

**UNIVERSIDAD NACIONAL DEL SANTA**

**FACULTAD DE CIENCIAS**

**ESCUELA PROFESIONAL DE BIOTECNOLOGÍA**



**UNS**  
UNIVERSIDAD  
NACIONAL DEL SANTA

**“Identificación y anotación de clústeres de genes biosintéticos relacionados con compuestos antitumorales, antimicrobianos y otros metabolitos bioactivos mediante análisis del genoma de *Streptomyces* sp. B-146”.**

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE  
LICENCIADO EN BIOTECNOLOGÍA**

**AUTOR:**

Bach. Zegarra Bermudez, Yoel Marino

**ASESOR:**

Dr. Castro Alvarado, Ángel Pablo

**CO-ASESORA:**

PhD. Flores Clavo, Rene

Nuevo Chimbote, Perú

2023-06-23

**UNIVERSIDAD NACIONAL DEL SANTA**

**FACULTAD DE CIENCIAS**

**ESCUELA PROFESIONAL DE BIOTECNOLOGÍA**



**UNS**  
UNIVERSIDAD  
NACIONAL DEL SANTA

**“Identificación y anotación de clústeres de genes biosintéticos  
relacionados con compuestos antitumorales, antimicrobianos y otros  
metabolitos bioactivos mediante análisis del genoma de *Streptomyces*  
sp. B-146”.**

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE  
LICENCIADO EN BIOTECNOLOGÍA**

Revisado y Aprobado por:

Dr. Castro Alvarado, Ángel Pablo

**ASESOR**

DNI: 06451148

ORCID: 0000-0002-5984-0587

PhD. Flores Clavo, Rene

**CO-ASESORA**

DNI: 44656347

ORCID: 0000-0002-4448-5981

**UNIVERSIDAD NACIONAL DEL SANTA**  
**FACULTAD DE CIENCIAS**  
**ESCUELA PROFESIONAL DE BIOTECNOLOGÍA**



**UNS**  
UNIVERSIDAD  
NACIONAL DEL SANTA

**“Identificación y anotación de clústeres de genes biosintéticos  
relacionados con compuestos antitumorales, antimicrobianos y otros  
metabolitos bioactivos mediante análisis del genoma de *Streptomyces*  
sp. B-146”.**

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE  
LICENCIADO EN BIOTECNOLOGÍA**

Revisado y Aprobado por el Jurado Evaluador:

Dr. Azáñero Díaz, Carlos Alberto  
**PRESIDENTE**

DNI: 18093785

ORCID: 0000-0002-5612-4104

Dr. Capa Robles, Willian Robert  
**SECRETARIO**

DNI: 32969285

ORCID: 0000-0003-3402-613X

Dr. Castro Alvarado, Ángel Pablo  
**INTEGRANTE**

DNI: 06451148

ORCID: 0000-0002-5984-0587

## ACTA DE CALIFICACIÓN DE LA SUTENTACIÓN DE LA TESIS

En el Distrito de Nuevo Chimbote, en la Universidad Nacional de Santa, en el Aula 1 de la Escuela de Posgrado de la UNS, siendo las 11:00 horas del día Viernes 23 de junio del 2023, dando cumplimiento a la Resolución N° 199-2023-UNS-FC se reunió el Jurado Evaluador presidido por Dr. Carlos Alberto Azañero Díaz, teniendo como miembros a Dr. William Robert Cape Robles (secretario) (a), y Dr. Angel Pablo Castro Alvarado (integrante), para la sustentación de tesis a fin de optar el título de Licenciado en Biotecnología, realizado por el, (la), (los) tesista (as) .....

Yoel Marino Zegana Bermudez (Cod. 0201023015), quien (es) sustentó (aron) la tesis intitulada: "Identificación y anotación de clústeres de genes biosintéticos relacionados con compuestos antimicrobianos, antimicrobianos y otros metabolitos bioactivos mediante análisis del genoma de Streptomyces sp. B-146."

Terminada la sustentación, el (la), (los) tesista (as)s respondió (ieron) a las preguntas formuladas por los miembros del jurado.

El Jurado después de deliberar sobre aspectos relacionados con el trabajo, contenido y sustentación del mismo y con las sugerencias pertinentes, declara la sustentación como Excelente asignándole un calificativo de Diecinueve (19) puntos, según artículo 111° del Reglamento General de Grados y Títulos vigente (Resolución N° 580-2022-CU.-R-UNS)

Siendo las 13:00 horas del mismo día se dio por terminado el acto de sustentación firmando los miembros del Jurado en señal de conformidad

Carlos Alberto Azañero Díaz  
Nombre: Carlos Alberto Azañero Díaz  
Presidente

William Robert Cape Robles  
Nombre: William Robert Cape Robles  
Secretario

Angel Pablo Castro Alvarado  
Nombre: Angel Pablo Castro Alvarado  
Integrante

Distribución: Integrantes J.E (3), tesis (1) y archivo (02).



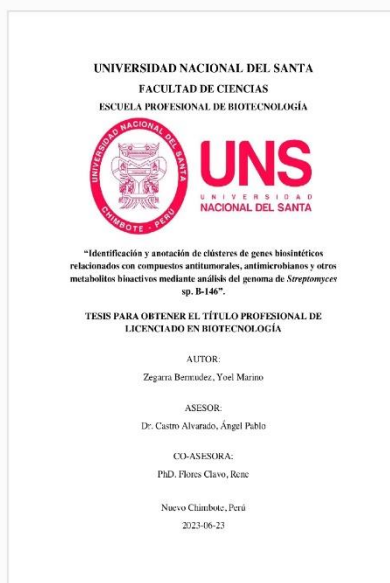


## Recibo digital

Este recibo confirma que su trabajo ha sido recibido por **Turnitin**. A continuación podrá ver la información del recibo con respecto a su entrega.

La primera página de tus entregas se muestra abajo.

Autor de la entrega: Yoel Marino Zegarra Bermudez  
Título del ejercicio: Revision 1  
Título de la entrega: Informe de Tesis  
Nombre del archivo: INFORME\_DE\_TESIS\_-\_ZEGARRA\_-\_UNS\_-\_CIICAM\_-\_ProCIENC...  
Tamaño del archivo: 3.72M  
Total páginas: 171  
Total de palabras: 38,599  
Total de caracteres: 239,737  
Fecha de entrega: 17-feb.-2024 08:54p. m. (UTC-0500)  
Identificador de la entrega: 2297409086





## Informe de Tesis

### INFORME DE ORIGINALIDAD

11%	10%	5%	3%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

### FUENTES PRIMARIAS

1	<a href="https://hdl.handle.net">hdl.handle.net</a> Fuente de Internet	2%
2	<a href="https://brcdonloads.vbi.vt.edu">brcdonloads.vbi.vt.edu</a> Fuente de Internet	1%
3	Submitted to Universidad Nacional Pedro Ruiz Gallo Trabajo del estudiante	1%
4	<a href="http://www.homd.org">www.homd.org</a> Fuente de Internet	<1%
5	<a href="http://www.theseed.org">www.theseed.org</a> Fuente de Internet	<1%
6	<a href="http://www.frontiersin.org">www.frontiersin.org</a> Fuente de Internet	<1%
7	<a href="http://brenda-enzymes.info">brenda-enzymes.info</a> Fuente de Internet	<1%
8	<a href="http://idus.us.es">idus.us.es</a> Fuente de Internet	<1%
9	<a href="http://www.biorxiv.org">www.biorxiv.org</a> Fuente de Internet	

## **DEDICATORIA**

A mi familia y amigos, quienes me han brindado su apoyo incondicional en todas las etapas de mi formación académica y profesional.

## **AGRADECIMIENTO**

A la PhD. Rene Flores Clavo, por brindarme la oportunidad de desarrollar este proyecto de tesis, por impartirme asesoramiento científico y por hacerme participe de su proyecto de investigación doctoral.

A la PhD.(c) Gladys Angelica Apaza Castillo, por su disposición a compartir sus conocimientos altamente especializados y brindarme la información base para la ejecución de este proyecto de tesis.

Al Dr. Ángel Castro Alvarado, por su asesoramiento científico y administrativo clave para la ejecución y culminación de este proyecto de tesis.

Al Centro de Investigación e Innovación en Ciencias Activas Multidisciplinarias (CIICAM) por su compromiso y contribución al desarrollo de la ciencia en el Perú.

A Concytec – Banco Mundial “Mejoramiento y ampliación de los Servicios del Sistema Nacional de Ciencia, Tecnología e Innovación Tecnológica” 8682-PE, a través de su unidad ejecutora ProCIENCIA [contrato número 190-2018], por el financiamiento de este proyecto.



## ÍNDICE DE CONTENIDOS

DEDICATORIA .....	vii
AGRADECIMIENTO .....	viii
RESUMEN .....	xiv
ABSTRACT .....	xv
I. INTRODUCCIÓN .....	16
II. MARCO TEÓRICO.....	18
2.1. <i>ACTINOBACTERIAS</i> .....	18
2.2. <i>STREPTOMYCES</i> .....	18
2.3. PRODUCTOS NATURALES .....	19
2.4. CLÚSTERES DE GENES BIOSINTÉTICOS .....	19
2.5. MINERACIÓN GENÓMICA.....	20
2.6. CLASIFICACIÓN DE LOS PRODUCTOS NATURALES .....	20
2.6.1. PÉPTIDOS NO RIBOSOMALES .....	20
2.6.2. POLICÉPTIDOS.....	21
2.6.3. PÉPTIDOS RIBOSÓMICAMENTE SINTETIZADOS Y MODIFICADOS POSTRADUCCIONALMENTE.....	23
2.6.4. TERPENOS.....	26
2.7. SISTEMAS CRISPR/Cas EN PROCARIOTAS.....	27
2.7.1. SISTEMAS CRISPR/Cas .....	27
2.7.2. FUNCIÓN DE LOS SISTEMAS CRISPR/Cas .....	28
2.7.3. DIVERSIDAD DE LOS SISTEMAS CRISPR/Cas.....	28
2.7.4. SISTEMAS CRISPR/Cas TIPO I, SUBTIPO E.....	29
III. MATERIALES Y MÉTODOS .....	31
3.1. MÉTRICAS DE LA SECUENCIA DEL GENOMA DE <i>Streptomyces</i> sp. B-146 .....	31
3.2. ANÁLISIS FILOGENÉTICOS Y FILOGENÓMICOS DE <i>Streptomyces</i> sp. B-146.....	32
3.3. ANOTACIÓN DEL GENOMA DE <i>Streptomyces</i> sp. B-146 .....	33
3.4. IDENTIFICACIÓN DE BGCs EN EL GENOMA DE <i>Streptomyces</i> sp. B-146.....	34
3.5. IDENTIFICACIÓN DE SISTEMAS CRISPR EN EL GENOMA DE <i>Streptomyces</i> sp. B-146. .....	34
IV. RESULTADOS Y DISCUSIÓN.....	35
4.1. DESCRIPCIÓN DE LAS MÉTRICAS DE LA SECUENCIA DEL GENOMA DE <i>Streptomyces</i> sp. B-146.....	35

4.2. RESULTADOS DE LOS ANÁLISIS FILOGENÉTICOS Y FILOGENÓMICOS DE <i>Streptomyces</i> sp. B-146.....	36
4.3. RESULTADOS DE LA ANOTACIÓN DEL GENOMA DE <i>Streptomyces</i> sp. B-146 .....	41
4.3.1. ANOTACIÓN DEL GENOMA DE <i>Streptomyces</i> sp. B-146 POR PROKKA .....	41
4.3.2. ANOTACIÓN FUNCIONAL DEL GENOMA DE <i>Streptomyces</i> sp. B-146 POR RAST49	
4.3.3. CATEGORIZACIÓN FUNCIONAL DEL GENOMA DE <i>Streptomyces</i> sp. B-146 POR BlastKOALA.....	57
4.4. RESULTADOS DE LA IDENTIFICACIÓN DE BGCS EN EL GENOMA DE <i>Streptomyces</i> sp. B-146 .....	61
4.4.1. BGCs PUTATIVOS CON BIOACTIVIDAD ANTIVIRAL, CITOTÓXICA Y ANTIBACTERIANA .....	67
4.4.2. BGCs CON ACTIVIDAD ANTIBACTERIANA .....	71
4.4.3. BGCs CON OTRAS BIOACTIVIDADES .....	74
4.5. RESULTADOS DE LA IDENTIFICACIÓN DE SISTEMAS CRISPR EN EL GENOMA DE <i>Streptomyces</i> sp. B-146.....	77
4.5.1. DESCRIPCIÓN DE LOS ARREGLOS CRISPR IDENTIFICADOS .....	77
4.5.2. ANOTACIÓN DE GENES Cas IDENTIFICADOS .....	81
V. CONCLUSIONES Y RECOMENDACIONES .....	88
VI. REFERENCIAS BIBLIOGRÁFICAS .....	89
VII. ANEXOS .....	129

## ÍNDICE DE TABLAS

<b>Cuadro 01:</b> Métricas de la secuencia final del genoma de <i>Streptomyces</i> sp. B146 generados por gVolante, CheckM y QUAST. ....	35
<b>Cuadro 02:</b> Principales características de los arreglos CRISPR detectados por Prokka, RAST y los servidores web CRISPRCasFinder y CRISPRminer. ....	79
<b>Cuadro 03:</b> Clúster de genes <i>Cas</i> detectados por los servidores web CRISPRCasFinder y CRISPRminer comparados con la anotación realizada por Prokka y RAST. ....	82
<b>Cuadro 04:</b> Anotación a través de Uniprot del clúster de genes <i>Cas</i> detectados por los servidores web CRISPRCasFinder y CRISPRminer. ....	86

## ÍNDICE DE FIGURAS

<b>Figura 01:</b> Construcción del árbol filogenético a partir de la secuencia parcial del gen ARNr 16S (MW562805, NCBI) de <i>Streptomyces</i> sp. B-146, mediante el método “Neighbor-Joining” usando el software Mega 11. ....	37
<b>Figura 02:</b> Árbol filogenético basado en la secuencia completa del gen ARNr 16S del genoma de <i>Streptomyces</i> sp. B-146 a través del servidor web TYGS. ....	38
<b>Figura 03:</b> Árbol filogenético basados en la secuencia completa del genoma de <i>Streptomyces</i> sp. B-146 a través del servidor web TYGS. ....	39
<b>Figura 04:</b> Árbol filogenético obtenido de los análisis filogenómicos basados en la secuencia completa del genoma de <i>Streptomyces</i> sp. B-146 a través del servidor web autoMLST. ....	40
<b>Figura 05:</b> Visualización del genoma de <i>Streptomyces</i> sp. B-146, destacando la anotación de CDSs realizada por Prokka. ....	42
<b>Figura 06:</b> Categorización funcional del genoma de <i>Streptomyces</i> sp. B-149, resaltando la cobertura de los subsistemas y la abundancia de las categorías reportadas por RAST. ....	50
<b>Figura 07:</b> Anotación y categorización funcional por BlastKOALA de las secuencias de proteínas detectadas en el genoma de <i>Streptomyces</i> sp. B-146. ....	58
<b>Figura 08:</b> BGCs identificados en el genoma de <i>Streptomyces</i> sp. B-146 por antiSMASH 6.1.1. ....	63
<b>Figura 09:</b> BGCs identificados contrastando la anotación realizada con AntiSMASH 6.1.1, ARTS 2.0, PRISM 4.4.5, RiPPMiner-Genome y BAGEL4. ....	64
<b>Figura 10:</b> Identificación y anotación de un subconjunto de genes del smBGC22 de <i>Streptomyces</i> sp. B146 implicado en la biosíntesis de Xiamicina A. ....	69
<b>Figura 11:</b> Identificación y anotación de un subconjunto de genes del smBGC22 de <i>Streptomyces</i> sp. B146 implicado en la biosíntesis de Pactamidas. ....	71
<b>Figura 12:</b> Identificación y anotación del smBGC17 de <i>Streptomyces</i> sp. B146 implicado en la biosíntesis de Terpenoide idéntico a Albaflavenona. ....	73
<b>Figura 13:</b> Identificación y anotación del smBGC20 de <i>Streptomyces</i> sp. B146 implicado en la biosíntesis de un lantipéptido idéntico a SapB. ....	76
<b>Figura 14:</b> Visualización mediante Proksee de la localización de los sistemas CRISPR identificados en el genoma de <i>Streptomyces</i> sp. B-146. ....	78

## ÍNDICE DE ANEXOS

<b>Anexo 1.</b> Anotación de CDS del genoma de <i>Streptomyces</i> sp. B-146 relacionados a genes ARG usando PROKKA. ....	129
<b>Anexo 2.</b> Anotación de CDS relacionados a subsistemas implicados en la resistencia a antibióticos y síntesis de productos naturales por RAST. ....	137
<b>Anexo 3.</b> Anotación de CDS del genoma de <i>Streptomyces</i> sp. B-146 relacionados a genes ARG usando RAST. ....	139
<b>Anexo 4.</b> Identificación de BGCs putativos en el genoma de <i>Streptomyces</i> sp. B-146 usando el servidor web antiSMASH versión 6.1.1. ....	141
<b>Anexo 5.</b> Clasificación de BGCs detectados en el genoma de <i>Streptomyces</i> sp. B-146 según los servidores web antiSMASH versión 6.1.1, ARTS 2.0, PRISM 4.4.5, RiPPMiner-Genome, BAGEL4 y NaPDoS. ....	146
<b>Anexo 6.</b> Anotación de los genes contenidos en el smBGC22 detectado en el genoma de <i>Streptomyces</i> sp. B-146 a través de PROKKA, BlastKOALA, RAST, antiSMASH 6.1.1, UniProt y MIBiG. ....	151
<b>Anexo 7.</b> Anotación de los genes contenidos en el smBGC17 detectado en el genoma de <i>Streptomyces</i> sp. B-146 a través de PROKKA, BlastKOALA, RAST, antiSMASH 6.1.1, UniProt Y MIBiG. ....	160
<b>Anexo 8.</b> Anotación de los genes contenidos en el smBGC20 detectado en el genoma de <i>Streptomyces</i> sp. B-146 a través de PROKKA, BlastKOALA, RAST, antiSMASH, BAGEL 6.1.1, UniProt Y MIBiG. ....	164
<b>Anexo 9.</b> Identificación de arreglos CRISPR en el genoma de <i>Streptomyces</i> sp. B-146 usando el servidor web RAST. ....	168
<b>Anexo 10.</b> Identificación de arreglos CRISPR en el genoma de <i>Streptomyces</i> sp. B-146 usando el servidor web CRISPRCasFinder. ....	170
<b>Anexo 11.</b> Identificación de arreglos CRISPR en el genoma de <i>Streptomyces</i> sp. B-146 usando el servidor web CRISPRminer. ....	172

## RESUMEN

Los genomas de las bacterias pertenecientes al género *Streptomyces*, aisladas de ambientes extremos, son de particular interés para descubrir nuevos metabolitos secundarios bioactivos con una gran variedad de aplicaciones terapéuticas. En ese contexto, se ha observado un crecimiento significativo en la disponibilidad de nuevas herramientas bioinformáticas de alto rendimiento que posibilitan análisis genómicos rápidos y precisos. En esta investigación, hemos evaluado el genoma de *Streptomyces* sp. B-146, aislada previamente de las salinas de Bayobar en el norte de Perú. Nuestro enfoque abarcó la identificación y anotación de clústeres de genes biosintéticos (BGCs) relacionados con compuestos antitumorales, antibacterianos, antivirales y con otras bioactividades. Además, evaluamos el genoma en busca de sistemas CRISPR/Cas y llevamos a cabo análisis filogenéticos y filogenómicos. Los resultados de nuestra evaluación revelaron que *Streptomyces* sp. B-146 alberga hasta 39 BGCs putativos, la mayoría de los cuales comparten similitudes inferiores al 80% con otros BGCs conocidos. La anotación manual de cuatro BGCs específicos demostró que esta bacteria tiene el potencial genético para biosintetizar Terpenoides (Xiamicina A y Albaflavenona), Macrolactamas (Pactamidas) y Lantipéptidos (SapB), los cuales se caracterizan por exhibir actividades antivirales, antibacterianas y citotóxicas. Adicionalmente, identificamos con éxito varios genes de resistencia, principalmente transportadores de membrana, que podrían desempeñar un papel fundamental en la exportación de los compuestos bioactivos biosintetizados. Los análisis de filogenia indican que *Streptomyces* sp. B-146 está filogenéticamente relacionada a *Streptomyces olivaceus*. Finalmente, también reportamos la identificación de un sistema CRISPR/Cas Subtipo I-E incompleto.

**Palabras Clave:** *Streptomyces* sp. B-146, BGCs, Antimicrobianos, Antitumorales, Antivirales, Análisis Filogenéticos, Análisis Filogenómicos, sistemas CRISPR/cas.

## ABSTRACT

The genomes of bacteria belonging to the genus *Streptomyces*, isolated from extreme environments, are of particular interest to discover new bioactive secondary metabolites with a wide variety of therapeutic applications. In this context, significant growth has been observed in the availability of new high-throughput bioinformatics tools that enable rapid and accurate genomic analyses. In this research, we have evaluated the genome of *Streptomyces* sp. B-146, previously isolated from the Bayobar salt flats in northern Peru. Our approach encompassed the identification and annotation of biosynthetic gene clusters (BGCs) related to antitumor, antibacterial, antiviral, and other bioactivities compounds. Furthermore, we screened the genome for CRISPR/Cas systems and performed phylogenetic and phylogenomic analyses. The results of our evaluation revealed that *Streptomyces* sp. B-146 hosts up to 39 putative BGCs, most of which share less than 80% similarities with other known BGCs. Manual annotation of four specific BGCs demonstrated that this bacterium has the genetic potential to biosynthesize Terpenoids (Xiamycin A and Albaflavenone), Macrolactams (Pactamides) and Lantipeptides (SapB), which are characterized by exhibiting antiviral, antibacterial and cytotoxic activities. Additionally, we successfully identified several resistance genes, mainly membrane transporters, which could play a key role in the export of the biosynthesized bioactive compounds. Phylogeny analyzes indicate that *Streptomyces* sp. B-146 is phylogenetically related to *Streptomyces olivaceus*. Finally, we also report the identification of an incomplete CRISPR/Cas Subtype I-E system.

**Keywords:** *Streptomyces* sp. B-146, BGCs, Antimicrobials, Antitumorals, Antivirals, Phylogenetic Analysis, Phylogenomic Analysis, CRISPR/cas systems.



## I. INTRODUCCIÓN

Los microorganismos son productores notables de una amplia diversidad de productos naturales que actúan como agentes farmacéuticos y terapéuticos en el tratamiento de enfermedades relevantes, permitiendo mejorar significativamente la salud y el bienestar de los humanos (Ramírez-Rendon *et al.*, 2022). En ese sentido, *Streptomyces*, el género bacteriano más ubicuo en entornos ambientales, tienen la capacidad prolífica para producir productos naturales diversos y valiosos con actividades biológicas significativas en los sectores de la medicina, el medio ambiente, la industria alimentaria y la agronomía (Alam *et al.*, 2022). Como resultado, los fármacos disponibles en los sistemas de vigilancia de la salud, en su mayoría derivan de principios activos producidos por bacterias del género *Streptomyces* (Sivalingam *et al.*, 2019).

Actualmente, es urgente descubrir y/o desarrollar nuevos antibióticos y medicamentos contra el cáncer, debido al rápido crecimiento de la resistencia a los antibióticos y a los agentes quimioterapéuticos contra el cáncer (Alam *et al.*, 2022). La resistencia a antibióticos, causado por el uso masivo e inadecuado de antibióticos, se ha convertido ya en uno de los principales desafíos que enfrenta la medicina moderna en todo el mundo (Andersson *et al.*, 2020; Ghosh *et al.*, 2020). Asimismo, el cáncer, una de las principales causas de muerte en todo el mundo, requiere de nuevos medicamentos sin o con menores efectos secundarios y que sean capaces de evadir o revertir la resistencia a la quimioterapia (Bukowski *et al.*, 2020; Sivalingam *et al.*, 2019).

A pesar de que las bacterias del género *Streptomyces* suelen ser los microorganismos biosintetizadores más prolíficos y, en consecuencia, un recurso valioso para el descubrimiento de nuevos productos naturales con bioactividades antimicrobianas o antitumorales; a menudo resulta imposible detectarlos a través de los métodos tradicionales de screening bioquímico (Ward & Allenby, 2018; Lee *et al.*, 2020b). Esto debido a que estas bacterias suelen inactivar la biosíntesis de sus productos naturales en condiciones de cultivo axénicas debido a la falta de estímulos ambientales (Lee *et al.*, 2020b). Para sortear este problema se recurre a la minería del genoma, un método computacional para la detección y anotación automática de genes biosintéticos físicamente colocados en los cromosomas denominados “Clústeres de genes biosintéticos” (BGCs) (Sivalingam *et al.*, 2019; Russell & Truman, 2020).

El análisis bioinformático de los BGCs, responsables de la biosíntesis de los productos naturales (también llamados metabolitos secundarios), brindan mucha información sobre la clase de compuestos que codifican y, en algunos casos, sus estructuras precisas (Jensen, 2026; Tran *et al.*, 2019).

Por todo lo anteriormente mencionado, planteamos llevar a cabo la mineración del genoma completo de *Streptomyces* sp. B-146; una cepa bacteriana aislada de las salinas de Bayobar en la costa norte del Perú y cuyo extracto crudo ha mostrado actividad antibacteriana contra *Escherichia coli* ATCC 10536, *Staphylococcus aureus* ATCC 6538 y *Acinetobacter baumannii* AC-972 MDR; y también actividad antiproliferativa contra las líneas celulares de cáncer U251, MCF7 y NCI-H460 (Flores-Clavo *et al.*, 2021). Para ello, proponemos como objetivo identificar y anotar clústeres de genes biosintéticos relacionados con compuestos antitumorales, antimicrobianos y otros metabolitos bioactivos mediante análisis del genoma de *Streptomyces* sp. B-146. Esto nos dará la oportunidad de predecir el descubrimiento de nuevos productos naturales con propiedades terapéuticas únicas y entender su función biológica teórica, un punto de partida para escalar las averiguaciones a nivel de laboratorio y verificar sus actividades terapéuticas putativas.

## II. MARCO TEÓRICO

### 2.1. ACTINOBACTERIAS

Las *actinobacterias* son bacterias gram-positivas que se caracterizan por su elevado contenido GC en sus genomas, conforman uno de los filos bacterianos más extensos y se distribuyen de forma ubicua en diversos ecosistemas terrestres y acuáticos (Barka *et al.*, 2015). Constituyen 425 géneros distribuidos en seis clases, 46 órdenes y 79 familias (Salam *et al.*, 2020). Dicha abundancia les permite prosperar en una amplia gama de hábitats extremos; lo que posibilita su aislamiento de entornos singulares como suelos salinos alcalinos, sedimentos de aguas profundas, aguas termales, entre otros (Sheng *et al.*, 2016).

### 2.2. STREPTOMYCES

Los *Streptomyces* son un género muy extendido dentro del phylum *Actinobacteria* y, en los últimos años, gran parte de la atención se ha centrado en el aislamiento de especies de *Streptomyces* de los hábitats ambientales más extremos del mundo (Zhao *et al.*, 2016a; Sivalingam *et al.*, 2019). A septiembre de 2021, el género *Streptomyces* alberga 673 especies con un nombre correcto y válidamente publicado según la base de datos LPSN (Parte *et al.*, 2020).

El secuenciamiento del ADN de varios *Streptomyces* mostró que estas bacterias tienen cromosomas con tamaños que oscilan entre 6 y 12 Mpb y que tienen altos contenidos de GC de alrededor del 70% (Romero-Rodríguez *et al.*, 2015).

El ciclo de vida de los *Streptomyces* comprende tres fases de desarrollo claramente definidas: hifas vegetativas, hifas aéreas y esporas (Jones & Elliot, 2018). La reproducción de estas bacterias miceliales multicelulares se produce cuando las hifas aéreas especializadas se diferencian en cadenas de esporas (Hoskisson & van Wezel, 2019). Este proceso singular de diferenciación morfológica, único en bacterias Gram positivas, demanda un metabolismo altamente especializado y coordinado (Procópio *et al.*, 2012). En consecuencia, los *Streptomyces* exhiben un ciclo de vida notablemente complejo y una gran capacidad de producir productos naturales en abundancia (Hoskisson & van Wezel, 2019).

En ese sentido, las bacterias del género *Streptomyces* son reconocidas por su habilidad de biosintetizar una amplia gama de compuestos bioactivos, siendo responsables de la producción de alrededor de la mitad de los antibióticos usados clínicamente y otros compuestos farmacéuticamente útiles como antivirales, antifúngicos, antihelmínticos, anticancerígenos e inmunosupresores (Hoskisson y van Wezel, 2019; Lee *et al.*, 2020b). Por ende, los *Streptomyces* son los productores más prolíficos de productos naturales con propiedades bioactivas y también los mejor estudiados (Ward & Allenby, 2018).

### **2.3. PRODUCTOS NATURALES**

Los productos naturales son pequeñas moléculas orgánicas producidas principalmente por bacterias, hongos y plantas que a menudo se utilizan como agentes terapéuticos debido a su asombrosa diversidad estructural y su amplia gama de bioactividades (Jensen, 2016; Tomm *et al.*, 2019). Son el resultado de la exploración de la naturaleza a través de la evolución y producidos para llevar a cabo funciones biológicas específicas dentro y/o entre organismos con la finalidad de dar una ventaja selectiva al organismo productor (Grigalunas *et al.*, 2022). Su diversidad estructural es generada por enzimas biosintéticas codificadas generalmente en BGCs, que construyen andamios centrales y que realizan modificaciones periféricas que permite definir familias de productos naturales (Chavali & Rhee, 2018; Scott & Piel, 2019).

### **2.4. CLÚSTERES DE GENES BIOSINTÉTICOS**

Los BGCs son agrupaciones génicas que generalmente codifican vías completas que facilitan la biosíntesis, ensamblaje, modificación y regulación de los precursores de sus respectivos productos naturales (Lee *et al.*, 2020b). Dicha regulación está estrictamente controlada por sistemas reguladores complejos a niveles de transcripción y traducción para utilizar de manera efectiva los precursores que son suministrados por el metabolismo primario (Kim *et al.*, 2020). En otras palabras, los BGCs contienen genes codificantes de enzimas biosintéticas centrales responsables del primer paso metabólico en la biosíntesis de los productos naturales; genes codificantes de enzimas de "adaptación" encargados de modificar la molécula producida por las enzimas centrales; genes reguladores responsables de la activación de la transcripción de los BGCs y genes codificantes de transportadores de membrana necesarios para el eflujo de los productos naturales generados o genes de "resistencia" necesarios para la desintoxicación/autoprotección frente a las moléculas producidas (Crits-Christoph *et al.*, 2020; Kwon *et al.*, 2021).

Las bacterias del género *Streptomyces* portan usualmente entre 8 a 83 BGCs y miembros de una misma especie puede variar enormemente respecto a la cantidad y diversidad de BGCs que albergan (Belknap *et al.*, 2020).

## **2.5. MINERACIÓN GENÓMICA**

Con los avances en los métodos de secuenciación de genomas de alto rendimiento, varias estrategias *in silico* para la minería genómica han sido desarrolladas y aplicadas al mapeo de genomas de los *Streptomyces* (Lee *et al.*, 2020b). La minería genómica podría ser definido como un método computacional aplicado a la detección y anotación de BGCs a partir de datos de secuencias genómicas, mediante análisis de homología con BGCs conocidos (Ward & Allenby, 2018; Russell & Truman, 2020).

Para ser específicos, las herramientas bioinformáticas usadas en la mineración genómica son una nueva estrategia de descubrimiento de productos naturales que permite a los investigadores detectar y predecir computacionalmente las secuencias de los BGCs y sus vías biosintéticas (Li *et al.*, 2018; Mungan *et al.*, 2020).

En consecuencia, la mineración de genomas tiene el potencial de acelerar el descubrimiento de moléculas importantes, como agentes antimicrobianos y anticancerígenos (Russell y Truman, 2020). En ese sentido, la mineración de 1.110 genomas de *Streptomyces* ha revelado que los BGCs más comunes son los que biosintetizan Policétidos, Péptidos ribosómicamente sintetizados y postraduccionalmente modificados, Péptidos no ribosomales y Terpenos; seguidos Butirolactonas, Ladderanos, Arilpolienos, Melaninas, Fosfonatos, Sideróforos, Ectoínas, Oligosacáridos, Nucleósidos, Indoles, Aminoglucósidos/aminociclitoles, Fenazinas, Betalactámicos, entre otros (Belknap *et al.*, 2020). Asimismo, existen una gran cantidad de BGCs inexplorados y no caracterizados, cuyos metabolitos secundarios son desconocidos (Lee *et al.*, 2020a; AbuSara *et al.*, 2019).

## **2.6. CLASIFICACIÓN DE LOS PRODUCTOS NATURALES**

### **2.6.1. PÉPTIDOS NO RIBOSOMALES**

Los péptidos no ribosomales (NRP) son producidos por una amplia variedad de hongos y bacterias, tienen una diversidad estructural notable y constituyen una clase extendida de antibióticos potentes y otros fármacos importantes utilizados para enfermedades infecciosas, el cáncer o la inmunosupresión (Marahiel, 2009).

Para su biosíntesis las bacterias y hongos utilizan grandes enzimas multifuncionales, denominadas Péptido sintetasas no ribosomales (NRPS) (Finking & Marahiel, 2004). El ensamblaje de estos péptidos implica la incorporación de un pequeño conjunto de unidades monoméricas en un oligómero lineal, dichos monómeros son aminoácidos proteinogénicos, no proteinogénicos y otros ácidos carboxílicos (Fischbach & Walsh, 2006).

### **2.6.2. POLICÉTIDOS**

Los policétidos (PKs) representan la clase más grande y estructuralmente diversa de productos naturales con aplicaciones en medicina, agricultura y salud animal; comparten un proceso de biosíntesis común que emplea enzimas modulares llamadas Policétido sintetasas (PKS), que se componen de dominios enzimáticos que comparten secuencia y similitud funcional en todos los PKS conocidos (Yuzawa *et al.*, 2018).

La línea de ensamblaje de los PKS se encuentra entre las maquinarias de proteínas más complejas conocidas en la naturaleza, responsables de la biosíntesis de numerosos compuestos utilizados clínicamente (Nivina *et al.*, 2019). Diferentes tipos de interacciones proteína-proteína, intra e inter-módulos PKS, desempeñan un papel fundamental en el ciclo catalítico de un PKS multimodular (Robbins *et al.*, 2016). En *Streptomyces* se han identificado tres tipos de PKS (I, II y III), responsables de la biosíntesis de policétidos con bioactividades antibacterianas, antifúngicas, anticancerígenas, inmunosupresoras, antivirales, anti-colesterol, antiinflamatorias, entre otras (Risidian *et al.*, 2019).

#### **a. Policétidos Tipo I**

Los policétidos tipo 1 son biosintetizados por las enzimas Policétido sintetasas de tipo I (PKS T1) que actúan como fábricas moleculares para generar policétidos complejos a partir de precursores orgánicos simples y operan en organismos tan diversos como estreptomicetos, micobacterias, mixobacterias, cianobacterias, hongos, amebas y animales (Keatinge-Clay, 2012). Las PKS tipo I son enzimas multifuncionales organizadas en módulos, donde cada uno alberga un conjunto de dominios distintos entre sí, encargados de catalizar un ciclo de elongación de la cadena de policétidos; prototípicamente, el módulo de elongación de las PKS tipo I contiene como mínimo tres dominios: una aciltransferasa (AT), una proteína transportadora de acilo (ACP) y una  $\beta$ -cetoacil sintasa (KS) (Cheng *et al.*, 2009).

Las PKS tipo I se dividen en PKS de tipo I iterativos y PKS de tipo I modulares (o no iterativos) (Keatinge-Clay, 2012). Una PKS iterativa suele ser una proteína de un solo módulo que consta de un solo conjunto de dominios funcionales que es capaz de realizar múltiples rondas de extensión de cadena y procesamiento de  $\beta$ -ceto (Chen y Du, 2016).

En contraste, un PKS no iterativo son múltiples módulos secuenciales, cada módulo contiene los dominios de extensión y adaptación necesarios y son responsables de solo una ronda de elongación de la cadena y el posterior procesamiento de  $\beta$ -ceto, antes de transferir el policétido intermedio a un módulo posterior para el siguiente paso de procesamiento y elongación de la cadena (Zhan, 2009; Chen & Du, 2016).

Las T1PKS iterativas son inesperadamente abundantes y generalizadas en *Streptomyces*, lo que indica el enorme potencial para identificar y confirmar experimentalmente las BGC T1PKS iterativas para descubrir nuevos productos naturales de policétidos (Wang *et al.*, 2020a).

#### **b. Policétidos Tipo II**

Los policétidos tipo II son biosintetizados por enzimas monofuncionales denominadas Policétido Sintetas tipo II, cuyas enzimas centrales constan de dos subunidades de  $\beta$ -cetoacil sintasa (KSa y KSb) y una proteína transportadora de acilo (ACP), que trabajan juntas en la biosíntesis de pigmentos de esporas, antibióticos aromáticos y otros policétidos aromáticos (Komaki & Harayama, 2006). Los policétidos aromáticos, miembros importantes de la familia de los policétidos, son abundantes en las *actinobacterias* y tienen una amplia variedad de propiedades bioactivas, como actividades antibacterianas, antitumorales y antivirales (Zhan, 2009).

La biosíntesis de policétidos aromáticos por PKSs tipo II produce cadenas de policétidos muy diversas mediante la condensación secuencial de las unidades de inicio con unidades de extensión, seguida de reacciones de reducción, ciclación, aromatización y adaptación (Wang *et al.*, 2020b). La ciclación de la cadena de policétidos es un punto de control crítico en la biosíntesis de policétidos aromáticos, donde las aromatasas/ciclasas (ARO/CYCs) son un componente importante en las ciclaciones del primer anillo y/o aromatización de policétidos aromáticos (Ames *et al.*, 2011).



### **c. Policétido Tipo III**

Los policétidos tipo II son biosintetizados por las enzimas Policétido Sintetas tipo III, cuyas enzimas son homodímeros simples de cetosintetasas de un solo dominio, que contienen un único sitio activo en cada monómero que cataliza las reacciones de cebado, extensión y ciclación de forma iterativa para formar productos de policétidos (Katsuyama & Ohnishi, 2012; Yu *et al.*, 2012a).

Las PKS de tipo III son las estructuras más simples entre los otros tipos de PKS y, a diferencia de las PKS de tipo I y II, no utilizan ACP como un ancla para la producción del metabolito policétido, usando en este caso directamente los acil-CoA como sustratos para generar compuestos policétidos (Risidian *et al.*, 2019). Los PKS de tipo III desempeñan un papel fundamental en la biosíntesis de importantes metabolitos secundarios, que van desde moléculas de señalización hasta productos naturales bioactivos (Yu *et al.*, 2012a). Dichos productos naturales tienen funciones biológicas significativas y/o actividades farmacéuticas importantes (Katsuyama & Ohnishi, 2012). Por esta razón, los PKS de tipo III han atraído más atención debido a sus diversos productos, amplia distribución, estructuras relativamente simples y fácil manipulación genética (Yu *et al.*, 2012a).

### **2.6.3. PÉPTIDOS RIBOSÓMICAMENTE SINTETIZADOS Y MODIFICADOS POSTRADUCCIONALMENTE**

Los péptidos ribosómicamente sintetizados y modificados postraduccionalmente (RiPP) son una clase de productos naturales peptídicos cíclicos o lineales con una notable diversidad estructural y funcional (Zhong *et al.*, 2020). Inicialmente los RiPPs se biosintetizan a partir de un gen estructural como péptidos precursores de aproximadamente 20 - 110 residuos de longitud (Arnison *et al.*, 2013). Los péptidos precursores constan de un péptido líder N-terminal y un péptido central C-terminal (Zhong *et al.*, 2020).

El péptido líder contiene motivos que actúan como secuencias de reconocimiento para las enzimas de modificación postraducciona (enzimas PTM), las cuales unidas al péptido líder realizan una serie de modificaciones sobre el péptido central (Russell & Truman, 2020). Por tanto, los BGCs tipo RiPPs mínimamente deben albergar un péptido precursor y varias enzimas PTM (Zhong *et al.*, 2020).

Los RiPP tienen una serie de ventajas sobre los fármacos de moléculas pequeñas en términos de especificidad y afinidad por los objetivos, y sobre los fármacos basados en proteínas en términos de penetrabilidad celular, estabilidad y tamaño (Fu *et al.*, 2021).

Los recientes avances en genómica y biología sintética nos están facilitando el descubrimiento de una gran cantidad de nuevos productos naturales ribosomales, incluidos lantipéptidos, lazopéptidos, sactipéptidos, tiopéptidos, microviridinas, cianobactinas, péptidos lineales que contienen tiazol / oxazol, etc. (Zhong *et al.*, 2020).

#### **a. Lantipéptidos**

Los lantipéptidos son una familia en rápido crecimiento de productos naturales de péptidos policíclicos pertenecientes a los RiPPs, están ampliamente distribuidos en especies taxonómicamente distantes y sus sistemas biosintéticos y actividades biológicas son igualmente diversas (Zhang *et al.*, 2015). Los lantipéptidos son compuestos prometedores, principalmente en el tratamiento de infecciones bacterianas, pero también actúan como agentes antitumorales, antivirales y analgésicos (Geng & Smith, 2018).

Estos péptidos contienen múltiples enlaces cruzados de tioéter generados por la deshidratación de residuos de Ser y Thr a deshidroalanina y deshidrobutirina seguida de la adición de residuos de Cys para generar los aminoácidos entrecruzados de lantionina y metillantionina tioéter de los cuales los lantipéptidos derivan su nombre (van der Donk & Nair, 2014; Repka *et al.*, 2017). Después de la instalación de los enlaces cruzados de tioéter característicos, las enzimas de adaptación introducen modificaciones postraduccionales adicionales que son únicas para cada lantipéptido y que afinan sus actividades y / o estabilidad (Repka *et al.*, 2017).

Su péptido precursor comprende: de un péptido líder N-terminal, el cual es crucial para el reconocimiento enzimático, especialmente para las enzimas de modificación y actúa además como una señal de secreción para el exportador de lantipéptidos y de un péptido central C-terminal, el cual es el sitio diana para las modificaciones postraduccionales y contiene aminoácidos deshidratados y anillos de lantionina (Lagedroste *et al.*, 2021). Si bien los BGCs de lantipéptidos pueden identificarse por la presencia de genes que codifican enzimas características involucradas en el proceso de modificación postraducciona, localizar los péptidos precursores codificados dentro de estos grupos es un desafío debido a su corta longitud y alta variabilidad de secuencia, lo que limita la exploración a gran escala de la biosíntesis de lantipéptidos (Walker *et al.*, 2020).

El análisis de los BGCs tipo lantipéptidos de *Actinobacteria* sugiere que las rutas biosintéticas de los lantipéptidos y sus productos naturales son mucho más diversas de lo que se aprecia actualmente y contienen muchas modificaciones postraduccionales diferentes; Además, las lantionina sintetasas son mucho más diversas en secuencia y topología de dominio que los sistemas caracterizados actualmente (Zhang *et al.*, 2015).

Hay cuatro clases diferentes de lantipéptidos bien conocidas que se definen por los cuatro tipos diferentes de lantipéptido sintetasas (LanB y LanC para la clase I, LanM para la clase II, LanKC para la clase III y LanL para la clase IV) (Zhang *et al.*, 2012; Arison *et al.*, 2013; Yu *et al.*, 2013). Recientemente, se informó sobre un miembro de una nueva clase de lantipéptidos designados como clase V, que es biosintetizado por tres proteínas (LxmK, LxmX y LxmY); además, se han identificado BGCs de lantipéptidos de clase V en *Actinobacterias* (Xu *et al.*, 2020).

## **b. Lazopéptidos**

Los Lazopéptidos pertenecen a la superfamilia de productos naturales RiPPs y están definidos por un anillo de macrolactama N-terminal que está enhebrado por la cola C-terminal (Hegemann, 2020). Los Lazopéptidos contienen tres estructuras bien definidas: el anillo que es clave tanto para su estructura altamente compacta y estable como para su propiedad antibacteriana intrínseca; el bucle que se considera como la región de modificación adecuada; y la cola que sirve como mecanismo de autoprotección y para exportar péptido lazo fuera de las células (Zhao *et al.*, 2016b). Según la presencia, el número y la ubicación de los puentes disulfuro, existen cuatro clases de péptidos de lazo: la clase I contiene dos enlaces disulfuro; La clase II no contiene enlaces disulfuro; la clase III tiene un enlace disulfuro formado entre el anillo macrocíclico y la cola; y la clase IV tienen un enlace disulfuro formado en la cola (Cheng & Hua, 2020).

La mineración genómica ha revelado que un gran número de BGCs tipo lazopéptidos se distribuyen sobre los genomas de *Proteobacterias* y *Actinobacterias*, pero en la mayoría de los casos son BGCs crípticos (Kodani & Unno, 2020). A la fecha, la mayoría de los lazopéptidos clase II se descubrieron en *Proteobacterias*, mientras que los lazopéptidos clase I, III y IV en su mayoría se han aislado de *Actinobacterias* (Cheng & Hua, 2020). Sorprendentemente, la estructura de los lazopéptidos confiere resistencia a la degradación térmica y proteolítica, por lo que gozan de excelentes propiedades farmacocinéticas (Liu *et al.*, 2021).

