, S., Eckshtain-Levi, N., Sharma, P., Reed, A., Yang, S., Tian, L., Lefevre, C. T., Menguy, N., Du, L., Monteil, C. L., & Vinatzer, B. A. (2022). Ice nucleation in a Gram-positive bacterium isolated from precipitation depends on a polyketide synthase and non-ribosomal peptide synthetase. *ISME Journal*, 16(3), 890–897. <https://doi.org/10.1038/s41396-021-01140-4>
41. Hecht, S. M. (2000). Bleomycin: New perspectives on the mechanism of action. *Journal of Natural Products*, 63(1), 158–168. <https://doi.org/10.1021/np990549f>
42. Tedesco, D., & Haragsim, L. (2012). Cyclosporine: A Review. *Journal of Transplantation*, 2012, 1–7. <https://doi.org/10.1155/2012/230386>
43. Xue, J., Lou, X., Ning, D., Shao, R., & Chen, G. (2023). Mechanism and treatment of  $\alpha$ -amanitin poisoning. *Archives of Toxicology*, 97(1), 121–131. <https://doi.org/10.1007/s00204-022-03396-x>
44. Kaniusaite, M., Tailhades, J., Marschall, E. A., Goode, R. J. A., Schittenhelm, R. B., & Cryle, M. J. (2019). A proof-reading mechanism for non-proteinogenic amino acid incorporation into glycopeptide antibiotics. In *Chemical Science* (Vol. 10, Issue 41, pp. 9466–9482). <https://doi.org/10.1039/c9sc03678d>
45. Schmidt, Y., Van Der Voort, M., Crusemann, M., Piel, J., Josten, M., Sahl, H. G., Miess, H., Raaijmakers, J. M., & Gross, H. (2014). Biosynthetic origin of the antibiotic cyclocarbamate brabantamide A (SB-253514) in Plant-associated *pseudomonas*. *ChemBioChem*, 15(2), 259–266. <https://doi.org/10.1002/cbic.201300527>
46. Hennessy, R. C., Phippen, C. B. W., Nielsen, K. F., Olsson, S., & Stougaard, P. (2017). Biosynthesis of the antimicrobial cyclic lipopeptides nunamycin and nunapeptin by *Pseudomonas fluorescens* strain In5 is regulated by the LuxR-type transcriptional regulator NunF. *MicrobiologyOpen*, 6(6), 1–14. <https://doi.org/10.1002/mbo3.516>
47. Nguyen, D. D., Melnik, A. V., Koyama, N., Lu, X., Schorn, M., Fang, J., Aguinaldo, K., Lincecum, T. L., Ghequire, M. G. K., Carrion, V. J., Cheng, T. L., Duggan, B. M., Malone, J. G., Mauchline, T. H., Sanchez, L. M., Kilpatrick, A. M., Raaijmakers, J. M., De Mot, R., Moore, B. S., ... Dorrestein, P. C. (2016). Indexing the *Pseudomonas* specialized metabolome enabled the discovery of poaeamide B and the bananamides. *Nature Microbiology*, 2(October), 1–10. <https://doi.org/10.1038/nmicrobiol.2016.197>
48. Li, W., Rokni-Zadeh, H., De Vleeschouwer, M., Ghequire, M. G. K., Sinnaeve, D., Xie, G. L., Rozenski, J., Madder, A., Martins, J. C., & De Mot, R. (2013). The Antimicrobial Compound Xantholysin Defines a New Group of *Pseudomonas* Cyclic Lipopeptides. *PLoS ONE*, 8(5). <https://doi.org/10.1371/journal.pone.0062946>
49. Scholz-Schroeder, B. K., Soule, J. D., & Gross, D. C. (2003). The *sypA*, *sypB*, and *sypC* synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptin by *Pseudomonas syringae* pv. *syringae* B301D. *Molecular Plant-Microbe Interactions*, 16(4), 271–280. <https://doi.org/10.1094/MPMI.2003.16.4.271>
50. Fischer, D., Gessner, G., Fill, T. P., Barnett, R., Tron, K., Dornblut, K., Kloss, F., Stallforth, P., Hube, B., Stefan H. Heinemann, C. H., Kirstin Scherlach, & Brunke, S. (2019). Disruption of Membrane Integrity by the Bacterium-Derived. *Antimicrob Agents Chemother*, June, 1–20. <https://doi.org/10.1128/AAC.00707-19>
51. Miao, V., Coëffet-LeGal, M. F., Brian, P., Brost, R., Penn, J., Whiting, A., Martin, S., Ford, R., Parr, I., Bouchard, M., Silva, C. J., Wrigley, S. K., & Baltz, R. H. (2005). Daptomycin biosynthesis in *Streptomyces roseosporus*: Cloning and analysis of the gene cluster and



- revision of peptide stereochemistry. *Microbiology*, 151(5), 1507–1523. <https://doi.org/10.1099/mic.0.27757-0>
52. Rötting, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., & Kohlbacher, O. (n.d.). NRSPredictor2—a web server for predicting NRPS adenylation domain specificity. <https://doi.org/10.1093/nar/gkr323>
  53. Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Research*, 46(W1), W200–W204. <https://doi.org/10.1093/nar/gky448>
  54. Rausch, C., Hoof, I., Weber, T., Wohlleben, W., & Huson, D. H. (2007). Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evolutionary Biology*, 7, 1–15. <https://doi.org/10.1186/1471-2148-7-78>
  55. Bloudoff, K., & Schmeing, T. M. (2017). Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1865(11), 1587–1604. <https://doi.org/10.1016/j.bbapap.2017.05.010>
  56. Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood: Education and Practice Edition*, 98(6), 236–238. <https://doi.org/10.1136/archdischild-2013-304340>
  57. Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 35(SUPPL. 1), 61–65. <https://doi.org/10.1093/nar/gkl842>
  58. Olson, R. D., Assaf, R., Brettin, T., Conrad, N., Cucinell, C., Davis, J. J., Dempsey, D. M., Dickerman, A., Dietrich, E. M., Kenyon, R. W., Kuscuglu, M., Lefkowitz, E. J., Lu, J., Machi, D., Macken, C., Mao, C., Niewiadomska, A., Nguyen, M., Olsen, G. J., ... Stevens, R. L. (2023). Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Research*, 51(D1), D678–D689. <https://doi.org/10.1093/nar/gkac1003>
  59. Baltz, R. H. (2017). Molecular beacons to identify gifted microbes for genome mining. *Journal of Antibiotics*, 70(5), 639–646. <https://doi.org/10.1038/ja.2017.1>
  60. Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., De Bruijn, I., Chooi, Y. H., Claesen, J., Coates, R. C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D. J., Fewer, D. P., Garg, N., Geiger, C., Gomez-Escribano, J. P., Greule, A., ... Glöckner, F. O. (2015). Minimum Information about a Biosynthetic Gene cluster. *Nature Chemical Biology*, 11(9), 625–631. <https://doi.org/10.1038/nchembio.1890>
  61. Letzel, A. C., Pidot, S. J., & Hertweck, C. (2014). Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC Genomics*, 15(1), 1–21. <https://doi.org/10.1186/1471-2164-15-983>
  62. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., ... Van Der Donk, W. A. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Natural Product Reports*, 30(1), 108–160. <https://doi.org/10.1039/c2np20085f>
  63. Shen, B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology*, 7(2), 285–295. [https://doi.org/10.1016/S1367-5931\(03\)00020-6](https://doi.org/10.1016/S1367-5931(03)00020-6)
  64. Rudolf, J. D., & Chang, C.-Y. (2020). Terpene synthases in disguise: enzymology, structure, and opportunities of non-canonical terpene synthases. <https://doi.org/10.1039/c9np00051h>
  65. Schwab, W., Fuchs, C., & Huang, F. C. (2013). Transformation of terpenes into fine chemicals. *European Journal of Lipid Science and Technology*, 115(1), 3–8. <https://doi.org/10.1002/ejlt.201200157>
  66. Ali, S. S., & Vidhale, N. N. (2016). Bacterial Siderophore and their Application : A review. *Int.J.Curr.Microbiol.App.Sci*, 2(January 2013), 303–312.
  67. Marahiel, M. A. (2009). Working outside the protein-synthesis rules: Insights into non-ribosomal peptide synthesis. *Journal of Peptide Science*, 15(12), 799–807. <https://doi.org/10.1002/psc.1183>

68. Sieber, S. A., & Marahiel, M. A. (2005). Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. *Chemical Reviews*, 105(2), 715–738. <https://doi.org/10.1021/cr0301191>
69. Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., Van Wezel, G. P., Medema, M. H., & Weber, T. (2021). AntiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Research*, 49(W1), W29–W35. <https://doi.org/10.1093/nar/gkab335>
70. Navarro-muñoz, J. C., Selem-mojica, N., Mallowney, M. W., Kautsar, S. A., Tryon, J. H., Parkinson, E. I., Santos, E. L. C. D. L., Yeong, M., Cruz-morales, P., Abubucker, S., Roeters, A., Lokhorst, W., Fernandez-guerra, A., Teresa, L., Cappelini, D., Goering, A. W., Thomson, R. J., Metcalf, W. W., Kelleher, N. L., ... Medema, M. H. (2020). A computational framework to explore large-scale biosynthetic diversity. *Nature Chemical Biology*, 16(January). <http://dx.doi.org/10.1038/s41589-019-0400-9>
71. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C., & Eddy, S. R. (2004). The Pfam protein families database. *Nucleic Acids Research*, 32(DATABASE ISS.), 138–141. <https://doi.org/10.1093/nar/gkh121>
72. Kautsar, S. A., Blin, K., Shaw, S., Weber, T., & Medema, M. H. (2021). BiG-FAM: The biosynthetic gene cluster families database. *Nucleic Acids Research*, 49(D1), D490–D497. <https://doi.org/10.1093/nar/gkaa812>
73. Katoh, K., Rozewicki, J., & Yamada, K. D. (2018). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160–1166. <https://doi.org/10.1093/bib/bbx108>
74. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
75. Rivera-Chávez, J., Ceapă, C. D., & Figueroa, M. (2022). Biological Dark Matter Exploration using Data Mining for the Discovery of Antimicrobial Natural Products #. *Planta Medica*, 88(9–10), 702–720. <https://doi.org/10.1055/a-1795-0562>
76. Stowell, E. A., Ehrenberger, M. A., Lin, Y. L., Chang, C. Y., & Rudolf, J. D. (2022). Structure-guided product determination of the bacterial type II diterpene synthase *Tpn2*. *Communications Chemistry*, 5(1), 1–8. <https://doi.org/10.1038/s42004-022-00765-6>
77. Pan, X., Du, W., Zhang, X., Lin, X., Li, F. R., Yang, Q., Wang, H., Rudolf, J. D., Zhang, B., & Dong, L. Bin. (2022). Discovery, Structure, and Mechanism of a Class II Sesquiterpene Cyclase. *Journal of the American Chemical Society*, 144(48), 22067–22074. <https://doi.org/10.1021/jacs.2c09412>
78. Pan, X., Du, W., Zhang, X., Lin, X., Li, F. R., Yang, Q., Wang, H., Rudolf, J. D., Zhang, B., & Dong, L. Bin. (2022). Discovery, Structure, and Mechanism of a Class II Sesquiterpene Cyclase. *Journal of the American Chemical Society*, 144(48), 22067–22074. <https://doi.org/10.1021/jacs.2c09412>
79. Ulloa-Muñoz, R., Olivera-Gonzales, P., Castañeda-Barreto, A., Villena, G. K., & Tamariz-Angeles, C. (2020). Diversity of endophytic plant-growth microorganisms from *Gentianaella weberbaueri* and *Valeriana pycnantha*, highland Peruvian medicinal plants. *Microbiological Research*, 233 (September 2019), 126413. <https://doi.org/10.1016/j.micres.2020.126413>
80. Ali, M., Ahmad, Z., Ashraf, M. F., & Dong, W. (2021). Maize endophytic microbial communities revealed by removing PCR and 16S rRNA sequencing and their synthetic applications to suppress maize banded leaf and sheath blight. *Microbiological Research*, 242. <https://doi.org/10.1016/j.micres.2020.126639>
81. Cordero, J., de Freitas, J. R., & Germida, J. J. (2020). Bacterial microbiome associated with the rhizosphere and root interior of crops in Saskatchewan, Canada. *Canadian Journal of Microbiology*, 66(1), 71–85. <https://doi.org/10.1139/cjm-2019-0330>
82. Liu, Y., Wang, R., Li, Y., Cao, Y., Chen, C., Qiu, C., Bai, F., Xu, T., Zhang, X., Dai, W., Zhao, J., & Cheng, C. (2017). High-throughput sequencing-based analysis of the composition and diversity of endophytic bacterial community in seeds of “Beijing” hybrid maize planted in China. *Plant Growth Regulation*, 81(2), 317–324. <https://doi.org/10.1007/s10725-016-0208-5>
83. Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D.,



- Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., ... Yeats, C. (2009). InterPro: The integrative protein signature database. *Nucleic Acids Research*, 37(SUPPL. 1), 211–215. <https://doi.org/10.1093/nar/gkn785>
84. Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., Antunes, R., Arganiska, J., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Chavali, G., Cibrian-Uhalte, E., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., ... Zhang, J. (2015). UniProt: A hub for protein information. *Nucleic Acids Research*, 43(D1), D204–D212. <https://doi.org/10.1093/nar/gku989>
  85. Davis, J. J., Gerdes, S., Olsen, G. J., Olson, R., Pusch, G. D., Shukla, M., Vonstein, V., Wattam, A. R., & Yoo, H. (2016). PATTyFams: Protein families for the microbial genomes in the PATRIC database. *Frontiers in Microbiology*, 7(FEB), 1–12. <https://doi.org/10.3389/fmicb.2016.00118>
  86. Ehmann, D. E., Shaw-Reid, C. A., Losey, H. C., & Walsh, C. T. (2000). The *EntF* and *EntE* adenylation domains of *Escherichia coli* enterobactin synthetase: Sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proceedings of the National Academy of Sciences of the United States of America*, 97(6), 2509–2514. <https://doi.org/10.1073/pnas.040572897>
  87. Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H. W., & Kropinski, A. M. (2008). Unifying classical and molecular taxonomic classification: analysis of the Podoviridae using BLASTP-based tools. *Research in Microbiology*, 159(5), 406–414. <https://doi.org/10.1016/j.resmic.2008.03.005>
  88. Goodman, H. M., Ecker, J. R., & Dean, C. (1995). The genome of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 92(24), 10831–10835. <https://doi.org/10.1073/pnas.92.24.10831>
  89. Burlando, B., & Cornara, L. (2014). Therapeutic properties of rice constituents and derivatives (*Oryza sativa* L.): A review update. *Trends in Food Science and Technology*, 40(1), 82–98. <https://doi.org/10.1016/j.tifs.2014.08.002>
  90. R, A. S., S, S., N, R., TS, G., Karthikeyan, M., Gnanasekaran, A., GK, C., & Basalingappa, K. M. (2018). *Solanum tuberosum* L: Botanical, Phytochemical, Pharmacological and Nutritional Significance. *International Journal of Phytomedicine*, 10(3), 115. <https://doi.org/10.5138/09750185.2256>
  91. Moshawih, S., Abdullah Juperi, R. N. A., Paneerselvam, G. S., Ming, L. C., Liew, K. Bin, Goh, B. H., Al-Worafi, Y. M., Choo, C. Y., Thuraisingam, S., Goh, H. P., & Kifli, N. (2022). General Health Benefits and Pharmacological Activities of *Triticum aestivum* L. *Molecules*, 27(6), 1–20. <https://doi.org/10.3390/molecules27061948>
  92. Kushwaha, U.K.S. (2016). Black Rice. In: *Black Rice*. Springer, Cham. [https://doi.org/10.1007/978-3-319-30153-2\\_2](https://doi.org/10.1007/978-3-319-30153-2_2)
  93. Kong L, Wang Y, Cao Y (2008) Determination of Myo-inositol and D-chiro-inositol in black rice bran by capillary electrophoresis with electrochemical detection. *J Food Compos*
  94. Demir, A., & Tipirdamaz, R. (2021). Halophytes as Medicinal Herbs. *Handbook of Halophytes*, 2279–2301. [https://doi.org/10.1007/978-3-030-57635-6\\_94](https://doi.org/10.1007/978-3-030-57635-6_94)
  95. Parmar, H. S., & Kar, A. (2008). Medicinal values of fruit peels from *Citrus sinensis*, *Punica granatum*, and *Musa paradisiaca* with respect to alterations in tissue lipid peroxidation and serum concentration of glucose, insulin, and thyroid hormones. *Journal of Medicinal Food*, 11(2), 376–381. <https://doi.org/10.1089/jmf.2006.010>
  96. Rai, R. V., Rajesh, P. S., & Kim, H. M. (2013). Medicinal use of *Coscinium fenestratum* (Gaertn.) Colebr.: An short review. *Oriental Pharmacy and Experimental Medicine*, 13(1), 1–9. <https://doi.org/10.1007/s13596-012-0094-y>
  97. Mathew, F., & Varghese, B. (2019). A review on medicinal exploration of *costus igneus*: the insulin plant. *International Journal of Pharmaceutical Sciences Review and Research*, 54(2), 51–57.
  98. Byambasuren, S.-E., Wang, J., Gaudel, G., & Junru Wang, C. (2019). Medicinal value of wolfberry (*Lycium barbarum* L.). *Plantsjournal.Com*, 7(4), 90–97. <https://www.plantsjournal.com/archives/2019/vol7issue4/PartB/7-3-24-473.pdf>
  99. Ali Esmail Al-Snafi, Hanaa Salman Khadem, Hussein Ali Al-Saedy, Ali M. Alqahtani, Gaber El-Saber Batiha, & Jafari-Sales Abolfazl. (2021). A review on *Medicago sativa*: A potential

- medicinal plant. International Journal of Biological and Pharmaceutical Sciences Archive, 1(2), 022–033. <https://doi.org/10.30574/ijbpsa.2021.1.2.0302>
100. Kim, J. H. (2018). Pharmacological and medical applications of *Panax ginseng* and ginsenosides: a review for use in cardiovascular diseases. Journal of Ginseng Research, 42(3), 264–269. <https://doi.org/10.1016/j.jgr.2017.10.004>
  101. Sahu, T., & Sahu, J. (2015). Cucumis Sativus (Cucumber): a Review on Its Pharmacological Activity. Journal of Applied Pharmaceutical Research, 3(1), 4–9. [www.japtronline.com](http://www.japtronline.com)
  102. González-Pérez, J. E., & Guerrero-Beltrán, J. Á. (2021). Tomatillo or husk tomato (*Physalis philadelphica* and *Physalis ixocarpa*): A review. Scientia Horticulturae, 288(April). <https://doi.org/10.1016/j.scienta.2021.110306>
  103. Ali, S. E., Yuan, Q., Wang, S., & Farag, M. A. (2021). More than sweet: A phytochemical and pharmacological review of sugarcane (*Saccharum officinarum* L.). Food Bioscience, 44(October). <https://doi.org/10.1016/j.fbio.2021.101431>
  104. Kumar, M., Tomar, M., Bhuyan, D. J., Punia, S., Grasso, S., Sá, A. G. A., Carciofi, B. A. M., Arrutia, F., Changan, S., Radha, Singh, S., Dhumal, S., Senapathy, M., Satankar, V., Anitha, T., Sharma, A., Pandiselvam, R., Amarowicz, R., & Mekhemar, M. (2021). Tomato (*Solanum lycopersicum* L.) seed: A review on bioactives and biomedical activities. Biomedicine and Pharmacotherapy, 142. <https://doi.org/10.1016/j.biopha.2021.112018>
  105. Wang, A. H. J., Hakoshima, T., van der Marel, G., van Boom, J. H., & Rich, A. (1984). AT base pairs are less stable than GC base pairs in Z-DNA: The crystal structure of d(m5CGTAm5CG). Cell, 37(1), 321–331. [https://doi.org/10.1016/0092-8674\(84\)90328-3DEGC](https://doi.org/10.1016/0092-8674(84)90328-3DEGC)
  106. Michael, R., Ranjan, A., Kumar, R. S., Pathak, P. K., & Trivedi, P. K. (2020). Light-regulated expression of terpene synthase gene, AtTPS03, is controlled by the bZIP transcription factor, HY5, in *Arabidopsis thaliana*. Biochemical and Biophysical Research Communications, 529(2), 437–443. <https://doi.org/10.1016/j.bbrc.2020.05.222>
  107. Grammbitter, G. L. C., Schmalhofer, M., Karimi, K., Shi, Y. M., Schöner, T. A., Tobias, N. J., Morgner, N., Groll, M., & Bode, H. B. (2019). An Uncommon Type II PKS Catalyzes Biosynthesis of Aryl Polyene Pigments. Journal of the American Chemical Society, 141(42), 16615–16623. <https://doi.org/10.1021/jacs.8b10776>
  108. Helfrich, E. J. N., Vogel, C. M., Ueoka, R., Schäfer, M., Ryffel, F., Müller, D. B., Probst, S., Kreuzer, M., Piel, J., & Vorholt, J. A. (2018). Bipartite interactions, antibiotic production and biosynthetic potential of the *Arabidopsis* leaf microbiome. Nature Microbiology, 3(8), 909–919. <https://doi.org/10.1038/s41564-018-0200-0>
  109. Mitter, B., Petric, A., Shin, M. W., Chain, P. S. G., Hauberg-Lotte, L., Reinhold-Hurek, B., Nowak, J., & Sessitsch, A. (2013). Comparative genome analysis of *Burkholderia phytofirmans* PsJN reveals a wide spectrum of endophytic lifestyles based on interaction strategies with host plants. Frontiers in Plant Science, 4(APR). <https://doi.org/10.3389/fpls.2013.00120>
  110. Worsley, S. F., Newitt, J., Rassbach, J., Batey, S. F. D., Holmes, N. A., Murrell, J. C., Wilkinson, B., & Hutchings, M. I. (2020). *Streptomyces* endophytes promote host health and enhance growth across plant species. Applied and Environmental Microbiology, 86(16), 1–17. <https://doi.org/10.1128/AEM.01053-20>
  111. Zhou, L., Song, C., Li, Z., & Kuipers, O. P. (2021). Antimicrobial activity screening of rhizosphere soil bacteria from tomato and genome-based analysis of their antimicrobial biosynthetic potential. BMC Genomics, 22(1), 1–14. <https://doi.org/10.1186/s12864-020-07346-8>
  112. Samantha J. Bann, R. D. B. and S. A. C. (2021). The tridecaptins non ribosomal peptides thatselectively target Gram-negative bacteria.pdf (pp. 538–551).
  113. Magana, M., Pushpanathan, M., Santos, A. L., Leanse, L., Fernandez, M., Ioannidis, A., Giulianotti, M. A., Apidianakis, Y., Bradfute, S., Ferguson, A. L., Cherkasov, A., Seleem, M. N., Pinilla, C., de la Fuente-Nunez, C., Lazaridis, T., Dai, T., Houghten, R. A., Hancock, R. E. W., & Tegos, G. P. (2020). The value of antimicrobial peptides in the age of resistance. The Lancet Infectious Diseases, 20(9), e216–e230. [https://doi.org/10.1016/S1473-3099\(20\)30327-3](https://doi.org/10.1016/S1473-3099(20)30327-3)

114. Bozhüyük, K. A., Micklefield, J., & Wilkinson, B. (2019). Engineering enzymatic assembly lines to produce new antibiotics. In *Current Opinion in Microbiology* (Vol. 51, pp. 88–96). <https://doi.org/10.1016/j.mib.2019.10.007>
115. Baltz, R. H. (2018). Synthetic biology, genome mining, and combinatorial biosynthesis of NRPS-derived antibiotics: a perspective. *Journal of Industrial Microbiology and Biotechnology*, 45(7), 635–649. <https://doi.org/10.1007/s10295-017-1999-8>
116. Bozhüyük, K. A., Micklefield, J., & Wilkinson, B. (2019). Engineering enzymatic assembly lines to produce new antibiotics. In *Current Opinion in Microbiology* (Vol. 51, pp. 88–96). <https://doi.org/10.1016/j.mib.2019.10.007>
117. Theobald, S., Vesth, T. C., & Andersen, M. R. (2019). Genus level analysis of PKS-NRPS and NRPS-PKS hybrids reveals their origin in Aspergilli. *BMC Genomics*, 20(1), 1–12. <https://doi.org/10.1186/s12864-019-6114-2>
118. Belknap, K. C., Park, C. J., Barth, B. M., & Andam, C. P. (2020). Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Scientific Reports*, 10(1), 1–9. <https://doi.org/10.1038/s41598-020-58904-9>
119. Komaki, H., Ichikawa, N., Hosoyama, A., Takahashi-Nakaguchi, A., Matsuzawa, T., Suzuki, K. ichiro, Fujita, N., & Gonoï, T. (2014). Genome based analysis of type-I polyketide synthase and nonribosomal peptide synthetase gene clusters in seven strains of five representative *Nocardia* species. *BMC Genomics*, 15(1), 1–11. <https://doi.org/10.1186/1471-2164-15-323>
120. Han, W., Wang, F., Zhang, L., Zhao, H., Zheng, Y., Sun, R., & Meng, L. (2023). Variations of soil bacterial community and denitrifier abundance with depth under different land-use types. *Journal of Soils and Sediments*, 0123456789. <https://doi.org/10.1007/s11368-023-03428-8>
121. Männle, D., McKinnie, S. M. K., Mantri, S. S., Steinke, K., Lu, Z., Moore, B. S., Ziemert, N., & Kayssera, L. (2020). Comparative Genomics and Metabolomics in the Genus *Nocardia*. *MSystems*, 5(3), 1–19.
122. Drogue, B., Sanguin, H., Borland, S., Prigent-Combaret, C., & Wisniewski-Dyé, F. (2014). Genome wide profiling of *Azospirillum lipoferum* 4B gene expression during interaction with rice roots. *FEMS Microbiology Ecology*, 87(2), 543–555. <https://doi.org/10.1111/1574-6941.12244>
123. Alzate Zuluaga, M. Y., Miras-Moreno, B., Monterisi, S., Roupheal, Y., Colla, G., Lucini, L., Cesco, S., & Pii, Y. (2022). Integrated Metabolomics and Morpho-Biochemical Analyses Reveal a Better Performance of *Azospirillum brasilense* over Plant-Derived Biostimulants in Counteracting Salt Stress in Tomato. *International Journal of Molecular Sciences*, 23(22). <https://doi.org/10.3390/ijms232214216>
124. Van Elsas, J. D., Turner, S., & Bailey, M. J. (2003). Horizontal gene transfer in the phytosphere. *New Phytologist*, 157(3), 525–537. <https://doi.org/10.1046/j.1469-8137.2003.00697>
125. Morris, J. J., Lenski, R. E., & Zinser, E. R. (2012). The black queen hypothesis: Evolution of dependencies through adaptive gene loss. *MBio*, 3(2). <https://doi.org/10.1128/mBio.00036-12>
126. Garbero, M., Pedranzani, H., Zirulnik, F., Molina, A., Pérez-Chaca, M. V., Vigliocco, A., & Abdala, G. (2011). Short-term cold stress in two cultivars of *Digitaria eriantha*: Effects on stress-related hormones and antioxidant defense system. *Acta Physiologiae Plantarum*, 33(2), 497–507. <https://doi.org/10.1007/s11738-010-0573-z>
127. Skinnider, M. A., Johnston, C. W., Gunabalasingam, M., Merwin, N. J., Kieliszek, A. M., MacLellan, R. J., Li, H., Ranieri, M. R. M., Webster, A. L. H., Cao, M. P. T., Pfeifle, A., Spencer, N., To, Q. H., Wallace, D. P., Dejong, C. A., & Magarvey, N. A. (2020). Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nature Communications*, 11(1), 1–9. <https://doi.org/10.1038/s41467-020-19986-1>

## 9. Anexo

### 9.1 Tablas del Anexo

Tabla I. Genomas de BAP empleados en este estudio.

ID del genoma	Nombre del genoma	Contigs	Tamaño	Contenido de GC (%)	CDS
217204.1	<i>Achromobacter insolitus</i> strain LCU2	1	6428890	65.2	6033
217203.16	<i>Achromobacter spanius</i> strain UQ283	1	6687823	63.8	6331
40215.113	<i>Acinetobacter junii</i> strain KCTC 42611	15	3208237	38.5	3031
1785128.14	<i>Acinetobacter lactucae</i> strain ME27	68	3885226	38.8	3675
1219382.4	<i>Acinetobacter nectaris</i> BB226	67	2576508	36.7	2550
1148157.47	<i>Acinetobacter oleivorans</i> strain HW32	14	3953258	38.5	3711
48296.519	<i>Acinetobacter pittii</i> strain HW4Z	27	3699487	38.7	3435
2707340.3	<i>Acinetobacter</i> sp. EKM10A	58	3977250	39.0	3812
2605268.3	<i>Acinetobacter</i> sp. FNA11	75	2664814	36.6	2622
1427535.3	<i>Actinomadura barringtoniae</i> strain GKU 128	76	11715108	70.5	10935
561175.3	<i>Actinomadura flavalba</i> DSM 45200	24	6167487	73.3	5699
1427538.3	<i>Actinomadura syzygii</i> strain	27	9698644	71.8	9188
33936.22	<i>Aeribacillus pallidus</i> strain P18	1	3833114	39.0	4258
2775288.3	<i>Aeromicrobium</i> sp. CFBP 8757 strain CFBP8757	25	3757853	71.2	3737
1736306.3	<i>Aeromicrobium</i> sp. Leaf245	20	3469640	71.6	3404
1736498.3	<i>Aeromicrobium</i> sp. Root236	1	3919167	69.0	3966
1736521.3	<i>Aeromicrobium</i> sp. Root344	1	3874254	68.8	3895
1736540.3	<i>Aeromicrobium</i> sp. Root472D3	4	3707866	71.3	3722
358.5	<i>Agrobacterium tumefaciens</i>	61	5122354	56.6	4883
1736274.3	<i>Arthrobacter</i> sp. Leaf145	29	4493998	63.2	4339
2830984.3	<i>Arthrobacter</i> sp. NtRootA2	3	4585189	63.1	4346
2744252.3	<i>Aureimonas</i> sp. ME7	60	4379078	66.9	4275
192.45	<i>Azospirillum brasilense</i> strain Az39	6	7391279	68.5	6699
192.5	<i>Azospirillum brasilense</i> strain Sp7	6	6587527	68.3	6081
137722.3	<i>Azospirillum</i> sp. B510	7	7599738	67.6	6784
293387.105	<i>Bacillus altitudinis</i> strain Ba1449	1	3718493	41.4	3856
293387.4	<i>Bacillus altitudinis</i> strain FD48	1	3751533	41.2	4010
293387.41	<i>Bacillus altitudinis</i> strain Lc5	69	3628062	41.4	3871
293387.11	<i>Bacillus altitudinis</i> strain RIT380	85	3972159	41.0	4365
1390.838	<i>Bacillus amyloliquefaciens</i> Bam1	1	3954399	45.9	4244
1390.788	<i>Bacillus amyloliquefaciens</i> strain GKT04	2	4149690	46.3	4332
1449088.3	<i>Bacillus amyloliquefaciens</i> subsp. <i>Plantarum</i> TrigoCor1448	1	3957904	46.5	3986
1155777.3	<i>Bacillus amyloliquefaciens</i> subsp. <i>Plantarum</i> YAU B9601-Y2	1	4242774	45.9	4278
1126211.3	<i>Bacillus amyloliquefaciens</i> Y2	1	4238624	45.9	4289