Los lazopéptidos tienen propiedades antimicrobianas, anticancerígenas, antivirales, de inhibición enzimática y de bloqueo de receptores, por lo que prometen efectos terapéuticos sobre infecciones fúngicas, el cáncer, el VIH, enfermedades gastrointestinales, la tuberculosis, el Alzheimer y enfermedades cardiovasculares (Cheng & Hua, 2020). Por todo ello, los lazopéptidos tienen un gran potencial en el desarrollo de nuevos fármacos (Liu *et al.*, 2021).

### **c. Tiopéptidos**

Los tiopéptidos son una subclase de RiPPs que tienen arquitecturas moleculares complejas debido a numerosas modificaciones postraduccionales y, además, muestran una gran variedad de bioactividades, incluyendo una potente actividad antimicrobiana (Luo *et al.*, 2016; Aggarwal *et al.*, 2021). Su biosíntesis comienza con la traducción ribosomal de un péptido precursor de 50-60 aminoácidos de longitud; su péptido líder, generalmente 30-40 residuos, sirve como motivo de reconocimiento para las Enzimas PTM, las cuales modifican el péptido central hasta convertirlo en un tiopéptido maduro (Vinogradov & Suga, 2020).

Los tiopéptidos maduros tienen tres componentes principales: un anillo macrocíclico central, decorado con varios aminoácidos, así como sustituyentes tiazol, oxazol y tiazolina; (2) un anillo de piridina central conectado a un anillo macrocíclico; (3) una cola que se extiende desde el anillo central de piridina; además, en algunos casos, los tiopéptidos tienen un anillo lateral secundario con restos químicos únicos (Chan & Burrows, 2021). El progreso durante la última década ha ampliado enormemente nuestro conocimiento de los tiopéptidos en lo que respecta a su diversidad estructural, funcional y biosintética (Vinogradov y Suga, 2020). Actualmente, gracias a la mineración genómica, es posible identificar nuevos tiopéptidos putativos codificados en BGCs de diferentes genomas bacterianos, especialmente en *Actinobacterias* (Aggarwal *et al.*, 2021).

### **2.6.4. TERPENOS**

Los Terpenos son hidrocarburos que consisten en unidades de isopreno de 5 carbonos (C<sub>5</sub>H<sub>8</sub>) como componente básico y cuyas “columnas vertebrales” pueden ser reorganizadas en estructuras cíclicas para generar Terpenoides (Mahizan *et al.*, 2019).

Los Terpenoides, también llamados Isoprenoides, son la familia más grande y estructuralmente diversa de productos naturales presentes en todos los dominios de la vida; sin embargo, los Terpenoides bacterianos son menos conocidos y están clasificados en dos familias principales y 55 subfamilias distintas (Rudolf *et al.*, 2020).

Las terpeno-sintasas (o terpeno-ciclasas) son enzimas clave en el proceso de síntesis de Terpenoides bacterianos a partir de los Terpenos (Zhang *et al.*, 2014). Estas se distribuyen ampliamente entre microorganismos y se han estudiado principalmente en miembros del género *Streptomyces*, revelando que existen al menos 10 tipos de terpeno-sintasas, las cuales están relacionadas con la producción de geosmina, 2-metilisoborneol, epi-isozizaene, 7-epi - $\alpha$ -eudesmol, epi-cubenol, cariolan-1-ol, ciclooctat-9-en-7-ol, isoafrianol, pentaleno y  $\alpha$ -amorfeno (Martín-Sánchez *et al.*, 2019).

Los BGC que sintetizan Terpenoides han revelado una gran cantidad de enzimas oxidativas previamente desconocidas, incluidos los citocromos P450, dioxigenasas no hemo dependientes de hierro y flavina-monooxigenasas (Cane & Ikeda, 2012).

## **2.7. SISTEMAS CRISPR/Cas EN PROCARIOTAS**

### **2.7.1. SISTEMAS CRISPR/Cas**

Los sistemas CRISPR/Cas proporcionan a los procariotas inmunidad adaptativa hereditaria contra elementos genéticos foráneos y están conformados por un Arreglo CRISPR (Agrupaciones de pequeñas repeticiones palíndromas interespaciados regularmente) y diversos genes Cas (Proteínas asociadas a los CRISPR) (Makarova *et al.*, 2015; Strich & Chertow, 2019).

#### **a. Arreglos CRISPR**

Un arreglo CRISPR, comprende pequeñas secuencias repetitivas de ADN típicamente de 20 a 40 bases (llamadas repeticiones); las cuales están separadas por secuencias hipervariables de ADN generalmente de 20 a 58 bases (llamados “Espaciadores”) (Edgar, 2007; Sorek *et al.*, 2008; Horvath *et al.*, 2008; Makarova *et al.*, 2015). Además, existe un tercer elemento denominado “Secuencia Líder”, que flanquea usualmente el extremo 5’ de los arreglos CRISPR y son escasamente conservadas y de tamaño hipervariable (Lillestøl *et al.*, 2006; Sorek *et al.*, 2008; Horvath *et al.*, 2009; Díez-Villaseñor *et al.*, 2010; Karimi *et al.*, 2018).

## **b. Genes Cas**

Los genes Cas que usualmente flanquean los arreglos CRISPR están organizados en uno o más operones (Shmakov *et al.*, 2020). Generalmente, están ubicados a cientos e incluso a miles de pb de un arreglo CRISPR, mayoritariamente en la región 5'; no obstante, pueden también ser encontrados al lado 3' o incluso en ambos lados de un arreglo CRISPR (Jansen *et al.*, 2002; Horvath *et al.*, 2009). Las proteínas codificadas por estos genes están involucradas en las diferentes etapas de la actividad de los sistemas CRISPR/Cas (Makarova *et al.*, 2015).

### **2.7.2. FUNCIÓN DE LOS SISTEMAS CRISPR/Cas**

La inmunidad mediada por CRISPR/Cas abarca tres etapas: adaptación, expresión e interferencia (Staals *et al.*, 2016). Durante la adaptación, los espaciadores, derivados de elementos genéticos foráneos, son incorporados en los arreglos CRISPR para proporcionar inmunidad contra invasores que albergan secuencias idénticas a los espaciadores adquiridos (Horvath *et al.*, 2009; Makarova *et al.*, 2015; Staals *et al.*, 2016; Amitai & Sorek, 2016). Durante la expresión, el arreglo CRISPR se transcribe como un pre-crRNA que es procesado por algunas proteínas Cas para generar ARN CRISPR (crRNA) maduros, los cuales junto a otras proteínas Cas forman complejos efectores de ribonucleoproteínas (RNP) (Brouns *et al.*, 2008; Amitai & Sorek, 2016; Devi *et al.*, 2022a). Tanto, en la adaptación como en la expresión, la secuencia Líder juega un rol fundamental debido a los elementos reguladores que contiene (Alkhnbashi *et al.*, 2016). Durante la interferencia, el complejo efector RNP identifica por apareamiento de bases las secuencias objetivo, las cuales son degradadas por proteínas o dominios con actividad nucleasa siempre que esté presente un Motivo Adyacente al Protoespaciador (PAM), cuya secuencia y ubicación varía entre subtipos CRISPR/Cas (Amitai & Sorek, 2016; Barrangou & Horvath, 2017).

### **2.7.3. DIVERSIDAD DE LOS SISTEMAS CRISPR/Cas**

A pesar de que los sistemas CRISPR-Cas tienen presencia generalizada en bacterias y arqueas, su arquitectura y composición proteica es extraordinariamente diversa y tienen una alta tasa de evolución tanto para los genes Cas como para el contenido único de sus espaciadores (Sorek *et al.*, 2008; Wiedenheft *et al.*, 2009; Makarova *et al.*, 2011; Koonin *et al.*, 2017).

Debido a su enorme diversidad, los distintos *loci* CRISPR-Cas generalmente usan diferentes componentes y mecanismos para llevar a cabo su función esencial; es decir, proporcionar inmunidad contra la invasión de elementos genéticos foráneos (Carte *et al.*, 2014; Devi *et al.*, 2022a).

No obstante, a pesar de su gran diversidad, análisis comparativos exhaustivos han revelado puntos comunes en la evolución de los sistemas CRISPR/Cas (Koonin *et al.*, 2017). En ese sentido, se ha implementado una estrategia computacional multifacética para la clasificación evolutiva de los sistemas CRISPR/Cas, que develan de momento 2 clases, 6 tipos y 33 subtipos; específicamente, la clase 1 abarca 16 subtipos distribuidos en los tipos I, III y IV y la clase 2 abarca 17 subtipos repartidos en los tipos II, V y VI (Makarova *et al.*, 2020). La clase 1 se caracteriza por tener complejos efectores RNP de múltiples subunidades; mientras, la clase 2 tiene complejos efectores RNP conformado por una sola proteína multidominio (Makarova *et al.*, 2015; Koonin *et al.*, 2017).

#### **2.7.4. SISTEMAS CRISPR/Cas TIPO I, SUBTIPO E**

Los sistemas CRISPR Tipo I abarcan siete subtipos (A-G) y se caracterizan por la presencia de genes que codifican el complejo Cascade (CRISPR-associated complex for antiviral defense), el cual desempeña un rol fundamental en el proceso de maduración de los crRNA, así como en el reconocimiento y unión a la secuencia objetivo; y un gen distintivo que codifica para la enzima Cas3, la cual se asocia al complejo Cascade para formar el complejo efector RNP encargado de la degradación de la secuencia objetivo en la etapa de interferencia (Makarova *et al.*, 2011; 2015; 2020).

El sistema CRISPR/cas subtipo I-E, detectado por primera vez en *E. coli* K12, consta de ocho genes Cas: Cas1 (integrasa), Cas2 (endoribonucleasa), Cas3 (nucleasa HD / helicasa DEAD-box) y cinco genes adicionales denominados CasABCDE (CasA, CasB, CasC, CasD, CasE) (Brouns *et al.*, 2008). Los genes Cas1 y Cas2 codifican proteínas que conforman el complejo de adaptación, cuya función es adquirir protoespaciadores del ADN invasor producto de la acción enzimática de RecBCD, Cas3 u otras enzimas e incorporándolos como nuevos espaciadores a los arreglos CRISPR del huésped (Nuñez *et al.*, 2014; Amitai & Sorek, 2016; Xue & Sashital, 2019).



Los genes *CasABCDE* codifican las subunidades del complejo Cascade, el cual está implicado en las etapas del procesamiento e interferencia y cuyas funciones específicas son: (1) generar crRNA maduros a partir del pre-crRNA y asociarse a estos crRNA para formar el complejo efector RNP; (2) reconocer la secuencia PAM del ADN invasor; (3) formar una estructura estable entre el ADN invasor y el complejo Cascade; y (4) formar una estructura llamada “R-loop” producto de la separación de las cadenas del ADN invasor y la hibridación del crRNA con la cadena complementaria del ADN invasor para facilitar el acoplamiento de Cas3 (Brouns *et al.*, 2008; Sinkunas *et al.*, 2011; Charpentier *et al.*, 2015; Liu & Doudna, 2020).

El gen *Cas3* codifica para una nucleasa implicada en la etapa de interferencia y cuyas funciones específicas son: (1) acoplarse a la estructura “R-loop” formada por la interacción del complejo Cascade, el crRNA y el ADN invasor; (2) escindir en varios sitios la cadena simple de ADN invasor en la estructura “R-loop” para facilitar la liberación del complejo efector RNP; (3) escindir la cadena simple complementaria al crRNA luego del desacoplamiento del complejo efector RNP; y finalmente, (4) liberarse para escindir otras moléculas de ADN invasor (Sinkunas *et al.*, 2011; Charpentier *et al.*, 2015; Liu & Doudna, 2020; Makarova *et al.*, 2020).

### III. MATERIALES Y MÉTODOS

#### 3.1. MÉTRICAS DE LA SECUENCIA DEL GENOMA DE *Streptomyces* sp. B-146

La secuencia completa del genoma de *Streptomyces* sp. B-146 fue proporcionada por el departamento de Biotecnología del Centro de investigación e innovación en ciencias activas multidisciplinares (CIICAM), con sede en Chiclayo, Perú. Brevemente, previo a esta investigación, el ADN genómico extraído de la cepa *Streptomyces* sp. B-146 fue secuenciado en las plataformas de secuenciamiento Illumina y PacBio. Los “*shorts reads*” de la plataforma Illumina fueron trimados usando el programa Trimmomatic 0.39 (Bolger, 2014), que consiste en la remoción de “*Illumina adapters*” y secuencias con “*Phread quality score*” menor a 24, para conservar únicamente secuencias de alta calidad. La corrección de los “*long reads*” de la plataforma PacBio se realizó con los “*Illumina reads*” trimados usando Colormap (Haghshenas *et al.*, 2016).

Los “*shorts reads*” (Illumina) y “*long reads*” (PacBio) corregidos fueron usados para obtener ensamblajes híbridos más precisos mediante Unicycler (Wick *et al.*, 2017). Para visualizar los alineamientos de los contigs se usó Icarus (Mikheenko *et al.*, 2016). Asimismo, se realizó el ensamblaje de *nov*o de “*contigs*”, el “*Scaffolding*” y el cierre de “*gaps*” a partir de los “*shorts reads*” trimados de Illumina usando Platanus version 1.2.4 (Kajitani *et al.*, 2014). El “*Scaffolding*” de *nov*o de “*contigs*” y el cierre de “*gaps*” a partir de los “*long reads*” corregidos de PacBio se realizó con SSPACE-LongRead version 1-1 (Boetzer & Pirovano, 2014) y LR\_Gapcloser (Xu *et al.*, 2019), respectivamente. Para descartar la presencia de plásmidos se usó Circlator (Hunt *et al.*, 2015).

Finalmente, la última versión del “*scaffold*” obtenido fue usado para determinar el genoma más próximo usando Blast 2.9.0+ (Altschul *et al.*, 1990). El genoma más próximo que correspondía a *S. olivaceus* strain KLBMP 5084 (CP016795), fue usado como referencia para cerrar el genoma de *Streptomyces* sp. B-146 usando Medusa (Medlock *et al.*, 2020). Para pulir el “*draft genome*”, corrigiendo bases y errores de ensamblaje, se usó Pilon 1.24.jar (Walker *et al.*, 2014). La evaluación de calidad e integridad de la secuencia del genoma completo de *Streptomyces* sp. B-146 se realizó con Quast 5.0.2 (Gurevich *et al.*, 2013) y CheckM v1.2.0 (Parks *et al.*, 2015) y gVolante (Nishimura *et al.*, 2017).

### 3.2. ANÁLISIS FILOGENÉTICOS Y FILOGENÓMICOS DE *Streptomyces* sp. B-146

Para los análisis filogenéticos se recuperó la secuencia parcial del gen *ARNr 16S* de *Streptomyces* sp. B-146 depositada en GenBank del NCBI (MW562805). La secuencia recuperada fue sometida a un proceso de búsqueda de secuencias próximas a través del servicio "Identify" en la base de datos EZBiocloud (Yoon *et al.*, 2017), para determinar, seleccionar y recuperar las "Type strain" más cercanas a *Streptomyces* sp. B-146. Además, se seleccionó y recuperó la secuencia del gen *ARNr 16S* de *Kitasatospora kazusensis* (NR041539) para ser usada como secuencia fuera de grupo. Todas las secuencias recuperadas fueron sometidas a análisis filogenéticos usando MEGA versión 11.0 (Tamura *et al.*, 2021); el cual tiene incorporado el software Clustal W (Thompson *et al.*, 1994), encargado de ejecutar los alineamientos para consensuar un tamaño único de secuencias, a partir de las cuales, se construyeron los árboles filogenéticos. La matriz de distancias evolutivas se calculó utilizando los parámetros del método "Maximum Composite Likelihood" (Yang, 1994) y el árbol filogenético se construyó a partir de las distancias evolutivas calculadas por el método "Neighbor-joining" (Saitou & Nei, 1987), con valores de "bootstrap" de 1000 repeticiones (Felsenstein, 1985).

Los análisis filogenómicos se llevaron a cabo a partir de la secuencia completa del genoma de *Streptomyces* sp. B-146. Para ello se recurrió al servidor web autoMLST (*Automated Multi-Locus Species Tree*) (Alanjary *et al.*, 2019), el cual realiza un análisis de secuencia de múltiples *loci* para producir árboles de especies de alta resolución; resaltando, además, su potencial biosintético de productos naturales.

Los 79 genes (relacionados a la síntesis de proteínas, la biosíntesis de aminoácidos, la biosíntesis de cofactores, el metabolismo de energía, el metabolismo del ADN, entre otros), fueron determinados de forma automática por el servidor web. Además, se seleccionó manualmente las secuencias genómicas más próximas que pertenecían a "type strain" de especies validadas. Se incluyó la secuencia de *Kitasatospora purpeofusca* NRRL B1817 (GCF000718025) como secuencia fuera de grupo.

Adicionalmente, se recurrió al servidor web TYGS (Type (Strain) Genome Server) (Meier-Kolthoff & Göker, 2019), para realizar análisis taxonómicos a partir de genomas completos. Este servidor cuenta con una base de datos de genomas pertenecientes a "type strain" de especies validadas. La construcción de árboles se ejecuta a partir de todo el genoma, como también a partir de los genes *ARNr 16s* contenidos en el genoma evaluado.

A las secuencias seleccionadas automáticamente por el servidor web, se le adicionó manualmente las secuencias de *Streptomyces flavovariabilis* NRRL B-16367 (GCF\_000725785), *Streptomyces iakyrus* NRRL-ISP 5482 (GCF\_000717055), *Streptomyces azureus* ATCC 14921 (GCA\_001270025) y *Streptomyces pactum* ATCC 27456 (GCF\_01031615) para facilitar la comparación de los árboles filogenéticos generados con los obtenidos por autoMLST y MEGA. Se incluyó la secuencia de *Microbispora rosea* ATCC 12950 (GCA\_900156315) como secuencia fuera de grupo. Los árboles filogenéticos generados los servidores web autoMLST y TYGS fueron visualizados y editados usando el servidor web iTOL (Letunic & Bork, 2021).

### **3.3. ANOTACIÓN DEL GENOMA DE *Streptomyces* sp. B-146**

Además del ensamblaje de novo del genoma completo, el equipo CIICAM realizó la anotación del genoma completo de *Streptomyces* sp. B-146 usando Prokka 1.14.6, un “pipeline” que alberga varios programas encargados de localizar genes de rRNA, genes de tRNA, CDSs y otras características genómicas, de traducir los CDS detectados, de buscar homólogos a proteínas predichas y generar archivos de salida estándar para aplicaciones posteriores (Seemann, 2014).

Como complemento a los análisis genómicos realizados por el equipo de investigadores de CIICAM, hemos realizado la anotación funcional del genoma completo de *Streptomyces* sp. B-146 usando RAST Seedviewer versión 2.0 (Aziz *et al.*, 2008; Overbeek *et al.*, 2014; Brettin *et al.*, 2015), el cual puede predecir subsistemas representados en el genoma en base a las funciones de las características genómicas identificadas y anotadas (genes de ARNt, ARNr y codificantes de proteínas). Finalmente, los archivos de salida de Prokka que contenían las secuencias traducidas de los CDSs fueron usados como archivos de entrada en BlasKoala versión 2.2 (Kanehisa *et al.*, 2016a), para realizar la categorización funcional usando la base de datos KEGG (*Kyoto Encyclopedia of Genes and Genomes*) (Kanehisa *et al.*, 2016b). Para ejecutar los análisis se asignó el grupo taxómico de *Streptomyces* sp. B146 “Prokaryotes, Bacteria, Actinobacteria (Taxonomy ID: 1883)” y seleccionó la base de datos de genes KEGG “Genus\_prokaryotes”. Adicionalmente, hemos generado un mapa interactivo del genoma completo de *Streptomyces* sp. B-146 a partir de los resultados de anotación arrojados por el Prokka a través del servidor web Proksee (<https://proksee.ca/>).

### **3.4. IDENTIFICACIÓN DE BGCs EN EL GENOMA DE *Streptomyces* sp. B-146**

La identificación y anotación de los BGCs contenidos en el genoma completo de *Streptomyces* sp. B-146 se realizó con el servidor web antiSMASH versión 6.1.1 (Blin *et al.*, 2021). Como archivo de entrada se usó los resultados obtenidos por Prokka. Los servidores web PRISM 4.4.5 (Skinnider *et al.*, 2017), ARTS 2.0 (Mungan *et al.*, 2020) y NaPDos (Ziemert *et al.*, 2012) fueron usados como complemento a antiSMASH 6.1.1.

Adicionalmente, se usaron los servidores web BAGEL4 (van Heel *et al.*, 2018) y RiPPMiner-Genome (Agrawal *et al.*, 2021) para identificar y anotar específicamente BGCs tipo RiPPs y compararlos con la data obtenida por antiSMASH 6.1.1. Para contrastar la anotación de los genes contenidos en los BGCs identificados por los servidores web mencionados, se usó la data obtenida de los análisis realizados en Prokka (Seemann, 2014) y BlastKOALA (Kanehisa *et al.*, 2016<sup>a</sup>). Para deslindar dudas o incongruencias en la anotación recurrió a la base de datos del UniProt la cual contiene alrededor de 190 millones secuencias de proteínas anotadas con información funcional (UniProt Consortium, 2023).

### **3.5. IDENTIFICACIÓN DE SISTEMAS CRISPR EN EL GENOMA DE *Streptomyces* sp. B-146.**

La identificación de sistemas CRISPR/Cas se realizó evaluando el genoma completo de *Streptomyces* sp. B-146 recurriendo a los servidores web CRISPRCasFinder (Couvin *et al.*, 2018) y CRISPRminer (Zhang *et al.*, 2018a). La importancia de detectar arreglos CRISPR y genes Cas en nuevos genomas bacterianos radica en la posibilidad de descubrir sistemas CRISPR/Cas novedosos. Producto de ello, ahora disponemos de técnicas moleculares de edición genómica sencillas y altamente eficientes basadas en el mecanismo molecular específico de sistemas CRISPR/Cas. Dichas técnicas actualmente son indispensables en la manipulación genética de microorganismos, especialmente de aquellos con capacidad biosintetizadora de productos naturales como los *Streptomyces* y otras *Actinobacterias* (Tong *et al.*, 2019).

#### IV. RESULTADOS Y DISCUSIÓN

##### 4.1. DESCRIPCIÓN DE LAS MÉTRICAS DE LA SECUENCIA DEL GENOMA DE *Streptomyces* sp. B-146

La secuencia final del genoma de *Streptomyces* sp. B-146, proporcionado por CIICAM, consta de un “*scaffold*” de 8,759,250 pb, con un contenido de G+C de 72.31% (Figura 01). No se detectó la presencia de plásmidos. La integridad del genoma y la contaminación se estimó en 100.00% y 0,00%, respectivamente. Valores adicionales de integridad y estadísticas de longitud y composición se muestran en la Cuadro 01.

**Cuadro 01:** Métricas de la secuencia final del genoma de *Streptomyces* sp. B146 generados por gVolante, CheckM y QUAST.

Características	Valores
<b>Evaluación de integridad</b>	
“ <i>core genes</i> ” consultados	1,579
“ <i>core genes</i> ” detectados	1,576 (completos) + 2 (parciales) (99.94%)
“ <i>core genes</i> ” no detectados	1 (0.06%)
Promedio de ortólogos por “ <i>core genes</i> ”	1.00
“ <i>core genes</i> ” con más de 1 ortólogo	0.32%
Integridad	100.00%
Contaminación	0.00%
<b>Estadística de longitud y composición</b>	
Secuencias detectadas	1
Longitud total (nt)	8,759,250
Longitud de secuencias N <sub>50</sub> (nt)	8,759,250
Longitud de secuencias N <sub>75</sub> (nt)	8,759,250
Conteo de secuencias L <sub>50</sub>	1
Conteo de secuencias L <sub>75</sub>	1
Composición de bases (A)	13.82 %
Composición de bases (T)	13.87 %
Composición de bases (G)	36.09%
Composición de bases (C)	36.21 %

«Continuación»

Composición de bases (Otros)	0.00 %
Contenido de G+C	72.31%
Número de “gaps” ( $\geq 5$ N's)	0
Número de non-ACGTN (nt)	0
Número de “Ns” por 100 kpb	0

FUENTE: Elaboración propia con datos proporcionados por el departamento de Biotecnología de CIICAM.

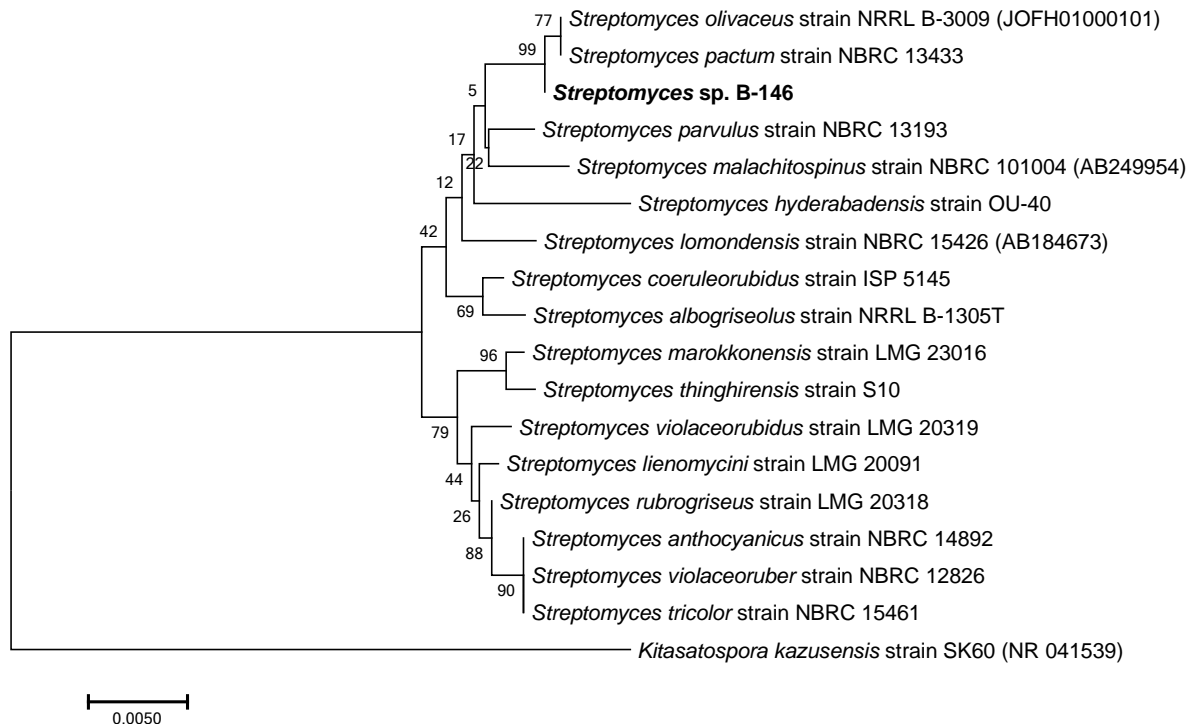
## 4.2. RESULTADOS DE LOS ANÁLISIS FILOGENÉTICOS Y FILOGENÓMICOS DE *Streptomyces* sp. B-146

El gen *ARNr 16S* es ampliamente usado en filogenética bacteriana, en la delineación de especies y en estudios de microbiomas (Hassler *et al.*, 2022). En consecuencia, elaboramos un árbol filogenético a partir de secuencias alineadas de 1,283 pb derivadas de la secuencia parcial de gen *ARNr 16S* de *Streptomyces* sp. B-146 (GenBank: MW562805), junto con otras 18 secuencias filogenéticamente cercanas y una secuencia filogenéticamente distante, recuperadas de EZBiocloud. Los resultados del árbol filogenético revelaron que *Streptomyces* sp. B-146 es próxima a *S. olivaceus* y *S. pactum* (Figura 01).

EZBiocloud es una base de datos integrada que está representada por una cantidad relativamente pequeña de secuencias genómicas y secuencias de genes *ARNr 16S* de alta calidad para la identificación y clasificación taxonómica de Bacterias y Arqueas nivel de especie, subespecie, género y/o filotipo (Yoon *et al.*, 2017). Resultando ser la base de datos de referencia y más precisa para la identificación taxonómica tanto a nivel de género como de especie, comparada con las bases de datos de Greengenes y Silva (Park & Won, 2018).

Es importante resaltar que es preferible realizar análisis filogenéticos a partir de genes de *ARNr 16S* de secuencia completa para mejorar la resolución taxonómica y filogenética (Zhang *et al.*, 2022). No obstante, si el objetivo es lograr filogenias precisas a nivel de especie, lo que en sí mismo es un gran desafío, no es conveniente basarse únicamente en el ADNr 16S debido a que desafortunadamente adolece de baja resolución a nivel de especie a causa de la alta conservación de sus secuencias (Alanjary *et al.*, 2019). Por lo mencionado, adicionalmente se recurrió a los servidores web TYGS y autoMLST para realizar análisis de filogenia complementarios de mayor precisión y también análisis filogenómicos, a partir del genoma completo de *Streptomyces* sp. B-146.





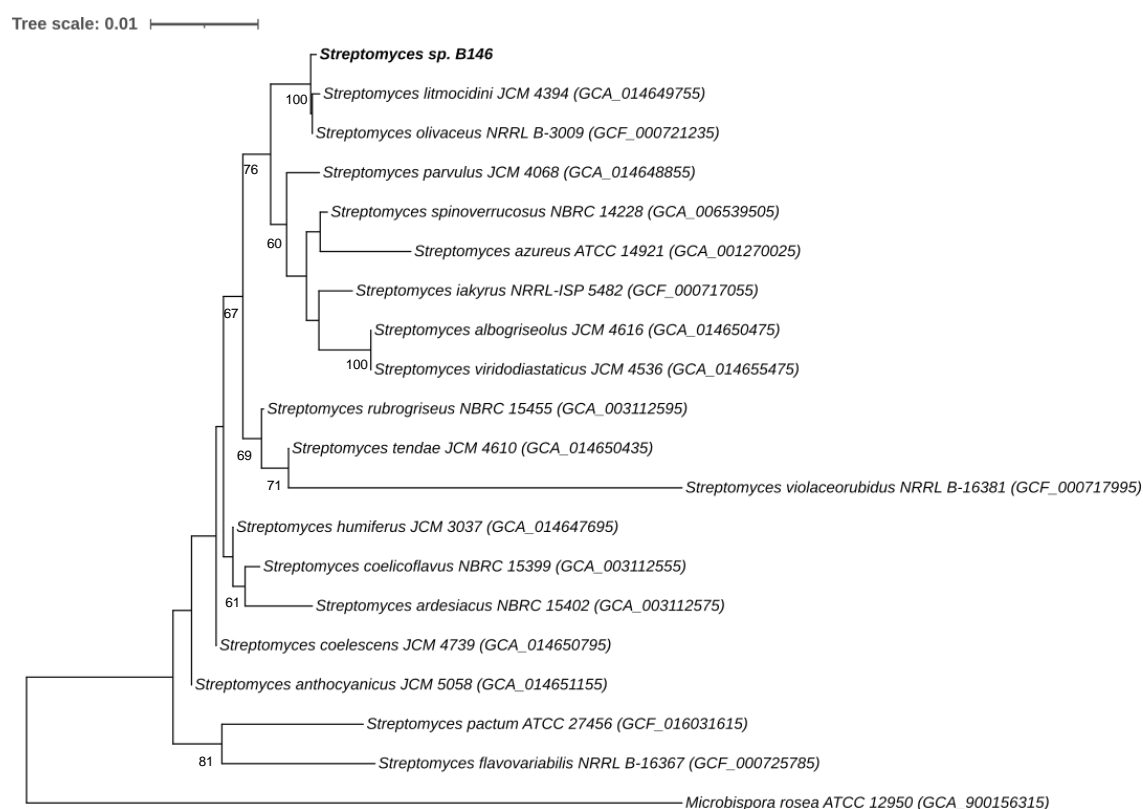
**Figura 01:** Construcción del árbol filogenético a partir de la secuencia parcial del gen ARNr 16S (MW562805, NCBI) de *Streptomyces* sp. B-146, mediante el método “Neighbor-Joining” usando el software Mega 11.

El servidor web TYGS es una herramienta de alto rendimiento para la taxonomía de procariotas basada en genomas, que conecta con una gran base de datos de información genómica, taxonómica y de nomenclatura en constante crecimiento; capaz de inferir filogenias a escala genómica a partir de secuencias genómicas de “Type strain” próximas y cuyos resultados son accesibles en su propia página web (Meier-Kolthoff & Göker, 2019). Esta plataforma arroja resultados que incluyen dos árboles filogenéticos, los cuales son contruidos a partir de secuencias de *ADNr 16S* y de genomas completos, respectivamente.

Para la construcción del árbol filogenético basado en genes *ARNr 16S*, se calcula las distancias evolutivas entre las secuencias completas de los genes *ARNr 16S* extraídos del genoma de consulta y una lista de secuencias conformada de preferencia por genes *ARNr 16S* extraídos de genomas secuenciados de las “Type strain” más próximas; sin embargo, si no son lo suficientemente próximas, se opta por añadir a la lista secuencias completas y validadas de genes *ARNr 16S* de “Type strain” próximas que carecen de secuencias genómicas publicadas (Meier-Kolthoff & Göker, 2019).

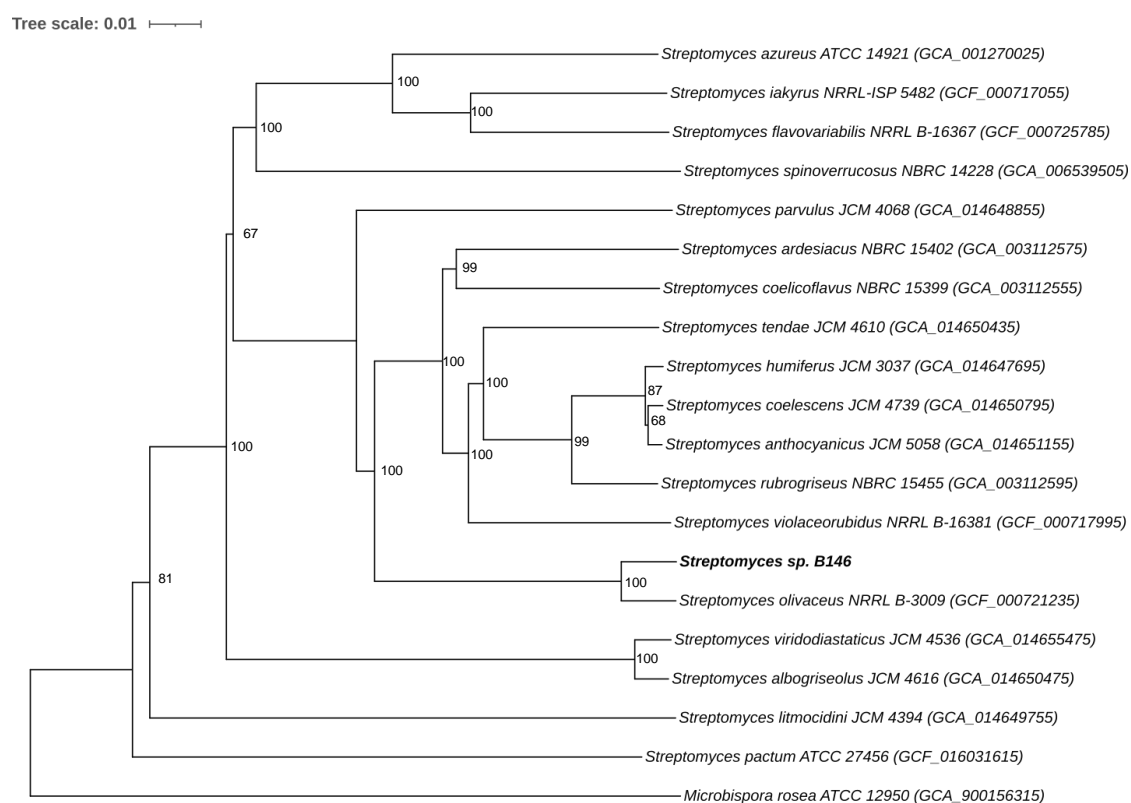
Los resultados del árbol filogenético demuestran que *Streptomyces* sp. B-146 es próxima a *S. olivaceus* y *S. litmocidini* (Figura 02). No obstante, dicha deducción puede ser cuestionada debido a que el número de copias de genes *ARNr 16S* entre los genomas es variable y sufren de heterogeneidad intragenómica; por lo que existe el riesgo de confundir las métricas de diversidad y, por ende, las inferencias filogenéticas podrían ser poco fiables (Hassler *et al.*, 2022). Por ello, recurrimos a la construcción del árbol filogenético a partir de la secuencia completa del genoma de *Streptomyces* sp. B-146, llevada a cabo por el servidor web TYGS bajo un enfoque filogenómico.

Para su construcción se calcula las distancias evolutivas entre el genoma de consulta y los genomas completos de las “*Type strain*” próximas disponibles en la base de datos de TYGS (Meier-Kolthoff & Göker, 2019). Los resultados demuestran nuevamente que *Streptomyces* sp. B-146 es próxima a *S. oliváceos* (Figura 03). Dicha interpretación resulta ser más fiable y precisa.



**Figura 02:** Árbol filogenético basado en la secuencia completa del gen ARNr 16S del genoma de *Streptomyces* sp. B-146 a través del servidor web TYGS.

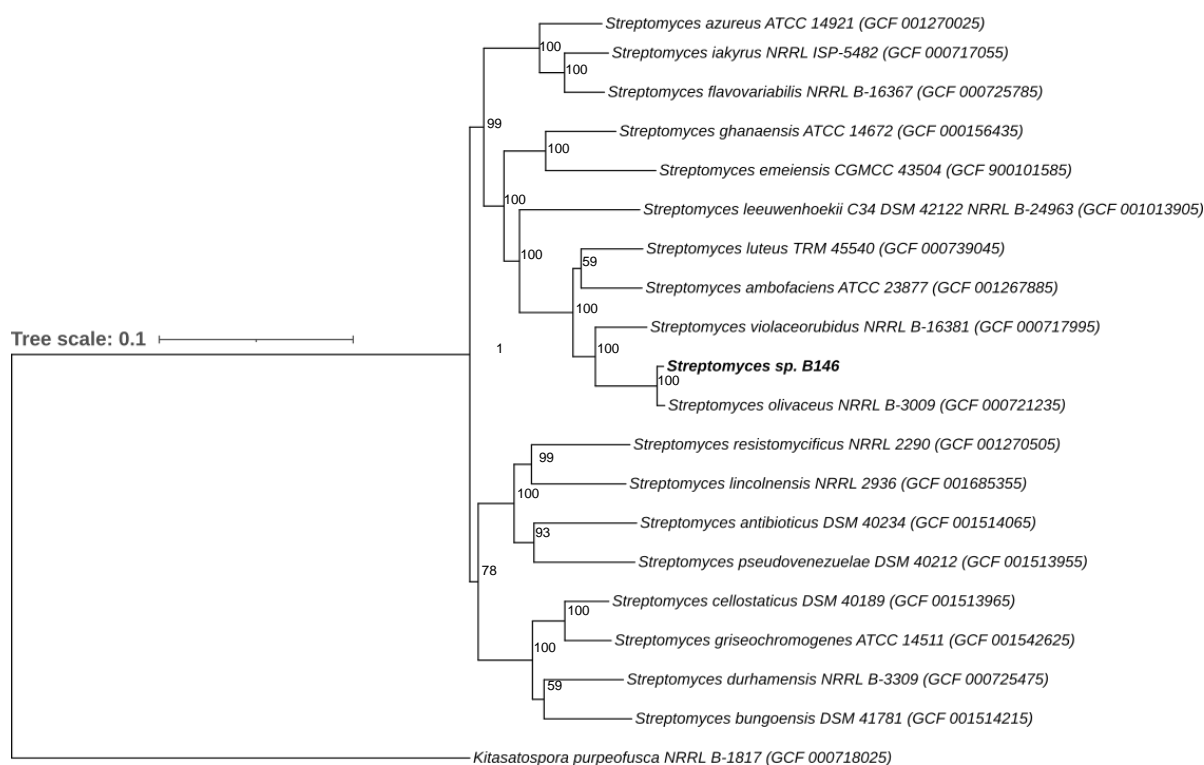
Cabe recalcar que la cercanía filogenética entre *Streptomyces* sp. B-146 y *S. pactum* inferida por análisis de filogenia en MEGA 11, no se pudo corroborar usando el servidor web TYGS debido que su base de datos no contaba con genomas de “Type strain” de *S. pactum*. También, cabe precisar que los cálculos de alta precisión de las distancias evolutivas utilizadas para la inferencia de ambos árboles filogenéticos se obtuvieron a través del método GBDP (*Genome BLAST Distance Phylogeny*), el cual fue desarrollado para determinar filogenias en procariotas utilizando la comparación de genomas completos (Henz *et al.*, 2005).



**Figura 03:** Árbol filogenético basados en la secuencia completa del genoma de *Streptomyces* sp. B-146 a través del servidor web TYGS.

Adicionalmente, para tener una visión más amplia de la identificación taxonómica de *Streptomyces* sp. B-146 se realizaron análisis de filogenia a través del servidor web autoMLST; el cual utiliza un enfoque de análisis de múltiples secuencias de genes de diferentes loci (MLSA, *Multi-Locus Sequence Analysis*) para producir árboles a nivel de especie de alta resolución que incluyen anotaciones útiles, como designaciones de clados de especies y recuentos de metabolitos secundarios para ayudar a la prospección de productos naturales (Alanjary *et al.*, 2019). Con este enfoque nos planteamos potenciar los análisis filogenéticos previamente obtenidos.

Los resultados demuestran una vez más que *Streptomyces* sp. B-146 es filogenéticamente próxima a *S. olivaceus* (Figura 04). No obstante, al igual que los análisis filogenómicos llevados a cabo en TYGS, no evaluó la cercanía filogenómica con *S. pactum*, debido a que su base de datos tampoco contiene una secuencia de referencia “Type strain”.



**Figura 04:** Árbol filogenético obtenido de los análisis filogenómicos basados en la secuencia completa del genoma de *Streptomyces* sp. B-146 a través del servidor web autoMLST.

Considerando todos los resultados se puede asumir que *Streptomyces* sp. B-146 podría convertirse en una especie nueva dentro del género *Streptomyces* y próxima a *S. oliváceos*. Además, resulta evidente que la identificación taxonómica de cepas bacterianas es difícil y compleja. Por ello, para realizar filogenias más precisas que contribuyan a la delineación filogenética de nuevas especies se recomienda utilizar secuencias de genomas completos o múltiples secuencias de genes, incluyendo, además del gen *ARNr 16S*, otros genes ARN ribosomales (ARNr 5S o ARNr 23S), uno o más genes codificantes de proteínas altamente conservados y/o genes característicos presentes exclusivamente entre las cepas de una categoría taxómica (Hassler *et al.*, 2022).

Es importante resaltar que autoMLST genera resultados adicionales relacionados con el recuento de BGCs presentes en los genomas de las especies bacterianas filogenéticamente cercanas a *Streptomyces* sp. B-146. Los resultados indican que *S. olivaceus*, la especie bacteriana más próxima a *Streptomyces* sp. B-146, contiene más de 28 BGCs productores de metabolitos secundarios.

Los cuales de mayor a menor abundancia son: NRPSs, PKSs, RiPPs, Terpenos, Sideróforos y Ectoínas. Pero también contiene una cantidad considerable de BGCs híbridos. Todo esto nos da una prospección de los BGCs que podría contener el genoma *Streptomyces* sp. B146. Aunque, debemos recordar que incluso cepas de una misma especie de *Streptomyces* pueden variar ampliamente respecto a la cantidad y variedad de BGCs que albergan (Belknap *et al.*, 2020). Mas adelante se realizarán análisis más profundos para determinar el contenido de BGCs en el genoma de *Streptomyces* sp. B-146.

#### **4.3. RESULTADOS DE LA ANOTACIÓN DEL GENOMA DE *Streptomyces* sp. B-146**

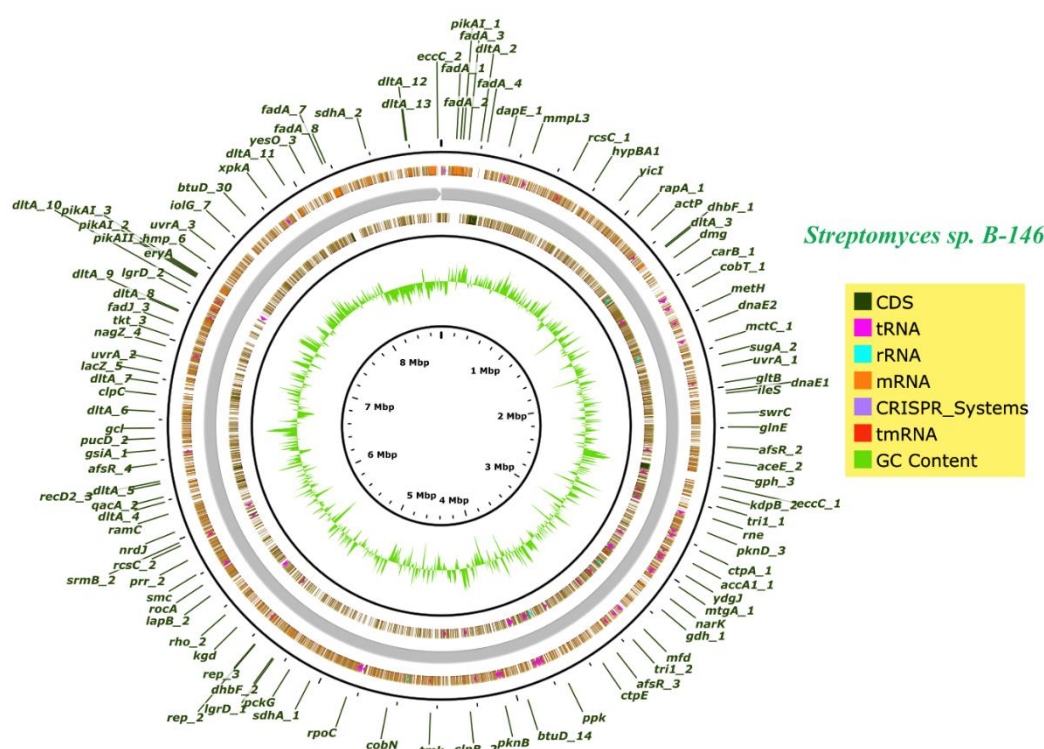
##### **4.3.1. ANOTACIÓN DEL GENOMA DE *Streptomyces* sp. B-146 POR PROKKA**

La anotación del genoma de *Streptomyces* sp. B-146 fue realizada por el equipo científico de CIICAM. Ellos usaron el “*pipeline*” Prokka, una herramienta muy versátil que depende de una base de datos pequeña y bien seleccionada para ejecutar anotaciones rápidas y de alta calidad (Kremer *et al.*, 2016; Ruiz-Perez *et al.*, 2021). Debido a que el genoma de *Streptomyces* sp. B-146 fue ensamblado de *novo*, se diseñó y optimizó un flujo de trabajo único y personalizado para llevar a cabo dicha anotación. Esto en consonancia con la recomendación de llevar a cabo las anotaciones evitando la configuración predeterminada y optar por configuraciones y bases de datos personalizadas para aumentar su rendimiento y obtener resultados más precisos (Kremer *et al.*, 2016); especialmente cuando se trata de genomas nuevos o divergentes (Ruiz-Perez *et al.*, 2021).

Con este enfoque, Prokka reportó un total 7,487 CDSs. El porcentaje de anotación fue de alrededor del 47% (3,519/7,487), de los cuales un 61.5% (2,163/3,519) contenían identificadores COG (*Clusters of Orthologous Genes*) y un 61.2% (2,153/3,519) contenían asignaciones “*EC Numbers*” (*Enzyme Commission numbers*) (Figura 05). A modo de comparación, el porcentaje de anotación promedio obtenido en genomas de *Escherichia coli* usando Prokka es del 75% (Ruiz-Perez *et al.*, 2021).

A primera vista, podría parecer que la anotación realizada en *Streptomyces* sp. B-146 fue deficiente; sin embargo, es importante tener en cuenta que las bases de datos fuente actualmente disponibles no cuentan con un repertorio extenso de proteínas caracterizadas del género *Streptomyces* en comparación con especies bacterianas más conocidas como *E. coli*. Por otro lado, aproximadamente un 53% (3,968/7,487) de los genes codificantes en *Streptomyces* sp. B-146 fueron clasificadas como “Proteínas hipotéticas”.

Respecto a la clasificación “Proteínas hipotéticas”, existe el inconveniente de discernir entre las proteínas que carecen de anotación y las proteínas que si tienen coincidencias con proteínas reportadas como hipotéticas en la base de datos usada; esta situación provoca que el usuario pierda un contexto biológico más amplio para descubrir la funcionalidad de los CDS evaluados (Shaffer *et al.*, 2020; Ruiz-Perez *et al.*, 2021).



**Figura 05:** Visualización del genoma de *Streptomyces* sp. B-146, destacando la anotación de CDSs realizada por Prokka.

#### a. Detección de genes de resistencia a antibióticos por Prokka

Una de las mayores ventajas de Prokka sobre otras herramientas de anotación es su capacidad de detectar módulos adicionales de resistencia a antibióticos (Ruiz-Perez *et al.*, 2021).

Para aprovechar dicha particularidad, se filtraron los CDSs cuyos productos contenían la palabra clave “*resistance*”, obteniéndose un total de 59 CDSs relacionados a Genes de Resistencia a Antibióticos (ARG, *Antibiotic Resistance Genes*). Los genes ARG están ampliamente distribuidos en bacterias productoras de antibióticos, como las *Actinobacterias*, debido a la coevolución entre la biosíntesis de antibióticos y los mecanismos de autoprotección frente a los antibióticos producidos (D'Costa *et al.*, 2007).

Estos genes, con fines de autoprotección, deben expresarse oportunamente, ya sea antes o simultáneamente con la biosíntesis del antibiótico y dicha expresión puede estar estrechamente ligada a la expresión de los genes que biosintetizan los antibióticos o ser inducidos por antibióticos afines o por moléculas intermedias de sus vías biosintéticas (Mak *et al.*, 2014). No obstante, además de desempeñar un papel crucial en la autoprotección, los genes ARG presentes en bacterias ambientales y comensales pueden actuar como un gran reservorio para que las bacterias patógenas los adquieran mediante la transferencia horizontal de genes (von Wintersdorff *et al.*, 2016)

La diseminación de genes ARG a través de la transferencia horizontal de genes, como la aparición de nuevos genes ARG producto de mutaciones genéticas, constituyen el mecanismo más importante por el cual las bacterias adquieren resistencia a los antibióticos (Tao *et al.*, 2022). Dicho mecanismo implica un cambio genético permanente y se le conoce como “resistencia adquirida”; no obstante, existen otros dos mecanismos que favorecen la resistencia a antibióticos, denominados “resistencia intrínseca” y “resistencia adaptativa” (Arzanlou *et al.*, 2017). La resistencia intrínseca ocurre cuando la constitución bioquímica inherente de las bacterias limita la acción de los antibióticos (p. ej., la expresión constitutiva de Bombas de eflujo); mientras que la “resistencia adaptativa” ocurre cuando las bacterias se adaptan y sobreviven al estrés provocado por los antibióticos, alterando rápida y convenientemente sus transcriptomas (McDermott *et al.*, 2003; Arzanlou *et al.*, 2017).

Generalmente, los genes ARG en bacterias se clasifican según su mecanismo de acción, los más comunes son: Reducción de la permeabilidad en la membrana celular; Aumento de las bombas de eflujo de fármacos; Modificación o inactivación enzimática de los antibióticos; Modificación del sitio objetivo para evadir la acción de los antibióticos; Adquisición de vías metabólicas alternas a las inhibidas por el fármaco y Formación de biopelículas (McDermott *et al.*, 2003; Barbosa *et al.*, 2020; Varela *et al.*, 2021).

No obstante, a efectos prácticos, resulta conveniente agruparlos como: Genes ARG codificantes de Bombas de eflujo bacterianas (56/59) y Genes ARG codificantes de enzimas (3/59). A continuación, detallamos los mismos:

#### ❖ **Genes ARG codificantes de Bombas de Eflujo Bacterianas**

Las Bombas de eflujo son proteínas transportadoras que se ubican en la membrana celular bacteriana y el periplasma y funcionan en primera línea para proteger a las bacterias contra antibióticos endógenos y no endógenos disminuyendo su concentración intracelular (Nag & Mehra, 2021; Huang *et al.*, 2022). Es decir, pueden reconocer varios antibióticos, con diferentes estructuras moleculares y clases de acción, que ingresan a las bacterias a través del espacio periplásmico o la membrana célula y los expulsan activamente para evitar que lleguen a sus ubicaciones de destino (Nishino *et al.*, 2021).

Estas bombas de eflujo pueden ser agrupadas principalmente en la superfamilia ABC (*ATP-Binding Cassette*), la superfamilia MFS (*Major Facilitator Superfamily*), la superfamilia MATE (*Multidrug and Toxic Compound Extrusion*), la superfamilia RND (*Resistance Nodulation and Cell Division*), la superfamilia SMR (*Small Multidrug Resistance*) y la superfamilia PACE (*Proteobacterial Antimicrobial Compound Efflux*) (Hassan *et al.*, 2018; Huang *et al.*, 2022).

Este tipo de transportadores se encuentran tanto en bacterias Grampositivas y gramnegativas, así como en células eucariotas (Van Bambeke *et al.*, 2000; Webber & Piddock, 2003). En procariotas, con pocas excepciones, están codificados cromosómicamente y presentan una organización altamente conservada tanto a nivel genético como proteico (Blanco *et al.*, 2016). En *Streptomyces* se ha reportado una amplia gama de bombas de eflujo, probablemente debido a la necesidad de proteger sus células contra antibióticos producidos por sí mismos o por otros microorganismos; principalmente de tipo MFS y ABC que confieren resistencia a múltiples fármacos estructuralmente no relacionados (Nag & Mehra, 2021; 2022; Lei *et al.*, 2023).

#### • **Genes ARG pertenecientes a la superfamilia MFS**

La superfamilia de proteínas MFS son transportadores de membrana que se encuentran de forma ubicua en bacterias, arqueas y eucariotas (Pao *et al.*, 1998). Estos transportadores comparten varias secuencias de aminoácidos altamente conservadas que desempeñan funciones esenciales durante el transporte de compuestos (Kumar *et al.*, 2020).



Asimismo, tienen un amplio espectro de sustratos, abarcando iones orgánicos e inorgánicos, péptidos cortos, lípidos, azúcares y muchos otros compuestos (Pao *et al.*, 1998; Yan, 2015). En bacterias estos transportadores actúan como bombas de eflujo de múltiples fármacos, incluyendo antibióticos y contribuyen decisivamente a la supervivencia, adaptación y patogenicidad bacteriana (Pasqua *et al.*, 2021). A continuación, describimos 31 CDSs relacionados a la superfamilia de proteínas MFS.

– **Genes de resistencia *mdtD***

Se detectaron 10 CDSs relacionados al gen *mdtD* (*YegB*), que conforma un clúster génico denominado *mdtABCD* en *E. coli* y codifica un transportador tipo MFS que junto con otras proteínas del mismo clúster confieren resistencia a múltiples fármacos como a los antibióticos novobiocina y desoxicolato (Baranova & Nikaido, 2002).

– **Genes de resistencia *Stp***

Se detectaron ocho CDSs relacionados a gen *stp* que codifica para una proteína que actúa como una bomba de eflujo en *M. tuberculosis* y contribuye a la resistencia intrínseca a la espectinomicina y la tetraciclina (Ramón-García *et al.*, 2007). Según la base de datos del UniProt, estas proteínas de resistencia a múltiples fármacos pertenecen a la familia de proteínas EmrB de la superfamilia MFS (UniProt Consortium, 2023).

– **Genes de resistencia *abaF***

Cuatro CDSs fueron relacionados al gen *AbaF*, que codifica una proteína que actúa como una bomba de eflujo tipo MFS y confiere resistencia a fosfomicina en *A. baumannii* y posiblemente también este implicado en la expulsión de otras moléculas de defensa que expelen los huéspedes a los que infecta (Sharma *et al.*, 2017).

– **Genes de resistencia *qacA***

Tres CDSs fueron relacionados al gen *QacA* que codifica para un transportador de membrana tipo MFS con 14 segmentos transmembranales y gracias a que interactúa directamente con compuestos estructuralmente diversos ofrece resistencia a una amplia gama de antibióticos tipo lipofílicos, catiónicos, monovalentes o divalentes en *Staphylococcus aureus* (Paulsen *et al.*, 1996; Brown & Skurray, 2001).

– **Genes de resistencia *mdtH***

Dos CDSs estuvieron relacionados al gen *MdtH* (*yceL*) que codifica para una bomba de eflujo de múltiples fármacos tipo MFS y confiere resistencia a tetraciclina y antibióticos de quinolona como norfloxacin, enoxacin, ciprofloxacina, norfloxacina y ofloxacina en *E. coli* (Yu *et al.*, 2020).

– **Genes de resistencia *bmr3***

Se detectaron dos CDSs relacionados al gen *bmr3* que codifica para un transportador de membrana de múltiples fármacos que confiere resistencia a puromicina, tosufloxacina y norfloxacina en *B. subtilis* (Ohki & Murata, 1997). Este transportador denominado como Proteína 3 en la base de datos del UniProt, pertenece a la familia de proteínas EmrB de la superfamilia MFS (UniProt Consortium, 2023).

– **Genes de resistencia *bcR***

Un CDS fue relacionado al gen *bcR* que codifica para un transportador de membrana de múltiples fármacos tipo MFS que confiere resistencia a los antibióticos biciclomicina, tetraciclina, fosfomicina, sulfatiazol y kanamicina en *E. coli* (Bentley *et al.*, 1993; Vedantam *et al.*, 1998; Yamada *et al.*, 2006; Nishino & Yamaguchi, 2001).

– **Genes de resistencia *fsR***

Se detectó un CDS relacionado al gen *fsR* que codifica para una proteína con 12 segmentos transmembranales tipo MFS que funciona como una bomba de eflujo de fármacos y confiere resistencia a antibióticos como la fosmidomicina, tetraciclina y cloranfenicol en *E. coli* y en otras bacterias gramnegativas (Fujisaki *et al.*, 1996; Messiaen *et al.*, 2011; UniProt Consortium, 2023).

• **Genes ARG pertenecientes a la superfamilia ABC**

La superfamilia de proteínas ABC son abundantes en los tres dominios de la vida y están ampliamente descritas en la base de datos Pfam (*Protein families*) (Srikant, 2020). Gran parte de las proteínas ABC son transportadores y se caracterizan por la presencia de dominios NBD (*Nucleotide-Binding Domain*) y dominios TMD (*transmembrane domains*), codificados en polipéptidos independientes o fusionados en un solo polipéptido (Davidson *et al.*, 2008; Srikant, 2020).

Los dominios NBD son responsables de unirse e hidrolizar los ATP para liberar la energía necesaria para mover sus sustratos a través de las membranas (Srikant, 2020).

Los transportadores ABC en bacterias se pueden dividir en importadores y exportadores; los importadores median la absorción de nutrientes esenciales como azúcares, iones metálicos, vitaminas, aminoácidos, entre otros; y los exportadores median la expulsión de toxinas antimicrobianas, lípidos de la membrana externa y fármacos, por lo que están relacionados con la resistencia a antibióticos (Beis, 2015; Tanaka *et al.*, 2018; Srikant, 2020). A continuación, describimos 25 CDSs relacionados a la superfamilia de proteínas ABC.

– **Genes de autorresistencia *drrAB***

Se detectaron 11 y ocho CDSs relacionados a los genes de autorresistencia *drrA* y *drrB*, respectivamente. Ambos genes pertenecen a un operón denominado *drrAB* en *Streptomyces* y codifican proteínas de membrana que funcionan conjuntamente como un transportador ABC multiespecífico gracias a los dominios NBD de *drrA* y a los dominios TMD y a la poliespecificidad de los residuos aromáticos de *drrB*; lo que facilita la salida de múltiples fármacos, como los antibióticos anticancerígenos daunorrubicina / doxorrubicina, con la finalidad de mantener una concentración fisiológica subinhibidora constante dentro de las células productoras (Rao & Kaur, 2008; Srinivasan *et al.*, 2010; Li *et al.*, 2014; Brown *et al.*, 2017).

– **Genes de resistencia *lnrL***

Se detectaron tres CDSs relacionados al gen *lnrL* que pertenece a un operón denominado *lnrLMN*, el cual codifica para un transportador ABC asociado con la resistencia a antibióticos tipo polienos como la linearmicina y anfotericina B en *Bacillus subtilis* (Stubbendieck *et al.*, 2017; Revilla-Guarinos *et al.*, 2020).

– **Genes de resistencia *tmrB***

Se detectó un CDS relacionado al gen *tmrB* que codifica para una proteína ABC ligada a la membrana, la cual confiere resistencia a la tunicamicina en *B. subtilis* (Noda *et al.*, 1992).

– **Genes de resistencia *yheI***

Se detectó un CDS relacionado al gen *yheI* que forma parte de un operón y que codifica para una de las dos subunidades de un heterodímero tipo transportador ABC que funciona como una bomba de eflujo, confiriendo resistencia a múltiples antibióticos, como azitromicina, cloranfenicol, eritromicina, gentamicina, entre otros en *B. subtilis* (Torres *et al.*, 2009).

– **Genes de resistencia *bmrA***

Se detectó un CDS relacionado al gen *bmrA* (*YvcC*) que codifica para un transportador ABC de múltiples fármacos y que confiere resistencia a antibióticos como a la cervimicina C en *B. subtilis* (Steinfels *et al.*, 2004; Krügel *et al.*, 2010).

❖ **Genes ARG codificantes de enzimas bacterianas**

• **Genes de resistencia *arnA***

Dos CDSs fueron relacionados al gen *ArnA* (*Pmrl*), el cual pertenece al operón *pmrHFIJKLM* y codifica una enzima que contiene dos dominios con distintas funciones catalíticas: un dominio deshidrogenasa y un dominio transformilasa y cuya función es modificar la estructura del lípido A con el azúcar pentosa Ara4N (4-amino-4-deoxy-L-arabinose), generando resistencia a péptidos antimicrobianos catiónicos y a antibióticos polimixinas (Gatzeva-Topalova *et al.*, 2004; Fischer *et al.*, 2015).

• **Genes de resistencia *gyrBR***

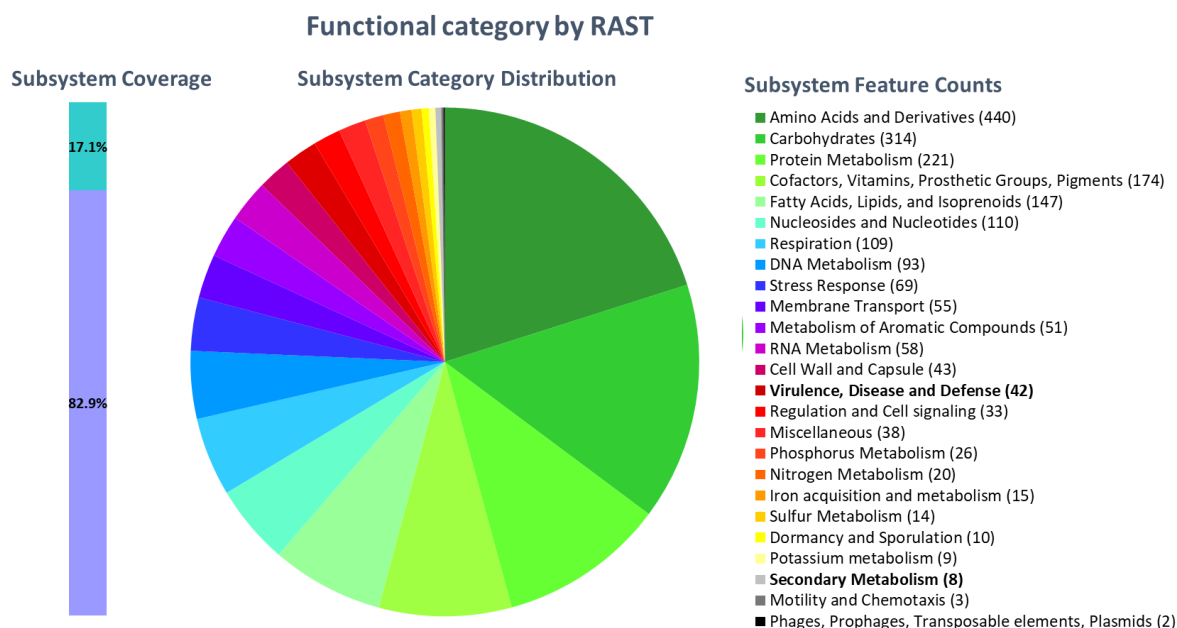
Se detectó un CDSs relacionado al gen *gyrBR* que codifica para una ADN girasa resistente a aminocumarinas (novobiocina, cumermicina A<sub>1</sub> y clorobiocina) y cuya expresión es inducida por la reducción del superenrollamiento del ADN, producto de la inhibición de la ADN girasa sensible (*gyrBS*) de expresión constitutiva en presencia de aminocumarinas, para contrarrestar el efecto negativo de este tipo de antibióticos autoproducidos en *Streptomyces* (Thiara & Cundliffe, 1989; Thiara & Cundliffe, 1993; Schmutz *et al.*, 2003). Este CDS fue también anotado por RAST y clasificado en el subsistema “*Resistance to fluoroquinolones*”, el cual es descrito más adelante.

#### 4.3.2. ANOTACIÓN FUNCIONAL DEL GENOMA DE *Streptomyces* sp. B-146 POR RAST

Análisis realizado por RAST, mostró una visión general de las características biológicas codificadas en el genoma de *Streptomyces* sp. B-146. Esta plataforma reportó un total de 7,885 CDSs. El porcentaje de anotación fue alrededor del 73.70% (5,811/7,885), de los cuales un 23.07% (1,341/5,811) tenían identificadores “*EC numbers*” y un 2.89% (168/5,811) identificadores FIGfam; mientras que un 26.30% (2,074/7,885) no tuvieron coincidencias y fueron clasificadas como “proteínas hipotéticas”. Comparativamente, los resultados fueron ligeramente inferiores a los porcentajes promedio de anotación obtenidos al evaluar 100 genomas de *E. coli*, donde ~79,7% de las proteínas fueron anotadas, de las cuales ~25,8 % tenían identificadores “*EC numbers*” y ~2,1% identificadores FIGfam (Ruiz-Perez *et al.*, 2021).

Respecto a la calidad del genoma, RAST estimó en tres los posibles genes faltantes en regiones intergenicas superiores a 2 kpb. Además, se detectaron automáticamente 2 “advertencias”, referido a posibles problemas de calidad, específicamente a “superposiciones convergentes” superiores a 50 pb entre dos genes orientados de forma contraria en cadenas opuestas. A parte de lo mencionado, no detectó mayores inconvenientes respecto a la calidad del genoma de *Streptomyces* sp. B-146.

Por otro lado, RAST clasificó a 1,351 CDSs anotados, dentro de alguno de los subsistemas reportados, logrando una cobertura del 17.1% (1,351/7,885). Estos subsistemas fueron agrupados y reportados en categorías y subcategorías según su funcionalidad. Cada categoría estuvo representada por la suma de todos los CDSs asignados a cada uno de los subsistemas que la componían. A continuación, mencionamos las categorías más abundantes: Aminoácidos y derivados con 440 CDSs; Carbohidratos con 314 CDSs; Metabolismo de proteínas con 221 CDSs; Cofactores, vitaminas, grupos prostéticos y pigmentos con 174 CDSs; Ácidos grasos, lípidos e isoprenoides con 147 CDSs; Nucleósidos y nucleótidos con 110 CDSs y Respiración con 109 CDSs; además de otras 17 categorías representadas por 582 CDSs (Figura06). Cabe resaltar que algunos CDSs fueron clasificados en más de un subsistema.



**Figura 06:** Categorización funcional del genoma de *Streptomyces* sp. B-149, resaltando la cobertura de los subsistemas y la abundancia de las categorías reportadas por RAST.

#### a. Detección de subsistemas por RAST

##### ❖ Categoría de Virulencia, Enfermedad y Defensa

Esta categoría contiene subsistemas involucrados en la resistencia a antibióticos (Anexo 2):

##### • Subsistema de tolerancia a Colicina E2

A este subsistema se le asignó un CDS relacionado a *CbrC* (*YieJ*) cuya sobreexpresión en *E. coli* K-12 causa tolerancia a la Colicina E2, una endonucleasa de ADN que actúa como un antibiótico proteico desplegado por las propias *E. coli* para eliminar cepas competidoras sensibles (Cariss *et al.*, 2010; Schaller & Nomura, 1976; Cohen-Khait *et al.*, 2021).

##### • Subsistema de resistencia a fluoroquinolonas

Este subsistema fue representado por dos CDSs relacionados a *gyrA* y *gyrB*, los cuales codifican para las subunidades A y B de la ADN-girasa, una topoisomerasa tipo II que confiere resistencia a las fluoroquinolonas en *M. tuberculosis* producto de mutaciones en zonas “hot spot” de los genes que la codifican (Al-Mutairi *et al.*, 2011; Pantel *et al.*, 2012).

Coincidentemente, el CDS relacionado a la subunidad B, también fue anotado por Prokka como *gyrB*. Información adicional del gen *gyrB* fue brindada anteriormente.

- **Subsistema Beta-Lactamasa**

Este subsistema fue representado por dos CDSs relacionados a enzimas beta-lactamasas de clase B y C, respectivamente. Las beta-lactamasas están ampliamente distribuidas en bacterias grampositivas y gramnegativas, median la resistencia a antibióticos  $\beta$ -lactámicos y uno de sus esquemas de clasificación vigente las agrupa en clase A, B, C y D de acuerdo con su secuencia de aminoácidos y motivos conservados (Ambler, 1980; Bush & Jacoby, 2010; Bush, 2013). Las metalo- $\beta$ -lactamasas (MBL) de clase B, dependientes de zinc, son un conjunto diverso de enzimas, miembros de una superfamilia de metalohidrolasas ampliamente distribuidas, que catalizan la hidrólisis de un amplio espectro de sustratos incluyendo prácticamente todos los antibióticos  $\beta$ -lactámicos con la excepción de los monobactámicos (Palzkill, 2013; Tooke *et al.*, 2019). Las  $\beta$ -lactamasas de clase C (*ampC*) están ampliamente distribuidas en Gramnegativas, especialmente en patógenos oportunistas, y pueden actuar y/o hidrolizar las penicilinas, cefalosporinas, cefamicinas, oximiinocefalosporinas y monobactámicos (Jacoby, 2009; Tooke *et al.*, 2019).

- ❖ **Categoría de Metabolismo Secundario**

Esta categoría contiene subsistemas relacionados a la biosíntesis de productos naturales (Anexo 2):

- **Subsistema Lantionina sintetas**

Este subsistema fue representado por tres genes relacionados a las enzimas LanC, LanL y a una O-metiltransferasa; las cuales están implicadas en la biosíntesis de lantipéptidos. Los lantipéptidos son péptidos policíclicos estructuralmente diversos que pueden actuar como antibióticos, antitumorales, antivirales o analgésicos (Zhang *et al.*, 2015; Geng & Smith, 2018). Los cuales son categorizados en cinco clases (I-V), de acuerdo con el tipo de enzimas (lantipéptido sintetetas) que forman sus característicos anillos lantionina; es decir, los lantipéptidos biosintetizados por LanB y LanC pertenecen a la clase I, por LanM a la clase II, por LanKC a la clase III, por LanL a la clase IV y por LxmK, LxmX y LxmY a la clase V (Zhang *et al.*, 2012; Arison *et al.*, 2013; Yu *et al.*, 2013; Xu *et al.*, 2020).

Los genes *LanC* tienen una baja identidad de secuencia entre sí de alrededor del 20% - 30% y codifican para una familia de enzimas ciclasas LanC de aproximadamente 400 aminoácidos; las cuales participan en la formación de enlaces tioéter para lograr la topología de anillo adecuada en los lantipéptidos clase I (Knerr & van der Donk, 2012; Repka *et al.*, 2017).

Los genes *LanL* están muy extendidos en la naturaleza y codifican lantipéptido sintetasa LanL, unas enzimas trifuncionales que constan de tres dominios catalíticos, un dominio pSer/pThr liasa N-terminal, un dominio central Ser/Thr quinasa y un dominio C-terminal de lantionina ciclasa que actúan juntos en la deshidratación y ciclación de los lantipéptidos clase IV (Goto *et al.*, 2010; Goto *et al.*, 2011; Knerr & van der Donk, 2012). Los genes que codifican enzimas O-metiltransferasas están implicadas en la biosíntesis de lantipéptidos; uno de estos genes, denominado *LanS*, codifica para una enzima auxiliar asociada a la biosíntesis de lantipéptidos de clase I en *S. olivaceus* (Acedo *et al.*, 2019).

- **Subsistema de Síntesis de Microcina modificada con Tiazol-oxazol (TOMM)**

Este subsistema estuvo representado por un gen relacionado a la proteína fusión denominada Ciclodeshidratasa (proteína C) / Scaffold de acoplamiento (proteína D). Esta enzima, implicada en la síntesis de microcinas, está ampliamente distribuida en procariotas y consta de un dominio heterociclasa y un dominio de unión a péptidos, por lo que muestra actividad de heterociclación e incluso es capaz de realizar reacciones de ciclodeshidratación dependiente de ATP durante la biosíntesis de TOMMs (Lee *et al.*, 2008; McIntosh & Schmidt, 2010; Dunbar *et al.*, 2012).

Los TOMMs son una clase de RiPPs debido a que son péptidos estructural y funcionalmente diversos producidos en los ribosomas con heterociclos instalados postraduccionalmente derivados de residuos de cisteína, serina y treonina (Melby *et al.*, 2011). Y debido a que se generan a partir de una gran variedad de secuencias peptídicas precursoras, así como de combinaciones desconocidas de enzimas de modificación, representan una fuente potencial de nuevos productos naturales con distintas bioactividades (Cox *et al.*, 2015).



## **b. Detección de genes de resistencia a antibióticos por RAST**

Para identificar CDSs relacionados a genes ARG en los datos de anotación proporcionados RAST, se realizó un filtrado usando la palabra clave “*Resistance*”. Tras la filtración se identificó y seleccionó 10 CDSs relacionados a genes ARG (Anexo 3). Dichos CDS fueron clasificados de la siguiente manera:

### **❖ Proteínas relacionadas a Bombas de eflujo tipo MFS**

Tres CDSs fueron registrados como Bombas de eflujo bacteriano y clasificados como miembros de la superfamilia de transportadores MFS. Los cuales son analizados a continuación:

#### **• Transportador de la Familia Bcr/CflA**

Un CDS (1,272 pb) fue anotado como un “Transportador de la familia Bcr/CflA”. Los miembros de esta familia de transportadores de membrana tipo MFS, como la proteína Bcr en *E. coli*, Flor en *Salmonella typhimurium* DT104 y CmlA en *Pseudomonas*, confieren resistencia a múltiples fármacos tales como el Cloranfenicol, Florfenicol y Biciclovina mediante el transporte activo de estos compuestos fuera de la célula (Smith *et al.*, 2009). Coincidentemente, Prokka anotó este CDS como un gen *Bcr*, el cual fue descrito anteriormente. Adicionalmente, RAST clasificó a este CDS dentro del subsistema “*Copper homeostasis*”, por lo que probablemente también participa en el transporte de iones de cobre.

#### **• Transportador de la Familia CmIV**

Un CDS (1,272 pb) fue anotado como miembro de la “familia de proteínas CmIV”. Esta familia de proteínas ha sido reportada en bacterias del género *Lactobacillus*, donde actúan como bombas de eflujo tipo MFS y confieren resistencia a Cloranfenicol (Moradi *et al.*, 2022). Asimismo, han sido reportadas en *Streptomyces coelicolor* y *Streptomyces venezuelae*, donde han mostrado resistencia a múltiples fármacos como el Cloranfenicol y la Ciprofloxacina (Nag & Mehra, 2021).

#### **• Transportador MFS**

Un CDS (1,533 pb) fue anotado como “Bomba de eflujo resistente a múltiples fármacos”.

Contrastando con los datos proporcionados por Prokka, encontramos que este CDS fue anotado como un gen *LfrA*, cuya proteína actúa como “Bomba de eflujo a múltiples fármacos”. El gen *lfrA* en *Mycobacterium smegmatis* codifica un transportador MFS que actúa como una bomba de eflujo para la salida activa de varias quinolonas, principalmente fluoroquinolonas (Liu *et al.*, 1996; Li *et al.*, 2004). Con UniProt se verificó que el CDS en cuestión guardaba una identidad del 99.2% con un “transportador MFS” (A0A1D8SL40) de *Streptomyces olivaceus* (UniProt Consortium, 2023). Tomando en conjunto las anotaciones realizadas, se concluye que este CDS pertenece a la superfamilia de transportadores MFS y posiblemente confiere resistencia a antibióticos.

#### ❖ **Proteínas relacionadas a Bombas de eflujo tipo ABC**

Tres CDSs fueron registrados como Bombas de eflujo bacteriano y clasificados como miembros de la superfamilia de transportadores ABC. En concreto, un CDS (1,062 pb) fue anotado como “Proteína ATP-Binding”; mientras que otros dos CDSs de 825 y 849 pb, respectivamente, fueron anotados como “Proteínas Permeasas”, las cuales se caracterizan por contener dominios TMD (Wen *et al.*, 2011). Dicha anotación coincide en términos generales con los datos brindados por Prokka (Ver Anexo 1 y Anexo 3).

Específicamente, según Prokka el CDS anotado como “Proteína ATP-Binding” estaría relacionado al polipéptido *drdA* que contiene dominios NBD; mientras que los CDSs anotados como “Proteínas Permeasas” estarían relacionados al polipéptido *drdB* que alberga dominios TMD. Ambos polipéptidos conformarían un transportador transmembranal (información detallada anteriormente). En contraste, RAST relacionó al CDS anotado como “Proteína ATP-Binding” y al CDS de 825 pb anotado como “Proteína Permeasa” con el transportador transmembranal *OtrC* tipo ABC de *Streptomyces rimosus*, que confiere resistencia y autoprotección celular a múltiples fármacos (Yu *et al.*, 2012b). En definitiva, tomando en conjunto los datos, podemos inferir que estos tres CDSs codificarían para dos subunidades de un complejo proteico que actúa como una Bomba de eflujo y que confiere resistencia a Oxitetraciclina, Daunorrubicina, Doxorrubicina, entre otros fármacos.

#### ❖ **Proteínas relacionadas a Bombas de eflujo SMR**

Dos CDSs (321 y 324 pb) fueron anotados como Bombas de eflujo y clasificados como miembros de la superfamilia de transportadores SMR.

Los transportadores SMR se componen de pequeñas subunidades de 100-140 aminoácidos, se caracterizan por albergar cuatro alfa hélices transmembranales y se ensamblan en homodímeros o heterodímeros para funcionar como Bombas de eflujo bacterianos (Bay *et al.*, 2008; Burata *et al.*, 2022).

En general, transportan solutos con carga positiva, generalmente antisépticos y biocidas ambientales, a través de las membranas bacterianas (Kermani *et al.*, 2020; Burata *et al.*, 2022). No obstante, algunos de estos transportadores también confieren resistencia a una amplia gama de antibióticos (Srinivasan *et al.*, 2009). Además, se ha demostrado que los transportadores SMR amplifican la resistencia a antibióticos debido a que muchos de sus genes están vinculados físicamente a genes ARG (Pal *et al.*, 2015; Wales & Davies, 2015; Kermani *et al.*, 2020).

En contraste, Prokka anotó a los CDSs en cuestión como genes *gdx*, codificantes de proteínas “Exportadoras de Guanidinio”. Los genes *gdx*, anteriormente conocidos como genes *SugE*, tienen como sustrato principal a compuestos de Guanidinio y de Amonios Cuaternarios (Slipski *et al.*, 2020). No obstante, también confieren resistencia a antibióticos como la tetraciclina y el cloranfenicol (Cruz *et al.*, 2013). Los análisis en UniProt mostraron que el CDS de 321 pb tenía una identidad del 95.3% con un “Transportador SugE” (A0A7L4Y067) de *Streptomyces* sp. QHH-9511; mientras que el CDS de 324 pb tenía una identidad del 94.4% con una “Proteína sugE” (A0A1M5XEF3) de *Streptomyces* sp. 3214.6 (UniProt Consortium, 2023). Por lo tanto, ambos CDSs estarían relacionados a transportadores SMR que confieren resistencia a un amplio rango de compuestos antibióticos y no antibióticos.

#### ❖ **Proteínas relacionadas a Bombas de eflujo tipo RND**

Un CDS (3,231 pb) fue anotado como un “Transportador de eflujo de múltiples fármacos tipo RND, resistente a la Acriflavina”. Los transportadores RND son un componente proteico de los complejos de eflujo tripartitos en bacterias Gramnegativas (Delmar *et al.*, 2014). Es decir, los transportadores RND, ubicados en la membrana interna, deben interactuar con una proteína adaptadora periplásmica y un canal de membrana externa para conformar un complejo tripartito que abarca tanto la membrana interna, el periplasma y la membrana externa (Li *et al.*, 2015).

Los transportadores de eflujo RND se caracterizan por exportar fármacos y cationes tóxicos, por lo que usualmente confieren resistencia a múltiples antibióticos (Wieczorek *et al.*, 2008; Li *et al.*, 2015; Plé *et al.*, 2022). Y aunque se encuentran principalmente en bacterias Gramnegativas, también han sido reportados en bacterias Grampositivas (Chalut, 2016).

En contraste, Prokka anotó este CDS como un gen *SwrC*; que codifica para una proteína necesaria en la “motilidad de enjambre” en *Bacillus subtilis* (Kearns *et al.*, 2004). No obstante, estas proteínas también funcionan como transportadores de membrana en dichas bacterias (Shi *et al.*, 2014). Para esclarecer la anotación realizada por Prokka, se realizó una búsqueda en UniProt usando las palabras “Swarming motility protein SwrC”. Con los resultados obtenidos se pudo verificar que las proteínas SwrC (O31501) de *B. subtilis* pertenecen a la familia de proteínas RND y confieren resistencia a la Acriflavina y autorresistencia a la surfactina, un antibiótico lipopéptido surfactante (UniProt Consortium, 2023).

Asimismo, se recurrió a UniProt para realizar una anotación directa del CDS en cuestión. Los resultados indicaron que dicho CDS tenía una identidad del 94.6% con un transportador de la familia MMPL (A0A7K3GFC4) de *Streptomyces* sp. SID5789 (UniProt Consortium, 2023). Los transportadores MMPL son una subclase de transportadores RND involucrados en la exportación de compuestos lipídicos, la secreción de sideróforos y la resistencia a una amplia variedad de antibióticos como la Clofazimina y bedaquilina (Chalut, 2016; Ma *et al.*, 2020). Por ende, podemos concluir que este CDS codifica para un transportador RND en *Streptomyces* sp. B-146 y podría estar involucrado en la secreción de sideróforos y la autorresistencia de péptidos antimicrobianos tipo RiPPs.

#### ❖ **Proteínas de protección ribosomal contra antibióticos**

Un CDS de 228 pb fue anotado como “Proteína de protección ribosomal resistente a la tetraciclina tipo Tet”. Las Proteínas de Protección Ribosomal (*RPP*, *Ribosome protection Protein*), son un grupo de proteínas que protegen a los ribosomas bacterianos de los antibióticos y actualmente se clasifican en: Proteínas tipo Tet, que median la resistencia a tetraciclinas; Proteínas tipo Fus, que median la resistencia al Ácido Fusídico; y Proteínas ABC-F, que median la resistencia a diversos antibióticos, incluidos macrólidos, fenicoles, entre otros (Ero *et al.*, 2021).

Las RPP tipo Tet se encuentran ampliamente distribuidas en bacterias grampositivas y gramnegativas, incluyendo a especies del género *Streptomyces* (Connell *et al.*, 2003; Doyle *et al.*, 1991; Dittrich & Schrempf, 1992). Se han reportado al menos 11 proteínas Tet; las cuales han sido agrupadas en: Grupo-1 (TetM, TetO, TetS, TetW, Tet32 y Tet36); Grupo-2 (TetBP, OtrA, y Tet) y Grupo-3 (TetQ y TetT); siendo las proteínas TetO y TetM las más conocidas (Connell *et al.*, 2003; Thaker *et al.*, 2010). A nivel molecular, las proteínas tipo Tet interaccionan de manera reiterativa con los ribosomas, sin introducir cambios permanentes, para rescatar el aparato de traducción de la inhibición por tetraciclinas y permitir que los tRNAs puedan unirse al sitio A del complejo ribosomal para que la síntesis de proteínas funcione correctamente (Connell *et al.*, 2003; Ero *et al.*, 2021). Las tetraciclinas son un grupo de antibióticos de amplio espectro que constan de una estructura central de cuatro anillos, inicialmente descubiertas en *Streptomyces aureofaciens* y *Streptomyces rimosus* (Nelson & Levy, 2011).

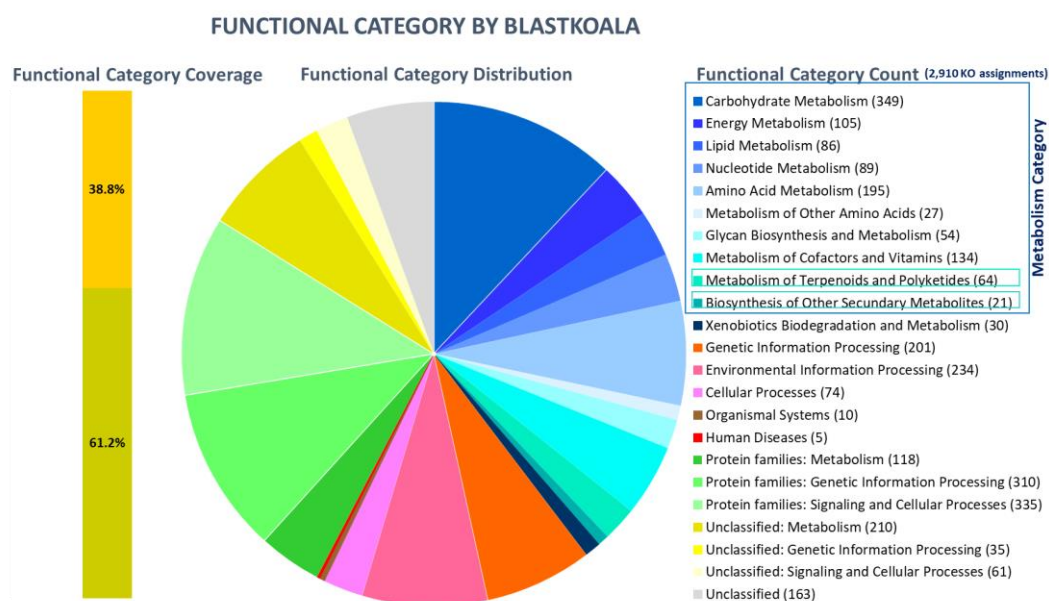
Para contrastar la información obtenida, se recurrió a UniProt para realizar la anotación directa del CDS en cuestión. Los resultados mostraron que el CDS tenía un motivo de 19 aminoácidos con 100% de identidad con una región del extremo C-terminal de la “Proteína OtrA” (A0A7K2JM59) de *Streptomyces* sp. SID5910 (UniProt Consortium, 2023). Por tanto, podemos concluir que este CDS podría estar relacionado con el grupo de proteínas RPP tipo Tet y, en consecuencia, podría conferir resistencia a antibióticos del grupo de las tetraciclinas autoproducidos por *Streptomyces* sp. B-146.

#### **4.3.3. CATEGORIZACIÓN FUNCIONAL DEL GENOMA DE *Streptomyces* sp. B-146 POR BlastKOALA**

El servidor web BlastKOALA realiza simultáneamente anotaciones automáticas y una caracterización funcional de las secuencias de aminoácidos para inferir su función biológica mediante la asignación de identificadores KO (*KEGG Orthology*); para lo cual recurre a un conjunto de datos de pangenoma no redundante generado a partir de las bases de datos contenidas en KEGG (Kanehisa *et al.*, 2016<sup>a</sup>; Kanehisa, 2017). KEGG es un recurso bioinformático de base de datos integradas para la interpretación biológica de secuencias genómicas, especialmente para descubrir funciones moleculares de genes y proteínas (Kanehisa *et al.*, 2014; Kanehisa *et al.*, 2016<sup>b</sup>). Los identificadores KO (también llamados “*K numbers*”), están asociados con referencias y datos de secuencia de proteínas caracterizadas experimentalmente (Kanehisa, 2017).

BlastKOALA anotó un total de 2,902 secuencias de aminoácidos de las 7,487 identificadas en el genoma de *Streptomyces* sp. B-146, representando una cobertura de 38.8% (Figura 07). A las 2,902 secuencias de aminoácidos anotadas se le asignó 1,837 “*K Numbers*”, lo que significa que varios “*K numbers*” fueron asignados a más de una secuencia de aminoácidos. Además, a ocho secuencias de aminoácidos se les asignó doble “*K numbers*”, por lo que BlastKOALA contabilizó un total de 2,910 asignaciones de identificadores KO (Figura 07).

Respecto a la categorización funcional, BlastKOALA determinó que el 39.7% (1,154 / 2,910) de todas las asignaciones de identificadores KO pertenecen a categorías funcionales relacionadas con el metabolismo. Incluyendo 64 asignaciones a la categoría funcional “Metabolismo de Terpenoides y Poliquétidos” y 21 asignaciones a la categoría funcional “Biosíntesis de otros metabolitos secundarios” (Figura 07). Ambas categorías vinculadas con la biosíntesis de diversos productos naturales.



**Figura 07:** Anotación y categorización funcional por BlastKOALA de las secuencias de proteínas detectadas en el genoma de *Streptomyces* sp. B-146.

Es importante destacar la detección de proteínas implicadas en el metabolismo de terpenoides y policétidos en el genoma de *Streptomyces* sp. B-146, que son dos clases distintas de productos naturales ampliamente diversas. Aunque la mayoría de los terpenoides conocidos son biosintetizados por plantas, cada vez más se descubren en bacterias, especialmente en especies del género *Streptomyces* (Martín-Sánchez *et al.*, 2019).

En contraste, los policétidos son biosintetizados por una amplia gama de organismos, incluyendo bacterias, hongos, plantas, protistas, insectos, moluscos y esponjas, siendo las bacterias *Streptomyces* una fuente especialmente prolífica de esta clase de compuestos naturales (Risidian *et al.*, 2019).

#### **a. Reconstrucción Vías KEGG, Módulos KEEG y Jerarquías BRITE**

BlastKOALA hace posible reconstruir Vías KEGG, Módulos KEEG y Jerarquías BRITE (*Biomolecular Relations in Information Transmission and Expression*) mediante la asignación de “*K numbers*”, lo cual permite inferir funciones de alto nivel codificado en el genoma de un organismo (Kanehisa *et al.*, 2016<sup>a</sup>). Para lograrlo, BlastKOALA recurre a las bases de datos “*KEGG PATHWAY*”, “*KEEG MODULE*”, y “*KEGG BRITE*”, respectivamente.