1392.242	<i>Bacillus anthracis strain RIT375</i>	62	5677918	34.9	6044
1452.35	<i>Bacillus atrophaeus strain GQJK17</i>	1	4325818	43.3	4507
1396.282	<i>Bacillus cereus strain 21155</i>	5	6258015	34.9	6438
1397.17	<i>Bacillus circulans strain PK3_109</i>	1	5175203	35.6	5309
1397.18	<i>Bacillus circulans strain PK3_138</i>	1	5274417	35.7	5399
483913.168	<i>Bacillus inaquosorum strain CCSR02</i>	1	4136264	44.1	4293
1402.438	<i>Bacillus licheniformis strain ME35</i>	80	4252220	46.1	4692
492670.14	<i>Bacillus methylotrophicus strain B26</i>	26	3868756	46.6	3914
72360.25	<i>Bacillus mojavensis strain PS17</i>	1	4083759	43.9	4316
1408.517	<i>Bacillus pumilus strain EB130</i>	1	3614840	41.8	3973
1408.521	<i>Bacillus pumilus strain HW16</i>	33	3645212	41.3	3768
1408.101	<i>Bacillus pumilus strain TUAT1</i>	1	3723433	41.4	3899
561879.153	<i>Bacillus safensis strain HW4</i>	19	3832643	41.3	4034
561879.67	<i>Bacillus safensis strain PgKB20</i>	1	3786744	41.7	3890
2233998.3	<i>Bacillus sp. COPE52</i>	1	5806513	35.1	6129
2231056.4	<i>Bacillus sp. CR71</i>	1	5914775	34.7	6270
393065.4	<i>Bacillus sp. E25</i>	1	5915788	34.7	6307
2708091.3	<i>Bacillus sp. EKM104P</i>	30	3722775	41.3	3941
1683607.3	<i>Bacillus sp. EKM213B</i>	61	4198232	45.8	4400
1683635.3	<i>Bacillus sp. EKM420B</i>	83	3908635	46.5	4047
1683642.3	<i>Bacillus sp. EKM601B</i>	73	4199360	45.8	4408
2585772.3	<i>Bacillus sp. HNA3</i>	1	3929648	46.5	4065
1409.8	<i>Bacillus sp. strain Co1-6</i>	1	3922431	46.6	3997
2448811.3	<i>Bacillus sp. WR11</i>	1	4202080	43.5	4444
1423.139	<i>Bacillus subtilis SG6</i>	1	4079669	43.8	4197
1423.765	<i>Bacillus subtilis strain 30VD-1</i>	18	3957760	43.8	4108
1423.1052	<i>Bacillus subtilis strain AMR1</i>	1	4142132	43.5	4453
1423.1814	<i>Bacillus subtilis strain Dcl1</i>	46	4294389	43.2	4604
1423.444	<i>Bacillus subtilis strain DKU_NT_02</i>	1	4014255	43.6	4401
1423.445	<i>Bacillus subtilis strain DKU_NT_03</i>	1	4196031	43.3	4634
1423.434	<i>Bacillus subtilis strain ge25</i>	1	4065429	43.8	4206
1423.414	<i>Bacillus subtilis strain GQJK2</i>	1	4072961	43.8	4249
1423.1718	<i>Bacillus subtilis strain SH1</i>	2	4253166	43.6	4485
1423.189	<i>Bacillus subtilis strain UD1022</i>	1	4025326	43.9	4172
1428.933	<i>Bacillus thuringiensis strain rc6</i>	96	5789681	34.9	6231
492670.948	<i>Bacillus velezensis strain AK-0</i>	1	3969447	46.5	4017
492670.5	<i>Bacillus velezensis strain AL7</i>	1	3894709	46.6	3946
492670.755	<i>Bacillus velezensis strain BT2.1</i>	33	3889067	46.4	3998
492670.713	<i>Bacillus velezensis strain BY6</i>	2	3905529	46.6	3921
492670.275	<i>Bacillus velezensis strain DKU_NT_04</i>	3	4328187	45.2	4743
492670.968	<i>Bacillus velezensis strain EB14</i>	1	4007882	46.5	4581
492670.754	<i>Bacillus velezensis strain HD2.2</i>	41	3874124	46.4	3984
492670.167	<i>Bacillus velezensis strain JYYP2</i>	1	3929789	46.5	4020
492670.1072	<i>Bacillus velezensis strain KOF112</i>	2	3929792	46.5	3975
492670.405	<i>Bacillus velezensis strain LP16S</i>	8	3219607	46.2	3261
492670.1113	<i>Bacillus velezensis strain MEJ05</i>	21	3886050	46.4	4000
492670.973	<i>Bacillus velezensis strain Mr12</i>	2	4001152	46.5	4248
492670.84	<i>Bacillus velezensis strain Pm9</i>	1	3890670	46.7	3993



492670.1092	<i>Bacillus velezensis</i> strain Q12	1	4182261	46.1	4333
492670.183	<i>Bacillus velezensis</i> strain T20E-257	2	3900066	46.7	3980
492670.1093	<i>Bacillus velezensis</i> strain US1	2	4132553	46.0	4278
492670.1114	<i>Bacillus velezensis</i> strain WB	1	3896799	46.7	3955
492670.441	<i>Bacillus velezensis</i> strain WRN014	1	4063541	46.3	4179
492670.32	<i>Bacillus velezensis</i> strain ZF2	1	3929773	46.5	4021
492670.177	<i>Bacillus velezensis</i> strain ZL918	1	3922711	46.5	4017
492670.1256	<i>Bacillus velezensis</i> Yao	1	3951864	46.6	4076
179879.13	<i>Burkholderia anthina</i> strain XXVI	100	7692989	67.0	7604
406425.4	<i>Burkholderia cenocepacia</i> MCO-3	3	7971389	66.6	7330
95486.304	<i>Burkholderia cenocepacia</i> strain CR318	3	7664894	66.8	7497
983594.16	<i>Burkholderia cepacia</i> ATCC 25416	4	8605945	66.6	8573
886913.3	<i>Burkholderia cepacia</i> Bu72	6	7423506	66.9	6957
28095.11	<i>Burkholderia gladioli</i> strain ATCC 25417	4	9312825	67.3	8985
28095.58	<i>Burkholderia gladioli</i> strain CGB10	3	8423203	68.1	7972
1176514.8	<i>Burkholderia glumae</i> AU6208	3	6530819	68.4	6473
626418.3	<i>Burkholderia glumae</i> BGR1	6	7284683	67.9	6491
1176492.5	<i>Burkholderia glumae</i> LMG 2196 = ATCC 33617	4	6820727	68.2	6653
337.31	<i>Burkholderia glumae</i> strain 257sh-1	4	6691685	68.2	6591
337.37	<i>Burkholderia glumae</i> strain 336gr-1	5	6879338	68.1	6712
337.38	<i>Burkholderia glumae</i> strain 411gr-6	4	6747595	68.3	6826
337.39	<i>Burkholderia glumae</i> strain 957856-41-c	4	6948206	68.1	6829
337.3	<i>Burkholderia glumae</i> strain GX	5	6866154	68.1	6904
337.33	<i>Burkholderia glumae</i> strain HN	3	6560085	68.3	6423
337.32	<i>Burkholderia glumae</i> strain HN1	5	6680415	68.1	6700
1229205.11	<i>Burkholderia phenoliruptrix</i> BR3459a	3	7651131	63.2	7462
391038.7	<i>Burkholderia phymatum</i> STM815	4	8676562	62.3	8063
398527.4	<i>Burkholderia phytofirmans</i> PsJN	3	8214658	62.3	8038
416344.3	<i>Burkholderia</i> sp. KJ006	4	6629912	67.2	6153
95485.12	<i>Burkholderia stabilis</i> strain EB159	5	9093480	66.0	9010
371608.7	<i>Candidatus Streptomyces philanthi</i> strain 23Af2	1	7341895	72.4	8483
253.169	<i>Chryseobacterium indologenes</i> strain PgBE177	1	5008467	36.4	4641
1932669.4	<i>Chryseobacterium</i> sp. JV274 strain F3_Prior	1	5227721	36.1	4978
2015076.3	<i>Chryseobacterium</i> sp. T16E-39 strain T16E-39	1	4872888	35.2	4572
164546.7	<i>Cupriavidus taiwanensis</i> strain LMG 19424	3	6476522	67.0	5752
164546.61	<i>Cupriavidus taiwanensis</i> SWF66294	3	6782863	66.9	6355
2035.5	<i>Curtobacterium flaccumfaciens</i> MEB126	92	3684420	70.9	3430

138532.17	<i>Curtobacterium flaccumfaciens</i> pv. <i>flaccumfaciens</i> strain Cff1037	2	3781436	70.9	3701
138532.12	<i>Curtobacterium flaccumfaciens</i> pv. <i>flaccumfaciens</i> strain P990	4	3934341	70.4	4672
69373.7	<i>Curtobacterium pusillum</i> strain AA3	2	4541048	70.4	4656
1905847.5	<i>Curtobacterium</i> sp. BH-2-1-1	1	3795948	71.4	3727
2744253.3	<i>Curtobacterium</i> sp. ME12	87	3707131	70.9	3672
2795488.3	<i>Curtobacterium</i> sp. YC1	2	3378526	71.6	3291
682863.4	<i>Delftia</i> sp. Lp-1	7	6271367	67.0	5832
1736374.3	<i>Devosia</i> sp. Leaf420	16	4219601	60.7	4142
1736229.3	<i>Devosia</i> sp. Leaf64	24	4252528	60.4	4198
1736423.3	<i>Devosia</i> sp. Root105	21	5850260	65.4	5778
1736531.3	<i>Devosia</i> sp. Root413D1	14	5851492	65.4	5738
1736537.3	<i>Devosia</i> sp. Root436	16	3919126	63.8	3914
1736575.3	<i>Devosia</i> sp. Root635	24	3816678	64.1	3754
1736587.3	<i>Devosia</i> sp. Root685	5	4397626	61.5	4241
2806348.3	<i>Devosia</i> sp. SL43	2	4264986	62.7	4222
1736266.3	<i>Duganella</i> sp. Leaf126	36	5315014	65.1	4632
1736227.3	<i>Duganella</i> sp. Leaf61	33	6231257	63.5	5411
1736471.3	<i>Duganella</i> sp. Root1480D1	50	6603920	62.0	6140
1736489.3	<i>Duganella</i> sp. Root198D2	51	6542676	62.3	6036
1736518.3	<i>Duganella</i> sp. Root336D2	39	6579467	62.3	6074
471854.5	<i>Dyadobacter fermentans</i> DSM 18053	1	6967790	51.5	5956
1736295.3	<i>Dyadobacter</i> sp. Leaf189	11	6065072	47.0	5303
61645.437	<i>Enterobacter asburiae</i> strain HW15	28	4487493	56.1	4253
61645.336	<i>Enterobacter asburiae</i> strain ME44	93	4519362	56.0	4313
881260.67	<i>Enterobacter bugandensis</i> strain ME13	83	4647965	56.1	4437
881260.66	<i>Enterobacter bugandensis</i> strain ME34	100	4652007	56.1	4422
1812935.2	<i>Enterobacter cloacae</i> complex 'Hoffmann cluster IV' strain 3D9	23	4653802	56.2	4396
550.1524	<i>Enterobacter cloacae</i> strain NFPP41	14	5014465	54.4	4815
158836.552	<i>Enterobacter hormaechei</i> strain EKM10E strain Not applicable	53	4758480	55.6	4554
158836.553	<i>Enterobacter hormaechei</i> strain EKM11E	43	4756423	55.7	4545
1812935.8	<i>Enterobacter roggkampii</i> strain ED5	2	4702851	56.0	4372
1216470.3	<i>Enterobacter</i> sp. SST3	63	4626853	56.1	4257
1683712.3	<i>Enterobacteriaceae</i> bacterium EKM102V	70	4687176	56.6	4628
1351.1815	<i>Enterococcus faecalis</i> strain CECT 7121	17	2901597	37.3	2921
1352.12494	<i>Enterococcus faecium</i> strain LB12	34	2714058	37.7	2755
1352.1249	<i>Enterococcus faecium</i> strain LB15	59	3159195	37.6	3244

1352.12493	<i>Enterococcus faecium</i> strain LB16	49	3158735	37.6	3230
1352.12489	<i>Enterococcus faecium</i> strain LB22	62	3159757	37.6	3244
1352.12492	<i>Enterococcus faecium</i> strain LB23	44	3156575	37.6	3225
1603299.3	<i>Erwinia</i> sp. ErVv1	72	4719019	54.4	4608
986.28	<i>Flavobacterium johnsoniae</i> strain FJ001	1	5855069	34.0	5212
986.7	<i>Flavobacterium johnsoniae</i> strain GSE09	1	5109718	34.3	4410
1587524.6	<i>Flavobacterium</i> sp. MEB061	80	5747109	33.6	4921
1894971.3	<i>Jishengella</i> sp. AZ1-13	83	5963223	71.9	5605
1566297.3	<i>Klebsiella</i> sp. NFIX53	24	5431045	58.0	5278
2202822.3	<i>Kosakonia</i> sp. MH5	52	5759387	53.9	5756
1683645.3	<i>Lactococcus</i> sp. EKM101L	98	2711771	34.9	2905
1683646.3	<i>Lactococcus</i> sp. EKM102L	84	2707623	34.8	2891
1683647.3	<i>Lactococcus</i> sp. EKM201L	94	2709521	34.9	2902
1683649.3	<i>Lactococcus</i> sp. EKM203L	77	2705547	34.8	2892
1683650.3	<i>Lactococcus</i> sp. EKM501L	86	2708771	34.9	2893
1245.191	<i>Leuconostoc mesenteroides</i> strain DPC 7261	56	3345650	40.3	3421
2724526.3	<i>Leuconostoc</i> sp. DB-1	3	2067165	37.7	2169
1082933.13	<i>Mesorhizobium amorphae</i> CCNWGS0123	5	7343952	62.9	7374
278148.9	<i>Mesorhizobium ciceri</i> biovar biserrulae strain WSM1284	2	6880454	62.5	6926
765698.19	<i>Mesorhizobium ciceri</i> biovar biserrulae WSM1271	2	6690028	62.6	6763
28104.15	<i>Mesorhizobium huakuii</i> strain 583	4	8449246	62.5	9118
1777867.5	<i>Mesorhizobium jarvisii</i> strain ATCC 700743	3	7201057	62.9	7537
935546.5	<i>Mesorhizobium loti</i> NZP2037	2	7481739	62.8	7703
381.111	<i>Mesorhizobium loti</i> strain 582	3	8330599	62.5	8632
381.7	<i>Mesorhizobium loti</i> strain TONO	4	8452151	62.6	8688
2744515.3	<i>Mesorhizobium</i> sp. 113-1-2	1	7938383	62.8	8046
2744517.3	<i>Mesorhizobium</i> sp. 113-3-9	1	7592533	62.9	7617
2777475.3	<i>Mesorhizobium</i> sp. J8	1	6690860	63.8	6658
2744523.3	<i>Mesorhizobium</i> sp. L-8-10	1	8451969	64.2	8419
2744522.3	<i>Mesorhizobium</i> sp. L-8-3	1	8445954	64.2	8388
2483402.3	<i>Mesorhizobium</i> sp. NZP2234	1	6749717	63.1	6805
2483403.3	<i>Mesorhizobium</i> sp. NZP2298	1	7336816	62.8	7427
1736371.3	<i>Methylophilus</i> sp. Leaf414	9	3028624	48.2	2949
1736373.3	<i>Methylophilus</i> sp. Leaf416	8	2940342	48.2	2823
1736383.3	<i>Methylophilus</i> sp. Leaf459	16	2948005	48.2	2819
104336.7	<i>Microbacterium foliorum</i> strain 122	1	3744056	67.9	3664
1587523.3	<i>Microbacterium</i> sp. ME1108Y	43	3669952	68.0	3553
1270.65	<i>Micrococcus luteus</i> strain TYU2	2	2488099	73.0	2339
1530042.3	<i>Micromonospora costii</i> strain CS1-	20	7182425	73.0	6926
515350.1	<i>Micromonospora endophytica</i> (Xie et al. 2001) Li et al. 2019 strain NBRC 109090	1	6586761	70.9	6553