La base de datos “*KEGG PATHWAY*” contiene diagramas de rutas bioquímicas debidamente caracterizadas, conocidas como Vías KEGG; incluyendo a la mayoría de las vías metabólicas y vías regulatorias, lo cual permite comprender mejor la interacción molecular de genes y proteínas (Ogata *et al.*, 1999; Kanehisa, 2017). Paralelamente, la base de datos “*KEGG MODULE*” contiene los denominados Módulos KEEG, que son unidades funcionales modulares de las Vías KEGG definidas por un conjunto de “*K numbers*” (Kanehisa, 2017). Mientras que la base de datos “*KEEG BRITE*” constituye una colección de sistemas jerárquicos que categorizan genes y proteínas en base a sus funciones y relaciones biológicas en Categorías Funcionales, denominadas Jerarquías BRITE; para identificar cómo se relacionan y contribuyen a diferentes procesos biológicos y rutas metabólicas (Kanehisa *et al.*, 2002; Kanehisa, 2017). A continuación, se describen los principales hallazgos reportados por BlastKOALA en el genoma de *Streptomyces* sp. B-146 con relación a las bases de datos mencionadas.

#### **❖ Biosíntesis de Policétidos**

BlastKOALA detectó proteínas de *Streptomyces* sp. B-146 involucradas en la biosíntesis de policétidos. Los análisis realizados en la base de datos KEEG BRITE reportó la jerarquía funcional “Proteínas de biosíntesis de Policétidos” gracias a la asignación de 26 “*K numbers*”. Asimismo, la base de datos KEEG PATHWAY reportó rutas metabólicas implicadas en la biosíntesis de los policétidos tipo antibióticos Ansamicinas, Vancomicinas, Tetraciclinas y Enediynas.

No obstante, la base de datos KEEG MODULE solo llegó a detectar módulos incompletos de vías metabólicas implicadas en la biosíntesis de Tetraciclinas/Oxitetraciclinas y de Clortetraciclinas.

Las Tetraciclinas, Oxitetraciclinas y Clortetraciclinas, denominados globalmente como “Tetraciclinas”, son antibióticos que exhiben actividad contra una amplia gama de microorganismos, incluidas bacterias grampositivas y gramnegativas, clamidias, micoplasmas, rickettsias y parásitos protozoarios (Chopra & Roberts, 2001). Las tetraciclinas inicialmente fueron descubiertas en especies del género *Streptomyces* (Nelson & Levy, 2011). Por ello, es altamente probable que *Streptomyces* sp. B-146 sea capaz de biosintetizar esa clase de antibióticos.

Además, es preciso recordar que la anotación realizada por RAST, detectó un CDS probablemente implicado en la autorresistencia a las Tetraciclinas. Por tanto, se refuerza la probabilidad de que *Streptomyces* sp. B-146 tenga el repertorio génico necesario para biosintetizar Tetraciclinas o moléculas similares.

#### ❖ Biosíntesis de Terpenoides

Los análisis mediante la base de datos KEEG PATHWAY reportó rutas metabólicas implicadas en la biosíntesis de Terpenoides como los Sesquiterpenoides, Triterpenoides y Carotenoides. Además, del Isopreno, unidad básica de todos los Terpenoides. No obstante, la base de datos KEEG MODULE llegó a detectar módulos completos e incompletos de vías metabólicas implicadas en la biosíntesis de Terpenoides de 5 carbonos (Isopreno) y de Terpenoides de 10-20 carbonos como los Sesquiterpenoides.

Los Terpenoides pueden ser biosintetizados por plantas, hongos y bacterias a través de las vías metabólicas del Mevalonato y No-Mevalonato (Mahizan *et al.*, 2019; Rudolf *et al.*, 2020). Según BlastKOALA, ambas vías metabólicas estarían presentes en la bacteria *Streptomyces* sp. B-146. No obstante, solo la vía metabólica del Mevalonato ha sido ampliamente descrita en especies del género *Streptomyces* (Kuzuyama, 2017).

En general, solo los Terpenoides de origen vegetal han sido ampliamente caracterizados y han demostrado bioactividades anticancerígenas, antimicrobianas, antiinflamatorias, antioxidantes y antialérgicas (Rudolf *et al.*, 2020; Masyita *et al.*, 2022). En contraste, los Terpenoides producidos por bacterias recién están siendo identificados, especialmente en bacterias grampositivas como *Streptomyces* y otros *Actinomycetes* (Cane & Ikeda, 2012).



No obstante, sean descrito algunos Terpenoides en *Streptomyces* como la Geosmina que participa en esporulación y la diseminación de esporas (Becher *et al.*, 2020); y la Albaflavenona que exhibe propiedades antimicrobianas (Gürtler *et al.*, 1994).

#### ❖ Biosíntesis de antimicrobianos

La base de datos KEEG PATHWAY reportó rutas metabólicas implicadas en la biosíntesis de varios metabolitos secundarios que actúan como agentes antimicrobianos. Entre otros la Validamicina, Estreptomina, Cefalosporina, Novobiocina, Neomicina, Gentamicina, Kanamicina, Carbapenems y Monobactams. Sin embargo, la base de datos KEEG MODULE, con respecto a lo anteriormente mencionado, solo llegó a detectar un módulo incompleto de la vía metabólica implicada en la biosíntesis de Validamicina A. Adicionalmente, llegó a detectar módulos incompletos de vías metabólicas involucrados en la biosíntesis de los antibióticos Cremeomicina y Ciclooctatina.

#### ❖ Resistencia a antimicrobianos

La base de datos KEEG BRITE reportó la jerarquía funcional “genes de resistencia a antimicrobianos” gracias a la asignación de 16 “*K numbers*”. Asimismo, la base de datos KEEG PATHWAY reportó rutas metabólicas implicadas en la resistencia a betalactámicos, Vancomicina y péptidos antimicrobianos catiónicos. Consistentemente, la base de datos KEEG MODULE llegó a detectar un módulo incompleto de la vía metabólica implicada en la resistencia a los betalactámicos. Además, reportó tres módulos incompletos de rutas metabólicas que confieren resistencia a múltiples fármacos que implica la actuación de bombas de eflujo. Por tanto, estos hallazgos son consistentes con los datos reportados por Prokka y RAST respecto a los genes de resistencia a antibióticos codificados en el genoma de *Streptomyces* sp. B-146.

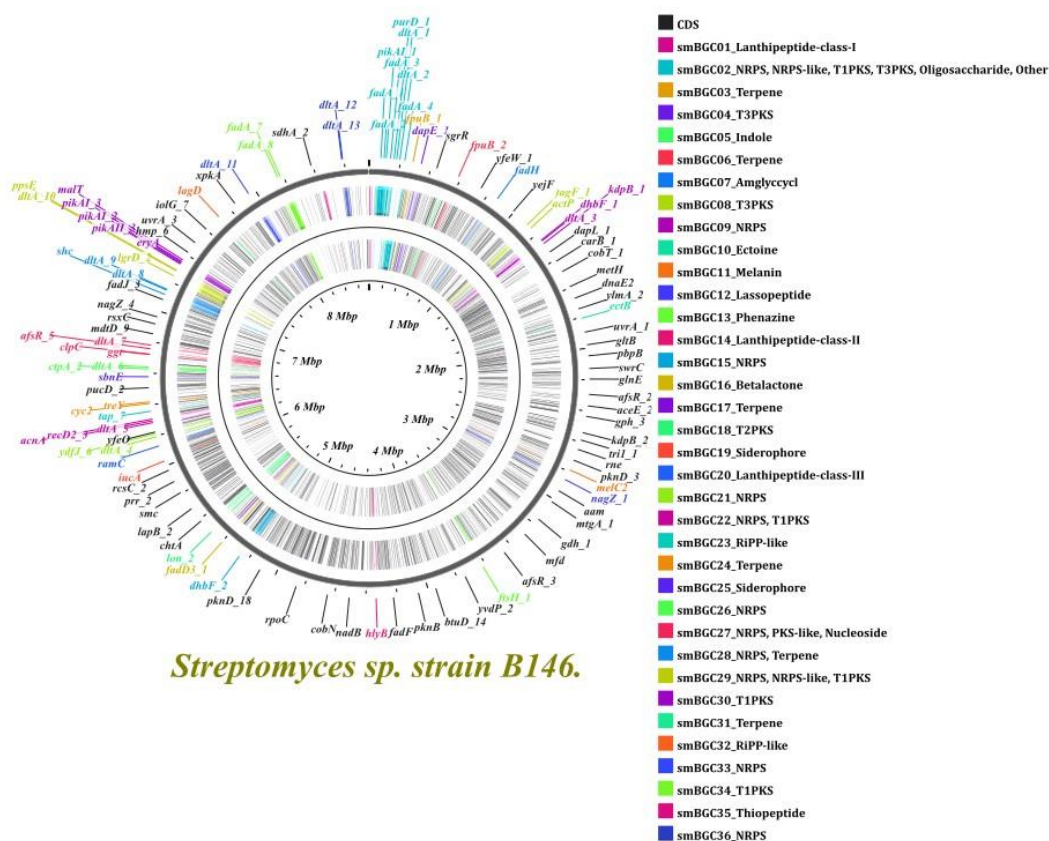
### 4.4. RESULTADOS DE LA IDENTIFICACIÓN DE BGCS EN EL GENOMA DE *Streptomyces* sp. B-146

El servidor web antiSMASH versión 6.1.1 detectó un total de 36 BGCs putativos en el genoma de *Streptomyces* sp. B-146 (Figura 08, Anexo 4). Los cuales en términos generales han sido clasificados como RiPPs (7), NRPSs (6), PKSs (5), Terpenos (5), BGCs híbridos (5), Sideróforos (2), Amglyccycl (1), Beta-lactonas (1), Ectoínas (1), Fenazinas (1), Indoles (1) y Melaninas (1).

Dichos resultados son consistentes con la clasificación de los 37 BGCs detectados en el genoma de *Streptomyces olivaceus* SCSIO T05, una especie filogenéticamente cercana a *Streptomyces* sp. B-146; los cuales también fueron clasificados como RiPPs (7), NRPSs (6), BGCs híbridos (6), PKSs (5), Terpenos (5), Sideróforos (2), Amglyccycl (1), Beta-lactonas (1), Ectoínas (1), Fenazinas (1), Indoles (1) y Melaninas (1) (Zhang *et al.*, 2019). En cambio, los 17 BGCs detectados en el genoma de *Streptomyces lannensis* T1317-0309, una cepa filogenéticamente lejana, fueron clasificados como Terpenos (3), RiPPs (3), BGCs híbridos (3), NRPSs (2), PKSs (2), Sideróforos (2), Ectoína (1) y Melanina (1) (Dahal *et al.*, 2020). Mostrando una variación en la cantidad y el patrón de la clasificación general de los BGCs detectados.

En general, los BGCs detectados en diferentes especies del género *Streptomyces* son clasificados principalmente como NRPSs, RiPPs, Terpenos, PKSs y BGCs híbridos (Arnison *et al.*, 2013; Nicault *et al.*, 2020). Siendo los BGCs híbridos más comunes del tipo PKS-NRPS, especialmente T1PKS-NRPS seguido de T1PKS-otros KS y Terpenos-NRPS (Nicault *et al.*, 2020; Belknap *et al.*, 2020). No obstante, pueden diferir enormemente respecto a la cantidad y diversidad, incluso si son miembros de una misma especie o si las cepas son taxonómicamente cercanas (Belknap *et al.*, 2020; Komaki *et al.*, 2020).

La extensa diversidad y abundancia de BGCs en el género *Streptomyces*, con un enfoque particular en BGCs híbridos, enriquecen enormemente la variedad de metabolitos secundarios biosintetizados, lo que a su vez puede facilitar el descubrimiento de nuevos productos terapéuticos (Belknap *et al.*, 2020). Además, la detección de BGCs con similitudes inferiores al 20% con respecto a BGCs de productos naturales conocidos, indica que estos BGCs podrían biosintetizar nuevos productos naturales (Gosse *et al.*, 2019). En ese sentido, nuestros resultados revelan que *Streptomyces* sp. B-146 contiene seis BGCs con similitud inferiores al 20% y otros cuatro BGCs que no muestran similitud con ningún BGC conocido. Asimismo, tan solo 14 BGCs presentan similitudes superiores al 70% (Anexo 4). Por tanto, existe grandes oportunidades de descubrir nuevos productos naturales con bioactividades terapéuticas únicas en *Streptomyces* sp. B-146.



**Figura 08:** BGCs identificados en el genoma de *Streptomyces sp. B-146* por antiSMASH 6.1.1.

Si bien antiSMASH es actualmente la herramienta más utilizada para la detección y caracterización de BGCs en bacterias (Blin *et al.*, 2021); en esta investigación hemos explorado otros servidores web automáticos de libre acceso, tales como ARTS 2.0, PRISM 4.4.5, BAGEL4, RiPPMiner-Genome y NaPDos (Anexo 5).

El servidor web ARTS 2.0 fue usado por su capacidad de realizar anotaciones adicionales sobre los genes contenidos en los BGCs detectados previamente por antiSMASH 5.1.1 (Mungan *et al.*, 2020). En este contexto, ARTS 2.0 reportó y clasificó de forma similar los 36 BGCs identificados por antiSMASH 6.1.1 y adicionalmente identificó un BGC clasificado como NRPS (Anexo 5). Este BGC adicional también fue detectado por RiPPMiner-Genome y clasificado como Linaridina, un pequeño grupo de péptidos lineales deshidratados perteneciente a la superfamilia RiPPs (Mo *et al.*, 2017). Para abordar esta contradicción, recurrimos a la anotación realizada por ambas herramientas bioinformáticas sobre los genes presentes en el mencionado BGC.

Luego del análisis respectivo, constatamos que ARTS 2.0 en conjunción con antiSMASH 5.1.1 detectaron siete genes biosintéticos relacionados con la producción de metabolitos secundarios tipo NRPS. Mientras que RiPPMiner-Genome tan solo anotó un gen con tres dominios metiltransferasa posiblemente relacionado con la producción de RiPPs. Por lo tanto, debido a que este BGC contiene muchos más genes involucrados en la biosíntesis de metabolitos secundarios tipo NRPS, en la Figura 09, lo hemos representado como un BGC putativo de tipo NRPS denominado “NRPS (ARTS)”.

**Figura 09:** BGCs identificados contrastando la anotación realizada con AntiSMASH 6.1.1, ARTS 2.0, PRISM 4.4.5, RiPPMiner-Genome y BAGEL4.

Cabe resaltar que estos cuatro BGCs también fueron reportados por antiSMASH 6.1.1, ARTS 2.0 y PRISM 4.4.5; y su clasificación coincide con la establecida por estas herramientas bioinformáticas.

El servidor web PRISM 4.4.5 tiene la capacidad de detectar hasta 22 distintos BGCs principalmente de tipo PKS, NRPS y RiPPs, pero no terpenos (Skinnider *et al.*, 2017). En ese sentido, dicho servidor web reveló la presencia de 24 BGCs, de los cuales 22 fueron clasificados de manera consistente y concordante con los resultados obtenidos a través de antiSMASH 6.1.1 y ARTS 2.0 (Anexo 5). No obstante, hay que resaltar que a diferencia de antiSMASH 6.1.1, PRISM 4.4.5 no tiene la capacidad de identificar BGCs híbridos extensos y biosintéticamente diversos. Como resultado, en algunas ocasiones, los BGCs híbridos extensos pueden ser reportados como dos o más BGCs aparentemente independientes (Anexo 5, BGC 02 y BGC 32).

Asimismo, PRISM 4.4.5 reportó otros dos BGCs clasificados respectivamente como PKS y NRPS; los cuales no fueron detectados por los otros servidores web a excepción de NaPDoS (Anexo 5, Figura 09). Respecto al BGC clasificado como PKS (Anexo 5, BGC 11), PRISM 4.4.5 detectó cinco dominios característicos implicados en la biosíntesis de los policétidos, incluyendo un dominio KS (*Ketosynthase*), que también fue reportado por NaPDoS. Por lo tanto, en este BGC, no se observa discrepancia alguna entre los resultados obtenidos. En virtud de esto, en la Figura 09 lo hemos representado como un BGC putativo de tipo PKS denominado “PKS (PRISM)”.

Por otro lado, en relación con el BGC clasificado como NRPS (Anexo 5, BGC 11), PRISM 4.4.5 detectó cuatro dominios, que incluyen un dominio “*Acyl adenylating Enzyme*”, dos dominios “*Thiolation*” y un dominio KS, el cual también fue reportado por NaPDoS. Al respecto, es importante señalar que los dominios centrales implicados en la biosíntesis de los NRPs son los dominios A (*Adenylation*), C (*Condensation*) y PCPs o T (*Peptidyl carrier protein* o *Thiolation*); mientras que los dominios centrales involucrados en la biosíntesis de los policétidos son los dominios AT (*Acetylation*), KS y ACP (*Acyl carrier protein*) (Khabthani *et al.*, 2021).

Por lo tanto, este BGC contiene al menos dos de los tres dominios centrales relacionados con la biosíntesis de los NRPs y su clasificación como un NRPS resulta ser la más adecuada.

No obstante, debido a la presencia de un dominio central implicado en la biosíntesis de policétidos, no se descarta la posibilidad de que se trate de un BGC híbrido tipo NRPS/PKS. Por lo que se requieren análisis adicionales para confirmar esta posibilidad. En esta investigación, lo hemos representado en la Figura 09 como un BGC putativo de tipo NRPS denominado “NRPS (PRISM)”.

El servidor web BAGEL4 fue desarrollado exclusivamente para detectar BGCs de tipo RiPPs usando una base de datos de *core* péptidos y/o a través la detección de motivos por el modelo HMM (*hidden Markov models*) (van Hell *et al.*, 2018). Este servidor reportó únicamente tres BGCs tipo RiPPs, una Zoocina A y dos Lantipéptidos.

Cabe destacar que el BGC clasificado como Zoocina A fue reportado por antiSMASH 6.1.1 y ARTS 2.0 como un BGC híbrido de tipo NRPS/Terpeno. En contraste, la clasificación de los otros dos BGCs como lantipéptidos coincidió con la clasificación realizada por las demás herramientas bioinformáticas empleadas.

Con el fin de obtener una mejor comprensión del BGC con clasificación diferente, se comparó la anotación realizada por BAGEL4 y antiSMASH 6.1.1 de los genes comprendidos en dicho BGC (datos no mostrados). Tras los análisis respectivos, observamos que BAGEL4 anotó 13 genes, uno de ellos denominado como “Zoocina A”. Por otro lado, antiSMASH 6.1.1 anotó 39 genes, 10 de los cuales estarían involucrados directamente en la biosíntesis de terpenos y otros 10 en la biosíntesis de NRPs; mientras que un gen adicional denominado “Aminotransferasa PigE” estaría relacionado tanto con la biosíntesis de terpenos como de NRPs.

Asimismo, el gen identificado como “Zoocina A” por BAGEL4 fue anotado por antiSMASH 6.1.1 como “Glicil-glicina endopeptidasa”. Aclarando, “Zoocina A” es una peptidoglicano-hidrolasa cuyo dominio catalítico tiene alto grado de similaridad con la región N-terminal de las enzimas “Glicil-glicina endopeptidasas” (Akesson *et al.*, 2007). Tomando en conjunto todo lo anteriormente mencionado, podemos concluir que este BGC claramente alberga grupos de genes directamente vinculados con la biosíntesis de Terpenos y NRPs; sin embargo, no se detectó evidencia contundente para demostrar la biosíntesis de Zoocina A.

Finalmente, el servidor web NaPDoS a través de un mecanismo rápido permite extraer y clasificar dominios KS (*Ketosynthase*) característicos de genes involucrados en la biosíntesis de PKs y dominios C (*condensation*) propios de genes implicados en la biosíntesis de NRPs; para lo cual recurre a una base de datos interna de genes biosintéticos caracterizados experimentalmente y a un esquema de clasificación basado en filogenias, que a su vez resulta particularmente útil para hacer predicciones más amplias y/o descubrir nuevas arquitecturas enzimáticas o nuevos mecanismos de ensamblaje de metabolitos secundarios (Ziemert *et al.*, 2012).

Los análisis aquí realizados, han detectado 43 dominios PKS y 39 dominios NRPS, distribuidos en los 17 BGCs de tipo PKS y NRPS detectados en esta investigación (Anexo 5). Asimismo, llego a detectar un dominio KS en una región del genoma de *Streptomyces* sp. B-146 donde las demás herramientas bioinformáticas no han reportado la presencia de BGCs (datos no mostrados). Por tanto, es importante destacar, que NaPDoS proporciona anotaciones precisas que resultan particularmente valiosas en análisis profundos de BGCs tipo NRPS y PKS detectados por otras herramientas bioinformáticas.

#### **4.4.1. BGCs PUTATIVOS CON BIOACTIVIDAD ANTIVIRAL, CITOTÓXICA Y ANTIBACTERIANA**

##### **a. Anotación del BGC putativo para la biosíntesis de Xiamicina A y Pactamidas.**

El smBGC22 putativo de *Streptomyces* sp. B-146 fue reportado simultáneamente por antiSMASH 6.1.1, ARTS 2.0, PRISM 4.4.5 y NaPDoS como un BGC híbrido tipo NRPS/PKS. No obstante, antiSMASH 6.1.1 brindó una clasificación más precisa al identificarlo como un híbrido NRPS/T1PKS (Anexo 5). Dicha clasificación se debe a que el ORF18 del smBGC22 que tiene una identidad del 99.49% a nivel de aminoácidos con la proteína PtmA de *Streptomyces pactum*, que se caracteriza por albergar módulos NRPS/PKS (Saha *et al.*, 2017). No obstante, el smBGC22 también alberga ORFs anotados como Terpeno-ciclasas, enzimas clave en la biosíntesis de Terpenoides. Específicamente, los ORF6 y ORF7 mostraron identidades del 99.5% y del 93.98% con las enzimas terpeno-ciclasas XiaI y XiaH de *Streptomyces* sp. SCSIO 02999, las cuales están involucradas en la biosíntesis del terpeno Xiamicina A (Anexo 6, Figura 10).

Luego de realizar alineamientos de secuencia y la anotación manual descubrimos que el smBGC22 contiene al menos seis ORFs adicionales a los 36 ORFs detectados por antiSMASH 6.1.1. Estos ORFs (-06, -05, -04, -03, -02, y -01) fueron localizados upstream del primer ORF detectado por antiSMASH 6.1.1 (Figura 10).

La anotación manual de los 42 ORFs consignados al smBGC22 fue realizada mediante UniProt y posteriormente comparada con la anotación realizada por PROKKA, BlastKOALA, RAST, antiSMASH 6.1.1 y MIBiG (Anexo 6). Dicha comparación nos permitió detectar dos BGCs previamente caracterizados con una similitud del 100%. Estos BGCs corresponden a la Xiamicina A (BGC0000666) de *Streptomyces* sp. SCSIO 02999 y a Pactamidas (BGC0001726) de *Streptomyces pactum*.

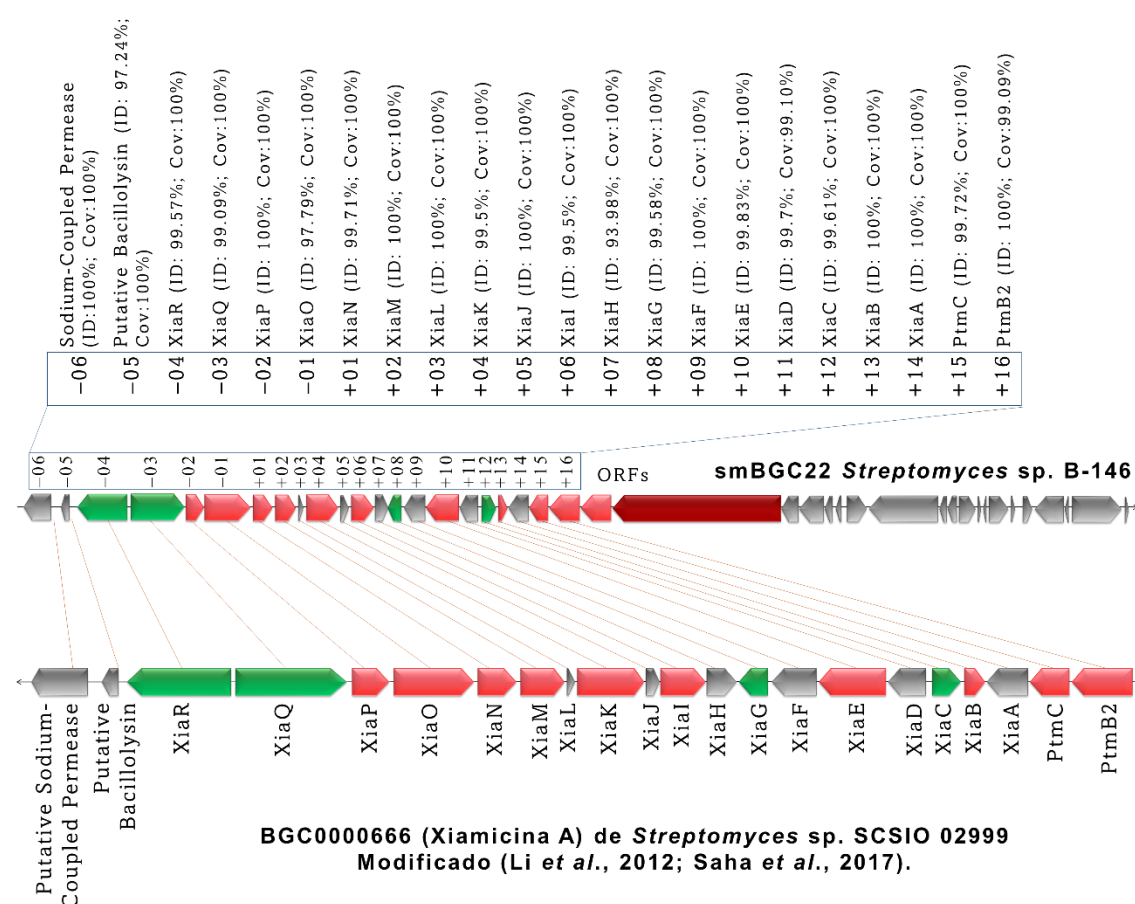
Respecto a la Xiamicina A, antiSMASH 6.1.1 reportó una similitud inicial del 72% con el BGC0000666 de *Streptomyces* sp. SCSIO 02999 (Anexo 4); no obstante, dicha similitud se elevó al 100% luego de la inclusión de los seis ORFs anteriormente mencionados (Anexo 6). El BGC0000666 reportado por MIBiG contiene 22 genes caracterizados, los cuales están representados en el smBGC22 de *Streptomyces* sp. B-146 con una identidad de entre 93 y 100% (Figura 10). No obstante, es importante aclarar que solo 18 genes (*xiaA* - *xiaR*) están directamente implicados en el proceso de biosíntesis de Xiamicina A, mientras que los otros cuatro no intervienen en su producción (Li *et al.*, 2012). Los genes *xiaE*, *xiaF*, *xiaG*, *xiaN* y *xiaP* estarían implicados en la biosíntesis de terpenos, los genes *xiaA*, *xiaB*, *xiaI*, *xiaJ*, *xiaK*, *xiaL*, *xiaM* y *xiaO* codificarían oxidorreductasas y los genes *xiaC*, *xiaG*, *xiaH*, *xiaQ* y *xiaR* estarían involucrados en la regulación y resistencia (Li *et al.*, 2012; Saha *et al.*, 2017).

La Xiamicina A fue descubierta por primera vez en *Streptomyces* sp. GT20021503, una bacteria endófitas de la planta de manglar *Bruguiera gymnorrhiza* (Ding *et al.*, 2010). Bioquímicamente, la Xiamicina A es un indolosesquiterpeno pentacíclico, perteneciente a la familia de productos naturales denominados indolosesquiterpenoides, los cuales eran desconocidos en procariotas hasta el descubrimiento de la Xiamicina A (Ding *et al.*, 2011; Meng *et al.*, 2015). A partir de entonces, la Xiamicina A y varios compuestos análogos han sido detectados en *Streptomyces* sp. HKI0595, *Streptomyces* sp. #HK18 y *Streptomyces xinghaiensis* NRRL B-24674T (Ding *et al.*, 2011; Kim *et al.*, 2016; Chen *et al.*, 2018).



Los indolosesquiterpenoides se distinguen por su complejidad estructural y su emergente y diversa actividad biológica (Munda *et al.*, 2023). En ese sentido, la Xiamicina A destaca por sus propiedades antivirales, especialmente contra el VIH, el virus del Herpes simple 1 (VHS-1) y el SARS-CoV-2 (Ding *et al.*, 2010; Meng *et al.*, 2015; Muhammad *et al.*, 2022).

Asimismo, se caracteriza por su baja citotoxicidad y por actuar contra bacterias grampositivas y gramnegativas, algunas de las cuales multirresistentes (Ding *et al.*, 2010; 2011; Meng *et al.*, 2015).



**Figura 10:** Identificación y anotación de un subconjunto de genes del smBGC22 de *Streptomyces* sp. B146 implicado en la biosíntesis de Xiamicina A.

Otros indolosesquiterpenoides con propiedades terapéuticas interesantes son la Xiamicina D por su potente efecto antiviral contra el virus de la diarrea epidémica porcina (PEDV) y la Xiamicina H por su notable actividad fungicida contra patógenos relevantes para la agricultura (Kim *et al.*, 2016; Pfaffenbach *et al.*, 2019).

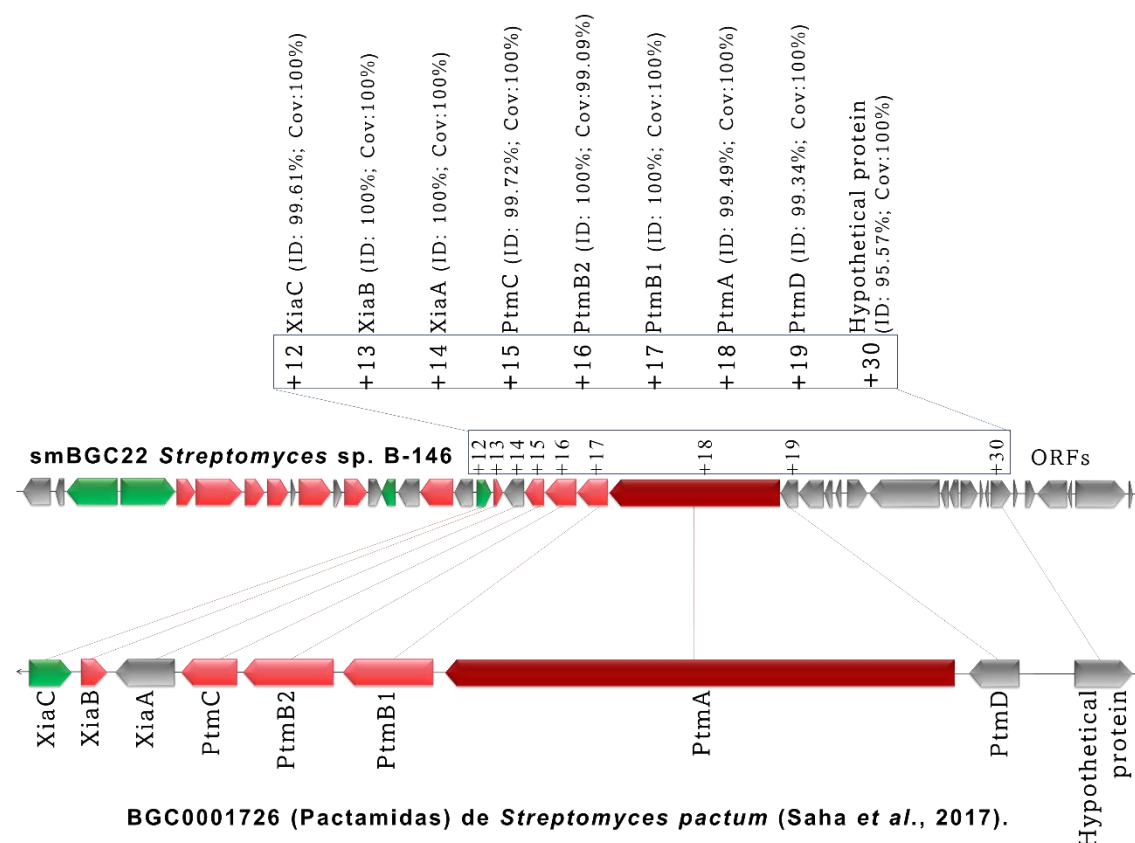
Por todo lo mencionado y dada su relevancia como agentes con propiedades bioactivas diversas, la Xiamicina A y algunos compuestos análogos han sido objeto de estudio por parte de la comunidad científica para desarrollar estrategias con diferentes enfoques para lograr su síntesis química total con fines terapéuticos (Meng *et al.*, 2015; Trotta *et al.*, 2017; Pfaffenbach *et al.*, 2019; Munda *et al.*, 2023).

Respecto a Pactamidas, el otro BGC detectado dentro del smBGC22, no fue reportado por antiSMASH 6.1.1 en el cuadro resumen de los BGCs putativos (Anexo 4). No obstante, los nueve genes caracterizados en el BGC0001726 de *Streptomyces pactum* estuvieron representados en el smBGC22 de *Streptomyces* sp. B146 con una identidad entre el 95 y 100% (Figura 11). Es importante aclarar que los genes *PtmA*, *PtmB1*, *PtmB2*, *PtmC* y *PtmD* son los que codifican a las enzimas responsables de la biosíntesis las Pactamidas (A – F) en *Streptomyces pactum*; mientras que los genes *xiaA*, *xiaB*, *xiaC* y el Orf(-1) que codifica para una “Proteína hipotética”, no estarían involucrados en la biosíntesis de estas moléculas (Saha *et al.*, 2017).

La anotación realizada por PROKKA, BlastKOALA, RAST, antiSMASH 6.1.1 y UniProt de los ORFs 15 – 19 del smBGC22 de *Streptomyces* sp. B146 es consistente con la anotación reportada en MIBiG (Anexo 6). En consecuencia, la función teórica de los genes representados por los ORFs 15 – 19 en el proceso de biosíntesis de las Pactamidas sería la siguiente: el ORF18 (*PtmA*) codificaría una enzima NRPS/PKS responsable de generar el precursor de las Pactamidas a partir del tetramato de polieno; los ORF17 (*PtmB1*), ORF16 (*PtmB2*) y ORF15 (*PtmC*) codificarían enzimas encargadas de someter al precursor a una cascada de ciclización reductiva para la formación de las Pactamidas C, E y A; mientras que el ORF19 (*PtmD*) codificaría para una hidroxilasa que probablemente junto a otras enzimas estarían involucradas en la formación de Pactamidas B, D y F a partir de Pactamida A (Saha *et al.*, 2017).

Las Pactamidas fueron descubiertas en *Streptomyces pactum* SCSIO 02999 y se caracterizan por su potente efecto citotóxico contra líneas celulares cancerígenas de humanos, especialmente Pactamida A que mostró actividad citotóxica con valores IC<sub>50</sub> entre 0.24–0.51 mM (Saha *et al.*, 2017). Las Pactamidas pertenecen a clase de productos naturales denominados macrolactamas de tetramato policíclico (PTM) que se derivan de una vía metabólica híbrida PKS/NRPS (Saha *et al.*, 2017; Li *et al.*, 2019).

Los PTMs exhiben una amplia gama de actividades biológicas antifúngicas, antibacterianas, antiprotozoarias, antiulcerosas, antivirales, citotóxicas e inductoras de apoptosis, y como tal, tienen un potencial significativo en aplicaciones medicinales y agrícolas (Zhang *et al.*, 2016; Saha *et al.*, 2017). Además, cabe resaltar que los BGCs responsables de la biosíntesis de los PTMs son altamente conservados y distribuyen ampliamente entre las bacterias; no obstante, en su mayoría son BGCs crípticos, es decir, permanecen en silencio (Saha *et al.*, 2017).



**Figura 11:** Identificación y anotación de un subconjunto de genes del smBGC22 de *Streptomyces* sp. B146 implicado en la biosíntesis de Pactamidas.

#### 4.4.2. BGCs CON ACTIVIDAD ANTIBACTERIANA

##### a. Anotación del BGC putativo para la biosíntesis Albaflavenona

El smBGC17 putativo de *Streptomyces* sp. B-146, similar a Albaflavenone, fue reportado únicamente por antiSMASH 6.1.1 y ARTS 2.0, debido a que las demás herramientas bioinformáticas usadas no tienen la capacidad de detectar BGCs tipo Terpenos (Anexo 5). Por tanto, en ambos casos fue clasificado como un Terpeno debido a que el ORF10 fue identificado como una enzima Terpeno sintasa/ciclasa (Figura 12).

Según antiSMASH 6.1.1, el smBGC17 alberga 19 ORFs y tiene una similaridad del 100% con el BGC0000660 de *Streptomyces coelicolor* A3(2) implicado en la biosíntesis de Albaflavenone (Anexo 4). No obstante, el BGC0000660 contiene solo dos genes caracterizados (*sco5222* y *sco5223*), los cuales están representados en los ORFs 10 y 11 del smBGC17, con una identidad del 92.88% y el 93.28%, respectivamente (Figura 12).

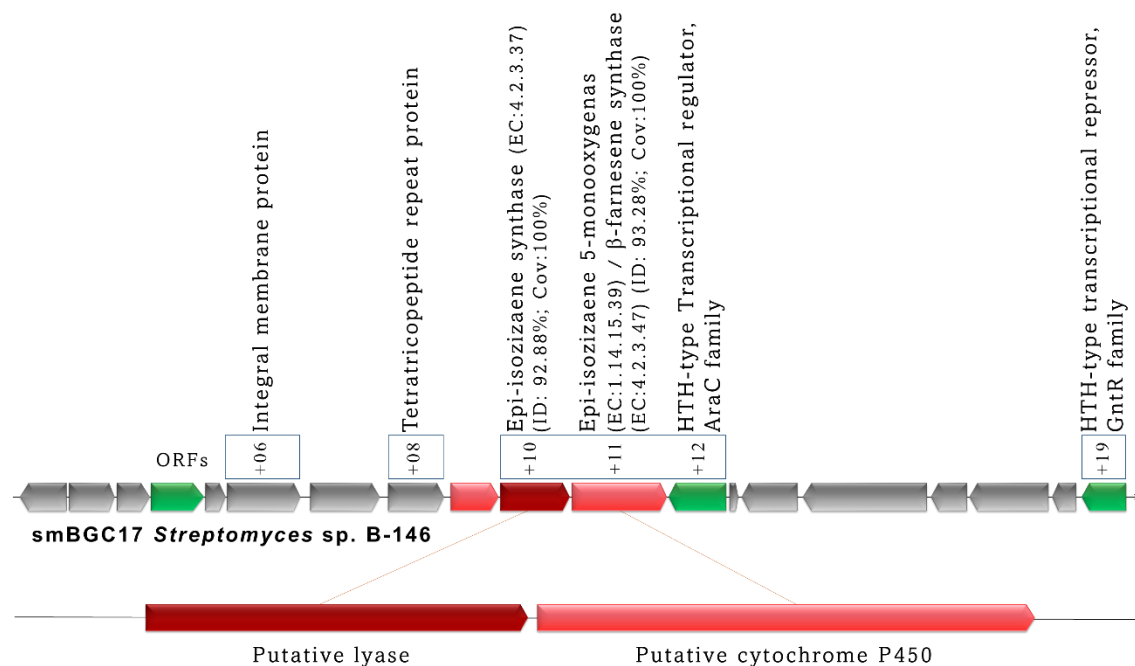
Al igual que el operón conformado por los genes *sco5222* y *sco5223* en *Streptomyces coelicolor* A3(2), los ORFs 10 y 11 comparten una superposición transcripcional de cuatro nucleótidos (ATGA) entre sus terminales 3' y 5', respectivamente (Zhao *et al.*, 2008; Lin & Cane, 2009).

La anotación realizada por PROKKA, BlastKOALA, RAST, antiSMASH 6.1.1 y UniProt de los ORFs 10 – 11 del smBGC17 de *Streptomyces* sp. B146 coincide con la anotación reportada en MIBiG (Anexo 7). Además, dichos anotadores identificaron a un Regulador transcripcional tipo HTH de la familia AraC en el ORF12 y a un Represor transcripcional tipo HTH de la familia GntR en el ORF19, con una identidad según UniProt del 94.1% y 99.2%, respectivamente (Anexo 7; Figura 12). Los cuales podrían estar involucrados en la transcripción del operón putativo conformado por los ORF10 y el ORF11.

En dicho operón putativo, el ORF10 codificaría para una enzima Epi-isozizaeno sintasa responsable de catalizar la ciclación del difosfato de farnesilo para generar un hidrocarburo sesquiterpénico denominado epi-isozizaeno (Zhao *et al.*, 2008; Lin & Cane, 2009). Mientras que el ORF11 codificaría para la enzima bifuncional Epi-isozizaeno 5-monooxigenasa / Beta-farneseno sintasa, encargada de llevar a cabo dos oxidaciones alílicas secuenciales; es decir, primero lleva a cabo una oxidación no estereoespecífica del epi-isozizaeno para generar una mezcla epímerica de dos albaflavenoles, los cuales sirven como sustratos para una segunda oxidación que los convierte en Albaflavenonas (Zhao *et al.*, 2009; Lin & Cane, 2009).

Además, el smBGC17 contiene ORFs que codificarían para proteínas que también han sido detectadas a nivel *in silico* alrededor de operones de Albaflavenona en diferentes especies de *Streptomyces*. Específicamente, los ORFs 06 y 08 codificarían para una Proteína integral de membrana (PIM) y para una Proteína repetida tetratricopeptido (PRT), con identidades según Uniprot del 83.6% y 98.5%, respectivamente.

La PIM podría estar involucrada en la exportación de Albaflavenona fuera de la célula; mientras que la PRT podría participar en la formación de un complejo enzimático entre Epi-isozizaeno sintasa y Epi-isozizaeno 5-monooxigenasa/Beta-farneseno sintasa, lo que permitiría la canalización de epi-isozizaeno entre ambas enzimas para lograr una producción eficiente de Albaflavenona (Figura12) (Moody *et al.*, 2012).



**BGC0000660 (Albaflavenone) de *Streptomyces coelicolor* A3(2) (Zhao *et al.*, 2008).**

**Figura 12:** Identificación y anotación del smBGC17 de *Streptomyces* sp. B146 implicado en la biosíntesis de Terpenoide idéntico a Albaflavenona.

La secuencia de aminoácidos del ORF10 contiene los característicos dominios conservados de unión a  $Mg^{2+}$  DDRHD y NDLCSPKE, responsables de la actividad catalítica de todas las enzimas Epi-isozizaeno sintasas en *Streptomyces* (Lin & Cane, 2009; Moody *et al.*, 2012). Asimismo, la secuencia de aminoácidos del ORF11, al igual que la secuencia de aminoácidos de la enzima Citocromo P450 de *Streptomyces coelicolor* A3(2), presenta dos dominios de unión a  $Mg^{2+}$  (DDLTLALLE y DDNGD); los cuales son responsables de la actividad bifuncional de este tipo de enzimas y cuya variación en algunas especies de *Streptomyces* ha conducido a la pérdida de su distintiva actividad bifuncional (Moody *et al.*, 2012).

Por tanto, dada la enorme similitud entre el operón putativo de *Streptomyces* sp. B-146 y el operón caracterizado de *Streptomyces coelicolor* A3(2), es altamente probable que *Streptomyces* sp. B-146 sea capaz de biosintetizar un terpenoide similar a Albaflavenona.

La Albaflavenona bioquímicamente es un sesquiterpeno tricíclico, un grupo importante de productos naturales que exhiben actividades biológicas interesantes y una amplia diversidad estructural (Le Bideau *et al.*, 2017).

Inicialmente, fue aislado de *Streptomyces albidoflavus*; no obstante, también ha sido detectado en *Streptomyces diastaticus*, *Streptomyces coelicofavus* y *Streptomyces rubrogriseus* (Gürtler *et al.*, 1994; Komaki *et al.*, 2018). Este metabolito se caracteriza por su naturaleza volátil y olorosa; además de exhibir propiedades antibacterianas interesantes (Gürtler *et al.*, 1994)

#### **4.4.3. BGCs CON OTRAS BIOACTIVIDADES**

##### **a. Anotación del BGC putativo para la biosíntesis SapB**

El smBGC20 de *Streptomyces* sp. B-146 fue detectado simultáneamente por antiSMASH 6.1.1, ARTS 2.0, PRISM 4.4.5, RiPPMiner-Genome y BAGEL4 como un BGC tipo Lantipéptido (RiPPs). No obstante, cabe resaltar que antiSMASH 6.1.1 brindó una clasificación más certera al identificarlo como un BGC tipo Lantipéptido de clase III (Anexo 5). Dicha clasificación se debe a que el ORF06 como del smBGC22 fue identificado como una Serina/treonina proteína kinasa, una enzima clave en la biosíntesis de lantipéptidos clase III, conocidas de forma genérica como lantipéptido sintasas LanKC (Zhang *et al.*, 2012).

Según antiSMASH 6.1.1 el smBGC20 alberga 15 ORFs y tiene una similaridad del 100% con el BGC0000551 de *Streptomyces coelicolor* A3(2) implicado en la biosíntesis del lantipéptido SapB (Anexo 4). El BGC0000551 contiene cinco genes caracterizados denominados *RamC*, *RamS*, *RamA*, *RamB* y *RamR* (Kodani *et al.*, 2004); no obstante, solo los cuatro primeros están reportados en MIBiG. Estos cuatro genes están representados en los ORFs 06, 07, 08 y 09 del smBGC20 de *Streptomyces* sp. B-146 con una identidad a nivel de aminoácidos entre el 68 y 95% (Figura 13).

Cabe aclarar que no fue posible establecer la identidad a nivel de aminoácidos entre el ORF10 y el gen *RamR*, debido a la falta de información acerca de las secuencias de nucleótidos y aminoácidos de este último. Sin embargo, sí llegamos a determinar la identidad a nivel de aminoácidos en un 43.69% y un 42.86% con relación a dos genes reguladores implicados en la biosíntesis de los lantipéptidos SRO15-2212 de *Streptomyces filamentosus* (BGC0000553) y AmfS de *Streptomyces griseus* (BGC0000496) (Anexo 8) (Ueda *et al.*, 2002; Kersten *et al.*, 2011).

Por lo tanto, es presumible que el ORF10 esté directamente implicado en la regulación de la transcripción de los genes responsables de la biosíntesis y transporte fuera de la célula de un lantipéptido idéntico a SapB en *Streptomyces* sp. B-146.

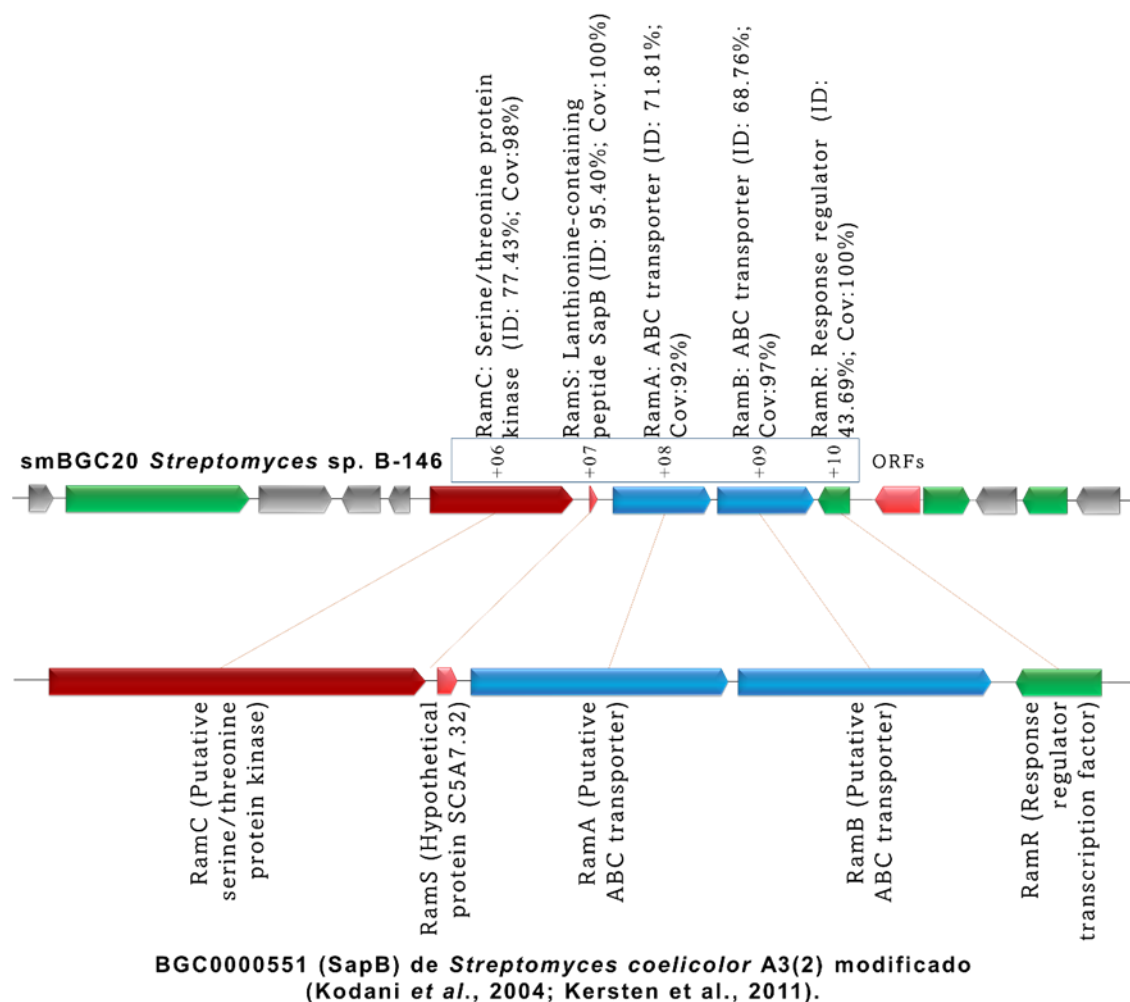
La anotación realizada por PROKKA, BlastKOALA, RAST, BAGEL4, antiSMASH 6.1.1 y UniProt de los ORFs 06 – 10 del smBGC20 de *Streptomyces* sp. B146 concuerda con la anotación reportada en MIBiG (Anexo 8). Por lo tanto, podemos predecir la función de las proteínas codificadas en los ORFs anteriormente mencionados. Es decir, el ORF07 (*RamS*) codificaría al péptido precursor de un lantipéptido idéntico a SapB; el ORF06 (*RamC*) codificaría a la enzima lantipéptido sintasa responsable de las modificaciones postraduccionales del péptido precursor; mientras que los ORF08 (*RamA*) y ORF09 (*RamB*) codificarían a dos transportadores ABC encargados de exportar fuera de la célula al lantipéptido recién biosintetizado; y por último, el ORF10 (*RamR*) codificaría para una proteína de unión al ADN capaz de regular la transcripción de los genes *RamC*, *RamS*, *RamA* y *RamB*, los cuales funcionarían como operón (Kodani *et al.*, 2004; Gaskell *et al.*, 2012).

SapB fue inicialmente descrito como un péptido de 2.027 kDa asociado a la formación del micelio aéreo en bacterias filamentosas como *Streptomyces coelicolor* (Willey *et al.*, 1991; Guijarro *et al.*, 1988). Este péptido funciona como una proteína tensioactiva que promueve la liberación y el crecimiento ascendente de los filamentos aéreos emergentes en bacterias del género *Streptomyces* (Capstick *et al.*, 2007; Gaskell *et al.*, 2012).

Bioquímicamente, es un lantipéptido de clase III (Gaskell *et al.*, 2012). Los lantipéptidos son RiPPs que caracterizan principalmente por sus anillos de lantionina y/o metillantionina que confieren estabilidad y ayudan en diversas actividades biológicas (Arias-Orozco *et al.*, 2021; Fu *et al.*, 2023). Este tipo de compuestos desempeñan una amplia variedad de actividades biológicas, desde antimicrobianas hasta antialodínicas (Repka *et al.*, 2017).

SapB comparte características estructurales distintivas con los lantibióticos, también conocidos como bacteriocinas de clase I (Kodani *et al.*, 2004; Gaskell *et al.*, 2012). Los lantibióticos son en realidad lantipéptidos que exhiben actividad antimicrobiana, especialmente contra bacterias gram positivas (Repka *et al.*, 2017; Geng & Smith, 2018).

Sin embargo, a diferencia de la mayoría de lantibióticos, SapB no parece tener actividad antimicrobiana; en cambio, exhibe propiedades surfactantes interesantes (Kodani *et al.*, 2004; Capstick *et al.*, 2007).



**Figura 13:** Identificación y anotación del smBGC20 de *Streptomyces* sp. B146 implicado en la biosíntesis de un lantipéptido idéntico a SapB.

Cabe resaltar, que en *Actinomycetes*, además de SapB, se han descrito y caracterizado estructuralmente otros lantibióticos de clase III como SapT, LabA1 y LabA2; asimismo, se han reportado una gran cantidad de clústeres de genes homólogos (Völler *et al.*, 2012). Dicha diversidad, muestra el enorme potencial para el descubrimiento de nuevos lantipéptidos con propiedades biotecnológicas únicas y/o la oportunidad de caracterizar nuevas enzimas biosintéticas de baja selectividad por sustratos para diversificar las estructuras y las bioactividades de los lantipéptidos a través de la biología sintética (Arias-Orozco *et al.*, 2021; Fu *et al.*, 2023).



## **4.5. RESULTADOS DE LA IDENTIFICACIÓN DE SISTEMAS CRISPR EN EL GENOMA DE *Streptomyces* sp. B-146**

### **4.5.1. DESCRIPCIÓN DE LOS ARREGLOS CRISPR IDENTIFICADOS**

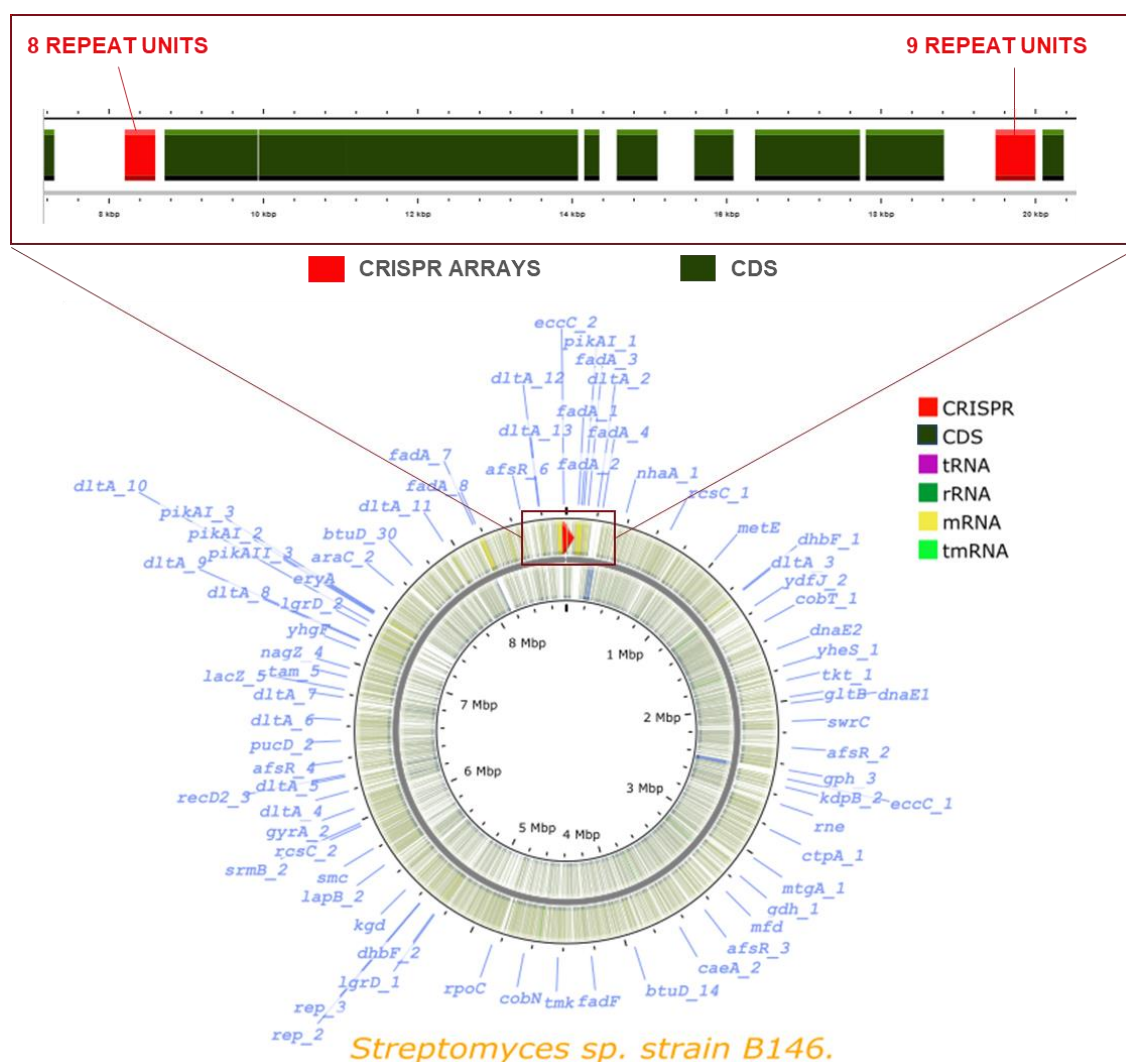
Tras la evaluación *in silico* del genoma de *Streptomyces* sp. B-146, las herramientas bioinformáticas Prokka y RAST detectaron dos arreglos CRISPR. Prokka reportó su ubicación en sentido positivo con tamaños de 396 y 516 pb con siete y nueve unidades de repetición, respectivamente. Los arreglos CRISPR fueron denominados Array\_01 y Array\_02, respectivamente. Prokka no reporta la ubicación, el tamaño y las secuencias de las unidades de repetición ni de los espaciadores contenidos en los arreglos CRISPR detectados. Paralelamente, RAST reportó los mismos arreglos CRISPR, con ubicación y tamaño invariable para Array\_02. No obstante, le asignó un tamaño mayor a Array\_01 (457 pb) debido a que identificó ocho unidades de repetición en lugar de las siete reportadas por Prokka. Además, a diferencia de Prokka, brindó la ubicación, el tamaño (29 pb) y las secuencias específicas de las unidades de repetición. También brindó la ubicación y tamaño (32-33 pb) de los espaciadores, pero no sus secuencias específicas (Anexo 9).

Para recabar información más precisa de los sistemas CRISPR en el genoma de *Streptomyces* sp. B-146 se emplearon los servidores web especializados CRISPRCasFinder y CRISPRminer. Los cuales también detectaron los arreglos CRISPR Array\_01 y Array\_02, y brindaron información acerca de su ubicación y tamaño; así como, la ubicación, tamaño y secuencias específicas de las unidades de repetición y de los espaciadores contenidos en ambos arreglos CRISPR.

El arreglo CRISPR Array\_01 reportado por CRISPRminer contenía seis unidades de repetición abarcando un tamaño de tan solo 366 pb. Las unidades de repetición tenían un tamaño de 30 pb; mientras que los espaciadores tenían un tamaño de 31 pb (Anexo 11). En contraste, CRISPRCasFinder, al igual que RAST, reportó el arreglo CRISPR Array\_01 con ocho unidades de repetición abarcando un tamaño de 457 pb; no obstante, a diferencia de RAST, los tamaños de las unidades de repetición y los espaciadores coincidían con lo reportado por el servidor web CRISPRminer; es decir, 30 pb para las unidades de repetición y 31-32 pb para los espaciadores (Anexo 9).

Respecto al arreglo CRISPR Array\_02, al igual Prokka y RAST, el servidor web CRISPRCasFinder reportó la misma ubicación y tamaño (516 pb). Asimismo, las ubicaciones y tamaños de las unidades repetición y de los espaciadores fueron invariables.

Específicamente, 29 pb para las nueve unidades de repetición y 31-32 pb para los espaciadores. No obstante, el servidor web CRISPRminer reportó el arreglo CRISPR Array\_02 con un tamaño de 515 pb debido a que identificó las repeticiones en 28 pb en lugar de 29 pb y, en consecuencia, los espaciadores fueron reportados con tamaños de 32-33 pb (Anexo 9, Anexo 10, Anexo 11). Tomando en conjunto los datos, se concluye que *Streptomyces* sp. B-146 contiene dos arreglos CRISPR de 8 y 9 unidades de repetición (Figura 14). La Cuadro 02 resume las principales características de los arreglos CRISPR Array\_01 y Array\_02 del genoma de *Streptomyces* sp. B-146, reportadas por Prokka, RAST, CRISPRCasFinder y CRISPRminer.



**Figura 14:** Visualización mediante Proksee de la localización de los sistemas CRISPR identificados en el genoma de *Streptomyces* sp. B-146.

**Cuadro 02:** Principales características de los arreglos CRISPR detectados por Prokka, RAST y los servidores web CRISPRCasFinder y CRISPRminer.

Arreglos CRISPR						
ID	Servidor	Ubicación (Cadena +)	Longitud (pb)	Repeticiones		
				Unidades	Longitud (pb)	Secuencia consenso
ARRAY_01	PROKKA	8 203 - 8 598	396	7	-	-
	RAST	8 203 - 8 659	457	8	29	CTGCTCCCCGCGCCCGCGGGGATGGTCCC
	CRISPRCas Finder	8 203 - 8 660	458	8	30	CTGCTCCCCGCGCCCGCGGGGATGGTCCCG
	CRISPRminer	8 203 - 8 568	366	6	30	CTGCTCCCCGCGCCCGCGGGGATGGTCCCG
ARRAY_02	PROKKA	19 482 - 19 997	516	9	-	-
	RAST	19 482 - 19 997	516	9	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
	CRISPRCas Finder	19 482 - 19 997	516	9	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
	CRISPRminer	19 482 - 19 996	515	9	28	CTGCTCCCCGCACCCGCGGGGATGGTCCC

FUENTE: Elaboración propia.

Además, CRISPRminer también detectó la secuencia de un espaciador que tenía una identidad del 100% con otra región del genoma de *Streptomyces* sp. B-146 ubicado en la posición 8,536,769 – 8,536,796 pb, constituyendo un espaciador autodirigido o *self-targeting* (Anexo 11). El *self-targeting* detectado se dirige parcialmente contra el extremo 3' de un CDS en sentido negativo de 1,116 pb ubicado en la posición 8,536,786 – 8,535 671, catalogado como una proteína hipotética tanto por Prokka como RAST.

Los espaciadores contenidos en los arreglos CRISPR normalmente derivan de la incorporación de material genético invasivo; sin embargo, no todos parecen tener origen foráneo y, en cambio, podrían originarse en el propio genoma del huésped (Wimmer & Beisel, 2020; Devi *et al.*, 2022b). La presencia de este tipo de espaciadores denominados *self-targeting*, plantea una paradoja, ya que se esperaría que los *self-targeting* indujeran una respuesta autoinmune y la muerte celular; no obstante, podrían ayudar en la regulación génica o en la evolución del huésped y, potencialmente, conducir o respaldar nuevas funcionalidades (Wimmer & Beisel, 2020).

Además, los *self-targeting* han sido reportados en números significativos entre especies bacterianas y en la mayoría de los casos, su presencia conduce a la inactivación del sistema CRISPR-Cas o la remodelación del genoma (Devi *et al.*, 2022b). Por otro lado, no se pudo verificar la presencia de la secuencia líder en los arreglos CRISPR detectados, debido a que ninguna de las herramientas bioinformáticas usadas tiene la capacidad de identificarlas.

En ese sentido, es importante resaltar que es común encontrar arreglos CRISPR en cromosomas y plásmidos que carecen de secuencia líder; no obstante, algunos de ellos si se pueden transcribirse (Shah & Garrett, 2011). Incluso se ha reportado arreglos CRISPR en plásmidos que carecen a la vez de secuencia líder y genes Cas (Lillestøl *et al.*, 2009).

En concreto, los arreglos CRISPR que carecen de secuencia líder adyacente son inactivos en la adquisición de nuevos espaciadores por lo que suelen ser mucho más pequeños y defectuosos en la transcripción, quedando por esclarecer si las secuencias líderes son eliminadas o tan solo separadas por eventos de trasposición (Alkhnbashi *et al.*, 2016). Por tanto, sin la detección de las secuencias líderes se dificulta elucidar si los arreglos CRISPR son funcionalmente activos en el genoma de *Streptomyces* sp. B146.

Otro punto por destacar es que las repeticiones directas entre los dos arreglos CRISPR detectados son altamente conservados, aun cuando están separados por aproximada 10,882 pb. Además, los servidores web CRISPRminer y CRISPRCasFinder reportaron un grupo de genes Cas a 9,458 pb downstream del arreglo CRISPR Array\_02. Al respecto, sabemos que cuando un genoma alberga múltiples arreglos CRISPR con la misma secuencia de repetición directa, usualmente los genes Cas están únicamente adyacentes a un solo arreglo CRISPR (Jansen *et al.*, 2002). Tomando en conjunto esta información, existe la probabilidad de que los arreglos CRISPR y los genes Cas detectados formen parte de un único sistema CRISPR/Cas en el genoma de *Streptomyces* sp. B-146.

Tampoco fue posible esclarecer con fiabilidad la orientación del presumible sistema CRISPR/Cas detectado. Mientras los arreglos CRISPR fueron reportados en la cadena positiva, los genes Cas fueron reportados en la cadena negativa. En ese sentido, podría entenderse que los arreglos CRISPR y los genes Cas tendrían direccionalidad de transcripción opuesta. No obstante, la orientación de los arreglos CRISPR es intrínsecamente ambigua, lo que dificulta deducir la hebra a partir de la cual se generan los crRNA (Alkhnbashi *et al.*, 2016). Por ello, sería determinante detectar la secuencia líder para elucidar las cuestiones de direccionalidad de transcripción del presumible sistema CRISPR/Cas detectado.

#### **4.5.2. ANOTACIÓN DE GENES Cas IDENTIFICADOS**

Los genes Cas reportados en el genoma de *Streptomyces* sp. B-146 por CRISPRCasFinder y CRISPRminer, denominados Cas-S1, Cas-S2, Cas-S3, Cas-S4 y Cas-S5, fueron contrastados con la información proveída por Prokka y RAST (Table 03).

Tomando los datos en conjunto se deduce que los genes Cas detectados conforman un clúster de cinco genes y probablemente funcionen como un operón debido a que sus ORFs son contiguos y ligeramente superpuestos. Guardando similitudes con la forma de transcripción de sistemas CRISPR-Cas I-E, los cuales suelen estar codificados por un solo operón, abarcando los genes Cas1, Cas2 y Cas3 junto con los cinco genes del complejo Cascade (Makarova *et al.*, 2015). En la anotación de los cinco CDS hubo un consenso contundente entre las CRISPRCasFinder y CRISPRminer, excepto que esta última herramienta no detectó el CDS denominado Cas-S4. Prokka no anotó los cinco CDS consultados, mientras que RAST anotó únicamente los CDS Cas-S2 y Cas-S5 como Cas5e y “Diguanilate cyclase / Phosphodiesterase”, respectivamente.

**Cuadro 03:** Clúster de genes *Cas* detectados por los servidores web CRISPRCasFinder y CRISPRminer comparados con la anotación realizada por Prokka y RAST.

Servidor	Nombre	Ubicación (Cadena -)	Dominio Conservado	Predicción	Subtipo Cas
CRISPRCasFinder	Cas-S1	29,455 - 30,183	N. R	Cas6	Cas-I-E
	Cas-S2	30,180 - 31,025	N. R	Cas5	Cas-I-E
	Cas-S3	31,022 - 32,383	N. R	Cas7	Cas-I-E
	Cas-S4	32,383 - 32,976	N. R	Cse2	Cas-I-E
	Cas-S5	32,973 - 34,595	N. R	Cse1	Cas-I-E
CRISPRminer	Cas-S1	29,455 - 30,141	pfam08798	Cas6e	Cas-I-E
	Cas-S2	30,180 - 31,025	cd09645	Cas5	Cas-I-E
	Cas-S3	31,022 - 32,383	pfam09344	Cas7	Cas-I-E
	Cas-S4	N. R	N. R	N. R	N. R
	Cas-S5	32,973 - 34,595	cd09729	Cas8e	Cas-I-E
Prokka	Cas-S1	29,455 - 30,183	N. R	Hypothetical protein	N. R
	Cas-S2	30,180 - 31,025	N. R	Hypothetical protein	N. R
	Cas-S3	31,022 - 32,383	N. R	Hypothetical protein	N. R
	Cas-S4	32,380 - 32,976	N. R	Hypothetical protein	N. R
	Cas-S5	32,973 - 34,595	N. R	Hypothetical protein	N. R

«Continuación»

RAST	Cas-S1	29,455 - 30,183	N. R	Hypothetical protein	N. R
	Cas-S2	30,180 - 31,025	N. R	Cas5e	N. R
	Cas-S3	31,022 - 32,383	N. R	Hypothetical protein	N. R
	Cas-S4	32,380 - 32,976	N. R	Hypothetical protein	N. R
	Cas-S5	32,973 - 34,595	GGDEF & EAL domains	Diguanylate cyclase / phosphodiesterase with PAS/PAC sensors	N. R

---

FUENTE: Elaboración propia.

La anotación de Cas-S2 por RAST coincide con las anotaciones realizadas por CRISPRCasFinder y CRISPRminer. Mientras, que la anotación de Cas-S5 a simple vista no guarda relación con las anotaciones realizadas de ambos servidores web. No obstante, se sabe que las proteínas Cas8 son codificadas por diversos genes e incluso se ha especulado que Cas8 es un homólogo altamente derivado de Cas10, una proteína que contiene dominios “Palm” tipo ciclasas involucradas en el reconocimiento del ARN, y que de manera análoga a Cas8, representa a una subunidad grande del complejo efector de sistemas CRISPR/Cas Tipo III (Zhu & Ye, 2012; Makarova *et al.*, 2013; Makarova *et al.*, 2015). Asimismo, algunas Cas10 como Cmr2, contienen dominios “Palm” tipo ciclasas guardan relación dominios pertenecientes a las enzimas Nucleótido ciclasas clase III, que incluyen las adenilato-ciclasas, guanilato ciclasas y diguanilato ciclasas (Sinha & Sprang, 2006; Zhu & Ye, 2012). Por tanto, la anotación realizada por RAST del CDS Cas-S5 no es inconsistente con la posibilidad de que este gen codifique para una proteína que represente la subunidad grande (Cas8) del complejo Cascade.

Dicha posibilidad fue confirmada por Uniprot, cuyos resultados de anotación también que fueron consistentes con los resultados de anotación obtenidos por RAST, CRISPRCasFinder y CRISPRminer (Cuadro 04). Además, corroboró que los cinco CDS identificados (Cas-S1-Cas-S5) pertenecen al sistema CRISPR/Cas subtipo I-E y que ya han sido reportados en otras cepas del género *Streptomyces*. Esto concuerda con los datos obtenidos de un análisis de 46 genomas completos de *Streptomyces*, donde lograron identificar 29 clúster de genes Cas, de los cuales 26 pertenecían a sistemas CRISPR/Cas Subtipo I-E (Zhang *et al.*, 2018b).

Finalmente, luego de consensuar la anotación de los cinco CDS (Cas-S1-Cas-S5) realizada por las herramientas bioinformáticas anteriormente mencionadas y contrastarla con bibliografía, detectamos que los genes Cas del sistema CRISPR/Cas subtipo I-E reciben varios nombres, lo cual añade confusión a la denominación de estos genes (Makarova *et al.*, 2011).

Aclarando, los genes *Cas1*, *Cas2* y *Cas3* también han sido denominados como *ygbT*, *ygbF* y *ygcb*, respectivamente (Babu *et al.*, 2011). Mientras que los cinco genes Cascade (*CasA*, *CasB*, *CasC*, *CasD* y *CasE*), fueron previamente etiquetados como *Cse1* (Cse = “CRISPR system from *E. coli*”), *Cse2*, *Cse4*, *Cas5e* (*Cse5*) y *Cse3*, respectivamente (Haft *et al.*, 2005). Y posteriormente también fueron referidos como *ygcl*, *ygck*, *ygcJ*, *ygcI* y *ygcH* (Babu *et al.*, 2011).



Para evitar dicha confusión se planteó mantener las denominaciones iniciales de *Cas1*, *Cas2* y *Cas3*; y denominar a los genes Casacade como *Cas8* (*Cas8e* para el subtipo I-E); *Cas11*; *Cas7*; *Cas5* (*Cas5e* para el subtipo I-E); y *Cas6* (*Cas6e* para el subtipo I-E), respectivamente (Makarova *et al.*, 2011; 2020; Charpentier *et al.*, 2015). Por ende, los CDS Cas-S1, Cas-S2, Cas-S3, Cas-S4 y Cas-S5 codificarían para las proteínas Cas6, Cas5, Cas7, Cas11 y Cas8, respectivamente (Cuadro 04). Respecto a las cuales sabemos lo siguiente:

**La proteína Cas6** es una endoribonucleasa enzimáticamente activa que interactúa físicamente con las subunidades Cas7 y Cas11 dentro del complejo Cascade y cuya función principal es procesar y unirse a los crRNA maduros en las etapas consecutivas del procesamiento e interferencia (Jackson *et al.*, 2014; Charpentier *et al.*, 2015; Makarova *et al.*, 2020; Liu & Doudna, 2020).

**La proteína Cas5** es una subunidad no catalítica de unión al ARN que forma parte del complejo Cascade y cuya función es unirse a la región 5' del crRNA maduro, interactuar con la subunidad Cas8 y facilitar la separación de cadenas del ADN invasor en la etapa de interferencia (Makarova *et al.*, 2015; 2020; Koonin *et al.*, 2017; Liu & Doudna, 2020).

**Las proteínas Cas7** son seis subunidades no catalíticas con dominio de reconocimiento al ARN que interactúa con las subunidades Cas5, Cas6, Cas8 y Cas11 dentro del complejo Cascade y tienen la función de unirse a la región central de los crRNA maduros para conformar una estructura central llamada “columna vertebral” del complejo Cascade (Makarova *et al.*, 2015; 2020; Liu & Doudna, 2020).

**Las proteínas Cas11** son dos subunidades  $\alpha$ -helicoidales dentro del Complejo Cascade, cuya función es interactuar con las seis subunidades Cas7 y posiblemente con el crRNA, y, además, conectar a las subunidades Cas6 y Cas8 ubicadas en los extremos del complejo Cascade, para mantener una estructura estable y funcional en forma de arco (Agari *et al.*, 2008; Jackson *et al.*, 2014; Makarova *et al.*, 2015; 2020).

**La proteína Cas8** es una subunidad enzimáticamente inactiva que interactúa con las subunidades Cas5, Cas7 y Cas11 dentro del complejo Cascade y cuya función principal es reconocer las secuencias PAM para que el complejo Cascade pueda unirse al ADN invasor y facilitar el acceso de Cas3 en la etapa de interferencia (Hochstrasser *et al.*, 2014; Cass *et al.*, 2015; Makarova *et al.*, 2015; 2020; Koonin *et al.*, 2017).

**Cuadro 04:** Anotación a través de Uniprot del clúster de genes Cas detectados por los servidores web CRISPRCasFinder y CRISPRminer.

Anotación por Uniprot						Anotación consensuada
CDS	Accesión	Proteína	Identidad	Subtipo	Organismo	
Cas-S1	A0A7K3H0W5	Cas6/Cse3/ CasE	94.70%	Cas I-E	<i>Streptomyces</i> sp. SID5789	Cas6
Cas-S2	A0A7K3H0V7	Cas5/CasD	97.50%	Cas I-E	<i>Streptomyces</i> sp. SID5789	Cas5
Cas-S3	A0A7K3H0V3	Cas	96.20%	-	<i>Streptomyces</i> sp. SID5789	Cas7
Cas-S4	A0A7K3H1A2	Cse2/CasB	95.00%	Cas I-E	<i>Streptomyces</i> sp. SID5789	Cas11
Cas-S5	A0A7K3H0T4	Cse1/CasA	95.70%	Cas I-E	<i>Streptomyces</i> sp. SID5789	Cas8

FUENTE: Elaboración propia.

Por otro lado, no llegamos a detectar los genes *Cas1*, *Cas2* y *Cas3* en el genoma de *Streptomyces* sp. B-146, los cuales son claves en los sistemas CRISPR/Cas funcionalmente activos y dinámicos. Las razones pueden ser varias, desde la falta de sensibilidad de las herramientas bioinformáticas usadas hasta la pérdida parcial de los genes Cas. Respecto a esto último, se sabe que la incorporación de espaciadores *self-targeting* que generan autoinmunidad poco agresiva puede conducir a la inactivación del locus CRISPR/cas, entre otras causas, por la pérdida de los genes Cas (Stern *et al.*, 2010). En ese sentido, podemos hipotetizar que quizás en algún punto de la evolución de la bacteria *Streptomyces* sp. B-146 la presencia del *self-targeting* pudo haber conducido a la pérdida de los genes cas y, por ende, la inactivación del sistema CRISPR/cas.

## V. CONCLUSIONES Y RECOMENDACIONES

El análisis del genoma de *Streptomyces* sp. B-146 ha revelado que esta bacteria podría albergar hasta 39 BGCs putativos presumiblemente productores de metabolitos secundarios bioactivos relacionados con actividades antimicrobianas, antitumorales, antivirales y otras bioactividades. La mayoría de los BGCs putativos detectados mostraron similitudes menores al 80% con otros BGCs previamente caracterizados, lo que incrementa significativamente la posibilidad de identificar a nivel experimental productos naturales con propiedades terapéuticas únicas. Aquí hemos logrado anotar manualmente cuatro BGCs relacionados con la biosíntesis de los terpenoides Xiamicina A y Albaflavenona, de las macrolactamas Pactamidas y del lantipéptido SapB. Los cuales teóricamente exhibirían bioactividades antivirales, antibacterianas y citotóxicas. Asimismo, se han identificado una gran cantidad genes de resistencia, predominantemente transportadores de membrana, que podrían desempeñar un rol fundamental en el flujo de salida de los productos naturales biosintetizados por *Streptomyces* sp. B146. Tomando en conjunto estos hallazgos, se demuestra el potencial biosintético de *Streptomyces* sp. B-146 y, en consecuencia, se recomienda extrapolar la búsqueda de productos naturales a nivel de laboratorio, aprovechando las estrategias experimentales más recientes de ingeniería genética, ingeniería metabólica, biosíntesis combinatoria y de bioprocesos para detectar productos naturales con propiedades farmacológicas novedosas.

Los análisis de filogenia basado en genoma completo y genes de ARNr 16S han demostrado que *Streptomyces* sp. B-146 es filogenéticamente próxima a *Streptomyces olivaceus*; no obstante, se recomienda análisis adicionales posteriores para definir si estamos frente a una nueva especie dentro del género bacteriano *Streptomyces*.

Por último, con la evaluación del genoma de *Streptomyces* sp. B-146 también se ha detectado parte de un sistema CRISPR/Cas Subtipo I-E, compuesto de dos pequeños arreglos CRISPR y un grupo de genes Cas que codifican para todas las subunidades del complejo Cascade, implicado en la etapa de procesamiento e interferencia del sistema CRISPR/Cas. Sin embargo, no se llegó a detectar los genes *Cas1* y *Cas2*, implicados en la etapa de adaptación; y el gen *Cas3*, determinante en la etapa de interferencia de este sistema de inmunidad adaptativa en procariotas.

## VI. REFERENCIAS BIBLIOGRÁFICAS

- AbuSara, N.F., Piercey, B.M., Moore, M.A., Shaikh, A.A., Nothias, L.F., Srivastava, S.K., Cruz-Morales, P., Dorrestein, P.C., Barona-Gómez, F., & Tahlan, K. (2019). Comparative genomics and metabolomics analyses of clavulanic acid-producing *Streptomyces* species provides insight into specialized metabolism. *Frontiers in Microbiology*, *10*, 2550. Doi: 10.3389/fmicb.2019.02550. PMID: 31787949. PMCID: PMC6856088.
- Acedo, J.Z., Bothwell, I.R., An, L., Trouth, A., Frazier, C., & van der Donk, W.A. (2019). O-Methyltransferase-mediated incorporation of a  $\beta$ -amino acid in lanthipeptides. *Journal of the American Chemical Society*, *141*(42), 16790-16801. Doi: 10.1021/jacs.9b07396. PMID: 31568727. PMCID: PMC6812601.
- Agari, Y., Yokoyama, S., Kuramitsu, S., & Shinkai, A. (2008). X-ray crystal structure of a CRISPR-associated protein, Cse2, from *Thermus thermophilus* HB8. *Proteins*, *73*(4), 1063-7. Doi: 10.1002/prot.22224. PMID: 18798563.
- Aggarwal, E., Chauhan, S., & Sareen, D. (2021). Thiopeptides encoding biosynthetic gene clusters mined from bacterial genomes. *Journal of Biosciences*, *46*, 36. Doi: 10.1007/s12038-021-00158-2. PMID: 33952727.
- Agrawal, P., Amir, S., Deepak, Barua, D., & Mohanty, D. (2021). RiPPMiner-Genome: a web resource for automated prediction of crosslinked chemical structures of RiPPs by genome mining. *Journal of Molecular Biology*, *433*(11), 166887. Doi: 10.1016/j.jmb.2021.166887. PMID: 33972022.
- Akesson, M., Dufour, M., Sloan, G.L., & Simmonds, R.S. (2007). Targeting of *streptococci* by zoocin A. *FEMS Microbiology Letters*, *270*(1), 155-61. Doi: 10.1111/j.1574-6968.2007.00669.x. PMID: 17319876.
- Alam, K., Mazumder, A., Sikdar, S., Zhao, Y.M., Hao, J., Song, C., Wang, Y., Sarkar, R., Islam, S., Zhang, Y., & Li, A. (2022). *Streptomyces*: The biofactory of secondary metabolites. *Frontiers in Microbiology*, *13*, 968053. Doi: 10.3389/fmicb.2022.968053. PMID: 36246257. PMCID: PMC9558229.

- Alanjary, M., Steinke, K., & Ziemert, N. (2019). AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential. *Nucleic Acids Research*, 47(W1), W276-W282. Doi: 10.1093/nar/gkz282. PMID: 30997504; PMCID: PMC6602446.
- Alkhnabashi, O.S., Shah, S.A., Garrett, R.A., Saunders, S.J., Costa, F., & Backofen, R. (2016). Characterizing leader sequences of CRISPR loci. *Bioinformatics*, 32(17), i576-i585. Doi: 10.1093/bioinformatics/btw454. PMID: 27587677.
- Al-Mutairi, N.M., Ahmad, S., & Mokaddas, E. (2011). First report of molecular detection of fluoroquinolone resistance-associated gyrA mutations in multidrug-resistant clinical *Mycobacterium tuberculosis* isolates in Kuwait. *BMC Research Notes*, 4, 123. Doi: 10.1186/1756-0500-4-123. PMID: 21492420. PMCID: PMC3095995.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-10. Doi: 10.1016/S0022-2836(05)80360-2. PMID: 2231712.
- Ambler, R.P. (1980). The structure of beta-lactamases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 289(1036), 321-31. Doi: 10.1098/rstb.1980.0049. PMID: 6109327.
- Ames, B.D., Lee, M.Y., Moody, C., Zhang, W., Tang, Y., & Tsai, S.C. (2011). Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway. *Biochemistry*, 50(39), 8392-406. Doi: 10.1021/bi200593m. PMID: 21870821. PMCID: PMC3189504.
- Amitai, G., & Sorek, R. (2016). CRISPR-Cas adaptation: insights into the mechanism of action. *Nature Reviews Microbiology*, 14(2), 67-76. Doi: 10.1038/nrmicro.2015.14. PMID: 26751509.
- Andersson, D.I., Balaban, N.Q., Baquero, F., Courvalin, P., Glaser, P., Gophna, U., Kishony, R., Molin, S., & Tønjum, T. (2020). Antibiotic resistance: turning evolutionary principles into clinical reality. *FEMS Microbiology Reviews*, 44(2), 171-188. Doi: 10.1093/femsre/fuaa001. PMID: 31981358.

- Arias-Orozco, P., Inklaar, M., Lanooij, J., Cebrián, R., & Kuipers, O.P. (2021). Functional expression and characterization of the highly promiscuous Lanthipeptide Synthetase SyncM, enabling the production of Lanthipeptides with a broad range of ring topologies. *ACS Synthetic Biology*, 10(10), 2579-2591. Doi: 10.1021/acssynbio.1c00224. PMID: 34554737. PMCID: PMC8524650.
- Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G., Camarero, J.A., ..., Truman, A.W., Vederas, J.C., Walsh, C.T., Walton, J.D., Wenzel, S.C., Willey, J.M., & 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), 108-60. Doi: 10.1039/c2np20085f. PMID: 23165928. PMCID: PMC3954855.
- Arzanlou, M., Chai, W.C., & Venter, H. (2017). Intrinsic, adaptive, and acquired antimicrobial resistance in Gram-negative bacteria. *Essays in Biochemistry*, 61(1), 49-59. Doi: 10.1042/EBC20160063. PMID: 28258229.
- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., & Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, 9, 75. Doi: 10.1186/1471-2164-9-75. PMID: 18261238. PMCID: PMC2265698.
- Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., Phanse, S., Joachimiak, A., Koonin, E.V., Savchenko, A., Emili, A., Greenblatt, J., Edwards, A.M., & Yakunin, A.F. (2011). A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Molecular Microbiology*, 79(2), 484-502. Doi: 10.1111/j.1365-2958.2010.07465.x. PMID: 21219465. PMCID: PMC3071548.
- Baranova, N., & Nikaido, H. (2002). The baeSR two-component regulatory system activates transcription of the yegMNOB (mdtABCD) transporter gene cluster in *Escherichia coli* and increases its resistance to novobiocin and deoxycholate. *Journal of Bacteriology*, 184(15), 4168-76. Doi: 10.1128/JB.184.15.4168-4176.2002. PMID: 12107134. PMCID: PMC135214.

- Barbosa, F., Pinto, E., Kijjoa, A., Pinto, M., & Sousa, E. (2020). Targeting antimicrobial drug resistance with marine natural products. *International Journal of Antimicrobial Agents*, 56(1), 106005. Doi: 10.1016/j.ijantimicag.2020.106005. PMID: 32387480.
- Barka, E.A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Meier-Kolthoff, J.P., Klenk, H.P., Clément, C., Ouhdouch, Y., & van Wezel, G.P. (2015). Taxonomy, physiology, and natural products of *actinobacteria*. *Microbiology and Molecular Biology Reviews*, 80(1), 1-43. Doi: 10.1128/MMBR.00019-15. PMID: 26609051. PMCID: PMC4711186.
- Barrangou, R., & Horvath, P. (2017). A decade of discovery: CRISPR functions and applications. *Nature Microbiology*, 2, 17092. Doi: 10.1038/nmicrobiol.2017.92. PMID: 28581505.
- Bay, D.C., Rommens, K.L., & Turner, R.J. (2008). Small multidrug resistance proteins: a multidrug transporter family that continues to grow. *Biochimica et Biophysica Acta*, 1778(9), 1814-38. Doi: 10.1016/j.bbamem.2007.08.015. PMID: 17942072.
- Becher, P.G., Verschut, V., Bibb, M.J., Bush, M.J., Molnár, B.P., Barane, E., Al-Bassam, M.M., Chandra, G., Song, L., Challis, G.L., Buttner, M.J., & Flärdh, K. (2020). Developmentally regulated volatiles geosmin and 2-methylisoborneol attract a soil arthropod to *Streptomyces* bacteria promoting spore dispersal. *Nature Microbiology*, 5(6), 821-829. DOI: 10.1038/s41564-020-0697-x. PMID: 32251369.
- Beis, K. (2015). Structural basis for the mechanism of ABC transporters. *Biochemical Society Transactions*, 43(5), 889-93. Doi: 10.1042/BST20150047. PMID: 26517899.
- Belknap, K.C., Park, C.J., Barth, B.M., & Adam, C.P. (2020). Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Scientific Reports*, 10, 2003. Doi: 10.1038/s41598-020-58904-9.
- Bentley, J., Hyatt, L.S., Ainley, K., Parish, J.H., Herbert, R.B., & White, G.R. (1993). Cloning and sequence analysis of an *Escherichia coli* gene conferring bicyclomycin resistance. *Gene*, 127(1), 117-20. Doi: 10.1016/0378-1119(93)90625-d. PMID: 8486276.



- Blanco, P., Hernando-Amado, S., Reales-Calderon, J.A., Corona, F., Lira, F., Alcalde-Rico, M., Bernardini, A., Sanchez, M.B., & Martinez, J.L. (2016). Bacterial Multidrug Efflux Pumps: Much more than Antibiotic resistance determinants. *Microorganisms*, 4(1), 14. Doi: 10.3390/microorganisms4010014. PMID: 27681908. PMCID: PMC5029519.
- 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. Doi: 10.1093/nar/gkab335. PMID: 33978755. PMCID: PMC8262755.
- Boetzer, M., & Pirovano, W. (2014). SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*, 15, 211. Doi: 10.1186/1471-2105-15-211.
- Bolger, A.M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-20. Doi: 10.1093/bioinformatics/btu170. PMID: 24695404. PMCID: PMC4103590.
- Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Thomason, J.A. 3rd., Stevens, R., Vonstein, V., Wattam, A.R., & Xia, F. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*, 5, 8365. Doi: 10.1038/srep08365. PMID: 25666585. PMCID: PMC4322359.
- Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuys, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., & van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 321(5891), 960-4. Doi: 10.1126/science.1159689. PMID: 18703739. PMCID: PMC5898235.
- Brown, K., Li, W., & Kaur, P. (2017). Role of Aromatic and Negatively Charged Residues of DrrB in Multisubstrate Specificity Conferred by the DrrAB System of *Streptomyces peucetius*. *Biochemistry*, 56(13), 1921-1931. Doi: 10.1021/acs.biochem.6b01155. PMID: 28272881.