709876.7	<i>Micromonospora noduli</i> strain Lupac 07	56	7120389	71.1	6711
285676.35	<i>Micromonospora saelicesensis</i> strain GAR05	75	7063653	71.2	6747
285676.38	<i>Micromonospora saelicesensis</i> strain GAR06	60	6963058	71.2	6614
285676.36	<i>Micromonospora saelicesensis</i> strain Lupac 06	59	7060307	71.2	6712
285676.39	<i>Micromonospora saelicesensis</i> strain PSN13	41	7378898	71.1	7069
2783662.3	<i>Micromonospora</i> sp. ANENR4	66	6550221	72.9	6222
2604467.3	<i>Micromonospora</i> sp. AP08	53	6187705	73.4	6010
2824876.3	<i>Micromonospora</i> sp. C31	84	6446142	73.3	5814
2824877.3	<i>Micromonospora</i> sp. C32	37	6548084	72.9	6221
2824878.3	<i>Micromonospora</i> sp. C41	75	6820983	72.9	6398
2717028.3	<i>Micromonospora</i> sp. CMU55-4	68	6624929	73.0	6169
2824886.3	<i>Micromonospora</i> sp. D93	65	7235627	70.9	6954
648999.4	<i>Micromonospora</i> sp. L5	1	6962533	72.9	6283
2824899.3	<i>Micromonospora</i> sp. U21	99	7467210	71.6	7466
1914461.4	<i>Micromonospora terminaliae</i> strain TMS7	40	6699733	73.4	6488
528178.3	<i>Neorhizobium alkalisoli</i> strain DSM 21826	68	5869725	60.3	5939
399.197	<i>Neorhizobium galegae</i> 7g	3	6049186	61.7	6057
399.115	<i>Neorhizobium galegae</i> strain Ng37	4	6552719	61.5	6556
399.118	<i>Neorhizobium galegae</i> strain Ng58	64	6565777	61.5	6719
399.116	<i>Neorhizobium galegae</i> strain Ng87	6	6741517	61.0	7076
399.196	<i>Neorhizobium galegae</i> VafX2	2	6733441	61.5	8443
2712894.3	<i>Neorhizobium</i> sp. AL 9.2.2	20	4904276	60.3	4793
2060726.3	<i>Neorhizobium</i> sp. SOG26	3	4561173	61.2	4543
37330.29	<i>Nocardia nova</i> strain LGO-A14	21	7959976	67.9	7625
2802282.3	<i>Nocardia</i> sp. 2 strain LPG-2	76	8382712	68.0	7986
2675850.3	<i>Nocardia</i> sp. CT2-14	42	7376943	67.9	7114
1736458.3	<i>Nocardia</i> sp. Root136	14	7282860	67.3	6948
2769252.3	<i>Nocardia</i> sp. XZ_19_231	28	7432816	67.0	7098
2582905.3	<i>Nocardioides</i> sp. S-1144	1	4426066	73.5	4642
2824874.3	<i>Nocardiopsis</i> sp. B62	83	5980749	70.7	5521
349520.38	<i>Paenibacillus polymyxa</i> E681 strain type B	1	5416175	45.8	5073
1406.33	<i>Paenibacillus polymyxa</i> EB4 G3	1	5739603	45.8	5198
1406.49	<i>Paenibacillus polymyxa</i> strain CF05	1	5762608	45.5	5138
1406.193	<i>Paenibacillus polymyxa</i> strain ZF129	3	5820553	45.4	5362
2871175.6	<i>Paenibacillus</i> sp. DR312	1	7299545	45.8	6476
2823904.3	<i>Paenibacillus</i> sp. S02	1	6060529	45.6	5609
2823905.3	<i>Paenibacillus</i> sp. S25	1	5958851	45.7	5453
2026089.3	<i>Paenibacillus</i> sp. XY044	27	7869445	52.3	7713

93220.83	<i>Pandoraea pnomenus</i> strain TF-18	1	5499432	64.8	5137
549.211	<i>Pantoea agglomerans</i> strain CFBP13505	48	4863834	54.9	4832
549.332	<i>Pantoea agglomerans</i> strain Pa17-	96	4829682	55.0	4722
1891675.3	<i>Pantoea alhagistrain</i> LTYR-11Z	2	4316301	53.5	4297
553.286	<i>Pantoea ananatis</i> strain PNA_98_8	31	4924392	53.4	4983
470931.12	<i>Pantoea anthophila</i> strain	75	4687858	56.6	4633
1683695.3	<i>Pantoea</i> sp. EKM101V	61	4593280	56.9	4462
1683696.3	<i>Pantoea</i> sp. EKM103V	76	4687205	56.6	4639
2708058.3	<i>Pantoea</i> sp. EKM10T	48	4708129	55.2	4504
2708059.3	<i>Pantoea</i> sp. EKM20T	65	4884693	55.3	4731
2708060.3	<i>Pantoea</i> sp. EKM21T	36	4873035	55.1	4701
2708061.3	<i>Pantoea</i> sp. EKM22T	46	4873346	55.1	4698
470934.12	<i>Pantoea vagans</i> strain PaVv11	84	4850774	55.1	4684
1683697.3	<i>Pediococcus</i> sp. EKM201D	13	1863914	37.1	1835
1683698.3	<i>Pediococcus</i> sp. EKM202D	16	1863914	37.1	1836
363852.4	<i>Pedobacter ginsengisoli</i> strain T01R-27	1	5373360	37.8	4682
332999.9	<i>Pedobacter suwonensis</i> strain Fj_001	1	5689882	39.5	5186
1736430.3	<i>Pelomonas</i> sp. Root1217	32	6649903	67.4	6132
150123.4	<i>Plantibacter flavus</i> strain 251	1	4211682	69.2	4014
47877.19	<i>Pseudomonas amygdali</i> 35-1	4	6596445	58.3	6115
930166.114	<i>Pseudomonas brassicacearum</i> strain 3Re2-7	1	6738544	60.8	6382
930166.5	<i>Pseudomonas brassicacearum</i> strain DF41	1	6652396	60.5	5773
930166.18	<i>Pseudomonas brassicacearum</i> strain L13-6-12	1	6715909	60.9	6156
930166.17	<i>Pseudomonas brassicacearum</i> strain TM1A3	29	6692370	60.9	6170
86192.15	<i>Pseudomonas chlororaphis</i> subsp. <i>aurantiaca</i> strain 449	1	6962068	62.9	6355
86192.14	<i>Pseudomonas chlororaphis</i> subsp. <i>aurantiaca</i> strain M12	1	6730980	63.0	6147
86192.17	<i>Pseudomonas chlororaphis</i> subsp. <i>aurantiaca</i> strain M71	1	6807002	62.9	6205
86192.2	<i>Pseudomonas chlororaphis</i> subsp. <i>aurantiaca</i> strain Q16	1	6928549	62.8	6274
587851.16	<i>Pseudomonas chlororaphis</i> subsp. <i>aureofaciens</i> strain ChPhzTR18	1	6873200	62.9	6339
587851.13	<i>Pseudomonas chlororaphis</i> subsp. <i>aureofaciens</i> strain ChPhzTR36	1	6934077	62.7	6317
587851.17	<i>Pseudomonas chlororaphis</i> subsp. <i>aureofaciens</i> strain ChPhzTR38	1	6947716	62.8	6359
587851.8	<i>Pseudomonas chlororaphis</i> subsp. <i>aureofaciens</i> strain P2	1	7203062	62.8	6638

1513890.12	<i>Pseudomonas chlororaphis</i> subsp. <i>piscium</i> strain PCL1391	1	6870622	62.8	6176
1513890.17	<i>Pseudomonas chlororaphis</i> subsp. <i>piscium</i> strain ToZa7	1	7015602	62.6	6348
1513890.6	<i>Pseudomonas chlororaphis</i> subsp. <i>piscium</i> strain ZJU60	1	6818002	62.8	6289
43306.3	<i>Pseudomonas denitrificans</i> (nomen <i>rejciendum</i> ) strain BG1	1	6880508	65.3	6719
1037911.3	<i>Pseudomonas fluorescens</i> A506	1	5962570	60.0	5300
294.93	<i>Pseudomonas fluorescens</i> MEP34	58	5914240	60.3	5174
216595.4	<i>Pseudomonas fluorescens</i> SBW25	2	7147633	60.1	6256
1038924.3	<i>Pseudomonas fluorescens</i> SS101	1	6179543	60.0	5309
294.391	<i>Pseudomonas fluorescens</i> strain FR1	1	6869391	60.8	6451
294.127	<i>Pseudomonas fluorescens</i> strain PCL1751	1	6143950	60.4	5637
158627.16	<i>Pseudomonas graminis</i> strain PgKB30	1	5521192	60.7	4957
76758.16	<i>Pseudomonas orientalis</i> strain R4-35-08	1	5912532	60.7	5331
303.935	<i>Pseudomonas putida</i> PgBE89	1	6299263	60.1	5678
303.79	<i>Pseudomonas putida</i> strain BR-PH17	2	5854781	61.4	5508
303.234	<i>Pseudomonas putida</i> strain IAC-RBcr5	26	5760798	62.4	5249
303.818	<i>Pseudomonas putida</i> strain ITEM 17297	86	5112403	61.9	4743
303.2	<i>Pseudomonas putida</i> strain PC2	1	5808624	63.2	5262
303.913	<i>Pseudomonas putida</i> strain PgBE89	1	6299263	60.1	5678
303.65	<i>Pseudomonas putida</i> strain W5	1	6166183	62.7	5692
390235.5	<i>Pseudomonas putida</i> W619	1	5774330	61.4	5246
321846.6	<i>Pseudomonas simiae</i> strain	44	6105492	60.4	5682
2870840.3	<i>Pseudomonas</i> sp. DR208	2	6273349	61.0	5720
2871095.3	<i>Pseudomonas</i> sp. DR48	5	7251677	58.9	6828
1825979.3	<i>Pseudomonas</i> sp. p21	15	5547454	63.2	5168
2202823.3	<i>Pseudomonas</i> sp. Q1	91	6462713	60.4	6019
1573714.4	<i>Pseudomonas</i> sp. S04	1	6098669	60.9	5661
316.162	<i>Pseudomonas stutzeri</i> strain KMS 55	51	4611727	63.8	4428
192087.6	<i>Pseudomonas syringae</i> pv. <i>atrofaciens</i> strain LMG5095 strain <i>Triticum aestivum</i> (Aestivum Group)	1	6080544	59.0	5487
1262350.4	<i>Pseudomonas syringae</i> pv. <i>syringae</i> SM	2	6124102	58.7	5276
33069.5	<i>Pseudomonas viridiflava</i> LMCA8	73	5992488	59.3	5304
33069.78	<i>Pseudomonas viridiflava</i> strain KF485	3	6221425	59.1	5694

33069.81	<i>Pseudomonas viridiflava</i> strain SV1779	1	6106689	59.2	5567
33069.82	<i>Pseudomonas viridiflava</i> strain T1426	1	6075960	59.3	5537
2703885.9	<i>Rahnella aceris</i> strain AR20	89	5650676	52.4	5505
34038.27	<i>Rahnella aquatilis</i> H145	4	6148824	51.6	6105
1151116.3	<i>Rahnella aquatilis</i> HX2	2	5533124	52.2	5048
34038.13	<i>Rahnella aquatilis</i> strain WPL_8_1	18	5554461	52.2	5405
34038.12	<i>Rahnella aquatilis</i> strain ZF7	2	5536721	52.3	5361
2816248.3	<i>Rahnella</i> sp. H11b	68	5403853	51.9	5064
1914994.3	<i>Rahnella</i> sp. J11-6 strain J11-6	81	5012778	53.2	4749
2816246.3	<i>Rahnella</i> sp. L72c	72	5753075	51.4	5621
2855688.3	<i>Rahnella</i> sp. PD12R	33	5504954	53.2	5315
2292446.3	<i>Rahnella</i> sp. RcJ3	60	5419222	52.2	5252
574964.4	<i>Rahnella variigena</i> strain CIP	2	5499108	52.0	5211
1510570.5	<i>Rahnella victoriana</i> strain BRK18a	2	5563295	53.5	5236
305.879	<i>Ralstonia solanacearum</i> strain YQ	2	5822685	66.9	5479
508661.4	<i>Rhizobium herbae</i> strain HU44	28	5123125	61.4	5104
2775301.3	<i>Rhizobium</i> sp. CFBP 8752 strain CFBP8752	13	4888650	60.3	4764
690585.4	<i>Rhizobium vignae</i> strain CCBAU 05176	84	6343049	61.6	6191
398843.6	<i>Rhodococcus kyotonensis</i> strain KB10	41	5471075	65.2	5209
334542.26	<i>Rhodococcus qingshengii</i> strain RL1	3	6846621	62.4	6674
1587522.3	<i>Rhodococcus</i> sp. MEB064	81	4664255	67.0	4418
1646377.35	<i>Rouxiella badensis</i> 20GA0316	38	5188000	52.7	4943
1646377.29	<i>Rouxiella badensis</i> 21GA0654	38	5148315	52.7	4890
1646377.17	<i>Rouxiella badensis</i> strain Rb14-3	32	5019837	52.9	4785
1646377.6	<i>Rouxiella badensis</i> strain SER3	1	5082781	52.8	4850
82995.6	<i>Serratia grimesii</i> strain BXF1	1	5090820	52.8	4826
614.41	<i>Serratia liquefaciens</i> strain FG3	3	5991142	55.2	6047
615.446	<i>Serratia marcescens</i> strain 1274	1	5214984	59.8	5527
615.144	<i>Serratia marcescens</i> strain B3R3	2	5594610	59.1	5491
615.1213	<i>Serratia marcescens</i> strain BP2	1	5088390	59.8	4913
615.141	<i>Serratia marcescens</i> strain RSC-14 (KNU-AP07, KCTC 12887BP)	1	5127030	59.6	4900
82996.18	<i>Serratia plymuthica</i> strain 3Re4-18	1	5439574	56.2	5188
82996.17	<i>Serratia plymuthica</i> strain 3Rp8	1	5546041	56.1	5346
82996.45	<i>Serratia plymuthica</i> strain C-1	1	5695428	56.1	5572
82996.58	<i>Serratia plymuthica</i> strain UBCF_13	1	5455815	56.2	5241
61652.43	<i>Serratia rubidaea</i> H440	2	4955777	59.2	4838
61652.42	<i>Serratia rubidaea</i> H469	2	4955995	59.2	4839
61652.24	<i>Serratia rubidaea</i> strain NCTC10848	1	4995355	57.7	4926
2447890.3	<i>Serratia</i> sp. 1D1416	1	5552047	59.6	5527
2597702.3	<i>Serratia</i> sp. KUDC3025	1	5098052	58.0	4933

2744257.3	<i>Serratia</i> sp. ME47	82	5276078	59.4	5175
2697369.3	<i>Serratia</i> sp. NGAS9	1	5155009	59.4	4928
300181.31	<i>Serratia ureilytica</i> strain CC119	2	5213616	59.5	4999
2100533.3	<i>Sphingobacterium</i> sp. 5JN-11	75	4694399	42.8	4249
2557994.3	<i>Sphingobacterium</i> sp. CZ-2	1	3925976	40.7	3565
2152717.3	<i>Sphingobacterium</i> sp. M46	17	6853865	40.6	6191
2546448.3	<i>Sphingobacterium</i> sp. NYYP31	48	4720650	36.4	4322
1848973.6	<i>Sphingomonas coloccasiae</i> strain JCM 31229	23	6569909	65.6	6098
2775302.3	<i>Sphingomonas</i> sp. CFBP 13720 strain CFBP13720	27	3692396	66.0	3542
1735690.3	<i>Sphingomonas</i> sp. Leaf24	22	4001919	66.4	3761
267128.3	<i>Sphingopyxis granuli</i> strain TFA	1	4679853	66.2	4499
40324.1951	<i>Stenotrophomonas maltophilia</i> strain 1 strain TR1	63	4449516	66.6	4128
487698.12	<i>Stenotrophomonas pavani</i> strain ME55	74	4393944	67.3	4115
216778.48	<i>Stenotrophomonas rhizophila</i> strain CFBP 13503	35	4855517	67.3	4460
216778.5	<i>Stenotrophomonas rhizophila</i> strain IS26	42	4087665	66.4	3676
2705012.3	<i>Streptacidiphilus</i> sp. PB12-B1b	1	7672494	72.6	7007
1886.94	<i>Streptomyces albidoflavus</i> strain LGO-A16	1	7054925	73.4	6308
1886.95	<i>Streptomyces albidoflavus</i> strain LGO-A23	1	7073859	73.4	6316
1886.66	<i>Streptomyces albidoflavus</i> strain UYFA156	2	7131381	73.4	6583
1892.26	<i>Streptomyces anulatus</i> strain	1	8874447	71.5	8005
1892.31	<i>Streptomyces anulatus</i> strain YINM00001	2	8690773	71.6	7758
2495589.3	<i>Streptomyces brunneus</i> strain CR22	1	10393987	70.8	9803
67258.4	<i>Streptomyces cavourensis</i> strain 1AS2a	1	7600475	72.1	6788
1969.5	<i>Streptomyces chartreusis</i> strain WZS021	22	10404130	70.6	9876
1908.1	<i>Streptomyces globisporus</i> strain TFH56	3	7666521	71.5	7248
67304.8	<i>Streptomyces griseorubiginosus</i> strain 3E-1	1	9512378	70.9	8665
67304.1	<i>Streptomyces griseorubiginosus</i> strain BTU6	1	9226027	71.2	8426
33900.7	<i>Streptomyces murinus</i> strain Am1	1	8539264	71.8	7698
1169025.9	<i>Streptomyces pratensis</i> strain S10	2	8281466	70.5	7275
545123.5	<i>Streptomyces silaceus</i> strain ACCC40021	1	8625867	72.1	7813
1851167.4	<i>Streptomyces</i> sp. 11-1-2	2	11655206	70.9	10366

2748873.3	<i>Streptomyces sp. CB04723</i>	1	7872196	72.4	7224
2153485.3	<i>Streptomyces sp. endophyte_N2</i>	1	8428700	71.8	7505
2653200.3	<i>Streptomyces sp. GY16</i>	2	10175440	70.9	9635
2305221.3	<i>Streptomyces sp. KPB2</i>	1	8082236	72.2	7449
2609807.3	<i>Streptomyces sp. LBUM 1475</i>	1	9741971	71.6	9085
2609808.3	<i>Streptomyces sp. LBUM 1480</i>	1	9982980	71.5	9360
2609809.3	<i>Streptomyces sp. LBUM 1482</i>	1	9782455	71.6	9156
646637.3	<i>Streptomyces sp. M2</i>	1	8718751	70.7	8126
2750011.3	<i>Streptomyces sp. Rer75</i>	1	10474412	71.4	9062
2705439.3	<i>Streptomyces sp. S4.7</i>	3	7940295	70.7	7234
1849967.7	<i>Streptomyces sp. SAT1</i>	1	7472530	73.2	6669
2563602.3	<i>Streptomyces sp. SS52</i>	1	8184045	72.5	7604
2855836.3	<i>Streptomyces sp. WY228</i>	1	8286721	71.7	7509
2021410.4	<i>Streptomyces sp. XY006</i>	74	8251847	72.3	7602
2730915.3	<i>Streptomyces sp. Z423-1</i>	1	9460473	70.8	9016
339.49	<i>Xanthomonas campestris LMCP11</i>	47	5063135	65.3	4336
339.239	<i>Xanthomonas campestris strain HWA1</i>	53	5005491	65.4	4533

Tabla II. Tabla de genomas de BAP con secuencias homólogas al dominio C resultado de un análisis de BLAST.

ID de genoma	Nombre del genoma	País de aislamiento		Hospedero
1148157.47	<i>Acinetobacter oleivorans strain HW32</i>	Australia		Trigo
48296.519	<i>Acinetobacter pittii strain HW4Z</i>	Australia		Trigo
1427535.3	<i>Actinoadura barringtoniae strain GKU 128</i>	Tailandia		<i>Barringtonia acutangula</i>
561175.3	<i>Actinoadura flavaalba DSM 45200</i>	China		<i>Maytenus austroyunnanensis</i>
1427538.3	<i>Actinoadura syzygii strain GKU157</i>	Tailandia		<i>Syzygium cumini</i>
33936.22	<i>Aeribacillus pallidus strain PI8</i>	Japón		<i>Glycine max</i>
192.45	<i>Azospirillum brasilense strain Az39</i>	Argentina		<i>Triticum aestivum</i>
192.5	<i>Azospirillum brasilense strain Sp7</i>	Brasil		<i>Digitaria eriantha</i>
293387.105	<i>Bacillus altitudinis strain Ba1449</i>	China		<i>Citrus reticulata</i>
293387.4	<i>Bacillus altitudinis strain FD48</i>	India		<i>Oryza sativa</i>
293387.41	<i>Bacillus altitudinis strain Lc5</i>	India		Arroz negro
293387.11	<i>Bacillus altitudinis strain RIT380</i>	Puerto Rico		<i>Costus igneus</i>
1390.795	<i>Bacillus amyloliquefaciens strain Bam1</i>	China		Pepino
1390.788	<i>Bacillus amyloliquefaciens strain GKT04</i>	China		Platano
1449088.3	<i>Bacillus amyloliquefaciens subsp. plantarum TrigoCor1448</i>	Brasil		<i>Triticum aestivum</i>
1155777.3	<i>Bacillus amyloliquefaciens subsp. plantarum YAU B9601-Y2</i>	China		<i>Triticum aestivum</i>
1126211.3	<i>Bacillus amyloliquefaciens Y2</i>	No reportado		<i>Triticum aestivum</i>
1392.242	<i>Bacillus anthracis strain RIT375</i>	Puerto Rico		<i>Costus igneus</i>
1452.35	<i>Bacillus atrophaeus strain GQJK17</i>	China		<i>Lycium barbarum L.</i>
1396.282	<i>Bacillus cereus strain 21155</i>	China		<i>Angelica sinensis Dlies</i>
483913.168	<i>Bacillus inaquosorum strain</i>	Pakistán		<i>Cenchrus ciliaris</i>
1402.438	<i>Bacillus licheniformis strain ME35</i>	India		<i>Terminalia bellirica</i>
492670.14	<i>Bacillus methylotrophicus strain</i>	Canadá		Hierba de cambio

72360.25	<i>Bacillus mojavensis</i> strain PS17	Tayikistán	<i>Triticum aestivum</i>
1408.517	<i>Bacillus pumilus</i> strain EB130	Irán	<i>Coriandrum sativum</i>
1408.521	<i>Bacillus pumilus</i> strain HW16	Australia	Trigo
561879.153	<i>Bacillus safensis</i> strain HW4	Australia	Trigo
561879.67	<i>Bacillus safensis</i> strain PgKB20	Corea del sur	<i>Panax ginseng</i>
2238998.3	<i>Bacillus</i> sp. COPE52	México	<i>Rubus ulmifolius</i>
2231056.4	<i>Bacillus</i> sp. CR71	México	<i>Physalis ixocarpa</i>
393065.4	<i>Bacillus</i> sp. E25	México	<i>Physalis ixocarpa</i>
2708091.3	<i>Bacillus</i> sp. EKM104P	Canadá	<i>Burpless beauty cucumber</i>
1683607.3	<i>Bacillus</i> sp. EKM213B	Canadá	Melón cantalupo
1683635.3	<i>Bacillus</i> sp. EKM420B	Canadá	Sandía variedad dulce
1683642.3	<i>Bacillus</i> sp. EKM601B	Canadá	Okra Chino
258772.3	<i>Bacillus</i> sp. HWA3	China	<i>Glycine max</i>
1409.8	<i>Bacillus</i> sp. strain Co1-6	Egipto	<i>Calendula officinalis</i>
2448811.3	<i>Bacillus</i> sp. WR11	China	<i>Triticum aestivum</i> L.
1423.139	<i>Bacillus subtilis</i> SG6	China	<i>Luffa</i>
1423.765	<i>Bacillus subtilis</i> strain 30VD-1	No reportado	No reportado
1423.1052	<i>Bacillus subtilis</i> strain AMR1	India	<i>Oryza sativa</i>
1423.1814	<i>Bacillus subtilis</i> strain Dd1	India	<i>Curcuma longa</i>
1423.444	<i>Bacillus subtilis</i> strain DKU_NT_02	Corea del sur	<i>Glycine max</i>
1423.445	<i>Bacillus subtilis</i> strain DKU_NT_03	Corea del sur	<i>Glycine max</i>
1423.434	<i>Bacillus subtilis</i> strain qe25	China	<i>Panax ginseng</i>
1423.414	<i>Bacillus subtilis</i> strain GOJK2	China	<i>Lycium barbarum</i> L.
1423.1718	<i>Bacillus subtilis</i> strain SH1	Corea del sur	<i>Solanum lycopersicum</i>
1423.189	<i>Bacillus subtilis</i> strain UD1022	EUA	<i>Daucus carota</i> var. <i>sativa</i>
1428.933	<i>Bacillus thuringiensis</i> strain rc6	India	Arroz negro
492670.948	<i>Bacillus velezensis</i> strain AK-0	Corea del sur	ginseng
492670.5	<i>Bacillus velezensis</i> strain AL7	China	<i>Gossypium hirsutum</i>
492670.755	<i>Bacillus velezensis</i> strain BT2.1	Viet Nam	Árbol de fruta dragón
492670.713	<i>Bacillus velezensis</i> strain BY6	China	<i>Populus davidiana</i> x <i>Populus alba</i> var. <i>pyramidalis</i> (PdPap)
492670.275	<i>Bacillus velezensis</i> strain DKU_NT_04	Corea del sur	<i>Glycine max</i>
492670.968	<i>Bacillus velezensis</i> strain EB14	Canadá	<i>Populus x jackii</i>
492670.754	<i>Bacillus velezensis</i> strain HD2.2	Viet Nam	Planta de jitomate
492670.167	<i>Bacillus velezensis</i> strain JTYP2	China	<i>Crassulaceae</i>
492670.1072	<i>Bacillus velezensis</i> strain KOF112	Japón	<i>Vitis vinifera</i>
492670.405	<i>Bacillus velezensis</i> strain LP16S	EUA	Sorgo
492670.1113	<i>Bacillus velezensis</i> strain MEJ05	China	Trigo
492670.973	<i>Bacillus velezensis</i> strain Mr12	China	<i>Malus sieversii</i>
492670.84	<i>Bacillus velezensis</i> strain Pm9	China	<i>Malus sieversii</i>
492670.1092	<i>Bacillus velezensis</i> strain Q12	Irán	<i>Solanum tuberosum</i>
492670.183	<i>Bacillus velezensis</i> strain T20E-257	Corea del sur	<i>Solanum lycopersicum</i>
492670.1093	<i>Bacillus velezensis</i> strain US1	Irán	<i>Solanum lycopersicum</i>
492670.1114	<i>Bacillus velezensis</i> strain WB	China	Sandía
492670.441	<i>Bacillus velezensis</i> strain WRN014	China	<i>Oryza sativa</i>
492670.32	<i>Bacillus velezensis</i> strain ZF2	China	<i>Cucumis sativus</i>
492670.177	<i>Bacillus velezensis</i> strain ZL918	China	<i>Sagittaria sagittifolia</i>
406425.4	<i>Burkholderia cenocepacia</i> MCO-3	No reportado	<i>Zea mays</i>
95486.304	<i>Burkholderia cenocepacia</i> strain CR318	Canadá	<i>Zea mays</i>
886913.3	<i>Burkholderia cepacia</i> Bu72	No reportado	<i>Lupinus luteus</i>
28095.11	<i>Burkholderia gladioli</i> strain ATCC 25417	No reportado	<i>Gladiolus</i>
28095.58	<i>Burkholderia gladioli</i> strain CG810	China	<i>Saccharum officinarum</i>
1176514.8	<i>Burkholderia glumae</i> AU6208	No reportado	<i>Oryza sativa</i>
626418.3	<i>Burkholderia glumae</i> BGR1	No reportado	<i>Oryza sativa</i>
1176492.5	<i>Burkholderia glumae</i> LMG 2196 = ATCC 33617	Japón	<i>Oryza sativa</i>
337.31	<i>Burkholderia glumae</i> strain 257sh-	EUA	<i>Oryza sativa</i>
337.37	<i>Burkholderia glumae</i> strain 336gr-1	EUA	<i>Oryza sativa</i>
337.38	<i>Burkholderia glumae</i> strain 411gr-6	EUA	<i>Oryza sativa</i>