- Brown, M.H., & Skurray, R.A. (2001). Staphylococcal multidrug efflux protein QacA. *Journal of Molecular Microbiology and Biotechnology*, 3(2), 163-70. PMID: 11321569.
- Bukowski, K., Kciuk, M., & Kontek, R. 2020. Mechanisms of multidrug resistance in cancer chemotherapy. *International Journal of Molecular Sciences*, 21(9), 3233. Doi: 10.3390/ijms21093233. PMID: 32370233. PMCID: PMC7247559.
- Burata, O.E., Yeh, T.J., Macdonald, C.B., & Stockbridge, R.B. (2022). Still rocking in the structural era: A molecular overview of the small multidrug resistance (SMR) transporter family. *Journal of Biological Chemistry*, 298(10), 102482. Doi: 10.1016/j.jbc.2022.102482. PMID: 36100040. PMCID: PMC9574504.
- Bush, K. (2013). The ABCD's of  $\beta$ -lactamase nomenclature. *Journal of Infection and Chemotherapy*, 19(4), 549-59. Doi: 10.1007/s10156-013-0640-7. PMID: 23828655.
- Bush, K., & Jacoby, G.A. (2010). Updated functional classification of beta-lactamases. *Antimicrobial Agents and Chemotherapy*, 54(3), 969-76. Doi: 10.1128/AAC.01009-09. PMID: 19995920. PMCID: PMC2825993.
- Cane, D.E., & Ikeda, H. (2012). Exploration and mining of the bacterial terpenome. *Accounts of Chemical Research*, 45(3), 463-72. Doi: 10.1021/ar200198d. PMID: 22039990. PMCID: PMC3288161.
- Capstick, D.S., Willey, J.M., Buttner, M.J., & Elliot, M.A. (2007). SapB and the chaplins: connections between morphogenetic proteins in *Streptomyces coelicolor*. *Molecular Microbiology*, 64(3), 602-13. Doi: 10.1111/j.1365-2958.2007.05674.x. PMID: 17462011.
- Cariss, S.J., Constantinidou, C., Patel, M.D., Takebayashi, Y., Hobman, J.L., Penn, C.W., & Avison, M.B. (2010). YieJ (CbrC) mediates CreBC-dependent colicin E2 tolerance in *Escherichia coli*. *Journal of Bacteriology*, 192(13), 3329-36. Doi: 10.1128/JB.01352-09. PMID: 20418396. PMCID: PMC2897664.
- Carte, J., Christopher, R.T., Smith, J.T., Olson, S., Barrangou, R., Moineau, S., Glover, C.V. 3rd., Graveley, B.R., Terns, R.M., & Terns, M.P. (2014). The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in *Streptococcus thermophilus*. *Molecular Microbiology*, 93(1), 98-112. Doi: 10.1111/mmi.12644. PMID: 24811454. PMCID: PMC4095994.

- Cass, S.D., Haas, K.A., Stoll, B., Alkhnbashi, O.S., Sharma, K., Urlaub, H., Backofen, R., Marchfelder, & A., Bolt, E.L. 2015. The role of Cas8 in type I CRISPR interference. *Bioscience Reports*, 35(3), e00197. Doi: 10.1042/BSR20150043. PMID: 26182359. PMCID: PMC4613674.
- Chalut, C. (2016). MmpL transporter-mediated export of cell-wall associated lipids and siderophores in *mycobacteria*. *Tuberculosis (Edinburgh, Scotland)*, 100, 32-45. Doi: 10.1016/j.tube.2016.06.004. PMID: 27553408.
- Chan, D.C.K., & Burrows, L.L. (2021). Thiopeptides: antibiotics with unique chemical structures and diverse biological activities. *The Journal of Antibiotics (Tokyo)*, 74(3), 161-175. Doi: 10.1038/s41429-020-00387-x. PMID: 33349675.
- Charpentier, E., Richter, H., van der Oost, J., & White, M.F. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiology Reviews*, 39(3), 428-41. Doi: 10.1093/femsre/fuv023. PMID: 25994611. PMCID: PMC5965381.
- Chavali, A. K., & Rhee, S. Y. (2018). Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. *Briefings in bioinformatics*, 19(5), 1022–1034. Doi: 10.1093/bib/bbx020.
- Chen, H., & Du, L. (2016). Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Applied Microbiology and Biotechnology*, 100(2), 541-57. Doi: 10.1007/s00253-015-7093-0. PMID: 26549236. PMCID: PMC4706475.
- Chen, L.Y., Wang, X.Q., Wang, Y.M., Geng, X., Xu, X.N., Su, C., Yang, Y.L., Tang, Y.J., Bai, F.W., & Zhao, X.Q. (2018). Genome mining of *Streptomyces xinghaiensis* NRRL B-24674T for the discovery of the gene cluster involved in anticomplement activities and detection of novel xiamycin analogs. *Applied Microbiology and Biotechnology*, 102(22), 9549-9562. Doi: 10.1007/s00253-018-9337-2. PMID: 30232534.
- Cheng, C., & Hua, Z.C. (2020). Lasso peptides: heterologous production and potential medical application. *Frontiers in Bioengineering and Biotechnology*, 8, 571165. Doi: 10.3389/fbioe.2020.571165. PMID: 33117783. PMCID: PMC7549694.

- Cheng, Y.Q., Coughlin, J.M., Lim, S.K., & Shen, B. (2009). Type I polyketide synthases that require discrete acyltransferases. *Methods in Enzymology*, 459, 165-86. Doi: 10.1016/S0076-6879(09)04608-4. PMID: 19362640. PMCID: PMC4480786.
- Chopra, I., & Roberts, M. (2001). Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiology and Molecular Biology Reviews*, 65(2), 232-60. Doi: 10.1128/MMBR.65.2.232-260.2001. PMID: 11381101. PMCID: PMC99026.
- Cohen-Khait, R., Harmalkar, A., Pham, P., Webby, M.N., Housden, N.G., Elliston, E., Hopper, J.T.S., Mohammed, S., Robinson, C.V., Gray, J.J., & Kleanthous, C. (2021). Colicin-mediated transport of DNA through the iron transporter FepA. *mBio*, 12(5), e0178721. Doi: 10.1128/mBio.01787-21. PMID: 34544275. PMCID: PMC8546555.
- Connell, S.R., Tracz, D.M., Nierhaus, K.H., & Taylor, D.E. (2003). Ribosomal protection proteins and their mechanism of tetracycline resistance. *Antimicrobial Agents and Chemotherapy*, 47(12), 3675-81. Doi: 10.1128/AAC.47.12.3675-3681.2003. PMID: 14638464. PMCID: PMC296194.
- Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E.P.C., Vergnaud, G., Gautheret, D., & Pourcel, C. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research*, 46(W1), W246-W251. Doi: 10.1093/nar/gky425. PMID: 29790974. PMCID: PMC6030898.
- Cox, C.L., Doroghazi, J.R., & Mitchell, D.A. (2015). The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles. *BMC Genomics*, 16, 778. Doi: 10.1186/s12864-015-2008-0. PMID: 26462797. PMCID: PMC4603692.
- Crits-Christoph, A., Bhattacharya, N., Olm, M.R., Song, Y.S., Banfield, J.F. (2020). Transporter genes in biosynthetic gene clusters predict metabolite characteristics and siderophore activity. *Genome Research*, 31(2), 239–50. Doi: 10.1101/gr.268169.120. PMID: 33361114. PMCID: PMC7849407.

- Cruz, A., Micaelo, N., Félix, V., Song, J.Y., Kitamura, S., Suzuki, S., & Mendo, S. (2013). sugE: A gene involved in tributyltin (TBT) resistance of *Aeromonas molluscorum* Av27. *The Journal of General and Applied Microbiology*, 59(1), 39-47. Doi: 10.2323/jgam.59.47. PMID: 23518517.
- Cruz, P.G., Fribley, A.M., Miller, J.R., Larsen, M.J., Schultz, P.J., Jacob, R.T., Tamayo-Castillo, G., Kaufman, R.J., & Sherman, D.H. (2015). Novel lobophorins inhibit oral cancer cell growth and induce Atf4- and chop-dependent cell death in murine fibroblasts. *ACS Medicinal Chemistry Letters*, 6(8), 877-81. Doi: 10.1021/acsmchemlett.5b00127. PMID: 26288688. PMCID: PMC4538430.
- Dahal, R.H., Nguyen, T.M., Pandey, R.P., Yamaguchi, T., Sohng, J.K., Noh, J., Myung, S.W., & Kim, J. 2020. The genome insights of *Streptomyces lannensis* T1317-0309 reveals actinomycin D production. *The Journal of Antibiotics (Tokyo)*, 73(12), 837-844. Doi: 10.1038/s41429-020-0343-0. PMID: 32641781.
- Davidson, A.L., Dassa, E., Orelle, C., & Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and Molecular Biology Reviews*, 72(2), 317-64. Doi: 10.1128/MMBR.00031-07. PMID: 18535149. PMCID: PMC2415747.
- D'Costa, V.M., Griffiths, E., Wright, G.D. (2007). Expanding the soil antibiotic resistome: exploring environmental diversity. *Current Opinion in Microbiology*, 10(5), 481-9. Doi: 10.1016/j.mib.2007.08.009. PMID: 17951101.
- Delmar, J.A., Su, C.C., Yu, E.W. (2014). Bacterial multidrug efflux transporters. *Annual Review of Biophysics*, 43, 93-117. Doi: 10.1146/annurev-biophys-051013-022855. PMID: 24702006. PMCID: PMC4769028.
- Devi, V., Harjai, K., & Chhibber, S. (2022). CRISPR-Cas systems: role in cellular processes beyond adaptive immunity. *Folia microbiologica (Praha)*, 67(6), 837-850. Doi: 10.1007/s12223-022-00993-2. PMID: 35854181. PMCID: PMC9296112.
- Devi, V., Harjai, K., & Chhibber, S. (2022a). Self-targeting spacers in CRISPR-array: Accidental occurrence or evolutionarily conserved phenomenon. *Journal of Basic Microbiology*, 62(1), 4-12. Doi: 10.1002/jobm.202100514. PMID: 34904260.

- Díez-Villaseñor, C., Almendros, C., García-Martínez, J., & Mojica, F.J. (2010). Diversity of CRISPR loci in *Escherichia coli*. *Microbiology (Reading)*, 156(5), 1351-1361. Doi: 10.1099/mic.0.036046-0. PMID: 28206910.
- Ding, L., Maier, A., Fiebig, H.H., Lin, W.H., & Hertweck, C. (2011). A family of multicyclic indolosesquiterpenes from a bacterial endophyte. *Organic & Biomolecular Chemistry*, 9(11), 4029-31. Doi: 10.1039/c1ob05283g. PMID: 21528153.
- Ding, L., Münch, J., Goerls, H., Maier, A., Fiebig, H.H., Lin, W.H., & Hertweck, C. (2010). Xiamycin, a pentacyclic indolosesquiterpene with selective anti-HIV activity from a bacterial mangrove endophyte. *Bioorganic & Medicinal Chemistry Letters*, 20(22), 6685-7. Doi: 10.1016/j.bmcl.2010.09.010. PMID: 20880706.
- Dittrich, W., & Schrempf, H. (1992). The unstable tetracycline resistance gene of *Streptomyces lividans* 1326 encodes a putative protein with similarities to translational elongation factors and Tet(M) and Tet(O) proteins. *Antimicrobial Agents and Chemotherapy*, 36(5), 1119-24. Doi: 10.1128/AAC.36.5.1119. PMID: 1510403; PMCID: PMC188846.
- Doyle, D., McDowall, K.J., Butler, M.J., & Hunter, I.S. (1991). Characterization of an oxytetracycline-resistance gene, *otrA*, of *Streptomyces rimosus*. *Molecular Microbiology*, 5(12), 2923-33. Doi: 10.1111/j.1365-2958.1991.tb01852.x. PMID: 1809836.
- Dunbar, K.L., Melby, J.O., & Mitchell, D.A. (2012). YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nature Chemical Biology*, 8(6), 569-75. Doi: 10.1038/nchembio.944. PMID: 22522320. PMCID: PMC3428213.
- Edgar, R.C. (2007). PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics*, 8, 18. Doi: 10.1186/1471-2105-8-18. PMID: 17239253. PMCID: PMC1790904.
- Ero, R., Yan, X.F., & Gao, Y.G. (2021). Ribosome protection proteins-"new" players in the global arms race with antibiotic-resistant pathogens. *International Journal of Molecular Sciences*, 22(10), 5356. Doi: 10.3390/ijms22105356. PMID: 34069640. PMCID: PMC8161019.

- Felsenstein J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39(4), 783-791. Doi: 10.1111/j.1558-5646.1985.tb00420.x. PMID: 28561359.
- Finking, R., & Marahiel, M.A. (2004). Biosynthesis of nonribosomal peptides 1. *Annual Review of Microbiology*, 58, 453-88. Doi: 10.1146/annurev.micro.58.030603.123615. PMID: 15487945.
- Fischbach, M.A., & Walsh, C.T. (2006). Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chemical Reviews*, 106(8), 3468-96. Doi: 10.1021/cr0503097. PMID: 16895337.
- Fischer, U., Hertlein, S., & Grimm, C. (2015). The structure of apo ArnA features an unexpected central binding pocket and provides an explanation for enzymatic cooperativity. *Acta crystallographica. Section D, biological crystallography*, 71(Pt3), 687-96. Doi: 10.1107/S1399004714026686. PMID: 25760615.
- Flores Clavo, R., Ruiz Quiñones, N., Hernández-Tasco, Á.J., José Salvador, M., Tasca Gois Ruiz, A.L., de Oliveira Braga, L.E., Henrique Costa, J., Pacheco Fill, T., Arce Gil, Z.L., Serquen Lopez, L.M., & Fantinatti Garboggini, F. (2021). Evaluation of antimicrobial and antiproliferative activities of *Actinobacteria* isolated from the saline lagoons of northwestern Peru. *PLoS One*, 16(9), e0240946. Doi: 10.1371/journal.pone.0240946. PMID: 34495972. PMCID: PMC8425546.
- Fu, Y., Jaarsma, A.H., & Kuipers, O.P. (2021). Antiviral activities and applications of ribosomally synthesized and post-translationally modified peptides (RiPPs). *Cellular and Molecular Life Sciences*, 78(8), 3921-3940. Doi: 10.1007/s00018-021-03759-0. PMID: 33532865. PMCID: PMC7853169.
- Fu, Y., Xu, Y., Ruijne, F., & Kuipers, O.P. (2023). Engineering lanthipeptides by introducing a large variety of RiPP modifications to obtain new-to-nature bioactive peptides. *FEMS Microbiology Reviews*, 47(3), fuad017. Doi: 10.1093/femsre/fuad017. PMID: 37096385. PMCID: PMC10373908.
- Fujisaki, S., Ohnuma, S., Horiuchi, T., Takahashi, I., Tsukui, S., Nishimura, Y., Nishino, T., Kitabatake, M., & Inokuchi, H. (1996). Cloning of a gene from *Escherichia coli* that confers resistance to fosmidomycin as a consequence of amplification. *Gene*, 175(1-2), 83-7. Doi: 10.1016/0378-1119(96)00128-x. PMID: 8917080.

- Gaskell, A.A., Giovinazzo, J.A., Fonte, V., & Willey, J.M. (2012). Multi-tier regulation of the *Streptomyces* morphogenetic peptide SapB. *Molecular Microbiology*, 84(3), 501-15. Doi: 10.1111/j.1365-2958.2012.08041.x. PMID: 22486809.
- Gatzeva-Topalova, P.Z., May, A.P., & Sousa, M.C. (2004). Crystal structure of *Escherichia coli* ArnA (PmrI) decarboxylase domain. A key enzyme for lipid A modification with 4-amino-4-deoxy-L-arabinose and polymyxin resistance. *Biochemistry*, 43(42), 13370-9. Doi: 10.1021/bi048551f. PMID: 15491143. PMCID: PMC2680612.
- Geng, M., & Smith, L. (2018). Improving the attrition rate of Lanthipeptide discovery for commercial applications. *Expert Opinion on Drug Discovery*, 13(2), 155-167. Doi: 10.1080/17460441.2018.1410137. PMID: 29195488.
- Ghosh, D., Veeraraghavan, B., Elangovan, R., & Vivekanandan, P. (2020). Antibiotic resistance and epigenetics: more to it than meets the eye. *Antimicrobial Agents and Chemotherapy*, 64(2), e02225-19. Doi: 10.1128/AAC.02225-19. PMID: 31740560. PMCID: PMC6985748.
- Gosse, J.T., Ghosh, S., Sproule, A., Overy, D., Cheeptham, N., & Boddy, C.N. (2019). Whole genome sequencing and metabolomic study of cave *Streptomyces* isolates ICC1 and ICC4. *Frontiers in Microbiology*, 10, 1020. Doi: 10.3389/fmicb.2019.01020. PMID: 31134037. PMCID: PMC6524458.
- Goto, Y., Li, B., Claesen, J., Shi, Y., Bibb, M.J., & van der Donk, W.A. (2010). Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. *PLOS Biology*, 8(3), e1000339. Doi: 10.1371/journal.pbio.1000339. PMID: 20351769. PMCID: PMC2843593.
- Goto, Y., Okesli, A., & van der Donk, W.A. (2011). Mechanistic studies of Ser/Thr dehydration catalyzed by a member of the LanL lanthionine synthetase family. *Biochemistry*, 50(5), 891-8. Doi: 10.1021/bi101750r. PMID: 21229987. PMCID: PMC3031989.
- Grigalunas, M., Brakmann, S., & Waldmann, H. (2022). Chemical evolution of natural product structure. *Journal of the American Chemical Society*, 144(8), 3314-3329. Doi: 10.1021/jacs.1c11270. PMID: 35188375. PMCID: PMC8895405.



- Guijarro, J., Santamaria, R., Schauer, A., & Losick, R. (1988). Promoter determining the timing and spatial localization of transcription of a cloned *Streptomyces coelicolor* gene encoding a spore-associated polypeptide. *Journal of Bacteriology*, 170(4), 1895-901. Doi: 10.1128/jb.170.4.1895-1901.1988. PMID: 2450872. PMCID: PMC211047.
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072-1075. Doi: 10.1093/bioinformatics/btt086. PMID: 23422339. PMCID: PMC3624806.
- Gürtler, H., Pedersen, R., Anthoni, U., Christophersen, C., Nielsen, P.H., Wellington, E.M., Pedersen, C., & Bock, K. (1994). Albaflavenone, a sesquiterpene ketone with a zizaene skeleton produced by a *streptomycete* with a new rope morphology. *The Journal of Antibiotics (Tokyo)*, 47(4), 434-9. Doi: 10.7164/antibiotics.47.434. PMID: 8195043.
- Haft, D.H., Selengut, J., Mongodin, E.F., Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLOS Computational Biology*, 1(6), e60. Doi: 10.1371/journal.pcbi.0010060. PMID: 16292354. PMCID: PMC1282333.
- Haghshenas, E., Hach, F., Sahinalp S.C., & Chauve, C. (2016). CoLoRMap: correcting long reads by mapping short reads. *Bioinformatics*, 32(17), i545–i551. Doi: 10.1093/bioinformatics/btw463.
- Hassan, K.A., Liu, Q., Elbourne, L.D.H., Ahmad, I., Sharples, D., Naidu, V., Chan, C.L., Li, L., Harborne, S.P.D., Pokhrel, A., Postis, V.L.G., Goldman, A., Henderson, P.J.F., & Paulsen, I.T. (2018). Pacing across the membrane: the novel PACE family of efflux pumps is widespread in Gram-negative pathogens. *Research in Microbiology*, 169(7-8), 450-454. Doi: 10.1016/j.resmic.2018.01.001. PMID: 29409983. PMCID: PMC6195760.
- Hassler, H.B., Probert, B., Moore, C., Lawson, E., Jackson, R.W., Russell, B.T., & Richards, V.P. (2022). Phylogenies of the 16S rRNA gene and its hypervariable regions lack concordance with core genome phylogenies. *Microbiome*, 10(1), 104. Doi: 10.1186/s40168-022-01295-y. PMID: 35799218. PMCID: PMC9264627.

- Hegemann, J.D. (2020). Factors governing the thermal stability of lasso peptides. *Chembiochem*, 21(1-2), 7-18. Doi: 10.1002/cbic.201900364. PMID: 31243865.
- Henz, S.R., Huson, D.H., Auch, A.F., Nieselt-Struwe, K., & Schuster, S.C. (2005). Whole-genome prokaryotic phylogeny. *Bioinformatics*, 21(10), 2329-35. Doi: 10.1093/bioinformatics/bth324. PMID: 15166018.
- Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., & Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proceedings of the National Academy of Sciences of the United States of America*, 111(18), 6618-23. Doi: 10.1073/pnas.1405079111. PMID: 24748111. PMCID: PMC4020112.
- Horvath, P., Coûté-Monvoisin, A.C., Romero, D.A., Boyaval, P., Fremaux, C., & Barrangou, R. (2009). Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *International Journal of Food Microbiology*, 131(1), 62-70. Doi: 10.1016/j.ijfoodmicro.2008.05.030. PMID: 18635282.
- Horvath, P., Romero, D.A., Coûté-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., & Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *Journal of Bacteriology*, 190(4), 1401-12. Doi: 10.1128/JB.01415-07. PMID: 18065539. PMCID: PMC2238196.
- Hoskisson, P.A., & van Wezel, G.P. (2019). *Streptomyces coelicolor*. *Trends in Microbiology*, 27(5), 468-469. Doi: 10.1016/j.tim.2018.12.008. PMID: 30621999.
- Huang, L., Wu, C., Gao, H., Xu, C., Dai, M., Huang, L., Hao, H., Wang, X., & Cheng, G. (2022). Bacterial multidrug efflux pumps at the frontline of antimicrobial resistance: An overview. *Antibiotics (Basel)*, 11(4), 520. Doi: 10.3390/antibiotics11040520. PMID: 35453271. PMCID: PMC9032748.
- Hunt, M., Silva, N.D., Otto, T.D., Parkhill, J., Keane, J.A., & Harris, S.R. (2015). Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biology*, 16, 294. Doi: 10.1186/s13059-015-0849-0. PMID: 26714481. PMCID: PMC4699355.

- Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J., & Wiedenheft, B. (2014). Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science*, 345(6203), 1473-9. Doi: 10.1126/science.1256328. PMID: 25103409. PMCID: PMC4188430.
- Jacoby, G.A. (2009). AmpC beta-lactamases. *Clinical Microbiology Reviews*, 22(1), 161-82. Doi: 10.1128/CMR.00036-08. PMID: 19136439. PMCID: PMC2620637.
- Jansen, R., Embden, J.D., Gaastra, W., & Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, 43(6), 1565-75. Doi: 10.1046/j.1365-2958.2002.02839.x. PMID: 11952905.
- Jensen, P.R. (2016). Natural products and the gene cluster revolution. *Trends in Microbiology*, 24(12), 968-977. Doi: 10.1016/j.tim.2016.07.006. PMID: 27491886. PMCID: PMC5123934.
- Jones, S.E., & Elliot, M.A. (2018). 'Exploring' the regulation of *Streptomyces* growth and development. *Current Opinion in Microbiology*, 42, 25-30. Doi: 10.1016/j.mib.2017.09.009. PMID: 29024914.
- Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., Kohara, Y., Fujiyama, A., Hayashi, T., & Itoh, T. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, 24(8), 1384-95. Doi: 10.1101/gr.170720.113.
- Kanehisa, M. (2017). Enzyme annotation and metabolic reconstruction using KEGG. *Methods in Molecular Biology*, 1611, 135-145. Doi: 10.1007/978-1-4939-7015-5\_11. PMID: 28451977.
- Kanehisa, M., Goto, S., Kawashima, S., & Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Research*, 30(1), 42-6. Doi: 10.1093/nar/30.1.42. PMID: 11752249. PMCID: PMC99091.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*, 42(Database issue), D199-205. Doi: 10.1093/nar/gkt1076. PMID: 24214961. PMCID: PMC3965122.

- Kanehisa, M., Sato, Y., & Morishima, K. (2016<sup>a</sup>). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology*, 428(4), 726-731. Doi: 10.1016/j.jmb.2015.11.006. PMID: 26585406.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016<sup>b</sup>). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1), D457-62. Doi: 10.1093/nar/gkv1070. PMID: 26476454. PMCID: PMC4702792.
- Karimi, Z., Ahmadi, A., Najafi, A., & Ranjbar, R. (2018). Bacterial CRISPR regions: General features and their potential for epidemiological molecular typing studies. *The Open Microbiology Journal*, 12, 59-70. Doi: 10.2174/1874285801812010059. PMID: 29755603. PMCID: PMC5925864.
- Katsuyama, Y., & Ohnishi, Y. (2012). Type III polyketide synthases in microorganisms. *Methods in Enzymology*, 515, 359-77. Doi: 10.1016/B978-0-12-394290-6.00017-3. PMID: 22999182.
- Kearns, D.B., Chu, F., Rudner, R., & Losick, R. (2004). Genes governing swarming in *Bacillus subtilis* and evidence for a phase variation mechanism controlling surface motility. *Molecular Microbiology*, 52(2), 357-69. Doi: 10.1111/j.1365-2958.2004.03996.x. PMID: 15066026.
- Keatinge-Clay A.T. (2012). The structures of type I polyketide synthases. *Natural Product Reports*, 29(10), 1050-73. Doi: 10.1039/c2np20019h. PMID: 22858605.
- Kermani, A.A., Macdonald, C.B., Burata, O.E., Ben Koff, B., Koide, A., Denbaum, E., Koide, S., & Stockbridge, R.B. (2020). The structural basis of promiscuity in small multidrug resistance transporters. *Nature Communications*, 11(1), 6064. Doi: 10.1038/s41467-020-19820-8. PMID: 33247110. PMCID: PMC7695847.
- Kersten, R.D., Yang, Y.L., Xu, Y., Cimermancic, P., Nam, S.J., Fenical, W., Fischbach, M.A., Moore, B.S., & Dorrestein, P.C. (2011). A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nature Chemical Biology*, 7(11), 794-802. Doi: 10.1038/nchembio.684. PMID: 21983601. PMCID: PMC3258187.

- Khabthani, S., Rolain, J.M., Merhej, V. (2021). *In silico/in vitro* strategies leading to the discovery of new nonribosomal peptide and polyketide antibiotics active against human pathogens. *Microorganisms*, 9(11), 2297. Doi: 10.3390/microorganisms9112297. PMID: 34835423. PMCID: PMC8625390.
- Kim, S.H., Ha, T.K., Oh, W.K., Shin, J., & Oh, D.C. (2016). Antiviral indolosesquiterpenoid xiamycins C-E from a Halophilic *actinomycete*. *Journal of Natural Products*, 79(1), 51-8. Doi: 10.1021/acs.jnatprod.5b00634. PMID: 26698879.
- Kim, W., Hwang, S., Lee, N., Lee, Y., Cho, S., Palsson, B., & Cho, B.K. (2020). Transcriptome and translome profiles of *Streptomyces* species in different growth phases. *Scientific Data*, 7(1), 138. Doi: 10.1038/s41597-020-0476-9. PMID: 32385251. PMCID: PMC7210306.
- Knerr, P.J., & van der Donk, W.A. (2012). Discovery, biosynthesis, and engineering of lantipeptides. *Annual Review of Biochemistry*, 81, 479-505. Doi: 10.1146/annurev-biochem-060110-113521. PMID: 22404629.
- Kodani, S., & Unno, K. (2020). How to harness biosynthetic gene clusters of lasso peptides. *Journal of Industrial Microbiology & Biotechnology*, 47(9-10), 703-714. Doi: 10.1007/s10295-020-02292-6. PMID: 32705462.
- Kodani, S., Hudson, M.E., Durrant, M.C., Buttner, M.J., Nodwell, J.R., & Willey, J.M. (2004). The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene ramS in *Streptomyces coelicolor*. *Proceedings of the National Academy of Sciences of the United States of America*, 101(31), 11448-53. Doi: 10.1073/pnas.0404220101. PMID: 15277670. PMCID: PMC509221.
- Komaki, H., & Harayama, S. (2006). Sequence diversity of type-II polyketide synthase genes in *Streptomyces*. *Actinomycetologica*, 20(2), 42-48. Doi:10.3209/saj.20.42.
- Komaki, H., Sakurai, K., Hosoyama, A., Kimura, A., Igarashi, Y., & Tamura, T. (2018). Diversity of nonribosomal peptide synthetase and polyketide synthase gene clusters among taxonomically close *Streptomyces* strains. *Scientific Reports*, 8(1), 6888. Doi: 10.1038/s41598-018-24921-y. PMID: 29720592. PMCID: PMC5932044.

- Komaki, H., Sakurai, K., Hosoyama, A., Kimura, A., Trujilo, M.E., Igarashi, Y., & Tamura, T. (2020). Diversity of PKS and NRPS gene clusters between *Streptomyces abyssomicinicus* sp. nov. and its taxonomic neighbor. *The Journal of Antibiotics (Tokyo)*, 73(3), 141-151. Doi: 10.1038/s41429-019-0261-1. PMID: 31853029.
- Koonin, E.V., Makarova, K.S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology*, 37, 67-78. Doi: 10.1016/j.mib.2017.05.008. PMID: 28605718. PMCID: PMC5776717.
- Kremer, F.S., Eslabão, M.R., Dellagostin, O.A., & Pinto, L.D. (2016). Genix: a new online automated pipeline for bacterial genome annotation. *FEMS Microbiology Letters*, 363(23), fnw263. Doi: 10.1093/femsle/fnw263. PMID: 27856568.
- Krügel, H., Licht, A., Biedermann, G., Petzold, A., Lassak, J., Hupfer, Y., Schlott, B., Hertweck, C., Platzer, M., Brantl, S., & Saluz, H.P. (2010). Cervimycin C resistance in *Bacillus subtilis* is due to a promoter up-mutation and increased mRNA stability of the constitutive ABC-transporter gene bmrA. *FEMS Microbiology Letters*, 313(2), 155-63. Doi: 10.1111/j.1574-6968.2010.02143.x. PMID: 21077936.
- Kumar, S., Lekshmi, M., Parvathi, A., Ojha, M., Wenzel, N., & Varela, M.F. 2020. Functional and structural roles of the Major Facilitator Superfamily bacterial multidrug efflux pumps. *Microorganisms*, 8(2), 266. Doi: 10.3390/microorganisms8020266. PMID: 32079127. PMCID: PMC7074785.
- Kuzuyama, T. 2017. Biosynthetic studies on terpenoids produced by *Streptomyces*. *The Journal of Antibiotics (Tokyo)*, 70(7), 811-818. Doi: 10.1038/ja.2017.12. PMID: 28196976. PMCID: PMC5509993.
- Kwon, M.J., Steiniger, C., Cairns, T.C., Wisecaver, J.H., Lind, A.L., Pohl, C., Regner, C., Rokas, A., & Meyer, V. (2021). Beyond the biosynthetic gene cluster paradigm: genome-wide coexpression networks connect clustered and unclustered transcription factors to secondary metabolic pathways. *Microbiology Spectrum*, 9(2), e0089821. Doi: 10.1128/Spectrum.00898-21. PMID: 34523946. PMCID: PMC8557879.
- Lagedroste, M., Smits, S.H.J., & Schmitt, L. (2021). Importance of the leader peptide sequence on the lanthipeptide secretion level. *The FEBS Journal*, 288(14), 4348-4363. Doi: 10.1111/febs.15724. PMID: 33482024.

- Le Bideau, F., Kousara, M., Chen, L., Wei, L., & Dumas, F. (2017). Tricyclic sesquiterpenes from marine origin. *Chemical Reviews*, 117(9), 6110-6159. Doi: 10.1021/acs.chemrev.6b00502. PMID: 28379015.
- Lee, N., Hwang, S., Kim, J., Cho, S., Palsson, B., & Cho, B.K. (2020<sup>b</sup>). Mini review: Genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*. *Computational and Structural Biotechnology Journal*, 18, 1548-1556. ISSN: 2001-0370. Doi: 10.1016/j.csbj.2020.06.024.
- Lee, N., Kim, W., Hwang, S., Lee, Y., Cho, S., Palsson, B., & Cho, B.K. (2020<sup>a</sup>). Thirty complete *Streptomyces* genome sequences for mining novel secondary metabolite biosynthetic gene clusters. *Scientific Data*, 7, 55. Doi: 10.1038/s41597-020-0395-9.
- Lee, S.W., Mitchell, D.A., Markley, A.L., Hensler, M.E., Gonzalez, D., Wohlrab, A., Dorrestein, P.C., Nizet, V., & Dixon, J.E. (2008). Discovery of a widely distributed toxin biosynthetic gene cluster. *Proceedings of the National Academy of Sciences of the United States of America*, 105(15), 5879-84. Doi: 10.1073/pnas.0801338105. PMID: 18375757. PMCID: PMC2311365.
- Lei, Y., Asamizu, S., Ishizuka, T., & Onaka, H. (2023). Regulation of multidrug efflux pumps by TetR family transcriptional repressor negatively affects secondary metabolism in *Streptomyces coelicolor* A3(2). *Applied and Environmental Microbiology*, 89(3), e0182222. Doi: 10.1128/aem.01822-22. PMID: 36790176. PMCID: PMC10056966.
- Li, H., Zhang, Q., Li, S., Zhu, Y., Zhang, G., Zhang, H., Tian, X., Zhang, S., Ju, J., & Zhang, C. (2012). Identification and characterization of xiamycin A and oxiamycin gene cluster reveals an oxidative cyclization strategy tailoring indolosesquiterpene biosynthesis. *Journal of the American Chemical Society*, 134(21), 8996-9005. Doi: 10.1021/ja303004g. PMID: 22591327.
- Li, W., Sharma, M., & Kaur, P. (2014). The DrrAB efflux system of *Streptomyces peucetius* is a multidrug transporter of broad substrate specificity. *Journal of Biological Chemistry*, 289(18), 12633-46. Doi: 10.1074/jbc.M113.536136. PMID: 24634217. PMCID: PMC4007453.

- Li, X., Wang, H., Li, Y., & Du, L. (2019). Construction of a hybrid gene cluster to reveal coupled ring formation-hydroxylation in the biosynthesis of HSAF and analogues from *Lysobacter enzymogenes*. *Medchemcomm*, 10(6), 907-912. Doi: 10.1039/c9md00154a. PMID: 31303988. PMCID: PMC6595964.
- Li, X.Z., Plésiat, P., & Nikaido, H. (2015). The challenge of efflux-mediated antibiotic resistance in Gram-negative bacteria. *Clinical Microbiology Reviews*, 28(2), 337-418. Doi: 10.1128/CMR.00117-14. PMID: 25788514. PMCID: PMC4402952.
- Li, X.Z., Zhang, L., & Nikaido, H. (2004). Efflux pump-mediated intrinsic drug resistance in *Mycobacterium smegmatis*. *Antimicrobial Agents and Chemotherapy*, 48(7), 2415-23. Doi: 10.1128/AAC.48.7.2415-2423.2004. PMID: 15215089. PMCID: PMC434187.
- Li, Z., Zhu, D., & Shen, Y. (2018). Discovery of novel bioactive natural products driven by genome mining. *Database for Drug Research*, 12(6), 318-328. Doi: 10.5582/ddt.2018.01066. PMID: 30674766.
- Lillestøl, R.K., Redder, P., Garrett, R.A., & Brügger, K. (2006). A putative viral defence mechanism in archaeal cells. *Archaea*, 2(1), 59-72. Doi: 10.1155/2006/542818. PMID: 16877322. PMCID: PMC2685585.
- Lillestøl, R.K., Shah, S.A., Brügger, K., Redder, P., Phan, H., Christiansen, J., & Garrett, R.A. (2009). CRISPR families of the *crenarchaeal* genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Molecular Microbiology*, 72(1), 259-72. Doi: 10.1111/j.1365-2958.2009.06641.x. PMID: 19239620.
- Lin, X., & Cane, D.E. (2009). Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor*. Mechanism and stereochemistry of the enzymatic formation of epi-isozizaene. *Journal of the American Chemical Society*, 131(18), 6332-3. Doi: 10.1021/ja901313v. PMID: 19385616. PMCID: PMC2690639.
- Liu, J., Takiff, H.E., & Nikaido, H. (1996). Active efflux of fluoroquinolones in *Mycobacterium smegmatis* mediated by LfrA, a multidrug efflux pump. *Journal of Bacteriology*, 178(13), 3791-5. Doi: 10.1128/jb.178.13.3791-3795.1996. PMID: 8682782. PMCID: PMC232638.



- Liu, T., Ma, X., Yu, J., Yang, W., Wang, G., Wang, Z., Ge, Y., Song, J., Han, H., Zhang, W., Yang, D., Liu, X., & Ma, M. (2021). Rational generation of lasso peptides based on biosynthetic gene mutations and site-selective chemical modifications. *Chemical Science*, 12(37), 12353-12364. Doi: 10.1039/d1sc02695j. PMID: 34603665. PMCID: PMC8480316.
- Liu, T.Y., & Doudna, J.A. (2020). Chemistry of Class 1 CRISPR-Cas effectors: Binding, editing, and regulation. *Journal of Biological Chemistry*, 295(42), 14473-14487. Doi: 10.1074/jbc.REV120.007034. PMID: 32817336. PMCID: PMC7573268.
- Luo, X., Zambaldo, C., Liu, T., Zhang, Y., Xuan, W., Wang, C., Reed, S.A., Yang, P.Y., Wang, R.E., Javahishvili, T., Schultz, P.G., & Young, T.S. (2016). Recombinant thiopeptides containing noncanonical amino acids. *Proceedings of the National Academy of Sciences of the United States of America*, 113(13), 3615-20. Doi: 10.1073/pnas.1602733113. PMID: 26976568. PMCID: PMC4822594.
- Ma, S., Huang, Y., Xie, F., Gong, Z., Zhang, Y., Stojkoska, A., & Xie, J. (2020). Transport mechanism of *Mycobacterium tuberculosis* MmpL/S family proteins and implications in pharmaceutical targeting. *Journal of Biological Chemistry*, 401(3), 331-348. Doi: 10.1515/hsz-2019-0326. PMID: 31652116.
- Mahizan, N.A., Yang, S.K., Moo, C.L., Song, A.A., Chong, C.M., Chong, C.W., Abushelaibi, A., Lim, S.E., & Lai, K.S. (2019). Terpene derivatives as a potential agent against Antimicrobial Resistance (AMR) pathogens. *Molecules*, 24(14), 2631. Doi: 10.3390/molecules24142631. PMID: 31330955. PMCID: PMC6680751.
- Mak, S., Xu, Y., & Nodwell, J.R. (2014). The expression of antibiotic resistance genes in antibiotic-producing bacteria. *Molecular Microbiology*, 93(3), 391-402. Doi: 10.1111/mmi.12689. PMID: 24964724.
- Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., van der Oost, J., & Koonin, E.V. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology*, 9(6), 467-77. Doi: 10.1038/nrmicro2577. PMID: 21552286. PMCID: PMC3380444.

- Makarova, K.S., Wolf, Y.I., & Koonin, E.V. (2013). The basic building blocks and evolution of CRISPR-CAS systems. *Biochemical Society Transactions*, 41(6), 1392-400. Doi: 10.1042/BST20130038. PMID: 24256226. PMCID: PMC5898231.
- Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J., Terns, R.M., Terns, M.P., White, M.F., Yakunin, A.F., Garrett, R.A., van der Oost, J., Backofen, R., & Koonin, E.V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*, 13(11), 722-36. Doi: 10.1038/nrmicro3569. PMID: 26411297. PMCID: PMC5426118.
- Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Scott, D., Shah, S.A., Siksny, V., Terns, M.P., Venclovas, Č., White, M.F., Yakunin, A.F., Yan, W., Zhang, F., Garrett, R.A., Backofen, R., van der Oost, J., Barrangou, R., & Koonin, E.V. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, 18(2), 67-83. Doi: 10.1038/s41579-019-0299-x. PMID: 31857715. PMCID: PMC8905525.
- Marahiel, M.A. (2009). Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *Journal of Peptide Science*, 15(12), 799-807. Doi: 10.1002/psc.1183. PMID: 19827002.
- Martín-Sánchez, L., Singh, K.S., Avalos, M., van Wezel, G.P., Dickschat, J.S., & Garbeva, P. (2019). Phylogenomic analyses and distribution of terpene synthases among *Streptomyces*. *Beilstein Journal of Organic Chemistry*, 15, 1181-1193. Doi: 10.3762/bjoc.15.115. PMID: 31293665; PMCID: PMC6604706.
- Masyita, A., Mustika Sari, R., Dwi Astuti, A., Yasir, B., Rahma Rumata, N., Emran, T.B., Nainu, F., & Simal-Gandara, J. (2022). Terpenes and terpenoids as main bioactive compounds of essential oils, their roles in human health and potential application as natural food preservatives. *Food Chemistry: X*, 13, 100217. Doi: 10.1016/j.fochx.2022.100217. PMID: 35498985. PMCID: PMC9039924.
- McDermott, P.F., Walker, R.D., & White, D.G. (2003). Antimicrobials: modes of action and mechanisms of resistance. *International Journal of Toxicology*, 22(2), 135-43. Doi: 10.1080/10915810305089. PMID: 12745995.

- McIntosh, J.A., & Schmidt, E.W. (2010). Marine molecular machines: heterocyclization in cyanobactin biosynthesis. *Chembiochem*, 11(10), 1413-21. Doi: 10.1002/cbic.201000196. PMID: 20540059. PMCID: PMC3397156.
- Medlock, G.L., Moutinho, T.J., & Papin, J.A. (2020). Medusa: Software to build and analyze ensembles of genome-scale metabolic network reconstructions. *PLOS Computational Biology*, 16(4): e1007847. Doi: 10.1371/journal.pcbi.1007847.
- Melby, J.O., Nard, N.J., & Mitchell, D.A. (2011). Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. *Current Opinion in Chemical Biology*, 15(3), 369-78. Doi: 10.1016/j.cbpa.2011.02.027. PMID: 21429787. PMCID: PMC3947797.
- Meng, Z., Yu, H., Li, L., Tao, W., Chen, H., Wan, M., Yang, P., Edmonds, D.J., Zhong, J., & Li, A. (2015). Total synthesis and antiviral activity of indolosesquiterpenoids from the xiamycin and oridamycin families. *Nature Communications*, 6, 6096. Doi: 10.1038/ncomms7096. PMID: 25648883. PMCID: PMC4347019.
- Messiaen, A.S., Verbrugghen, T., Declerck, C., Ortmann, R., Schlitzer, M., Nelis, H., Van Calenbergh, S., & Coenye, T. (2011). Resistance of the *Burkholderia cepacia* complex to fosmidomycin and fosmidomycin derivatives. *International Journal of Antimicrobial Agents*, 38(3), 261-4. Doi: 10.1016/j.ijantimicag.2011.04.020. PMID: 21724375.
- Mikheenko, A., Valin, G., Prjibelski, A., Saveliev, V., & Gurevich, A. (2016). Icarus: visualizer for de novo assembly evaluation. *Bioinformatics*, 32(21), 3321-3323. Doi: 10.1093/bioinformatics/btw379. PMID: 27378299.
- Mo, T., Liu, W.Q., Ji, W., Zhao, J., Chen, T., Ding, W., Yu, S., & Zhang, Q. (2017). Biosynthetic insights into Linaridin natural products from genome mining and precursor peptide mutagenesis. *ACS Chemical Biology*, 12(6), 1484-1488. Doi: 10.1021/acscchembio.7b00262. PMID: 28452467.
- Moody, S.C., Zhao, B., Lei, L., Nelson, D.R., Mullins, J.G., Waterman, M.R., Kelly, S.L., & Lamb, D.C. (2012). Investigating conservation of the albaflavenone biosynthetic pathway and CYP170 bifunctionality in *Streptomyces*. *The FEBS Journal*, 279(9), 1640-9. Doi: 10.1111/j.1742-4658.2011.08447.x. PMID: 22151149.

- Moradi, J., Fathollahi, M., Halimi, S., Alvandi, A., Abiri, R., Vaziri, S., & Rezaei, A. (2022). Characterization of the resistome in *Lactobacillus* genomic sequences from the human gut. *Journal of Global Antimicrobial Resistance*, 30, 451-458. Doi: 10.1016/j.jgar.2022.05.014. PMID: 35643394.
- Muhammad, S., Qaisar, M., Iqbal, J., Khera, R.A., Al-Sehemi, A.G., Alarfaji, S.S., & Adnan M. (2022). Exploring the inhibitory potential of novel bioactive compounds from mangrove *actinomycetes* against nsp10 the major activator of SARS-CoV-2 replication. *Chemické Zvesti*, 76(5), 3051-3064. Doi: 10.1007/s11696-021-01997-x. PMID: 35103034. PMCID: PMC8791767.
- Munda, M., Nandi, R., Gavit, V.R., Kundu, S., Niyogi, S., & Bisai, A. (2022). Total syntheses of naturally occurring antiviral indolosesquiterpene alkaloids, xiamycins C-F via Csp3-H functionalization. *Chemical Science*, 13(39), 11666-11671. Doi: 10.1039/d2sc03479d. PMID: 36320384. PMCID: PMC9555729.
- Mungan, M.D., Alanjary, M., Blin, K., Weber, T., Medema, M.H., & Ziemert, N. (2020). ARTS 2.0: feature updates and expansion of the antibiotic resistant target seeker for comparative genome mining. *Nucleic Acids Research*, 48(W1), W546-W552. Doi: 10.1093/nar/gkaa374. PMID: 32427317. PMCID: PMC7319560.
- Nag, A., & Mehra, S. A (2021). Major facilitator superfamily (MFS) efflux pump, SCO4121, from *Streptomyces coelicolor* with roles in multidrug resistance and oxidative stress tolerance and its regulation by a MarR regulator. *Applied and Environmental Microbiology*, 87(7), e02238-20. Doi: 10.1128/AEM.02238-20. PMID: 33483304. PMCID: PMC8091613.
- Nag, A., & Mehra, S. (2022). Involvement of the SCO3366 efflux pump from *S. coelicolor* in rifampicin resistance and its regulation by a TetR regulator. *Applied Microbiology and Biotechnology*, 106(5-6), 2175-2190. Doi: 10.1007/s00253-022-11837-x. PMID: 35194656.
- Nelson, M.L., & Levy S.B. (2011). The history of the tetracyclines. *Annals of the New York Academy of Sciences journal*, 1241, 17-32. Doi: 10.1111/j.1749-6632.2011.06354.x. PMID: 22191524.