337.39	<i>Burkholderia glumae</i> strain 957856-41-c	EUA	<i>Oryza sativa</i>
337.3	<i>Burkholderia glumae</i> strain GX	China	<i>Oryza sativa</i>
337.33	<i>Burkholderia glumae</i> strain HN	China	<i>Oryza sativa</i>
337.32	<i>Burkholderia glumae</i> strain HN1	No reportado	<i>Oryza sativa</i>
95485.12	<i>Burkholderia stabilis</i> strain EB159	Corea del sur	ginsen
371608.7	<i>Candidatus Streptomyces philanthi</i> strain 23Af2	Alemania	<i>Phlonthus triangulum</i>
2035.5	<i>Curtobacterium flaccumfaciens</i> MEB126	EUA	<i>Arabidopsis thaliana</i>
2795488.3	<i>Curtobacterium</i> sp. YC1	China	<i>Nostoc flagelliforme</i>
1736266.3	<i>Duganella</i> sp. Leaf126	Suiza	<i>Arabidopsis thaliana</i>
1736471.3	<i>Duganella</i> sp. Root1480D1	Alemania	<i>Arabidopsis thaliana</i>
1736489.3	<i>Duganella</i> sp. Root198D2	Alemania	<i>Arabidopsis thaliana</i>
1736518.3	<i>Duganella</i> sp. Root336D2	Alemania	<i>Arabidopsis thaliana</i>
61645.437	<i>Enterobacter asbuniae</i> strain HW15	Australia	Trigo
61645.336	<i>Enterobacter asbuniae</i> strain ME44	India	<i>Coscinium fenestratum</i>
881260.67	<i>Enterobacter bugandensis</i> strain ME13	India	<i>Terminalia paniculata</i>
881260.66	<i>Enterobacter bugandensis</i> strain ME34	India	<i>Ventilago</i> sp.
1812935.2	<i>Enterobacter cloacae</i> complex 'Hoffmann cluster IV' strain 3D9	Nicaragua	<i>Zea nicaraguensis</i>
550.1524	<i>Enterobacter cloacae</i> strain	No reportado	No reportado
158836.552	<i>Enterobacter hormaechei</i> strain EKM10E strain Not applicable	Canadá	Pepino
158836.553	<i>Enterobacter hormaechei</i> strain EKM11E	Canadá	Pepino
1812935.8	<i>Enterobacter roggenkampii</i> strain ED5	China	<i>Saccharum officinarum</i>
1216470.3	<i>Enterobacter</i> sp. SST3	Jamaica	<i>Saccharum officinarum</i>
1683712.3	<i>Enterobacteriaceae</i> bacterium EKM102V	Canadá	Pepino
1603299.3	<i>Erwinia</i> sp. ErVv1	No reportado	No reportado
986.28	<i>Flavobacterium johnsoniae</i> strain FJ001	EUA	<i>Cucumis sativus</i>
1566297.3	<i>Klebsiella</i> sp. NFIX53	No reportado	No reportado
2202822.3	<i>Kosakonia</i> sp. MH5	Portugal	<i>Cicer arietinum</i>
2824876.3	<i>Micromonospora</i> sp. C31	Turquía	<i>Atriplex</i> sp.
37330.29	<i>Nocardia nova</i> strain LGO-A14	Brasil	<i>Citrus sinensis</i>
2802282.3	<i>Nocardia</i> sp. 2 strain LPG-2	Tailandia	<i>Musa</i> sp.
2675850.3	<i>Nocardia</i> sp. CT2-14	Tailandia	<i>Oryza sativa</i>
1736458.3	<i>Nocardia</i> sp. Root136	Alemania	<i>Arabidopsis thaliana</i>
2769252.3	<i>Nocardia</i> sp. XZ 19 231	China	<i>Hippophae rhamnoides</i>
2824874.3	<i>Nocardioopsis</i> sp. B62	Turquía	<i>Atriplex</i> sp.
1406.49	<i>Paenibacillus polymyxa</i> strain CF05	China	<i>Cryptomeria fortunei</i>
1406.193	<i>Paenibacillus polymyxa</i> strain ZF129	China	<i>Solanum tuberosum</i>
2823905.3	<i>Paenibacillus</i> sp. S25	Australia	<i>Lolium perenne</i>
549.211	<i>Pantoea agglomerans</i> strain CFBP13505	Francia	<i>Raphanus sativus</i> var. Flamboyant 5
549.332	<i>Pantoea agglomerans</i> strain Pa17-1	Canadá	<i>Pyrus communis</i>
470931.12	<i>Pantoea anthophila</i> strain EKM101A	Canadá	Pepino
1683695.3	<i>Pantoea</i> sp. EKM101V	Canadá	Pepino
1683696.3	<i>Pantoea</i> sp. EKM103V	Canadá	Pepino
2708058.3	<i>Pantoea</i> sp. EKM10T	Canadá	Pepino
2708059.3	<i>Pantoea</i> sp. EKM20T	Canadá	Melón cantalupo
2708060.3	<i>Pantoea</i> sp. EKM21T	Canadá	Melón cantalupo
2708061.3	<i>Pantoea</i> sp. EKM22T	Canadá	Melón cantalupo



470934.12	<i>Pantoea vagans</i> strain PaVv11	No reportado	No reportado
930166.114	<i>Pseudomonas brassicacearum</i> strain 3Re2-7	Alemania	<i>Solanum tuberosum</i>
930166.5	<i>Pseudomonas brassicacearum</i> strain DF41	Canadá	Canola
930166.18	<i>Pseudomonas brassicacearum</i> strain L13-6-12	Alemania	<i>Solanum tuberosum</i> cv Desiree
930166.17	<i>Pseudomonas brassicacearum</i> strain TM1A3	Suiza	<i>Solanum lycopersicum</i>
86192.15	<i>Pseudomonas chlororaphis</i> subsp. <i>aurantiaca</i> strain 449	Ucrania	<i>Zea mays</i>
86192.14	<i>Pseudomonas chlororaphis</i> subsp. <i>aurantiaca</i> strain M12	Serbia	<i>Zea mays</i>
86192.17	<i>Pseudomonas chlororaphis</i> subsp. <i>aurantiaca</i> strain M71	Italia	<i>Solanum lycopersicum</i>
86192.2	<i>Pseudomonas chlororaphis</i> subsp. <i>aurantiaca</i> strain Q16	Serbia	<i>Medicago sativa</i>
587851.16	<i>Pseudomonas chlororaphis</i> subsp. <i>aureofaciens</i> strain ChPhzTR18	Francia	<i>Solanum lycopersicum</i>
587851.13	<i>Pseudomonas chlororaphis</i> subsp. <i>aureofaciens</i> strain ChPhzTR36	Francia	<i>Solanum lycopersicum</i>
587851.17	<i>Pseudomonas chlororaphis</i> subsp. <i>aureofaciens</i> strain ChPhzTR38	Francia	<i>Solanum lycopersicum</i>
587851.8	<i>Pseudomonas chlororaphis</i> subsp. <i>aureofaciens</i> strain P2	Argelia	<i>Solanum tuberosum</i>
1513890.12	<i>Pseudomonas chlororaphis</i> subsp. <i>piscium</i> strain PCL1391	Suiza	<i>Solanum lycopersicum</i>
1513890.17	<i>Pseudomonas chlororaphis</i> subsp. <i>piscium</i> strain ToZa7	Grecia	<i>Solanum lycopersicum</i>
1513890.6	<i>Pseudomonas chlororaphis</i> subsp. <i>piscium</i> strain ZJU60	China	<i>Triticum aestivum</i>
43306.3	<i>Pseudomonas denitrificans</i> (nomen rejiciendum) strain BG1	No reportado	<i>Prosopis cineraria</i>
1037911.3	<i>Pseudomonas fluorescens</i> A506	EUA	<i>Pyrus communis</i>
294.93	<i>Pseudomonas fluorescens</i> MEP34	EUA	<i>Arabidopsis thaliana</i>
216595.4	<i>Pseudomonas fluorescens</i> SBW25	No reportado	Sugar beet, <i>Beta vulgaris</i>
1038924.3	<i>Pseudomonas fluorescens</i> SS101	Países bajos	<i>Triticum aestivum</i>
294.391	<i>Pseudomonas fluorescens</i> strain FR1	Argentina	<i>Digitalis</i>
76758.16	<i>Pseudomonas orientalis</i> strain R4-35-08	EUA	<i>Triticum</i> sp.
303.2	<i>Pseudomonas putida</i> strain PC2	China	<i>Pistacia chinensis</i> Bunge
303.913	<i>Pseudomonas putida</i> strain	No reportado	<i>Panax ginseng</i>
321846.6	<i>Pseudomonas simiae</i> strain MEB105	EUA	<i>Arabidopsis thaliana</i>
2870840.3	<i>Pseudomonas</i> sp. DR208	Corea del sur	Rizósfera de haba de soya
2202823.3	<i>Pseudomonas</i> sp. Q1	Portugal	<i>Lupinus angustifolius</i>
1573714.4	<i>Pseudomonas</i> sp. S04	Suiza	<i>Solanum tuberosum</i>
192087.6	<i>Pseudomonas syringae</i> pv. <i>atropaciens</i> strain LMG5095 strain <i>Triticum aestivum</i> (Aestivum)	Nueva Zelanda	<i>Triticum aestivum</i>
1262350.4	<i>Pseudomonas syringae</i> pv. <i>syringae</i> SM	EUA	<i>Triticum aestivum</i>
33069.5	<i>Pseudomonas viridiflava</i> LMCA8	EUA	<i>Arabidopsis thaliana</i>
33069.78	<i>Pseudomonas viridiflava</i> strain KF485	EUA	<i>Medicago sativa</i>
33069.81	<i>Pseudomonas viridiflava</i> strain SV1779	EUA	<i>Medicago sativa</i>
33069.82	<i>Pseudomonas viridiflava</i> strain T1426	EUA	<i>Medicago sativa</i>

1914994.3	<i>Rahnella</i> sp. J11-6 strain J11-6	China	<i>Oryza sativa</i>
305.879	<i>Ralstonia solanacearum</i> strain YQ	China	<i>Casuarina equisetifolia</i>
398843.6	<i>Rhodococcus kyotonensis</i> strain KB10	Corea del sur	<i>Arabidopsis thaliana</i>
334542.26	<i>Rhodococcus qingshengii</i> strain RL1	Alemania	<i>Eruca vesicaria</i> subsp. sativa
1587522.3	<i>Rhodococcus</i> sp. MEB064	EUA	<i>Arabidopsis thaliana</i>
1646377.6	<i>Rouxiiella badensis</i> strain SER3	México	Mora azul
82995.6	<i>Serratia gimesii</i> strain BXF1	Portugal	<i>Pinus pinaster</i>
614.41	<i>Serratia liquefaciens</i> strain FG3	Brasil	<i>Stachytarpheta glabra</i>
615.446	<i>Serratia marcescens</i> strain 1274	Brasil	<i>Aqave sisalana</i>
615.144	<i>Serratia marcescens</i> strain B3R3	China	<i>Zea mays</i>
615.1213	<i>Serratia marcescens</i> strain BP2	Brasil	<i>Jatropha curcas</i>
615.141	<i>Serratia marcescens</i> strain RSC-14 (KNU-AP07, KCTC 12887BP)	Corea del sur	<i>Solanum nigrum</i>
82996.18	<i>Serratia plymuthica</i> strain 3Re4-18	Alemania	<i>Solanum tuberosum</i>
82996.17	<i>Serratia plymuthica</i> strain 3Rp8	Alemania	<i>Brassica napus</i>
82996.45	<i>Serratia plymuthica</i> strain C-1	Corea del sur	<i>Capsicum annuum</i>
82996.58	<i>Serratia plymuthica</i> strain UBCF_13	No reportado	<i>Brassica juncea</i>
61652.24	<i>Serratia rubidaea</i> strain	No reportado	Hongo
2447890.3	<i>Serratia</i> sp. 1D1416	No reportado	<i>Euonymus japonicus</i>
2597702.3	<i>Serratia</i> sp. KUDC3025	Corea del sur	<i>Artemisia japonica</i> subsp. littoricola Kitam
2744257.3	<i>Serratia</i> sp. ME47	India	<i>Coscinium fenestratum</i>
2697369.3	<i>Serratia</i> sp. NGAS9	Tanzania	<i>Solanum tuberosum</i>
300181.31	<i>Serratia urelytica</i> strain CC119	EUA	Planta
40324.1951	<i>Stenotrophomonas maltophilia</i> strain 1 strain TR1	Túnez	<i>Solanum lycopersicum</i>
487698.12	<i>Stenotrophomonas pavonii</i> strain ME55	India	<i>Coix lacryma-jobi</i>
216778.48	<i>Stenotrophomonas rhizophila</i> strain CFBP 13503	Francia	<i>Raphanus sativus</i> var. Flamboyant 5
216778.5	<i>Stenotrophomonas rhizophila</i> strain IS26	Argelia	<i>Macrochloa tenacissima</i> L
1886.94	<i>Streptomyces albidoflavus</i> strain LGO-A16	Brasil	<i>Citrus sinensis</i>
1886.95	<i>Streptomyces albidoflavus</i> strain LGO-A23	Brasil	<i>Citrus sinensis</i>
1886.66	<i>Streptomyces albidoflavus</i> strain UYFA156	Uruguay	<i>Festuca arundinacea</i>
1892.26	<i>Streptomyces anulatus</i> strain	Nueva Zelanda	<i>Pseudocyphellaria dissimilis</i>
67258.4	<i>Streptomyces cavourensis</i> strain 1AS2a	Brasil	Rizósfera de <i>Triticum aestivum</i>
1969.5	<i>Streptomyces chartreusis</i> strain WZS021	China	<i>Saccharum officinarum</i>
1908.1	<i>Streptomyces globisporus</i> strain TFH56	Corea del sur	<i>Solanum lycopersicum</i>
67304.8	<i>Streptomyces griseorubiginosus</i> strain 3E-1	Corea del sur	<i>Artemisia vulgaris</i>
67304.1	<i>Streptomyces griseorubiginosus</i> strain BTU6	China	<i>Saccharum officinarum</i>
33900.7	<i>Streptomyces murinus</i> strain Am1	China	<i>Malus sieversii</i>
1169025.9	<i>Streptomyces pratensis</i> strain S10	China	<i>Solanum lycopersicum</i>
545123.5	<i>Streptomyces silaceus</i> strain ACCC40021	China	Rizósfera de alfalfa
1851167.4	<i>Streptomyces</i> sp. 11-1-2	Canadá	<i>Solanum tuberosum</i>
2748873.3	<i>Streptomyces</i> sp. CB04723	China	<i>Tripteris wilfordii</i>
2153485.3	<i>Streptomyces</i> sp. endophyte_N2	Reino Unido	<i>Arabidopsis thaliana</i>
2653200.3	<i>Streptomyces</i> sp. GY16	China	<i>Broussonetia papyrifera</i>