- Nicault, M., Tidjani, A.R., Gauthier, A., Dumarcay, S., Gelhaye, E., Bontemps, C., & Leblond, P. (2020). Mining the biosynthetic potential for specialized metabolism of a *Streptomyces* soil community. *Antibiotics (Basel)*, 9(5), 271. Doi: 10.3390/antibiotics9050271. PMID: 32456220; PMCID: PMC7277575.
- Nishimura, O., Hara, Y., & Kuraku, S. (2017). gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics*, 33(22), 3635-3637. Doi: 10.1093/bioinformatics/btx445. PMID: 29036533. PMCID: PMC5870689.
- Nishino, K., & Yamaguchi, A. (2001). Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. *Journal of Bacteriology*, 183(20), 5803-12. Doi: 10.1128/JB.183.20.5803-5812.2001. PMID: 11566977. PMCID: PMC99656.
- Nishino, K., Yamasaki, S., Nakashima, R., Zwama, M., & Hayashi-Nishino, M. (2021). Function and inhibitory mechanisms of multidrug efflux pumps. *Frontiers in Microbiology*, 12, 737288. Doi: 10.3389/fmicb.2021.737288. PMID: 34925258. PMCID: PMC8678522.
- Nivina, A., Yuet, K.P., Hsu, J., & Khosla, C. (2019). Evolution and diversity of assembly-line polyketide synthases. *Chemical Reviews*, 119(24), 12524-12547. Doi: 10.1021/acs.chemrev.9b00525. PMID: 31838842. PMCID: PMC6935866.
- Noda, Y., Yoda, K., Takatsuki, A., & Yamasaki, M. (1992). TmrB protein, responsible for tunicamycin resistance of *Bacillus subtilis*, is a novel ATP-binding membrane protein. *Journal of Bacteriology*, 174(13), 4302-7. Doi: 10.1128/jb.174.13.4302-4307.1992. PMID: 1624425. PMCID: PMC206213.
- Nuñez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., & Doudna, J.A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature Structural & Molecular Biology*, 21(6), 528-34. Doi: 10.1038/nsmb.2820. PMID: 24793649. PMCID: PMC4075942.
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 27(1), 29-34. Doi: 10.1093/nar/27.1.29. PMID: 9847135. PMCID: PMC148090.

- Ohki, R., & Murata, M. (1997). bmr3, a third multidrug transporter gene of *Bacillus subtilis*. *Journal of Bacteriology*, 179(4), 1423-7. Doi: 10.1128/jb.179.4.1423-1427.1997. PMID: 9023234. PMCID: PMC178848.
- Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., & Stevens, R. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Research*, 42 (Database issue), D206-14. Doi: 10.1093/nar/gkt1226. PMID: 24293654. PMCID: PMC3965101.
- Pal, C., Bengtsson-Palme, J., Kristiansson, E., & Larsson, D.G. (2015). Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC Genomics*, 16, 964. Doi: 10.1186/s12864-015-2153-5. PMID: 26576951. PMCID: PMC4650350.
- Palzkill, T. (2013). Metallo- $\beta$ -lactamase structure and function. *Annals of the New York Academy of Sciences journal*, 1277, 91-104. Doi: 10.1111/j.1749-6632.2012.06796.x. PMID: 23163348. PMCID: PMC3970115.
- Pantel, A., Petrella, S., Veziris, N., Brossier, F., Bastian, S., Jarlier, V., Mayer, C., & Aubry, A. (2012). Extending the definition of the GyrB quinolone resistance-determining region in *Mycobacterium tuberculosis* DNA gyrase for assessing fluoroquinolone resistance in *M. tuberculosis*. *Antimicrobial Agents and Chemotherapy*, 56(4), 1990-6. Doi: 10.1128/AAC.06272-11. PMID: 22290942. PMCID: PMC3318379.
- Pao, S.S., Paulsen, I.T., & Saier, M.H. Jr. (1998). Major facilitator superfamily. *Microbiology and Molecular Biology Reviews*, 62(1), 1-34. Doi: 10.1128/MMBR.62.1.1-34.1998. PMID: 9529885. PMCID: PMC98904.
- Park, S.C., & Won, S. (2018). Evaluation of 16S rRNA databases for taxonomic assignments using mock community. *Genomics and Informatics*, 16(4), e24. Doi: 10.5808/GI.2018.16.4.e24. PMID: 30602085. PMCID: PMC6440677.
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., & Tyson, G.W. (2014). Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25, 1043-1055. Doi: 10.1101/gr.186072.114.

- Parte, A.C., Sardà Carbasse, J., Meier-Kolthoff, J.P., Reimer, L.C., & Göker, M. (2020). List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology*, 70(11), 5607-5612. Doi: 10.1099/ijsem.0.004332. PMID: 32701423. PMCID: PMC7723251.
- Pasqua, M., Bonaccorsi di Patti, M.C., Fanelli, G., Utsumi, R., Eguchi, Y., Tirocco, R., Prosseda, G., Grossi, M., & Colonna, B. (2021). Host - bacterial pathogen communication: the wily role of the multidrug efflux pumps of the MFS family. *Frontiers in Molecular Biosciences*, 8, 723274. Doi: 10.3389/fmolb.2021.723274. PMID: 34381818. PMCID: PMC8350985.
- Paulsen, I.T., Brown, M.H., Littlejohn, T.G., Mitchell, B.A., & Skurray, R.A. (1996). Multidrug resistance proteins QacA and QacB from *Staphylococcus aureus*: membrane topology and identification of residues involved in substrate specificity. *Proceedings of the National Academy of Sciences of the United States of America*, 93(8), 3630-5. Doi: 10.1073/pnas.93.8.3630. PMID: 8622987; PMCID: PMC39662.
- Pfaffenbach, M., Bakanas, I., O'Connor, N.R., Herrick, J.L., & Sarpong, R. (2019). Total syntheses of Xiamycins A, C, F, H and Oridamycin A and preliminary evaluation of their anti-fungal properties. *Angewandte Chemie International Edition*, 58(43), 15304-15308. Doi: 10.1002/anie.201908399. PMID: 31419367. PMCID: PMC6791741.
- Plé, C., Tam, H.K., Vieira Da Cruz, A., Compagne, N., Jiménez-Castellanos, J.C., Müller, R.T., Pradel, E., Foong, W.E., Malloci, G., Ballée, A., Kirchner, M.A., Moshfegh, P., Herledan, A., Herrmann, A., Deprez, B., Willand, N., Vargiu, A.V., Pos, K.M., Flipo, M., & Hartkoorn, R.C. (2022). Pyridylpiperazine-based allosteric inhibitors of RND-type multidrug efflux pumps. *Nature Communications*, 13(1), 115. Doi: 10.1038/s41467-021-27726-2. PMID: 35013254. PMCID: PMC8749003.
- Procópio, R.E., Silva, I.R., Martins, M.K., Azevedo, J.L., & Araújo, J.M. (2012). Antibiotics produced by *Streptomyces*. *The Brazilian Journal of Infectious Diseases*, 16(5), 466-71. Doi: 10.1016/j.bjid.2012.08.014. PMID: 22975171.

- Ramírez-Rendon, D., Passari, A.K., Ruiz-Villafán, B., Rodríguez-Sanoja, R., Sánchez, S., & Demain, A.L. (2022). Impact of novel microbial secondary metabolites on the pharma industry. *Applied Microbiology and Biotechnology*, 106(5-6), 1855-1878. Doi: 10.1007/s00253-022-11821-5. PMID: 35188588. PMCID: PMC8860141.
- Ramón-García, S., Martín, C., De Rossi, E., & Aínsa, J.A. (2007). Contribution of the Rv2333c efflux pump (the Stp protein) from *Mycobacterium tuberculosis* to intrinsic antibiotic resistance in *Mycobacterium bovis* BCG. *Journal of Antimicrobial Chemotherapy*, 59(3), 544-7. Doi: 10.1093/jac/dkl510. PMID: 17242035.
- Rao, D.K., & Kaur, P. (2008). The Q-loop of DrrA is involved in producing the closed conformation of the nucleotide binding domains and in transduction of conformational changes between DrrA and DrrB. *Biochemistry*, 47(9), 3038-50. Doi: 10.1021/bi701699a. PMID: 18237140.
- Repka, L.M., Chekan, J.R., Nair, S.K., & van der Donk, W.A. (2017). Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chemical Reviews*, 117(8), 5457-5520. Doi: 10.1021/acs.chemrev.6b00591. PMID: 28135077. PMCID: PMC5408752.
- Revilla-Guarinos, A., Dürr, F., Popp, P.F., Döring, M., & Mascher, T. (2020). Amphotericin B specifically induces the two-component system LnrJK: Development of a novel whole-cell biosensor for the detection of amphotericin-like Polyenes. *Frontiers in Microbiology*, 11, 2022. Doi: 10.3389/fmicb.2020.02022. PMID: 32973732. PMCID: PMC7472640.
- Risdian, C., Mozef, T., & Wink, J. (2019). Biosynthesis of polyketides in *Streptomyces*. *Microorganisms*, 7(5), 124. Doi: 10.3390/microorganisms7050124. PMID: 31064143. PMCID: PMC6560455.
- Robbins, T., Liu, Y.C., Cane, D.E., & Khosla, C. (2016). Structure and mechanism of assembly line polyketide synthases. *Current Opinion in Structural Biology*, 41, 10-18. Doi: 10.1016/j.sbi.2016.05.009. PMID: 27266330. PMCID: PMC5136517.
- Romero-Rodríguez, A., Robledo-Casados, I., & Sánchez, S. (2015). An overview on transcriptional regulators in *Streptomyces*. *Biochimica et Biophysica Acta*, 1849(8), 1017-39. Doi: 10.1016/j.bbagr.2015.06.007. PMID: 26093238.



- Rudolf, J.D., Alsup, T.A., Xu, B., & Li, Z. (2021). Bacterial terpenome. *Natural Product Reports*, 38, 905-980. Doi: 10.1039/D0NP00066C.
- Ruiz-Perez, C.A., Conrad, R.E., & Konstantinidis, K.T. (2021). MicrobeAnnotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes. *BMC Bioinformatics*, 22(1), 11. Doi: 10.1186/s12859-020-03940-5. PMID: 33407081. PMCID: PMC7789693.
- Russell, A.H., & Truman, A.W. (2020). Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. *Computational and Structural Biotechnology Journal*, 18, 1838-1851. Doi: 10.1016/j.csbj.2020.06.032. PMID: 32728407. PMCID: PMC7369419.
- Saha, S., Zhang, W., Zhang, G., Zhu, Y., Chen, Y., Liu, W., Yuan, C., Zhang, Q., Zhang, H., Zhang, L., Zhang, W., & Zhang, C. (2017). Activation and characterization of a cryptic gene cluster reveals a cyclization cascade for polycyclic tetramate macrolactams. *Chemical Science*, 8(2), 1607-1612. Doi: 10.1039/c6sc03875a. PMID: 28451290. PMCID: PMC5361873.
- Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406-425. Doi: 10.1093/oxfordjournals.molbev.a040454.
- Salam, N., Jiao, J.Y., Zhang, X.T., & Li, W.J. (2020). Update on the classification of higher ranks in the phylum *Actinobacteria*. *International Journal of Systematic and Evolutionary Microbiology*, 70(2), 1331-1355. Doi: 10.1099/ijsem.0.003920. PMID: 31808738.
- Schaller, K., & Nomura, M. Colicin E2 is DNA endonuclease. 1976. *Proceedings of the National Academy of Sciences of the United States of America*, 73(11), 3989-93. Doi: 10.1073/pnas.73.11.3989. PMID: 1069283; PMCID: PMC431295.
- Schmutz, E., Mühlenweg, A., Li, S.M., & Heide, L. (2003). Resistance genes of aminocoumarin producers: two type II topoisomerase genes confer resistance against coumermycin A1 and clorobiocin. *Antimicrobial Agents and Chemotherapy*, 47(3), 869-77. Doi: 10.1128/AAC.47.3.869-877.2003. PMID: 12604514. PMCID: PMC149333.

- Scott, T.A., & Piel, J. (2019). The hidden enzymology of bacterial natural product biosynthesis. *Nature Reviews Chemistry*, 3(7), 404-425. Doi: 10.1038/s41570-019-0107-1. PMID: 32232178. PMCID: PMC7104373.
- Seemann T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068-9. Doi: 10.1093/bioinformatics/btu153. PMID: 24642063.
- Shaffer, M., Borton, M.A., McGivern, B.B., Zayed, A.A., La Rosa, S.L., Solden, L.M., Liu, P., Narrowe, A.B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M.C., Daly, R.A., Smith, G.J., Vik, D.R., Pope, P.B., Sullivan, M.B., Roux, S., & Wrighton, K.C. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research*, 48(16), 8883-8900. Doi: 10.1093/nar/gkaa621. PMID: 32766782. PMCID: PMC7498326.
- Shah, S.A., & Garrett, R.A. (2011). CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. *Research in Microbiology*, 162(1), 27-38. Doi: 10.1016/j.resmic.2010.09.001. PMID: 20863886.
- Sharma, A., Sharma, R., Bhattacharyya, T., Bhando, T., & Pathania, R. (2017). Fosfomycin resistance in *Acinetobacter baumannii* is mediated by efflux through a major facilitator superfamily (MFS) transporter-AbaF. *Journal of Antimicrobial Chemotherapy*, 72(1), 68-74. Doi: 10.1093/jac/dkw382. PMID: 27650185.
- Sheng, Q., Li, W.-J., Dastager, S.G., & Hozzein, W.N. (2016). Editorial: *Actinobacteria* in special and extreme habitats: diversity, function roles, and environmental adaptations. *Frontiers in Microbiology*, 7, 1415. Doi: 10.3389/fmicb.2016.01415. ISSN=1664-302X.
- Shi, L., Pigeonneau, N., Ventroux, M., Derouiche, A., Bidnenko, V., Mijakovic, I., & Noirot-Gros, M.F. (2014). Protein-tyrosine phosphorylation interaction network in *Bacillus subtilis* reveals new substrates, kinase activators and kinase cross-talk. *Frontiers in Microbiology*, 5, 538. Doi: 10.3389/fmicb.2014.00538. PMID: 25374563. PMCID: PMC4205851.
- Shmakov, S.A., Utkina, I., Wolf, Y.I., Makarova, K.S., Severinov, K.V., & Koonin, E.V. (2020). CRISPR arrays away from cas genes. *The CRISPR Journal*, 3(6), 535-549. Doi: 10.1089/crispr.2020.0062. PMID: 33346707. PMCID: PMC7757702.

- Sinha, S.C., & Sprang, S.R. (2006). Structures, mechanism, regulation and evolution of class III nucleotidyl cyclases. *Reviews of Physiology, Biochemistry and Pharmacology*, 157, 105-40. Doi: 10.1007/112\_0603. PMID: 17236651.
- Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *The EMBO Journal*, 30(7), 1335-42. Doi: 10.1038/emboj.2011.41. PMID: 21343909. PMCID: PMC3094125.
- Sivalingam, P., Hong, K., Pote, J., & Prabakar, K. (2019). Extreme environment *Streptomyces*: potential sources for new antibacterial and anticancer drug leads?. *International Journal of Microbiology*, 2019, 5283948. Doi: 10.1155/2019/5283948. PMID: 31354829. PMCID: PMC6636559.
- Skinneider, M.A., Merwin, N.J., Johnston, C.W., & Magarvey, N.A. (2017). PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Research*, 3(45), W49-W54. Doi: 10.1093/nar/gkx320. PMID: 28460067. PMCID: PMC5570231.
- Slipski, C.J., Jamieson, T.R., Zhanel, G.G., & Bay, D.C. (2020). Riboswitch-associated guanidinium-selective efflux pumps frequently transmitted on proteobacterial plasmids increase *Escherichia coli* biofilm tolerance to disinfectants. *Journal of Bacteriology*, 202(23), e00104-20. Doi: 10.1128/JB.00104-20. PMID: 32928929. PMCID: PMC7648145.
- Sorek, R., Kunin, V., & Hugenholtz, P. (2008). CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology*, 6(3), 181-6. Doi: 10.1038/nrmicro1793. PMID: 18157154.
- Srikant, S. (2020). Evolutionary history of ATP-binding cassette proteins. *FEBS Letters*, 594(23), 3882-3897. Doi: 10.1002/1873-3468.13985. PMID: 33145769.
- Srinivasan, P., Palani, S.N., & Prasad, R. (2010). Daunorubicin efflux in *Streptomyces peucetius* modulates biosynthesis by feedback regulation. *FEMS Microbiology Letters*, 305(1), 18-27. Doi: 10.1111/j.1574-6968.2010.01905.x. PMID: 20158521.

- Srinivasan, V.B., Rajamohan. G., & Gebreyes, W.A. (2009). Role of AbeS, a novel efflux pump of the SMR family of transporters, in resistance to antimicrobial agents in *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy*, 53(12), 5312-6. Doi: 10.1128/AAC.00748-09. PMID: 19770280; PMCID: PMC2786332.
- Staals, R.H., Jackson, S.A., Biswas, A., Brouns, S.J., Brown, C.M., & Fineran, P.C. (2016). Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nature Communications*, 7, 12853. Doi: 10.1038/ncomms12853. PMID: 27694798. PMCID: PMC5059440.
- Steinfels, E., Orelle, C., Fantino, J.R., Dalmas, O., Rigaud, J.L., Denizot, F., Di Pietro, A., & Jault, J.M. (2004). Characterization of YvcC (BmrA), a multidrug ABC transporter constitutively expressed in *Bacillus subtilis*. *Biochemistry*, 43(23), 7491-502. Doi: 10.1021/bi0362018. PMID: 15182191.
- Stern, A., Keren, L., Wurtzel, O., Amitai, G., & Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity?. *Trend in Genetics*, 26(8), 335-40. Doi: 10.1016/j.tig.2010.05.008. PMID: 20598393. PMCID: PMC2910793.
- Strich, J.R., & Chertow, D.S. (2019). CRISPR-Cas Biology and its application to infectious diseases. *Journal of Clinical Microbiology*, 57(4), e01307-18. Doi: 10.1128/JCM.01307-18. PMID: 30429256. PMCID: PMC6440769.
- Stubbendieck, R.M., & Straight, P.D. (2017). Linearmycins activate a two-Component signaling system involved in bacterial competition and biofilm morphology. *Journal of Bacteriology*, 199(18), e00186-17. Doi: 10.1128/JB.00186-17. PMID: 28461449. PMCID: PMC5573067.
- Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, 38(7), 3022-3027. Doi: 10.1093/molbev/msab120. PMID: 33892491. PMCID: PMC8233496.
- Tanaka, K.J., Song, S., Mason, K., & Pinkett, H.W. (2018). Selective substrate uptake: The role of ATP-binding cassette (ABC) importers in pathogenesis. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1860(4), 868-877. Doi: 10.1016/j.bbamem.2017.08.011. PMID: 28847505. PMCID: PMC5807212.

- Tao, S., Chen, H., Li, N., Wang, T., & Liang, W. (2022). The spread of antibiotic resistance genes in vivo model. *Canadian Journal of Infectious Diseases & Medical Microbiology*, 2022, 3348695. Doi: 10.1155/2022/3348695. PMID: 35898691. PMCID: PMC9314185.
- Thaker, M., Spanogiannopoulos, P., & Wright, G.D. (2010). The tetracycline resistome. *Cellular and Molecular Life Sciences*, 67(3), 419-31. Doi: 10.1007/s00018-009-0172-6. PMID: 19862477.
- Thiara, A.S., & Cundliffe, E. (1993). Expression and analysis of two gyrB genes from the novobiocin producer, *Streptomyces sphaeroides*. *Molecular Microbiology*, 8(3), 495-506. Doi: 10.1111/j.1365-2958.1993.tb01593.x. PMID: 8392138.
- Thiara, A.S., & Cundliffe, E. (1989). Interplay of novobiocin-resistant and -sensitive DNA gyrase activities in self-protection of the novobiocin producer, *Streptomyces sphaeroides*. *Gene*, 81(1), 65-72. Doi: 10.1016/0378-1119(89)90337-5. PMID: 2553539.
- Tomm, H.A., Ucciferri, L., & Ross, A.C. (2019). Advances in microbial culturing conditions to activate silent biosynthetic gene clusters for novel metabolite production. *Journal of Industrial Microbiology & Biotechnology*, 46(9-10), 1381-1400. Doi: 10.1007/s10295-019-02198-y. PMID: 31177350.
- Tong, Y., Weber, T., & Lee, S.Y. (2019). CRISPR/Cas-based genome engineering in natural product discovery. *Natural Product Reports*, 36(9), 1262-1280. Doi: 10.1039/c8np00089a. PMID: 30548045.
- Tooke, C.L., Hinchliffe, P., Bragginton, E.C., Colenso, C.K., Hirvonen, V.H.A., Takebayashi, Y., & Spencer, J. (2019).  $\beta$ -lactamases and  $\beta$ -lactamase inhibitors in the 21st century. *Journal of Molecular Biology*, 431(18), 3472-3500. Doi: 10.1016/j.jmb.2019.04.002. PMID: 30959050. PMCID: PMC6723624.
- Torres, C., Galián, C., Freiberg, C., Fantino, J.R., & Jault, J.M. (2009). The YheI/YheH heterodimer from *Bacillus subtilis* is a multidrug ABC transporter. *Biochimica et Biophysica Acta*, 1788(3), 615-22. Doi: 10.1016/j.bbamem.2008.12.012. PMID: 19167342.

- Tran, P.N., Yen, M.R., Chiang, C.Y., Lin, H.C., & Chen, P.Y. (2019). Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. *Applied Microbiology and Biotechnology*, 103(8), 3277-3287. Doi: 10.1007/s00253-019-09708-z. PMID: 30859257. PMCID: PMC6449301.
- Trotta, A.H. (2017). Toward a unified total synthesis of the xiamycin and oridamycin families of Indolosesquiterpenes. *The Journal of Organic Chemistry*, 82(24), 13500-13516. Doi: 10.1021/acs.joc.7b02623. PMID: 29171266. PMCID: PMC6585422.
- Ueda, K., Oinuma, K., Ikeda, G., Hosono, K., Ohnishi, Y., Horinouchi, S., & Beppu, T. (2002). AmfS, an extracellular peptidic morphogen in *Streptomyces griseus*. *Journal of Bacteriology*, 184(5), 1488-92. Doi: 10.1128/JB.184.5.1488-1492.2002. PMID: 11844785. PMCID: PMC134859.
- UniProt Consortium. 2023. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Research*, 51(D1), D523-D531. Doi: 10.1093/nar/gkac1052. PMID: 36408920. PMCID: PMC9825514.
- Van Bambeke, F., Balzi, E., & Tulkens, P.M. (2000). Antibiotic efflux pumps. *Biochemical Pharmacology*, 60(4), 457-70. Doi: 10.1016/s0006-2952(00)00291-4. PMID: 10874120
- van der Donk, W.A., & Nair, S.K. (2014). Structure and mechanism of lanthipeptide biosynthetic enzymes. *Current Opinion in Structural Biology*, 29, 58-66. Doi: 10.1016/j.sbi.2014.09.006. PMID: 25460269. PMCID: PMC4267917.
- van Heel, A.J., de Jong, A., Song, C., Viel, J.H., Kok, J., & Kuipers, O.P. (2018). BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Research*, 46(W1), W278-W281. Doi: 10.1093/nar/gky383. PMID: 29788290. PMCID: PMC6030817.
- Varela, M.F., Stephen, J., Lekshmi, M., Ojha, M., Wenzel, N., Sanford, L.M., Hernandez, A.J., Parvathi, A., & Kumar, S.H. (2021). Bacterial resistance to antimicrobial agents. *Antibiotics (Basel)*, 10(5), 593. Doi: 10.3390/antibiotics10050593. PMID: 34067579. PMCID: PMC8157006.

- Vedantam, G., Guay, G.G., Austria, N.E., Doktor, S.Z., & Nichols, B.P. (1998). Characterization of mutations contributing to sulfathiazole resistance in *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, 42(1), 88-93. Doi: 10.1128/AAC.42.1.88. PMID: 9449266. PMCID: PMC105461.
- Vinogradov, A.A., & Suga, H. (2020). Introduction to thiopeptides: Biological activity, biosynthesis, and strategies for functional reprogramming. *Cell Chemical Biology*, 27(8), 1032-1051. Doi: 10.1016/j.chembiol.2020.07.003. PMID: 32698017.
- Völler, G.H., Krawczyk, J.M., Pesic, A., Krawczyk, B., Nachtigall, J., & Süssmuth, R.D. (2012). Characterization of new class III lantibiotics--erythraeptin, avermipeptin and griseopeptin from *Saccharopolyspora erythraea*, *Streptomyces avermitilis* and *Streptomyces griseus* demonstrates stepwise N-terminal leader processing. *Chembiochem*, 13(8), 1174-83. Doi: 10.1002/cbic.201200118. PMID: 22556031.
- von Wintersdorff, C.J., Penders, J., van Niekerk, J.M., Mills, N.D., Majumder, S., van Alphen, L.B., Savelkoul, P.H., & Wolffs, P.F. (2016). Dissemination of antimicrobial resistance in microbial ecosystems through Horizontal Gene Transfer. *Frontiers in Microbiology*, 7, 173. Doi: 10.3389/fmicb.2016.00173. PMID: 26925045. PMCID: PMC4759269.
- Wales, A.D., & Davies, R.H. (2015). Co-selection of resistance to antibiotics, biocides and heavy metals, and its relevance to foodborne pathogens. *Antibiotics (Basel)*, 4(4), 567-604. Doi: 10.3390/antibiotics4040567. PMID: 27025641; PMCID: PMC4790313.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., & Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, 9(11), e112963. Doi: 10.1371/journal.pone.0112963.
- Walker, M.C., Eslami, S.M., Hetrick, K.J., Ackenhusen, S.E., Mitchell, D.A., van der & Donk, W.A. (2020). Precursor peptide-targeted mining of more than one hundred thousand genomes expands the lanthipeptide natural product family. *BMC Genomics*, 21(1), 387. Doi: 10.1186/s12864-020-06785-7. PMID: 32493223. PMCID: PMC7268733.

- Wang, B., Guo, F., Huang, C., & Zhao, H. (2020<sup>a</sup>). Unraveling the iterative type I polyketide synthases hidden in *Streptomyces*. *Proceedings of the National Academy of Sciences of the United States of America*, 117(15), 8449-8454. Doi: 10.1073/pnas.1917664117. PMID: 32217738. PMCID: PMC7165468.
- Wang, J., Zhang, R., Chen, X., Sun, X., Yan, Y., Shen, X., & Yuan, Q. (2020<sup>b</sup>). Biosynthesis of aromatic polyketides in microorganisms using type II polyketide synthases. *Microbial Cell Factories*, 19(1), 110. Doi: 10.1186/s12934-020-01367-4. PMID: 32448179. PMCID: PMC7247197.
- Ward, A.C., & Allenby, N.E. (2018). Genome mining for the search and discovery of bioactive compounds: the *Streptomyces* paradigm. *FEMS Microbiology Letters*, 365(24). Doi: 10.1093/femsle/fny240. PMID: 30265303.
- Webber, M. A., & Piddock, L. J. V. (2003). The importance of efflux pumps in bacterial antibiotic resistance. *Journal of Antimicrobial Chemotherapy*, 51(1), 9–11. Doi: 10.1093/jac/dkg050.
- Wen, P.C., & Tajkhorshid, E. (2011). Conformational coupling of the nucleotide-binding and the transmembrane domains in ABC transporters. *Biophysical Journal*, 101(3), 680-90. Doi: 10.1016/j.bpj.2011.06.031. PMID: 21806936. PMCID: PMC3145290.
- Wick, R.R., Judd, L.M., Gorrie, C.L., & Holt, K.E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology*, 13(6), e1005595. Doi: 10.1371/journal.pcbi.1005595.
- Wieczorek P, Sacha P, Hauschild T, Zórawski M, Krawczyk M & Tryniszewska E. (2008). Multidrug resistant *Acinetobacter baumannii*--the role of AdeABC (RND family) efflux pump in resistance to antibiotics. *Folia Histochemica et Cytobiologica*, 46(3), 257-67. Doi: 10.2478/v10042-008-0056-x. PMID: 19056528.
- Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W., & Doudna, J.A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*, 17(6), 904-12. Doi: 10.1016/j.str.2009.03.019. PMID: 19523907.



- Willey, J., Santamaria, R., Guijarro, J., Geistlich, M., & Losick, R. (1991). Extracellular complementation of a developmental mutation implicates a small sporulation protein in aerial mycelium formation by *S. coelicolor*. *Cell*, 65(4), 641-50. Doi: 10.1016/0092-8674(91)90096-h. PMID: 2032288.
- Wimmer, F., & Beisel, C.L. (2020). CRISPR-Cas systems and the paradox of self-targeting Spacers. *Frontiers in Microbiology*, 10, 3078. Doi: 10.3389/fmicb.2019.03078. PMID: 32038537. PMCID: PMC6990116.
- Xu, G.C., Xu, T.J., Zhu, R., Zhang, Y., Li, S.Q., Wang, H.W., & Li, J.T. (2019<sup>a</sup>). LR\_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience*, 8(1), giy157. Doi: 10.1093/gigascience/giy157. PMID: 30576505. PMCID: PMC6324547.
- Xu, M., Zhang, F., Cheng, Z., Bashiri, G., Wang, J., Hong, J., Wang, Y., Xu, L., Chen, X., Huang, S.X., Lin, S., Deng, Z., & Tao, M. (2020). Functional genome mining reveals a class V lanthipeptide containing a d-amino acid introduced by an F420 H<sub>2</sub>-dependent reductase. *Angewandte Chemie International Edition*, 59(41), 18029-18035. Doi: 10.1002/anie.202008035. PMID: 32648341.
- Xue, C., & Sashital, D.G. (2019). Mechanisms of Type I-E and I-F CRISPR-Cas Systems in *Enterobacteriaceae*. *EcoSal Plus*, 8(2), 10.1128/ecosalplus.ESP-0008-2018. Doi: 10.1128/ecosalplus.ESP-0008-2018. PMID: 30724156. PMCID: PMC6368399.
- Yamada, S., Awano, N., Inubushi, K., Maeda, E., Nakamori, S., Nishino, K., Yamaguchi, A., & Takagi, H. (2006). Effect of drug transporter genes on cysteine export and overproduction in *Escherichia coli*. *Applied and Environmental Microbiology*, 72(7), 4735-42. Doi: 10.1128/AEM.02507-05. PMID: 16820466. PMCID: PMC1489377.
- Yan, N. (2015). Structural Biology of the Major Facilitator Superfamily Transporters. *Annual Review of Biophysics*, 44, 257-83. Doi: 10.1146/annurev-biophys-060414-033901. PMID: 26098515.
- Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, 39(3), 306-14. Doi: 10.1007/BF00160154. PMID: 7932792.

- Yoon, S.H., Ha, S.M., Kwon, S., Lim, J., Kim, Y., Seo, H., & Chun, J. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International Journal of Systematic and Evolutionary Microbiology*, 67(5), 1613-1617. Doi: 10.1099/ijsem.0.001755. PMID: 28005526. PMCID: PMC5563544.
- Yu, D., Xu, F., Zeng, J., & Zhan, J. (2012<sup>a</sup>). Type III polyketide synthases in natural product biosynthesis. *IUBMB Life*, 64(4), 285-95. Doi: 10.1002/iub.1005. PMID: 22362498.
- Yu, L., Li, W., Li, Q., Chen, X., Ni, J., Shang, F., & Xue, T. (2020). Role of LsrR in the regulation of antibiotic sensitivity in avian pathogenic *Escherichia coli*. *Poultry Science*, 99(7), 3675-3687. Doi: 10.1016/j.psj.2020.03.064. PMID: 32616264. PMCID: PMC7597812.
- Yu, L., Yan, X., Wang, L., Chu, J., Zhuang, Y., Zhang, S., & Guo, M. (2012<sup>b</sup>). Molecular cloning and functional characterization of an ATP-binding cassette transporter OtrC from *Streptomyces rimosus*. *BMC Biotechnology*, 12, 52. Doi: 10.1186/1472-6750-12-52. PMID: 22906146. PMCID: PMC3533511.
- Yu, Y., Zhang, Q., & van der Donk, W.A. (2013). Insights into the evolution of lanthipeptide biosynthesis. *Protein Science*, 22(11), 1478-89. Doi: 10.1002/pro.2358. PMID: 24038659. PMCID: PMC3831664.
- Yuzawa, S., Backman, T.W.H., Keasling, J.D., & Katz, L. (2018). Synthetic biology of polyketide synthases. *Journal of Industrial Microbiology & Biotechnology*, 45(7), 621-633. Doi: 10.1007/s10295-018-2021-9. PMID: 29423743.
- Zhang, C., Ding, W., Qin, X., & Ju, J. (2019). Genome sequencing of *Streptomyces olivaceus* SCSIO T05 and activated production of Lobophorin CR4 via metabolic engineering and genome mining. *Marine Drugs*, 17(10), 593. Doi: 10.3390/md17100593. PMID: 31635159. PMCID: PMC6835275.
- Zhang, F., Zhao, S., Ren, C., Zhu, Y., Zhou, H., Lai, Y., Zhou, F., Jia, Y., Zheng, K., & Huang, Z. (2018<sup>a</sup>). CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions. *Communications Biology*, 1, 180. Doi: 10.1038/s42003-018-0184-6. PMID: 30393777. PMCID: PMC6208339.

- Zhang, G., Zhang, W., Saha, S., & Zhang, C. (2016). Recent advances in discovery, biosynthesis and genome mining of medicinally relevant polycyclic tetramate macrolactams. *Current Topics in Medicinal Chemistry*, 16(15), 1727-39. Doi: 10.2174/1568026616666151012112818. PMID: 26456464.
- Zhang, J., Li, X., Deng, Z., & Ou, H.Y. (2018<sup>b</sup>). Comparative analysis of CRISPR loci found in *Streptomyces* genome sequences. *Interdisciplinary Sciences*, 10(4), 848-853. Doi: 10.1007/s12539-018-0301-8. PMID: 30030786.
- Zhang, K., Lin, R., Chang, Y., Zhou, Q., & Zhang, Z. (2022). 16S-FASAS: an integrated pipeline for synthetic full-length 16S rRNA gene sequencing data analysis. *PeerJ*, 10, e14043. Doi: 10.7717/peerj.14043. PMID: 36172503. PMCID: PMC9511998.
- Zhang, Q., Doroghazi, J.R., Zhao, X., Walker, M.C., & van der Donk, W.A. (2015). Expanded natural product diversity revealed by analysis of lanthipeptide-like gene clusters in actinobacteria. *Applied and Environmental Microbiology*, 81(13), 4339-50. Doi: 10.1128/AEM.00635-15. PMID: 25888176. PMCID: PMC4475899.
- Zhang, Q., Yu, Y., Vélasquez, J.E., & van der Donk, W.A. (2012). Evolution of lanthipeptide synthetases. *Proceedings of the National Academy of Sciences of the United States of America*, 109(45), 18361-6. Doi: 10.1073/pnas.1210393109. PMID: 23071302. PMCID: PMC3494888.
- Zhang, X., Shang, G., Gu, L., & Shen, Y. (2014). Crystallization and preliminary X-ray diffraction analysis of the diterpene cyclooctatin synthase (CYC) from *Streptomyces* sp. LZ35. *Acta Crystallographica. Section F, Structural Biology Communications*, 70(Pt 3), 366-9. Doi: 10.1107/S2053230X14003100. PMID: 24598929. PMCID: PMC3944704.
- Zhao, B., Lei, L., Vassilyev, D.G., Lin, X., Cane, D.E., Kelly, S.L., Yuan, H., Lamb, D.C., & Waterman, M.R. (2009). Crystal structure of albaflavenone monooxygenase containing a moonlighting terpene synthase active site. *Journal of Biological Chemistry*, 284(52), 36711-36719. Doi: 10.1074/jbc.M109.064683. PMID: 19858213. PMCID: PMC2794785.

- Zhao, B., Lin, X., Lei, L., Lamb, D.C., Kelly, S.L., Waterman, M.R., & Cane, D.E. (2008). Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). *Journal of Biological Chemistry*, 283(13), 8183-9. Doi: 10.1074/jbc.M710421200. PMID: 18234666. PMCID: PMC2276382.
- Zhao, F., Qin, Y.H., Zheng, X., Zhao, H.W., Chai, D.Y., Li, W., Pu, M.X., Zuo, X.S., Qian, W., Ni, P., Zhang, Y., Mei, H., & He, S.T. (2016<sup>a</sup>). Biogeography and adaptive evolution of *Streptomyces* strains from saline environments. *Scientific Reports*, 6, 32718. Doi: 10.1038/srep32718. PMID: 27596681. PMCID: PMC5011734.
- Zhao, N., Pan, Y., Cheng, Z., & Liu, H. (2016<sup>b</sup>). Lasso peptide, a highly stable structure and designable multifunctional backbone. *Amino Acids*, 48(6), 1347-56. Doi: 10.1007/s00726-016-2228-x. PMID: 27074719.
- Zhong, Z., He, B., Li, J., & Li, Y.X. (2020). Challenges and advances in genome mining of ribosomally synthesized and post-translationally modified peptides (RiPPs). *Synthetic and Systems Biotechnology*, 5(3), 155-172. Doi: 10.1016/j.synbio.2020.06.002. PMID: 32637669. PMCID: PMC7327761.
- Zhu, X., & Ye, K. (2012). Crystal structure of Cmr2 suggests a nucleotide cyclase-related enzyme in type III CRISPR-Cas systems. *FEBS Letters*, 586(6), 939-45. Doi: 10.1016/j.febslet.2012.02.036. PMID: 22449983.
- Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E., & Jensen, P.R. (2012). The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One*, 7(3), e34064. Doi: 10.1371/journal.pone.0034064. PMID: 22479523. PMCID: PMC3315503.

## VII. ANEXOS

**Anexo 1.** Anotación de CDS del genoma de *Streptomyces* sp. B-146 relacionados a genes ARG usando PROKKA.

Nº	Start	End	Length bp	Strand	Gene	EC Number	COG	Product
<b>MFS Efflux Pumps</b>								
01	250,849	252,051	1203	+	<i>mdtD</i>	None	None	Putative multidrug resistance protein MdtD
02	556,972	558,735	1764	-	<i>mdtD</i>	None	None	Putative multidrug resistance protein MdtD
03	641,561	643,015	1455	+	<i>mdtD</i>	None	None	Putative multidrug resistance protein MdtD
04	1,690,744	1,692,363	1620	-	<i>mdtD</i>	None	None	Putative multidrug resistance protein MdtD
05	2,314,097	2,315,875	1779	+	<i>mdtD</i>	None	None	Putative multidrug resistance protein MdtD
06	3,352,029	3,353,468	1440	-	<i>mdtD</i>	None	None	Putative multidrug resistance protein MdtD

«Continuación»

07	3,866,655	3,869,153	2499	+	<i>mdtD</i>	None	None	Putative multidrug resistance protein MdtD
08	6,045,594	6,047,141	1548	-	<i>mdtD</i>	None	None	Putative multidrug resistance protein MdtD
09	6,850,119	6,852,590	2472	+	<i>mdtD</i>	None	None	Putative multidrug resistance protein MdtD
10	7,207,906	7,209,978	2073	-	<i>mdtD</i>	None	None	Putative multidrug resistance protein MdtD
11	2,312,355	2,313,803	1449	-	<i>Stp</i>	None	None	Multidrug resistance protein Stp
12	2,367,938	2,369,548	1611	-	<i>Stp</i>	None	None	Multidrug resistance protein Stp
13	2,534,464	2,535,918	1455	+	<i>Stp</i>	None	None	Multidrug resistance protein Stp
14	4,250,847	4,252,304	1458	+	<i>Stp</i>	None	None	Multidrug resistance protein Stp

«Continuación»

15	4,831,051	4,832,520	1470	+	<i>Stp</i>	None	None	Multidrug resistance protein Stp
16	6,172,501	6,174,144	1644	+	<i>Stp</i>	None	None	Multidrug resistance protein Stp
17	6,741,935	6,743,482	1548	+	<i>Stp</i>	None	None	Multidrug resistance protein Stp
18	7,780,154	7,781,629	1476	+	<i>Stp</i>	None	None	Multidrug resistance protein Stp
19	300,308	301,693	1386	-	<i>abaF</i>	None	None	Fosfomycin resistance protein AbaF
20	6,811,181	6,812,518	1338	+	<i>abaF</i>	None	None	Fosfomycin resistance protein AbaF
21	7,405,398	7,406,753	1356	-	<i>abaF</i>	None	None	Fosfomycin resistance protein AbaF
22	8,031,345	8,032,721	1377	-	<i>abaF</i>	None	None	Fosfomycin resistance protein AbaF

«Continuación»

23	3,873,854	3,875,368	1515	-	<i>qacA</i>	None	COG0477	Antiseptic resistance protein
24	6,219,640	6,221,250	1611	+	<i>qacA</i>	None	COG0477	Antiseptic resistance protein
25	8,010,901	8,012,319	1419	-	<i>qacA</i>	None	COG0477	Antiseptic resistance protein
26	3,691,365	3,692,663	1299	+	<i>mdtH</i>	None	None	Multidrug resistance protein MdtH
27	6,989,297	6,990,559	1263	+	<i>mdtH</i>	None	None	Multidrug resistance protein MdtH
28	4,525,777	4,527,378	1602	+	<i>bmr3</i>	None	None	Multidrug resistance protein 3
29	7,841,769	7,843,382	1614	-	<i>bmr3</i>	None	None	Multidrug resistance protein 3
30*	2,245,403	2,246,674	1272	-	<i>bcr</i>	None	None	Bicyclomycin resistance protein



«Continuación»

31	3,054,666	3,055,907	1242	-	<i>fsr</i>	None	None	Fosmidomycin resistance protein
----	-----------	-----------	------	---	------------	------	------	---------------------------------

### ABC Efflux Pumps

32	1,657,394	1,658,338	945	+	<i>drrA</i>	3.6.3.-	None	Daunorubicin/doxorubicin resistance ATP-binding protein DrrA
33*	2,220,902	2,221,933	1032	+	<i>drrA</i>	3.6.3.-	None	Daunorubicin/doxorubicin resistance ATP-binding protein DrrA
34	2,863,174	2,864,109	936	-	<i>drrA</i>	3.6.3.-	None	Daunorubicin/doxorubicin resistance ATP-binding protein DrrA
35	3,533,294	3,534,298	1005	-	<i>drrA</i>	3.6.3.-	None	Daunorubicin/doxorubicin resistance ATP-binding protein DrrA
36	3,834,135	3,835,103	969	+	<i>drrA</i>	3.6.3.-	None	Daunorubicin/doxorubicin resistance ATP-binding protein DrrA
37	4,094,218	4,095,249	1032	-	<i>drrA</i>	3.6.3.-	None	Daunorubicin/doxorubicin resistance ATP-binding protein DrrA

«Continuación»

38	4,192,133	4,193,152	1020	+	<i>drrA</i>	3.6.3.-	None	Daunorubicin/doxorubicin resistance ATP-binding protein DrrA
39	5,143,307	5,144,311	1005	+	<i>drrA</i>	3.6.3.-	None	Daunorubicin/doxorubicin resistance ATP-binding protein DrrA
40	5,234,302	5,235,255	954	+	<i>drrA</i>	3.6.3.-	None	Daunorubicin/doxorubicin resistance ATP-binding protein DrrA
41	6,188,033	6,189,190	1158	-	<i>drrA</i>	3.6.3.-	None	Daunorubicin/doxorubicin resistance ATP-binding protein DrrA
42	6,702,640	6,703,590	951	-	<i>drrA</i>	3.6.3.-	None	Daunorubicin/doxorubicin resistance ATP-binding protein DrrA
43	1,658,335	1,659,207	873	+	<i>drrB</i>	None	None	Daunorubicin/doxorubicin resistance ABC transporter permease protein DrrB
44*	2,221,930	2,222,754	825	+	<i>drrB</i>	None	None	Daunorubicin/doxorubicin resistance ABC transporter permease protein DrrB
45	3,532,473	3,533,297	825	-	<i>drrB</i>	None	None	Daunorubicin/doxorubicin resistance ABC transporter permease protein DrrB

«Continuación»

46*	4,093,373	4,094,221	849	-	<i>drrB</i>	None	None	Daunorubicin/doxorubicin resistance ABC transporter permease protein DrrB
47	4,193,149	4,193,937	789	+	<i>drrB</i>	None	None	Daunorubicin/doxorubicin resistance ABC transporter permease protein DrrB
48	5,144,308	5,145,165	858	+	<i>drrB</i>	None	None	Daunorubicin/doxorubicin resistance ABC transporter permease protein DrrB
49	5,261,301	5,262,122	822	+	<i>drrB</i>	None	None	Daunorubicin/doxorubicin resistance ABC transporter permease protein DrrB
50	6,187,215	6,188,045	831	-	<i>drrB</i>	None	None	Daunorubicin/doxorubicin resistance ABC transporter permease protein DrrB
51	3,485,794	3,486,567	774	+	<i>lnrL</i>	3.6.3.-	COG1131	Linearmycin resistance ATP-binding protein LnrL
52	3,939,369	3,940,280	912	-	<i>lnrL</i>	3.6.3.-	COG1131	Linearmycin resistance ATP-binding protein LnrL
53	4,342,699	4,343,682	984	+	<i>lnrL</i>	3.6.3.-	COG1131	Linearmycin resistance ATP-binding protein LnrL

«Continuación»

54	4,499,756	4,500,388	633	+	<i>tmrB</i>	None	None	Tunicamycin resistance protein
55	5,662,543	5,664,453	1911	-	<i>yheI</i>	3.6.3.-	COG1132	Putative multidrug resistance ABC transporter ATP-binding / permease protein YheI
56	7,723,241	7,724,995	1755	-	<i>bmrA</i>	3.6.3.-	COG1132	Multidrug resistance ABC transporter ATP-binding / permease protein BmrA
<b>Enzymes</b>								
57	7,249,094	7,250,041	948	+	<i>arnA</i>	None	COG0223	Bifunctional polymyxin resistance protein ArnA
58	7,938,108	7,939,055	948	-	<i>arnA</i>	None	COG0223	Bifunctional polymyxin resistance protein ArnA
59**	4,038,335	4,040,419	2085	+	<i>gyrBR</i>	5.6.2.2	None	DNA gyrase subunit B, novobiocin-resistant

FUENTE: Elaboración propia.