2609807.3	<i>Streptomyces sp. LBUM 1475</i>	Canadá	<i>Solanum tuberosum</i>
2609808.3	<i>Streptomyces sp. LBUM 1480</i>	Canadá	<i>Solanum tuberosum</i>
2609809.3	<i>Streptomyces sp. LBUM 1482</i>	Canadá	<i>Solanum tuberosum</i>
2750011.3	<i>Streptomyces sp. Rer75</i>	China	<i>Rehmannia glutinosa</i>
2705439.3	<i>Streptomyces sp. S4.7</i>	No reportado	<i>Leontopodium nivale</i>
1849967.7	<i>Streptomyces sp. SAT1</i>	China	<i>Adenophora trachelioides</i>
2563602.3	<i>Streptomyces sp. S552</i>	Viet Nam	<i>Phyllanthus urinaria L</i>
2855836.3	<i>Streptomyces sp. WY228</i>	China	<i>Arctium lappa L</i>
2021410.4	<i>Streptomyces sp. XY006</i>	China	<i>Camellia sinensis cv. Tieguanyin</i>
339.49	<i>Xanthomonas campestris LMCP11</i>	EUA	<i>Arabidopsis thaliana</i>
339.239	<i>Xanthomonas campestris strain HWA1</i>	Australia	Trigo

## 9.2 Figuras del Anexo

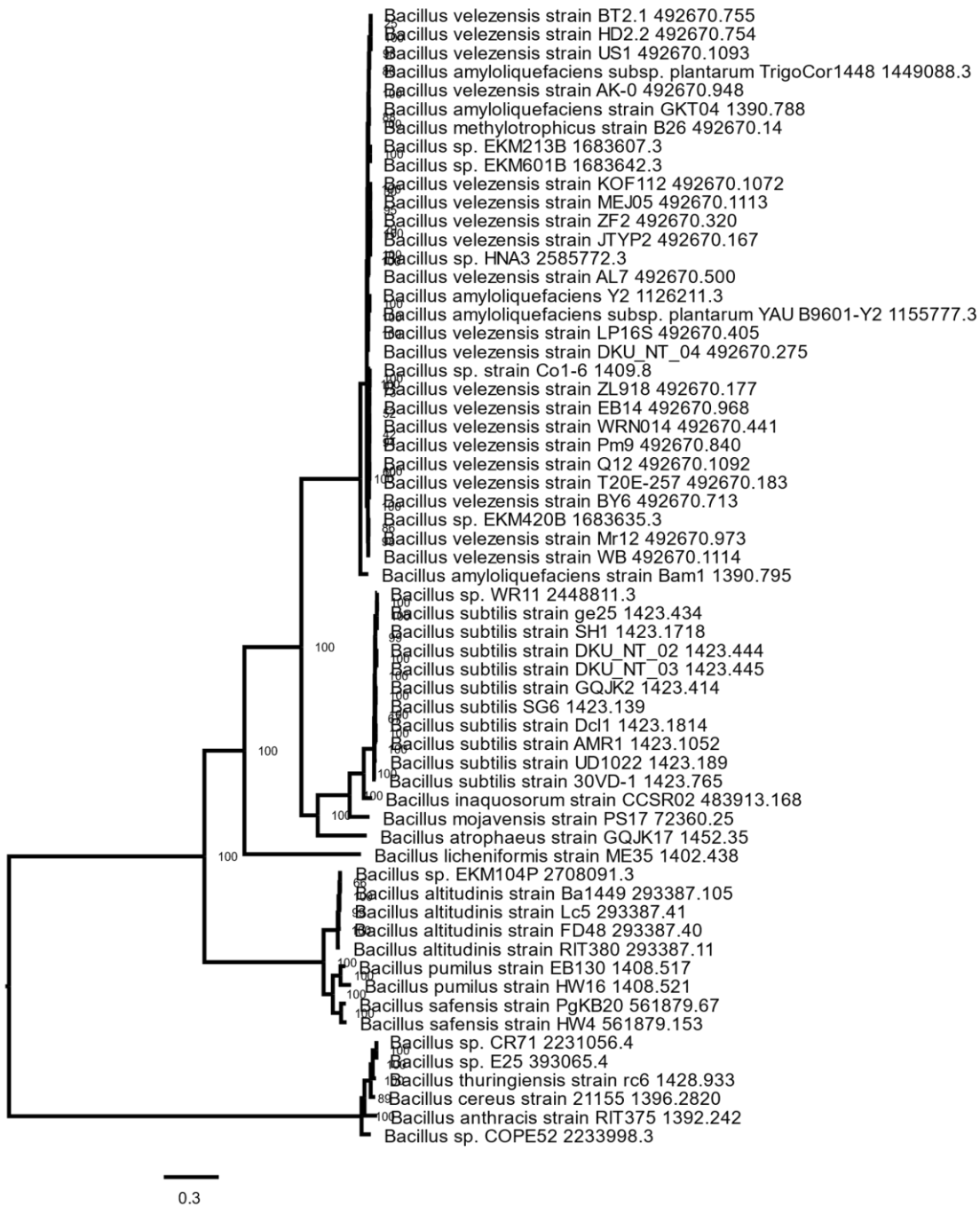


Figura I. Árbol de bacterias del género *Bacillus* resultantes de la búsqueda por BLAST. Para la búsqueda se usó un dominio de condensación perteneciente a una SPNR. Alineamiento realizado por MAFFT y analizado por RAXML en la plataforma BV-BRC usando el servicio "Construcción de árbol filogenético".



0.08

Figura II. Árbol de bacterias del género *Pseudomonas* resultantes de la búsqueda por BLAST. Para la búsqueda se usó un dominio de condensación perteneciente a una SPNR. Alineamiento realizado por MAFFT y analizado por RAxML en la plataforma BV-BRC usando el servicio "Construcción de árbol filogenético".

### 9.3 Secuencia 1

>Dominio\_C\_EntF

SQHLPLVAAQPGIWMAEKLSELPSAWSVAHYVELTGEVDSPLLARAVVAGLAQA  
DTLRMRFTEDNGEVWQWVDDALTFELPEIIDLRTNIDPHGTAQALMQADLQQD  
LRVDSGKPLVFHQLIQVADNRWYWYQRYHLLVDGFSFPAITRQIANIYCTWLR  
GEPTPASPFTPFADVVEEYQQYRESEAWQRDAAFWAEQRRQLPPPASLSPAPLP  
GRSASADILRLKLEFTDGEFRQLATQLSGVQRTDLALALAALWLGRLCNRMDYA  
AGFIFMRRLGSAALTATGPVLNVLPLGIHIAAQETLPELATRLAAQLKKMRRHQRY  
DAEQIVRDSGRAAGDEPLFGPVLNIKVFQDYQLDIPDVQAQTHTLATGPVNDLELA

LFPDVHGDLSIEILANKQRYDEPTLIQHAERLKMLIAQFAADPALLCGDVDIMLPG  
EYAQL

#### **9.4 Eventos y distinciones en los que se participó con el presente trabajo**

##### **Inauguración del Laboratorio de Microbiología del Instituto de Química**

El 3 de noviembre de 2021 se presentó la investigación en un cartel científico como medio ilustrativo de uno de los proyectos que se realizaba como parte de las líneas de investigación dentro del Laboratorio de Microbiología del Instituto de Química.

##### **Simposio Interno del Instituto de Química 2021**

Realizado el 9 de diciembre del 2021, se presentó el trabajo "Descubrimiento de NRPSs en bacterias asociadas a plantas para su uso biotecnológico y desarrollo de nuevos antibióticos" en modalidad de cartel en las instalaciones del Instituto de Química como uno de los primeros proyectos del laboratorio de Microbiología. (Figura III y IV)

##### **XXXIII Congreso Nacional de Bioquímica**

En la semana del 16 al 21 de Octubre de 2022 se presentó el trabajo "Plant-associated bacteria (PAB): Resources of specialized metabolites" en la modalidad de cartel (Figura V y VI) y en el mismo evento fue seleccionado para participar en la modalidad de Flash Talk (Figura VII y VIII) en la ciudad de Mérida, Yucatán.

##### **Tercer Congreso Virtual Latinoamericano LatinXChem**

El 3 de noviembre del 2022 se participó usando la red social de twitter en el tercer Congreso virtual Latinoamericano de LatinXChem en la modalidad de cartel con el trabajo "Plant-associated bacteria (PAB): Resources of specialized metabolites" (Figura V y IX)

## Simposio Interno del Instituto de Química 2021

Realizado el 13 de diciembre del 2022, se presentó el trabajo "Bacterias Asociadas a Plantas (PAB): Recursos de metabolitos especializados" (Figura V) en modalidad de cartel en las instalaciones del Instituto de Química donde se recibió la distinción como uno de los dos mejores trabajos de licenciatura de dicho evento (Figura X)



Figura III. Constancia de participación en el simposio interno del Instituto de Química 2021.

Trabajo titulado "Descubrimiento de NRPSs en bacterias asociadas a plantas para su uso biotecnológico y desarrollo de nuevos antibióticos"





# Descubrimiento de NRPSs en bacterias asociadas a plantas para su uso biotecnológico y desarrollo de nuevos antibióticos



**Abstracto:** El rápido desarrollo de la resistencia antimicrobiana (RAM) se debe en gran medida a que el descubrimiento y desarrollo de nuevos compuestos con carácter antibiótico se ha detenido casi por completo en los últimos años. Los péptidos no ribosomales (NRPs) son pequeños péptidos utilizados en la naturaleza por muchas bacterias como línea de defensa contra patógenos invasores para defenderse a sí mismas y a sus hospederos.

Las enzimas que los sintetizan pueden ser rastreadas en los genomas bacterianos por minería genómica. En este trabajo se propone una metodología completa, desde el muestreo y aislamiento de bacterias hasta la asociación de los metabolitos secundarios con sus respectivos clústeres de genes biosintéticos (BGCs). Hasta el momento se han logrado cultivar con éxito alrededor de 200 bacterias asociadas a plantas (BAPs) en condiciones de laboratorio. Así mismo los resultados de análisis bioinformáticos de aproximadamente 100 genomas bacterianos y secuencias de NRPSs han mostrado que este tipo de compuestos son una propuesta prometedora al momento de sugerir nuevos químicos para combatir la RAM.

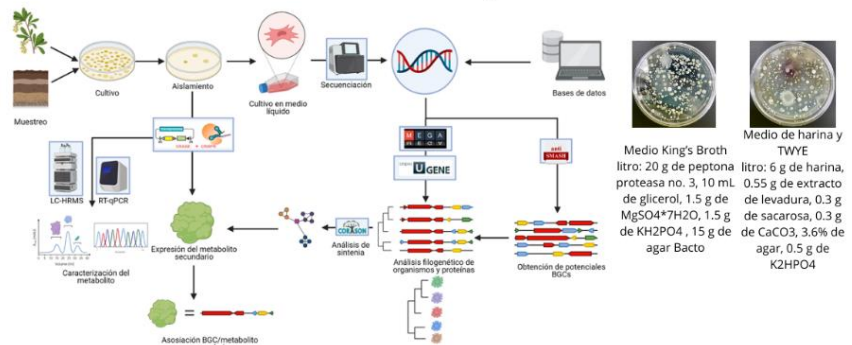
## Objetivos

El establecimiento de una metodología completa para el descubrimiento de metabolitos secundarios en genomas de BAPs.

### Objetivos particulares

- Cultivar y aislar exitosamente BAPs.
- Identificación de uno o más BGCs y caracterización de metabolitos secundarios asociados.
- Análisis bioinformáticos de NRPSs depositados en bases de datos y obtenidos por secuenciación genómica.

## Metodología



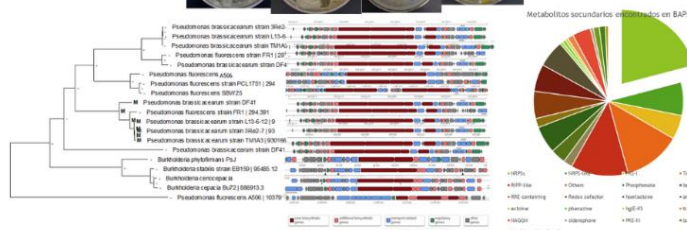
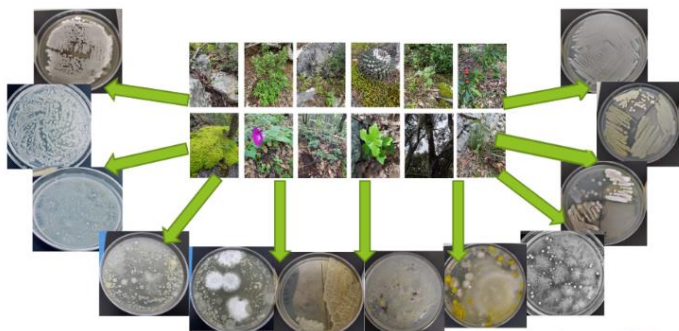
**Medio King's Broth**  
litro: 20 g de peptona, proteasa no. 3, 10 mL de glicerol, 1.5 g de MgSO4\*7H2O, 1.5 g de KH2PO4, 15 g de agar Bacto

**Medio de harina y TWYE**  
litro: 6 g de harina, 0.55 g de extracto de levadura, 0.3 g de sacarosa, 0.3 g de CaCO3, 3.6% de agar, 0.5 g de K2HPO4

## Resultados

Se cultivaron exitosamente muestras de plantas en los medios KB y de harina y TWYE, de estas placas se seleccionaron y aislaron alrededor de 200 colonias prometedoras para la producción de metabolitos secundarios.

### Criterios de selección de colonias



Se recolectaron alrededor de 100 genomas bases de datos públicas, se les realizó un análisis de minería genómica y aquellos que contenían NRPSs fueron analizados filogenéticamente, hubo una correspondencia en la evolución de las enzimas con el resto del genoma bacteriano, la mayoría de los metabolitos secundarios encontrados fueron NRPSs.

## Conclusiones

Se lograron cultivar y aislar exitosamente distintos tipos de BAPs en dos distintos medios, superando el reto de proveer las condiciones semejantes a su ecosistema de origen. Los análisis bioinformáticos señalan a los NRPSs como compuestos prometedores para su uso en el desarrollo de nuevos fármacos. Es necesario complementar este estudio con los genomas de las bacterias cultivadas y estudios de metabolómica.

## Agradecimientos

Investigación realizada gracias al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM IA201721. Agradezco a la DGAPA-UNAM la beca recibida.

### Referencias

Overmann, O., Abt, B., & Sikorski, J. (2017). Present and Future of Culturing Bacteria Changes may still occur before final publication. *Annual Review of Microbiology*, 71(July), 711-730. <https://doi.org/10.1146/annurev-micro-090816-093449>

Ceapá, C. D., Vázquez-Hernández, M., Rodríguez-Luna, S. D., Vázquez, A. P. C., Suárez, V. J., Rodríguez-Sanoja, R., Álvarez-Buylla, E. R., & Sánchez, S. (2018). Genome mining of *Streptomyces scabrisporus* NF3 reveals symbiotic features including genes related to plant interactions. *PLoS ONE*, 13(2), 1-27. <https://doi.org/10.1371/journal.pone.0192618>

Marahiel, M. A. (2009). Working outside the protein-synthesis rules: Insights into non-ribosomal peptide synthesis. *Journal of Peptide Science*, 15(12), 799-807. <https://doi.org/10.1002/psc.1183>

Recibimos tus donativos en:

**Fundación UNAM, A.C.**

**Banco:** Banamez **Cuenta:** 533019 **Sucursal:** 870

**Referencia:** 32101103

**CLABE interbancaria:** 0021800870053301

Figura IV. Cartel presentado en el Simposio Interno del Instituto de Química 2021





# Plant-associated bacteria (PAB): Resources of specialized metabolites



Reynaldo Villanueva Enríquez, Corina-Diana Ceapă  
Microbiology Laboratory, Chemistry Institute, National Autonomous University of Mexico (UNAM), Mexico City, corina.ceapa@iquimica.unam.mx, 55 5622 4420

## Introduction

The need to develop and introduce new chemical compounds with pharmaceutical properties has become a fundamental need to combat threats to human health of similar or greater impact than the pandemic caused by COVID-19. Such is the case of Antimicrobial Resistance, which recent studies show that it is the cause of death of 5 million people a year.(1) In nature, bacteria produce specialized metabolites as a mechanism of antagonism. Just the genus *Streptomyces* is the precursor of the 55% of all antibiotics currently used.(2) Plant-associated bacteria (PAB) help their host to grow in extreme temperature environments, low water availability and provide protection against pathogens, therefore, the medicinal properties that had been attributed to plants can in actuality be from the bacteria itself.(3) The exploration of natural products in bacteria is hampered by the need to cultivate and promote specific conditions for the production of metabolites. Genomic mining is a technique that seeks to associate a DNA fragment called a biosynthetic gene cluster (BGC) with its corresponding functional product. Together with the rapid advance in new sequencing technologies and the constant improvements in bioinformatics tools, genome mining allows us to explore a large number of specialized metabolites produced by bacteria even under laboratory conditions.

## Objetives

- Propose plant-associated bacteria as a source of discovery of natural products with biotechnological potential
- Propose a class of specialized metabolites for the development of compounds with biotechnological applications

## Methodology

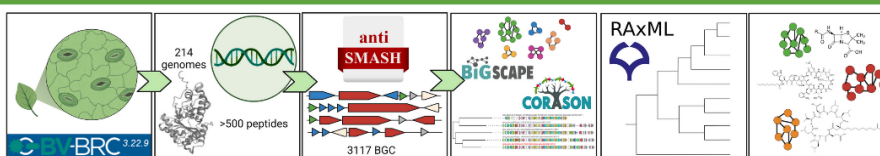


Fig1. PAB genomes were collected from BACTERIAL AND VIRAL BIOINFORMATICS RESOURCE CENTER (BV-BRC) database using search filters. Genomes were subjected to a gene search for homology with sequence of a condensation domain of the Enterobactin synthase component F (P11454) using the BLASTx tool. Precursor bacteria of the resulting peptides were analyzed by genome mining employing antiSMASH v.6.0 (4). Similarity networks were performed with BiG-SCAPE and genomic context analysis with CORASON. (5) Phylogenetic trees of bacterial genomes were made. BGC sequences were compared with the MiBIG database to associate them with a functional product and determine novelty in those that did not present coincidences.

## Results and Discussion

Genomic mining analyzes of 214 genomes revealed 3117 biosynthetic gene clusters (BGCs) of which 855 (27%) were related to NRPS, its conservation and abundance throughout all genera suggests potential for bioactivity (Fig 2). BiG-SCAPE classified 534 NRPS BGCs into 169 singletons and 85 Gene Cluster Families (GCFs) (Fig 3) of which of which 72 showed novelty and conservation among the same genera. The functional variety of the BGC already characterized ranges from antibiotics (Icosalide, Tridecaptin, Viscosin), biosurfactants (Lichenysin, Surfactin), metal uptake and delivery (Bacillibactin, Ornibactin, Streptobactin), and even the Tolaasin toxin produced by *Pseudomonas*. The *Serratia* sp. GCF (Fig 4) showed low similarity with two characterized compounds. It is probable that it resulted from a BGC duplication which acquired new functions and therefore interesting for further study

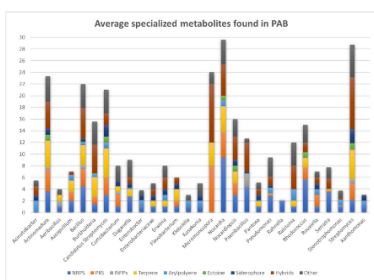


Fig 2. Average specialized metabolites found in each of the genera result of the homology search with BLASTx.

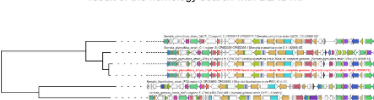


Fig 4. Analysis of the GCF of *Serratia* BGC with low similarity with vumbibactin and pseudomonine

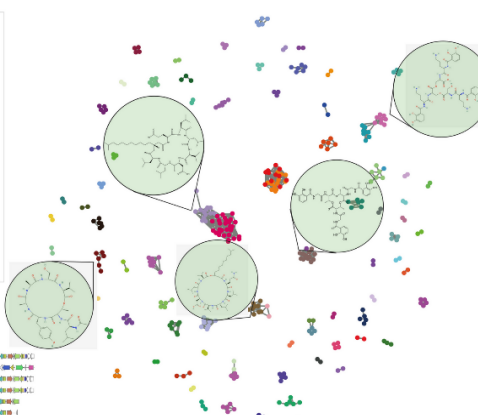


Fig 3. Similarity network analysis by BiG-SCAPE. The figure shows the largest GCFs with a similarity percentage greater than 70% with compounds in the MiBIG database, from left to right: Rhizomide, Surfactin, Lichenysin, Bacillibactin and Turnerbactin

## Conclusions

PAB proved to be a source of great diversity and high number of specialized metabolites, this likely due to their need to adapt to the various environmental conditions to which their hosts are subjected. The conservation and generation of new compounds with different bioactivities is essential to guarantee the holobiont survival. The NRPS have shown to have a great functional and structural diversity and bioinformatic tools facilitate their study for the discovery of new compounds with antimicrobial activity.

## Acknowledgements

This project was financed by PAPIIT (project number IA201721) and CONACYT (project number CF-2022/319816). Reynaldo Villanueva Enríquez received the a scholarship from PAPIIT and the internal undergraduate scholarship of the UNAM Chemistry Institute the support provided is appreciated.

- Murray, C., Ikuta, K., Sharara, F., Lancet, L. S.-T., & 2022, U. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis | Elsevier Enhanced Reader. In Elsevier (Vol. 399, Issue 10325, pp. 1–58). <https://www.sciencedirect.com/science/article/pii/S01406736210272409?oi=1&https://reader.elsevier.com/reader/sd/pii/S01406736210272409?token=A7DD5AAE0BD094D9F020A1FBC980C10BE19689624E9CBEB7264966A8698677303CAB3CF66B7D93C18779BAAF679152B8&originRegion=eu-west-1>
- Worsley, S. F., Newitt, J., Rassbach, J., Batey, S. F. D., Holmes, N. A., Murrell, J. C., Wilkinson, B., & Hutchings, M. I. (2020). *Streptomyces* endophytes promote host health and enhance growth across plant species. *Applied and Environmental Microbiology*, 86(16), 1–17. <https://doi.org/10.1128/AEM.01053-20>
- Garrido-Oter, R., Nakano, R. T., Dombrowski, N., Ma, K. W., McHardy, A. C., & Schulze-Lefert, P. (2018). Modular Traits of the Rhizobiales Root Microbiota and Their Evolutionary Relationship with Symbiotic Rhizobia. *Cell Host and Microbe*, 24(1), 155–167.e5.
- Roilm, L. A., Santos, F. C. M. dos, Chaves, L. L., Gonçalves, M. L. C. M., Freitas-Neto, J. L., Nascimento, A. L. da S. do, Soares-Sobrinho, J. L., Albuquerque, M. M. de, Lima, M. do C. A. de, & Roilm-Neto, P. J. (2020). antSMASH 6.0: improving cluster detection and comparison capabilities. In *Nature* (Vol. 388, pp. 539–547).

Figura V. Cartel presentado. Cartel para el XXXIII Congreso Nacional de Bioquímica, el tercer evento del Congreso virtual LatinXChem y en el simposio interno del IQ 2022.



Sociedad Mexicana de  
Bioquímica A.C.



SACRAMENTO NO. 413  
COL. INSURGENTES BORJA  
DEL BENITO JUÁREZ  
CP. 03100  
CDMX  
TEL. (55)66225742  
WWW.SMB.ORG.MX  
FACEBOOK:  
facebook.com/RedSMB

**MEMORIA DIRECTIVA 2021-2022**

**PRESIDENTA**

Dra. Estrella María Teresa Hernández Sotomayor

**VICEPRESIDENTE**

Dra. Agustín Guerrero Hernández

**SECRETARÍA TÉCNICA**

Dra. Bertha María Josefina González Pedraza

**SUBSECRETARÍA TÉCNICA**

Dra. María de Lourdes Olivé Correy

**SOCIOS FUNDADORES**

- Dra. Bárbara Arreguín Lozano
- Dra. Edmundo Cava Cuadilla
- Dra. Guillermo Corral Sánchez
- Dra. Joaquín Castañeda
- Dra. Carlos del Río Esteban
- Dra. Silvestre Fandi Fandi
- Dra. María García Hernández
- Dra. Jesús Guzmán García
- Dra. Jesús Kurián Rodríguez
- Dra. José Laporte García
- Dra. Guillermo Macosco Higuera
- Dra. Raúl Domínguez Vidaurri
- Dra. Silvia G. Pardo Castro
- Dra. Guillermo Salasán Álvarez

Se otorga la presente

**CONSTANCIA a:**

**Reynaldo Villanueva Enriquez**

Quien asistió y presentó el trabajo:

**Plant-associated bacteria (PAB): Resources of specialized metabolites**

Por: Reynaldo Villanueva Enriquez, Corina-Diana Ceapá

En la modalidad de cartel durante el  
XXXIII Congreso Nacional de Bioquímica  
18 - 21 de octubre de 2022, Mérida, Yucatán

Atentamente  
Por el Comité Organizador

Dra. Teresa Hernández Sotomayor  
Presidenta



Figura VI. Constancia de participación en XXXIII Congreso Nacional de Bioquímica. Presentación en la modalidad de cartel con el trabajo "Plant-associated bacteria (PAB): Resources of specialized metabolites"

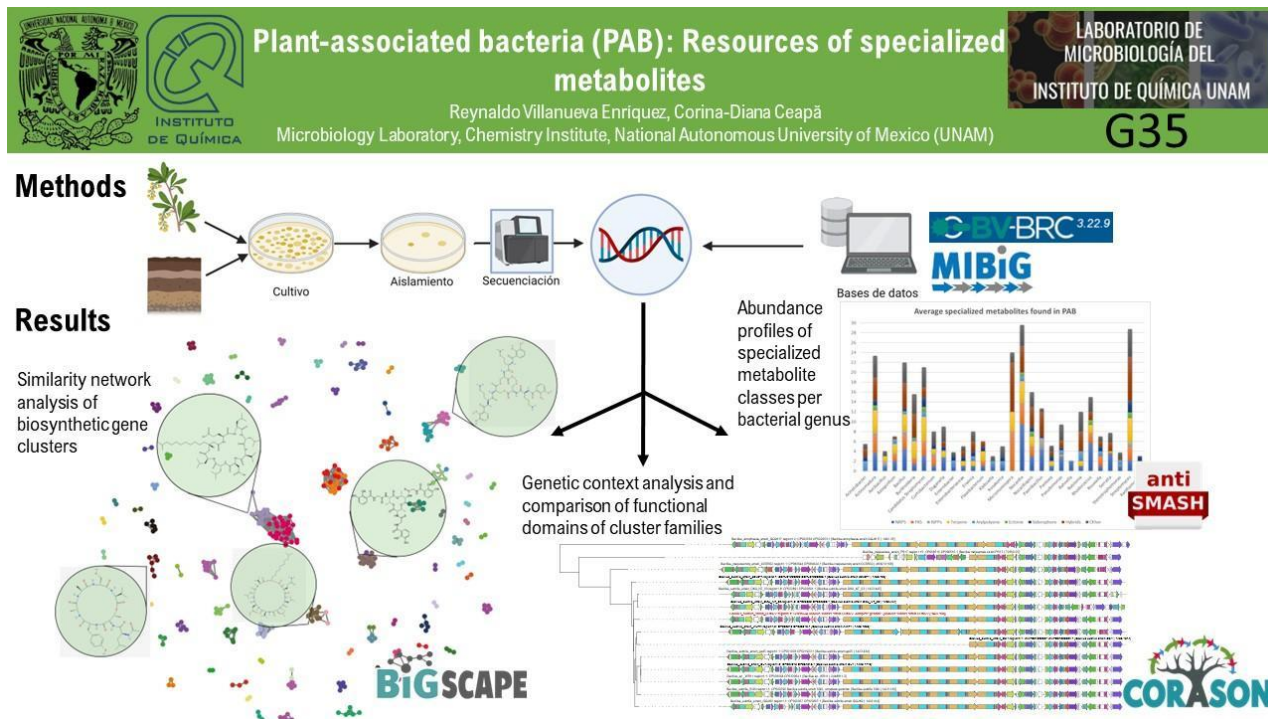


Figura VII. Diapositiva presentada en el XXXIII Congreso Nacional de Bioquímica. Empleada en la modalidad de Flash Talk.



Sociedad Mexicana de  
Bioquímica A.C.



SACRAMENTO NO. 413  
COL. INSURGENTES BORJA  
DEL BENITO JUÁREZ  
CP. 03100  
CDMX  
TEL. (55)56225742  
WWW.SMB.ORG.MX  
FACEBOOK  
facebook.com/RazSMB

MESA DIRECTIVA 2021-2023

PRESIDENTE

Dra. Soledad María Teresa Hernández Sotomayor

VICEPRESIDENTE

Dra. Agustín Guerrero Hernández

SECRETARÍA TESORERA

Dra. Bertha María Josefina González Padilla

SUBSECRETARÍA TESORERA

Dra. María de Lourdes Olmedo Cerezo

SOCIOS FUNDORES

Dra. Soledad Arreguín Lozano

Dra. Edmundo Calvo Coaralla

Dra. Guillermo Carragel Barahona

Dra. Joaquín Chantón

Dra. Carlos del Río Esteban

Dra. Silvana Frenk Pineda

Dra. María García Hernández

Dra. Jesús Guzmán García

Dra. Jesús Humberto Rodríguez

Dra. José Laguarda García

Dra. Guillermo Masera Velázquez

Dra. Raúl Orobaza Valiente

Dra. Efraín G. Pardo Castro

Dra. Guillermo Salazar Sureda

Se otorga la presente

**CONSTANCIA** a:

**Reynaldo Villanueva Enriquez**

Por la presentación de su trabajo:

**Plant-associated bacteria (PAB): Resources of specialized metabolites**

Por: Reynaldo Villanueva Enriquez, Corina-Diana Ceapă

En la modalidad

**Flash talks**

durante el XXXIII Congreso Nacional de Bioquímica  
16 - 21 de octubre de 2022, Mérida, Yucatán

Atentamente  
Por el Comité Organizador

Dra. Teresa Hernández Sotomayor  
Presidenta



Figura VIII. Constancia de participación en XXXIII Congreso Nacional de Bioquímica. Presentación en la modalidad de Flash Talk con el trabajo "Plant-associated bacteria (PAB): Resources of specialized metabolites"





Figura IX. Constancia de participación en tercer Congreso virtual Latinoamericano de LatinXChem. Presentación en la modalidad de cartel con el trabajo “Plant-associated bacteria (PAB): Resources of specialized metabolites”



Figura X. Constancia de distinción de uno de los dos mejores trabajos en la categoría de licenciatura en el simposio interno del IQ 2022.