\* CDS también anotados por RAST

\*\* CDS anotado y clasificado por RAST en el subsistema “Resistance to fluoroquinolones”.

**Anexo 2.** Anotación de CDS relacionados a subsistemas implicados en la resistencia a antibióticos y síntesis de productos naturales por RAST.

N <sup>o</sup>	Start	Stop	Strand	Length bp	Function	EC_number	Subsystem
<b>Antimicrobial Resistance (ARG)</b>							
1	3,029,254	3,030,006	-	753	Metal-dependent hydrolases of the beta-lactamase superfamily III	None	Beta-lactamase
2	8,127,255	8,128,037	+	783	Beta-lactamase class C and other penicillin binding proteins	None	Beta-lactamase
3*	4,038,335	4,040,419	+	2085	DNA gyrase subunit B	EC 5.99.1.3	Resistance to fluoroquinolones
4	4,040,462	4,043,056	+	2595	DNA gyrase subunit A	EC 5.99.1.3	Resistance to fluoroquinolones
5	3,404,967	3,405,500	-	534	Colicin E2 tolerance protein CbrC-like protein	None	Tolerance to colicin E2
<b>Natural Products synthesis</b>							
6	8,720	9,925	-	1206	Protein-L-isoaspartate O-methyltransferase	EC 2.1.1.77	Lanthionine Synthetases

«Continuación»

7	9,939	11,081	-	1143	Lanthionine biosynthesis cyclase LanC	None	Lanthionine Synthetases
8	6,132,982	6,135,630	+	2649	Lanthionine biosynthesis protein LanL	None	Lanthionine Synthetases
9	8,430,545	8,432,851	+	2307	TOMM biosynthesis cyclodehydratase (protein C) / TOMM biosynthesis docking scaffold (protein D)	None	Thiazole- oxazole- modified microcin (TOMM) synthesis

---

FUENTE: Elaboración propia.

\* CDS anotado también por PROKKA.

**Anexo 3.** Anotación de CDS del genoma de *Streptomyces* sp. B-146 relacionados a genes ARG usando RAST.

N	Start	Stop	Strand	Length (bp)	Function	Subsystem
<b>MFS Efflux Pumps</b>						
01	1,499,996	1,501,528	-	1533	Multidrug resistance efflux pump	None
02*	2,245,403	2,246,674	-	1272	Multidrug resistance transporter, Bcr/CflA family	Copper homeostasis
03	465,192	466,463	-	1272	Chloramphenicol resistance, MFS efflux pump => CmlV family	None
<b>ABC Efflux Pumps</b>						
04**	2,220,872	2,221,933	+	1062	Oxytetracycline resistance, ABC-type efflux pump ATP-binding component => Otr(C)	None
05*	2,221,930	2,222,754	+	825	Oxytetracycline resistance, ABC-type efflux pump permease component => Otr(C)	None
06*	4,093,373	4,094,221	-	849	Daunorubicin/doxorubicin resistance ABC transporter, permease protein	None

«Continuación»

SMR Efflux Pumps						
07	3,020,202	3,020,522	-	321	Small multidrug resistance family (SMR) protein	None
08	5,411,655	5,411,978	-	324	Small multidrug resistance family (SMR) protein	None
RND Efflux Pumps						
09	2,127,213	2,130,443	-	3231	RND multidrug efflux transporter; Acriflavin resistance protein	None
Ribosomal Protection proteins						
10	7,683,954	7,684,181	+	228	Tetracycline resistance, ribosomal protection type => Tet	None

FUENTE: Elaboración propia.

\* CDSs anotados por PROKKA

\*\* CDS anotado por PROKKA con codón de inicio en la posición 2,220,902 pb en lugar de los 2,220,872 pb registrado por RAST.



**Anexo 4.** Identificación de BGCs putativos en el genoma de *Streptomyces* sp. B-146 usando el servidor web antiSMASH versión 6.1.1.

smBGCs	Type	Position			BGC Similarity (%)	MIBiG ID	Reference strain (Accession)
		Start	End	Lenght			
01	Lanthipeptide Class I (RiPPs)	1	24,074	24,074	Unknown	None	None
02	Other, T1PKS, Oligosaccharide, NRPS- Like, T3PKS, NRPS	41,107	255,918	214,812	Lobophorin A (90%)	BGC0001183	<i>Streptomyces</i> sp. FXJ7.023 (JX306680)
03	Terpene	274,564	294,518	19,955	Ebelactone (5%)	BGC0001580	<i>Kitasatospora aburaviensis</i> (LT608336)
04	T3PKS	327,899	368,026	40,128	Germicidin (100%)	BGC0001454	<i>Streptomyces argillaceus</i> (LT989884)
05	Indole	505,150	526,277	21,128	5-isoprenylindole-3- carboxylate $\beta$ -D-glycosyl ester (28%)	BGC0001483	<i>Streptomyces</i> sp. RM-5-8 (KT895008)
06	Terpene	578,272	603,141	24,870	Isorenieratene (75%)	BGC0001456	<i>Streptomyces argillaceus</i> (LT989886)

«Continuación»

07	Amglyccycl	846,196	867,434	21,239	B-D- galactosylvalidoxylamine- A (22%)	BGC0000723	<i>Streptomyces hygroscopicus subsp. Limoneus</i> (DQ223652)
08	T3PKS	1,113,937	1,155,049	41,113	Herboxidiene (8%)	BGC0001065	<i>Streptomyces chromofuscus</i> (JN671974)
09	NRPS	1,226,411	1,279,185	52,775	Rimosamide (21%)	BGC0001760	<i>Streptomyces rimosus subsp. rimosus</i> ATCC 10970 (NZ_ANSJ01000026)
10	Ectoine	1,803,110	1,813,508	10,399	Ectoine (100%)	BGC0000853	<i>Streptomyces anulatus</i> (AY524544)
11	Melanin	2,783,163	2,793,789	10,627	Melanin (100%)	BGC0000910	<i>Streptomyces coelicolor</i> A3(2) (AL645882)
12	Lasso peptide (RiPPs)	2,854,106	2,876,627	22,522	Ssv-2083 (36%)	BGC0000579	<i>Streptomyces sviveus</i> ATCC 29083 (NZ_CM0000951)
13	Phenazine	3,602,742	3,623,182	20,441	Unknown	None	None
14	Lanthipeptide Class II (RiPPs)	4,320,836	4,349,008	28,173	Sbi-06990 a1 / sbi-06989 a2 (75%)	BGC0001229	<i>Streptomyces bingchenggensis</i> BCW-1 (CP002047)

«Continuación»

15	NRPS	5,227,634	5,280,721	53,088	Ishigamide (44%)	BGC0001623	<i>Streptomyces</i> sp. MSC090213JE08 (LC269948)
16	Betalactone	5,387,984	5,413,493	25,510	Julichrome Q3-3 / julichrome Q3-5 (22%)	BGC0002012	<i>Streptomyces afghaniensis</i> 772 (NZ_KE354310)
17	Terpene	5,414,606	5,434,311	19,706	Albaflavenone (100%)	BGC0000660	<i>Streptomyces coelicolor</i> A3(2) (AL645882)
18	T2PKS	5,484,866	5,557,420	72,555	Spore pigment (66%)	BGC0000271	<i>Streptomyces avermitilis</i> (AB070937)
19	Siderophore	6,027,742	6,037,976	10,235	Unknown	None	None
20	Lanthipeptide Class III (RiPPs)	6,122,982	6,145,630	22,649	Sapb (100%)	BGC0000551	<i>Streptomyces coelicolor</i> A3(2) (AL645882)
21	NRPS	6,170,020	6,219,416	49,397	Telomycin (29%)	BGC0001406	<i>Streptomyces canus</i> (KT881498)
22	T1PKS, NRPS	6,262,851	6,311,060	48,210	Xiamycin A (72%)	BGC0000666	<i>Streptomyces</i> sp. SCSIO 02999 (JQ812811)
23	RiPP-Like	6,356,461	6,367,586	11,126	Cyclothiazomycin (9%)	BGC0000603	<i>Streptomyces hygroscopicus</i> (FJ472825)

«Continuación»

24	Terpene	6,398,673	6,420,257	21,585	Geosmin (100%)	BGC0001181	<i>Streptomyces coelicolor</i> A3(2) (AL645882)
25	Siderophore	6,570,786	6,583,960	13,175	Unknown	None	None
26	NRPS	6,617,485	6,660,633	43,149	Diisonitrile antibiotic SF2768 (55%)	BGC0001574	<i>Streptomyces thioluteus</i> (KY427327)
27	PKS-Like, NRPS, Nucleoside	6,716,755	6,799,642	82,888	Caboxamycin (60%)	BGC0001444	<i>Streptomyces</i> sp. NTK 937 (JJOB01000001)
28	Terpene, NRPS	7,095,877	7,169,189	73,313	Hopene (92%)	BGC0000663	<i>Streptomyces coelicolor</i> A3(2) (AL645882)
29	NRPS, T1PKS, NRPS-Like	7,216,852	7,304,227	87,376	Mirubactin (78%)	BGC0000392	<i>Actinosynnema mirum</i> DSM 43827 (CP001630)
30	T1PKS	7,304,438	7,367,121	62,684	Divergolide A / divergolide B / divergolide C / divergolide D (79%)	BGC0001119	<i>Streptomyces</i> sp. HKI0576 (HF563079)
31	Terpene	7,680,551	7,701,579	21,029	Versipelostatin (5%)	BGC0001204	<i>Streptomyces versipellis</i> (LC006086)

«Continuación»

32	RiPP-Like	7,711,052	7,721,267	10,216	Informatipeptin (42%)	BGC0000518	<i>Streptomyces viridochromogenes</i> DSM 40736 (GG657757)
33	NRPS	7,927,228	7,976,771	49,544	Coelichelin (100%)	BGC0000325	<i>Streptomyces coelicolor</i> A3(2) (AL645882)
34	T1PKS	8,120,283	8,208,832	88,550	Elaiophylin (29%)	BGC0000053	Unknown. Unclassified (GP697151)
35	Thiopeptide (RiPPs)	8,420,545	8,446,013	25,469	Nocathiacin (4%)	BGC0000609	<i>Nocardia</i> sp. ATCC 202099 (GU564398)
36	NRPS	8,563,606	8,613,309	49,704	Kanamycin (1%)	BGC0000703	<i>Streptomyces kanamyceticus</i> (CP023699)

---

FUENTE: Elaboración propia.

**Anexo 5.** Clasificación de BGCs detectados en el genoma de *Streptomyces* sp. B-146 según los servidores web antiSMASH versión 6.1.1, ARTS 2.0, PRISM 4.4.5, RiPPMiner-Genome, BAGEL4 y NaPDoS.

smBGCs	BGCs Classification					
	antiSMASH	ARTS	PRISM	RiPPMiner-Genome	BAGEL	NaPDoS
01	Lanthipeptide Class I (RiPPs)	Lanthipeptide (RiPPs)	Class I Lantipeptide (RiPPs)	Lanthipeptide A (RiPPs)	None	None
02	Other, T1PKS, Oligosaccharide, NRPS-Like, T3PKS, NRPS	Other, oligosaccharide, T1PKS, T3PKS, NRPS,	PKS & PKS, NRPS *	None	None	KS Domains / C Domains
03	Terpene	Terpene	None	None	None	None
04	T3PKS	T3PKS	None	None	None	None
05	Indole	Indole	None	None	None	None
06	Terpene	Terpene	None	None	None	None
07	Amglyccycl	Amglyccycl	None	None	None	None

«Continuación»

08	T3PKS	T3PKS	None	None	None	None
09	NRPS	NRPS	NRPS	None	None	C Domains / KS Domains
10	Ectoine	Ectoine	Ectoine	None	None	None
A1	None	None	PKS	None	None	KS Domains
11	Melanin	Melanin	Melanin	None	None	None
12	Lasso peptide (RiPPs)	Lasso peptide (RiPPs)	Lasso peptide (RiPPs)	Lasso peptide (RiPPs)	None	None
13	Phenazine	Phenazine	None	None	None	None
14	Lanthipeptide Class II (RiPPs)	Lanthipeptide (RiPPs)	Class II lantipeptide (RiPPs)	Lanthipeptide B (RiPPs)	Lanthipeptide B (RiPPs)	None
A2	None	NRPS	None	Linaridin (RiPPs)	None	C Domains

«Continuación»

15	NRPS	NRPS	NRPS	None	None	C Domains
16	Betalactone	Betalactone	None	None	None	None
17	Terpene	Terpene	None	None	None	None
18	T2PKS	T2PKS	Pentangular polyphenol	None	None	KS Domains
A3	None	None	NRPS	None	None	KS Domains
19	Siderophore	Siderophore	None	None	None	None
20	Lanthipeptide Class III (RiPPs)	Lanthipeptide (RiPPs)	Class III/IV Lantipeptide (RiPPs)	Lanthipeptide C/D (RiPPs)	Lanthipeptide C (RiPPs)	None
21	NRPS	NRPS	NRPS	None	None	C Domains
22	T1PKS, NRPS	T1PKS, NRPS	PKS, NRPS	None	None	KS Domains / C Domains



«Continuación»

23	RiPP-Like	Bacteriocin (RiPPs)	None	None	None	None
24	Terpene	Terpene	None	None	None	None
25	Siderophore	Siderophore	None	None	None	None
26	NRPS	NRPS	NRPS	None	None	C Domains
27	PKS-Like, NRPS, Nucleoside	PKS-like, NRPS, Nucleoside	NRPS	None	None	C Domains
28	Terpene, NRPS	Terpene, NRPS	NRPS	None	Zoocin A (RiPPs)	C Domains
29	NRPS, T1PKS, NRPS-Like	NRPS, T1PKS	NRPS & PKS, NRPS *	None	None	C Domains
30	T1PKS	T1PKS	PKS, NRPS	None	None	KS Domains
31	Terpene	Terpene	None	None	None	None

«Continuación»

32	RiPP-Like	Bacteriocin (RiPPs)	None	None	None	None
33	NRPS	NRPS	NRPS	None	None	C Domains
34	T1PKS	T1PKS	PKS	None	None	KS Domains
35	Thiopeptide (RiPPs)	Thiopeptide	None	None	None	None
36	NRPS	NRPS	PKS/NRPS	None	None	C Domains

---

FUENTE: Elaboración propia.

**Nota:** La ubicación de referencia de los smBGCs (01 – 36) en el genoma de *Streptomyces* sp. B-146 fueron proporcionados por antiSMASH 6.1.1 (Anexo 4 y Figura 08). La ubicación de referencia de los smBGCs A1 (2,367,937pb -2,380,342pb) y A3 (5,813,667 - 5,823,247) fueron determinadas por PRISM 4.4.5. Mientras que la ubicación de referencia del smBGC A2 (4,643,226 - 4,703,014) fue determinada por ARTS 2.0.

\* PRISM 4.4.5, a diferencia de antiSMASH 6.1.1, no logro identificar a los smBGCs 02 y 29 como BGCs híbridos extensos y biosintéticamente diversos. Es decir, se limitó a reportar en cada caso dos clústeres independientes.

**Anexo 6.** Anotación de los genes contenidos en el smBGC22 detectado en el genoma de *Streptomyces* sp. B-146 a través de PROKKA, BlastKOALA, RAST, antiSMASH 6.1.1, UniProt y MIBiG.

ORFs	Annotation of smBGC_22 from <i>Streptomyces</i> sp. B-146					
	PROKKA	BlastKOALA	RAST	antiSMASH	UniProt	MIBiG
-6	Hypothetical protein	Solute: Na <sup>+</sup> symporter, SSS family	Sodium-solute symporter, putative	None	Sodium-coupled permease (89.1%)	Putative Sodium-coupled Permease <sup>(a)</sup> (100%)
-5	Hypothetical protein	None	Hypothetical protein	None	Uncharacterized protein (61%)	Putative Bacillolysin <sup>(a)</sup> (97.24%)
-4	Hypothetical protein	None	Hypothetical protein	None	HTH transcriptional regulator (66.1%)	XiaR: LuxR family transcriptional regulator <sup>(a)</sup> (99.57%)
-3	Hypothetical protein	None	Hypothetical protein	None	HTH transcriptional regulator (61.2%)	XiaQ: LuxR family transcriptional regulator <sup>(a)</sup> (99.09%)

«Continuación»

-2	Hypothetical protein	Geranylgeranyl diphosphate synthase, type I (2.5.1.10)	Hypothetical protein	None	Polyprenyl synthetase family protein (82.7%)	XiaP: Polyprenyl synthetase <sup>(a)</sup> (100%)
-1	Tryptophan 2,3-dioxygenase (1.13.11.11)	None	Hypothetical protein	None	FAD-binding domain-containing protein (68.4%)	XiaO: Epoxidase <sup>(a)</sup> [Monooxygenase] (97.79%)
1	(2E,6E) - farnesyl diphosphate synthase (2.5.1.10)	Geranylgeranyl diphosphate synthase, type I (2.5.1.10)	(2E,6E) - farnesyl diphosphate synthase (2.5.1.10)	Polyprenyl synthetase	Polyprenyl synthetase family protein (76.5%)	XiaN: Polyprenyl diphosphate synthase <sup>(a)</sup> (99.71%)
2	Cytochrome P450 FAS1 (1.14.-.-)	None	Putative cytochrome P450 hydroxylase	Cytochrome P450	Cytochrome P450 (82.5%)	XiaM: Carboxylase <sup>(a)</sup> [Cytochrome P450] (100%)
3	Ferredoxin-2	Ferredoxin	Ferredoxin	Ferredoxin	Ferredoxin (65.7%)	XiaL: Ferredoxin <sup>(a)</sup> (100%)

«Continuación»

4	2,4-dichlorophenol 6-monooxygenase (1.14.13.20)	2,4-Dichlorophenol 6-monooxygenase (1.14.13.20)	Hypothetical protein	Monooxygenase FAD-binding	FAD-dependent monooxygenase (77.3%)	XiaK: C,N- and N,N dimerization <sup>(a)</sup> [Aromatic ring hydroxylase] (99.5%)
5	Limonene-1,2-epoxide hydrolase (3.3.2.8)	Limonene-1,2-epoxide hydrolase (3.3.2.8)	Hypothetical protein	None	Limonene-1,2-epoxide hydrolase domain-containing protein (75.2%)	XiaJ: Limonene-1,2-epoxide hydrolase <sup>(a)</sup> (100%)
6	Flavin-dependent monooxygenase, oxygenase subunit HsaA (1.14.14.12)	3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione monooxygenase (1.14.14.12)	Acyl-CoA dehydrogenase	Acyl-CoA dehydrogenase type 2	Acyl-CoA dehydrogenase C-terminal domain-containing protein (79.6%)	XiaI: Oxidative terpene cyclase <sup>(a)</sup> [Indole oxygenase / acyl-CoA dehydrogenase] (99.5%)
7	Hypothetical protein	None	Hypothetical protein	None	Integral membrane protein (78%)	XiaH: Terpene cyclase <sup>(a)</sup> [Putative membrane protein] (93.98%)
8	Hypothetical protein	None	Transcriptional regulator, LuxR family	LuxR family DNA-binding response regulator	HTH transcriptional regulator (81.2%)	XiaG: LuxR family regulator <sup>(a)</sup> (99.58%)

«Continuación»

9	HMBPP synthase (1.17.7.3)	HMBPP synthase (1.17.7.3)	HMBPP synthase (1.17.7.3)	None	HMBPP synthase (1.17.7.3) (92.7%)	XiaF: HMBPP synthase <sup>(a)</sup> (100%)
10	DXP synthase (2.2.1.7)	DXP synthase (2.2.1.7)	DXP synthase (2.2.1.7)	DXP synthase	DXP synthase (2.2.1.7) (82.1%)	XiaE: DXP synthase <sup>(a)</sup> (99.83%)
11	HMBPP reductase (1.17.7.4)	HMBPP reductase (1.17.7.4)	HMBPP reductase (1.17.7.4)	None	HMBPP reductase (1.17.7.4) (83.5%)	XiaD: HMBPP reductase <sup>(a)</sup> (99.7%)
12	None	None	Hypothetical protein	IclR family transcriptional regulator	IclR family transcriptional regulator (74.4%)	XiaC: IclR family transcriptional regulator <sup>(a)</sup> (99.61%)
13	Putative flavin reductase (1.5.1.-)	None	Nitrilotriacetate monooxygenase component B (1.14.13.-)	Flavin reductase domain protein FMN-binding	Flavin reductase like domain-containing protein (80%)	XiaB: FMN-binding flavin reductase <sup>(a)</sup> (100%)

«Continuación»

14	Putative 3-ketosteroid-9-alpha-monooxygenase, oxygenase component (1.14.13.-)	None	Phenylpropionate dioxygenase and related ring-hydroxylating dioxygenases, large terminal subunit	None	Rieske-type oxygenase (96.2%)	XiaA: Rieske 2Fe-2S domain-containing oxygenase <sup>(a)</sup> (100%)
15	Putative NADP-dependent oxidoreductase YfmJ (1.-.-.-)	NADPH: quinone reductase (1.6.5.5)	Putative oxidoreductase	Crotonyl-CoA reductase / alcohol dehydrogenase	NADP-dependent oxidoreductase (97.4%)	PtmC: Alcohol Dehydrogenase <sup>(b)</sup> [Putative alcohol dehydrogenase zinc-binding domain protein] <sup>(a)</sup> (99.72%)
16	Phytoene desaturase (lycopene-forming) (1.3.99.31)	None	Putative FAD-dependent oxidoreductase	Phytoene desaturase (lycopene-forming)	Amine oxidase (95.8%)	PtmB2: Phytoene dehydrogenase <sup>(b)</sup> [Putative Ox3] <sup>(a)</sup> (100%)
17	Hypothetical protein	None	Putative FAD-dependent oxidoreductase	dehydrogenase	Amine oxidase (95%)	PtmB1: FAD-dependent oxidoreductase <sup>(b)</sup> (100%)

«Continuación»

18	D-alanine-D-alanyl carrier protein ligase (6.2.1.54)	Hybrid PKS/NRPS FtdB	Polyketide synthase modules and related proteins	dehydrogenase	Amino acid adenylation domain-containing protein (94.3%)	PtmA: hybrid PKS/NRPS <sup>(b)</sup> (99.49%)
19	Hypothetical protein	None	Fatty acid hydroxylase family (carotene hydroxylase / sterol desaturase)	Beta-ketoacyl synthase	Fatty acid hydroxylase domain-containing protein (92.4%)	PtmD: hydroxylase <sup>(b)</sup> (99.34%)
20	Hypothetical protein	None	Putative DNA helicase	None	DNA helicase IV (92.1%)	None
21	Hypothetical protein	None	Helicase, C-terminal: Type III restriction enzyme, res subunit: DEAD / DEAH box helicase, N-terminal	None	DEAD/DEAH box helicase (86.4%)	None
22	Hypothetical protein	None	Hypothetical protein	None	DNA replication and repair protein RecF (73.2%)	None



«Continuación»

23	Hypothetical protein	None	Hypothetical protein	None	Beta protein (90.4%)	None
24	ATP-dependent RecD-like DNA helicase (3.6.4.12)	None	Hypothetical protein	None	ATP-dependent RecD-like DNA helicase (3.6.4.12) (93.3%)	None
25	Hypothetical protein	None	Hypothetical protein	None	Putative membrane protein (85.3%)	None
26	Hypothetical protein	None	Putative regulatory protein	None	Histidine kinase/HSP90-like ATPase domain-containing protein (88.5%)	None
27	Hypothetical protein	None	Hypothetical protein	None	HTH cro/C1-type domain-containing protein (97.2%)	None

«Continuación»

28	Hypothetical protein	None	Small DUF397-containing protein	None	DUF397 domain-containing protein (95.1%)	None
29	Hypothetical protein	None	Possible Restriction / modification enzyme	None	Restriction endonuclease (96.8%)	None
30	Hypothetical protein	None	Hypothetical protein	None	Predicted protein	Hypothetical protein <sup>(b)</sup>
31	Hypothetical protein	None	Hypothetical protein	None	Transposase IS701-like DDE domain-containing protein (93.8%)	None
32	Hypothetical protein	None	Putative secreted protein	None	Transglycosylase SLT domain-containing protein (94.8%)	None

«Continuación»

33	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (2.5.1.7)	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (2.5.1.7)	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (2.5.1.7)	None	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (98.4%)	None
34	Hypothetical protein	None	Hypothetical protein	None	DUF4236 domain-containing protein (95.4%)	None
35	Aconitate hydratase A (4.2.1.3)	Aconitate hydratase (4.2.1.3)	Aconitate hydratase (4.2.1.3)	None	Aconitate hydratase (4.2.1.3) (97.8%)	None
36	Hypothetical protein	None	Hypothetical protein	None	DUF397 domain-containing protein (80.4%)	None

---

FUENTE: Elaboración propia.

(a) Anotación proporcionada por el BGC de Xiamicina A de *Streptomyces* sp. SCSIO 02999 (Accesión MIBiG: BGC0000666) (Li *et al.*, 2012).

(b) Anotación proporcionada por el BGC de Pactamidas de *Streptomyces pactum* (accesión MIBiG: BGC0001726) (Saha *et al.*, 2017).

**Anexo 7.** Anotación de los genes contenidos en el smBGC17 detectado en el genoma de *Streptomyces* sp. B-146 a través de PROKKA, BlastKOALA, RAST, antiSMASH 6.1.1, UniProt Y MIBiG.

ORFs	Annotation of smBGC_17 from <i>Streptomyces</i> sp. B-146					
	PROKKA	BlastKOALA	RAST	antiSMASH	UniProt	MIBiG
1	Hypothetical protein	None	Integral membrane protein	None	M50 family peptidase (92.9%)	None
2	SOS response-associated protein YedK (3.4.-.-)	None	Hypothetical protein	None	Abasic site processing protein (3.4.-.-) (95.2%)	None
3	Hypothetical protein	Uncharacterized protein	Hydrolase SCO5215, alpha/beta fold family	None	KANL3/Tex30 alpha/beta hydrolase-like domain-containing protein (89.6%)	None
4	Hypothetical protein	RNA polymerase sigma-70 factor, ECF subfamily	RNA polymerase ECF-type sigma factor	RNA polymerase, sigma-24 subunit, ECF subfamily	RNA polymerase sigma factor (93.7%)	None
5	Anti-sigma factor RsrA	None	Anti-sigma factor RsrA	None	Mycothioli system anti-sigma-R factor (93.2%)	None

«Continuación»

6	Hypothetical protein	None	Integral membrane protein	None	Integral membrane protein (86.3%)	None
7	Hypothetical protein	None	Putative lipoprotein	None	Lipoprotein (87.8%)	None
8	Hypothetical protein	None	Hypothetical protein	None	Tetratricopeptide repeat protein (98.5%)	None
9	Peptide deformylase (3.5.1.88)	Peptide deformylase (3.5.1.88)	Peptide deformylase (3.5.1.88)	Peptide deformylase	Peptide deformylase 4 (3.5.1.88) (95.8%)	None
10	Epi-isozizaene synthase (4.2.3.37)	None	Terpene cyclase	Terpene synthase/ cyclase metal-binding domain protein	Epi-isozizaene synthase (4.2.3.37) (92.9%)	Putative lyase <sup>(a)</sup>
11	Epi-isozizaene 5-monooxygenase/ (E)-beta-farnesene synthase (1.14.13.106)	None	Hypothetical protein	Cytochrome P450	Epi-isozizaene 5-monooxygenase/ Beta-farnesene synthase (1.14.15.39) (4.2.3.47) (93.3%)	Putative cytochrome P450 <sup>(a)</sup>

«Continuación»

12	HTH-type transcriptional regulator CdhR	None	Transcriptional regulator, AraC family	AraC family transcriptional regulator	Helix-turn-helix domain-containing protein (94.1%)	None
13	Hypothetical protein	None	Hypothetical protein	None	Secreted protein (92.1%)	None
14	Ribonucleoside-diphosphate reductase subunit beta (1.17.4.1)	Ribonucleoside-diphosphate reductase beta chain (1.17.4.1)	Ribonucleotide reductase of class Ia (aerobic), beta subunit (1.17.4.1)	None	Ribonucleoside-diphosphate reductase subunit beta (1.17.4.1) (96.2%)	None
15	Hypothetical protein	Ribonucleoside-diphosphate reductase alpha chain (1.17.4.1)	Ribonucleotide reductase of class Ia (aerobic), alpha subunit (1.17.4.1)	None	Ribonucleoside-diphosphate reductase (1.17.4.1) (93.4%)	None
16	Hypothetical protein	None	Acetyltransferase	None	Acetyltransferase (85.7%)	None
17	Mono-carboxylate transport permease protein	Solute:Na <sup>+</sup> symporter, SSS family	Na <sup>+</sup> /solute symporter	None	Sodium:solute symporter (96.1%)	None

«Continuación»

18	Hypothetical protein	None	Integral membrane protein	None	Integral membrane protein (84.6%)	None
19	HTH-type transcriptional repressor DasR	GntR family transcriptional regulator, nutrient-sensing system regulator	Predicted transcriptional regulator of N-Acetylglucosamine utilization, GntR family	GntR family transcriptional regulator	HTH-type transcriptional repressor DasR (99.2%)	None

---

FUENTE: Elaboración propia.

(a) Anotación proporcionada por el BGC de albaflavenone de *Streptomyces coelicolor* A3(2) (Accesión MIBiG: BGC0000660) (Zhao *et al.*, 2008).

**Anexo 8.** Anotación de los genes contenidos en el smBGC20 detectado en el genoma de *Streptomyces* sp. B-146 a través de PROKKA, BlastKOALA, RAST, antiSMASH, BAGEL 6.1.1, UniProt Y MIBiG.

ORFs	Annotation of smBGC_20 from <i>Streptomyces</i> sp. B-146						
	PROKKA	BlastKOALA	RAST	antiSMASH	BAGEL	UniProt	MIBiG
1	Hypothetical protein	None	Hypothetical protein	None	None	GIY-YIG domain-containing protein (79.6%)	None
2	Hypothetical protein	None	Hypothetical protein	Serine/threonine protein kinase	Serine/threonine-protein kinase PrkC	non-specific serine/threonine protein kinase, 2.7.11.1 (94.8%)	None
3	Hypothetical protein	None	Hypothetical protein	None	Putative uncharacterized protein YjhR	Phospholipase D-like domain-containing protein (94.4%)	None
4	Hypothetical protein	None	Hypothetical protein	None	None	Nucleotidyl transferase AbiEii/AbiGii toxin family protein (95.5%)	None
5	Hypothetical protein	None	Putative transcriptional regulator	None	None	Transcriptional regulator (91.5%)	None



«Continuación»

6	Putative SapB synthase (2.7.-.-)	Probable SapB synthase	Lanthionine biosynthesis protein LanL	Serine/ threonine protein kinase	None	Lantipeptide synthetase (88.9%)	RamC: Putative serine/ threonine protein kinase <sup>(a)</sup> (77.43%)
7	Lanthionine-containing peptide SapB precursor RamS	Lanthionine-containing peptide SapB	Hypothetical protein	None	SapB	Lanthionine-containing peptide SapB (100%)	RamS: Hypothetical protein SC5A7.32 <sup>(a)</sup> (95.40%)
8	ABC transporter ATP-binding protein RamA	ABC, subfamily B, bacterial RamA/AmfB	Hypothetical protein	ABC transporter related protein	ABC transporter ATP-binding protein RamA	ABC transporter ATP-binding protein (77.1%)	RamA: Putative ABC transporter <sup>(a)</sup> (71.81%)
9	ABC transporter ATP-binding protein RamB	ABC, subfamily B, bacterial RamB/AmfA	Heterodimeric efflux ABC transporter, permease/ATP-binding subunit 2	ABC transporter related protein	ABC transporter ATP-binding protein RamB	ABC domain-containing protein (84.8%)	RamB: Putative ABC transporter <sup>(a)</sup> (68.76%)
10	Response regulator RamR	Two-component system, NarL family, response regulator DesR	Two-component transcriptional response regulator, LuxR family	LuxR family DNA-binding response regulator	Response regulator RamR	DNA-binding response regulator (84.3%)	RamR: Response regulator transcription factor <sup>(b)</sup> (43.69%) Transcriptional regulator <sup>(c)</sup> (42.86%)

«Continuación»

11	Rhamnolipids biosynthesis 3-oxoacyl-[acyl-carrier-protein] reductase (1.1.1.100)	None	Oxidoreductase, short-chain dehydrogenase/reductase family	Short-chain dehydrogenase/reductase SDR	Monensin polyketide synthase putative ketoacyl reductase	Short-subunit dehydrogenase (90.2%)	None
12	HTH-type transcriptional regulator HdfR	None	Transcriptional regulator, LysR family	LysR family transcriptional regulator	HTH-type transcriptional regulator AlsR	LysR family transcriptional regulator (90.3%)	None
13	Peptidase E (3.4.13.21)	Dipeptidase E (3.4.13.21)	Alpha-aspartyl dipeptidase Peptidase E (3.4.13.21)	None	Peptidase E	Dipeptidase PepE (80.5%)	None
14	HTH-type transcriptional regulator XynR	None	Transcriptional regulator, IclR family	IclR family transcriptional regulator	HTH-type transcriptional regulator XynR	IclR family transcriptional regulator (86.1%)	None
15 <sup>(*)</sup>	Hypothetical protein	None	Undetected	None	None	Un-characterized protein (29.6%)	None

FUENTE: Elaboración propia.

(a) Anotación proporcionada por el BGC del lantipéptido SapB de *Streptomyces coelicolor* A3(2) (Accesión MIBiG: BGC0000551) (Kodani et al., 2004).

- (b) Anotación proporcionada por el BGC del lantipéptido SRO15-2212 de *Streptomyces filamentosus* NRRL 15998 (Accesión MIBiG: BGC0000553) (Kersten *et al.*, 2011).
- (c) Anotación proporcionada por el BGC del lantipéptido AmfS de *Streptomyces griseus* subsp. *griseus* NBRC 13350 (Accesión MIBiG: BGC0000496) (Ueda *et al.*, 2002).
- (\*) El ORF15 del smBGC20 de *Streptomyces* sp. no fue detectado por RAST, tampoco llegó a ser anotado por Prokka, BlastKOALA, antiSMASH 6.1.1 ni BAGEL4. Según la anotación manual con UniProt no guarda relación con ninguna proteína del género bacteriano *Streptomyces*, es más solo llegó a tener una identidad del 29.6% con una proteína no caracterizada de la planta medicinal *Jatropha curcas*.

**Anexo 9.** Identificación de arreglos CRISPR en el genoma de *Streptomyces* sp. B-146 usando el servidor web RAST.

Feature ID	Type	Start	Stop	Strand	Length (bp)	Sequence
ARRAY_01	Array	8203	8659	+	457	CTGCTCCCCGCGCCCGCGGGGATGGTCCC*
Repeat_01	Repeat	8203	8231	+	29	CTGCTCCCCGCGCCCGCAGGGATGGTCCC
Spacer_01	Spacer	8232	8263	+	32	-
Repeat_02	Repeat	8264	8292	+	29	CTGCTCCCCGCGCCCGCGGGGATGGTCCC
Spacer_02	Spacer	8293	8324	+	32	-
Repeat_03	Repeat	8325	8353	+	29	CTGCTCCCCGCGCCCGCGGGGATGGTCCC
Spacer_03	Spacer	8354	8385	+	32	-
Repeat_04	Repeat	8386	8414	+	29	CTGCTCCCCGCGCCCGCGGGGATGGTCCC
Spacer_04	Spacer	8415	8446	+	32	-
Repeat_05	Repeat	8447	8475	+	29	CTGCTCCCCGCGCCCGCGGGGATGGTCCC
Spacer_05	Spacer	8476	8507	+	32	-
Repeat_06	Repeat	8508	8536	+	29	CTGCTCCCCGCGCCCGCGGGGATGGTCCC
Spacer_06	Spacer	8537	8568	+	32	-
Repeat_07	Repeat	8569	8597	+	29	CTGCTCCCCGCGCCCGCGGGGATGGGTTT
Spacer_07	Spacer	8598	8630	+	33	-
Repeat_08	Repeat	8631	8659	+	29	CTACTGCGCGCCAGAGGCATCGTCCC

«Continuación»

ARRAY_02	Array	19482	19997	+	516	CTGCTCCCCGCACCCGCGGGGATGGTCCC*
Repeat_01	Repeat	19482	19510	+	29	CTGCTCCCCGCACCCGCGGGG <u>G</u> TGGTCCC
Spacer_01	Spacer	19511	19542	+	32	-
Repeat_02	Repeat	19543	19571	+	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
Spacer_02	Spacer	19572	19603	+	32	-
Repeat_03	Repeat	19604	19632	+	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
Spacer_03	Spacer	19633	19664	+	32	-
Repeat_04	Repeat	19665	19693	+	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
Spacer_04	Spacer	19694	19725	+	32	-
Repeat_05	Repeat	19726	19754	+	29	CTGCTCCCCGCACCCGCGGGGA <u>C</u> GGTCCC
Spacer_05	Spacer	19755	19786	+	32	-
Repeat_06	Repeat	19787	19815	+	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
Spacer_06	Spacer	19816	19847	+	32	-
Repeat_07	Repeat	19848	19876	+	29	CTGCTCCCCGCACCC <u>C</u> CGGGGATGGTCCC
Spacer_07	Spacer	19877	19908	+	32	-
Repeat_08	Repeat	19909	19937	+	29	CTGCTCC <u>T</u> CGCACCCGCGGGGATGGTCCC
Spacer_08	Spacer	19938	19968	+	31	-
Repeat_09	Repeat	19969	19997	+	29	CTGCTCCCCGCACCCGCGGGGATG <u>A</u> TCC <u>G</u>

FUENTE: Elaboración propia.

\* Consensus sequence.

**Anexo 10.** Identificación de arreglos CRISPR en el genoma de *Streptomyces* sp. B-146 usando el servidor web CRISPRCasFinder.

Feature ID	Type	Start	Stop	Length (bp)	Sequence
ARRAY_01	Array	8203	8660	458	CTGCTCCCCGCGCCCGCGGGGATGGTCCCG*
Repeat_01	Repeat	8203	8232	30	CTGCTCCCCGCGCCCGC <u>A</u> GGGATGGTCCCG
Spacer_01	Spacer	8233	8263	31	GCCGTAACAGGGCCGCGGTCCGCTGCTCGCC
Repeat_02	Repeat	8264	8293	30	CTGCTCCCCGCGCCCGCGGGGATGGTCCCG
Spacer_02	Spacer	8294	8324	31	AGAGTCGCTGGACCGAGCCGACTTGGATTGG
Repeat_03	Repeat	8325	8354	30	CTGCTCCCCGCGCCCGCGGGGATGGTCCCG
Spacer_03	Spacer	8355	8385	31	GCATCTTCGACCGGGGCCGGGCAGCGTTCAT
Repeat_04	Repeat	8386	8415	30	CTGCTCCCCGCGCCCGCGGGGATGGTCCCG
Spacer_04	Spacer	8416	8446	31	TCGGGGACTCGATGCTCAAGGGCGACTGGGT
Repeat_05	Repeat	8447	8476	30	CTGCTCCCCGCGCCCGCGGGGATGGTCCCG
Spacer_05	Spacer	8477	8507	31	TGACCGGCATCAACAAGGGCTCGGTCTCGAA
Repeat_06	Repeat	8508	8537	30	CTGCTCCCCGCGCCCGCGGGGATGGTCCCT
Spacer_06	Spacer	8538	8568	31	ACACGCAGGTTATCGAGACGGCCGTGCGGCA
Repeat_07	Repeat	8569	8598	30	CTGCTCCCCGCGCCCGCGGGGATGG <u>GTTT</u> G
Spacer_07	Spacer	8599	8630	32	GCGGGCGCGCAGGTAGAGGCGGTCAAGCGGCG
Repeat_08	Repeat	8631	8660	30	CT <u>A</u> CT <u>G</u> CGCGCGCC <u>A</u> G <u>A</u> GG <u>C</u> GAT <u>C</u> GTCCCT

«Continuación»

ARRAY_02	Array	19482	19997	516	CTGCTCCCCGCACCCGCGGGGATGGTCCC*
Repeat_01	Repeat	19482	19510	29	CTGCTCCCCGCACCCGCGGGG <u>G</u> TGGTCCC
Spacer_01	Spacer	19511	19542	32	AAGGACATCGACACGCTGACGTGGGACATCAG
Repeat_02	Repeat	19543	19571	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
Spacer_02	Spacer	19572	19603	32	GCGACGCCGAGCCACCAGGCGAGGGAGATGAT
Repeat_03	Repeat	19604	19632	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
Spacer_03	Spacer	19633	19664	32	ACGCGCCGGTCCACCGTCGTGCTCGGGGACAG
Repeat_04	Repeat	19665	19693	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
Spacer_04	Spacer	19694	19725	32	AGGATCCGACACCGCAGCGACAGACCCGTCCG
Repeat_05	Repeat	19726	19754	29	CTGCTCCCCGCACCCGCGGGGA <u>C</u> GGTCCC
Spacer_05	Spacer	19755	19786	32	GGCCACAGGACACCCATGCCAGCCCCCTCGG
Repeat_06	Repeat	19787	19815	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
Spacer_06	Spacer	19816	19847	32	ACCGGCGCCCCCGGCGGCGCAGCAGAGGACGG
Repeat_07	Repeat	19848	19876	29	CTGCTCCCCGCACCC <u>C</u> CGGGGATGGTCCC
Spacer_07	Spacer	19877	19908	32	GCCACAGCCGTACCCGCGCGGGACAGCTCAAC
Repeat_08	Repeat	19909	19937	29	CTGCTCC <u>T</u> CGCACCCGCGGGGATGGTCCC
Spacer_08	Spacer	19938	19968	31	TGCTGGCCGCTACGACGCGGCCCGCACGAGC
Repeat_09	Repeat	19969	19997	29	CTGCTCCCCGCACCCGCGGGGATG <u>A</u> TCC <u>G</u>

FUENTE: Elaboración propia.

\* Consensus sequence.

**Anexo 11.** Identificación de arreglos CRISPR en el genoma de *Streptomyces* sp. B-146 usando el servidor web CRISPRminer.

Feature ID	Type	Start	Stop	Length (bp)	Sequence
ARRAY_01	Array	8203	8568	366	CTGCTCCCCGCGCCCGCGGGGATGGTCCCG*
Repeat_01	Repeat	8203	8232	30	CTGCTCCCCGCGCCCGCAGGGATGGTCCCG
Spacer_01	Spacer	8233	8263	31	GCCGTAACAGGGCCGCGGTCCGCTGCTCGCC
Repeat_02	Repeat	8264	8293	30	CTGCTCCCCGCGCCCGCGGGGATGGTCCCG
Spacer_02	Spacer	8294	8324	31	AGAGTCGCTGGACCGAGCCGACTTGGATTGG
Repeat_03	Repeat	8325	8354	30	CTGCTCCCCGCGCCCGCGGGGATGGTCCCG
Spacer_03	Spacer	8355	8385	31	GCATCTTCGACCGGGGCCGGGCAGCGTTCAT
Repeat_04	Repeat	8386	8415	30	CTGCTCCCCGCGCCCGCGGGGATGGTCCCG
Spacer_04	Spacer	8416	8446	31	TCGGGGACTCGATGCTCAAGGGCGACTGGGT
Repeat_05	Repeat	8447	8476	30	CTGCTCCCCGCGCCCGCGGGGATGGTCCCG
Spacer_05	Spacer	8477	8507	31	TGACCGGCATCAACAAGGGCTCGGTCTCGAA
Repeat_06	Repeat	8508	8537	30	CTGCTCCCCGCGCCCGCGGGGATGGTCCCT
Spacer_06	Spacer	8538	8568	31	ACACGCAGGTTATCGAGACGGCCGTGCGGCA
ARRAY_02	Array	19482	19997	516	CTGCTCCCCGCACCCGCGGGGATGGTCCC*
Repeat_01	Repeat	19482	19510	29	CTGCTCCCCGCACCCGCGGGGGTGGTCCC
Spacer_01	Spacer	19511	19542	32	AAGGACATCGACACGCTGACGTGGGACATCAG
Repeat_02	Repeat	19543	19571	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC



«Continuación»

Spacer_02	Spacer	19572	19603	32	GCGACGCCGAGCCACCAGGCGAGGGAGATGAT
Repeat_03	Repeat	19604	19632	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
Spacer_03	Spacer	19633	19664	32	ACGCGCCGGTCCACCGTCGTGCTCGGGGACAG
Repeat_04	Repeat	19665	19693	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
Spacer_04	Spacer	19694	19725	32	AGGATCCGACACCGCAGCGACAGACCCGTCCG
Repeat_05	Repeat	19726	19754	29	CTGCTCCCCGCACCCGCGGGGA <u>C</u> GGTCCC
Spacer_05	Spacer	19755	19786	32	GGCCACAGGACACCCATGCCAGCCCCCTCGG**
Repeat_06	Repeat	19787	19815	29	CTGCTCCCCGCACCCGCGGGGATGGTCCC
Spacer_06	Spacer	19816	19847	32	ACCGGCGCCCCCGGCGGCGCAGCAGAGGACGG
Repeat_07	Repeat	19848	19876	29	CTGCTCCCCGCACCC <u>C</u> CGGGGATGGTCCC
Spacer_07	Spacer	19877	19908	32	GCCACAGCCGTACCCGCGCGGGACAGCTCAAC
Repeat_08	Repeat	19909	19937	29	CTGCTCC <u>T</u> CGCACCCGCGGGGATGGTCCC
Spacer_08	Spacer	19938	19968	31	TGCTGGCCGCTACGACGCGGCCCGCACGAGC
Repeat_09	Repeat	19969	19997	29	CTGCTCCCCGCACCCGCGGGGATG <u>A</u> TCC <u>G</u>

FUENTE: Elaboración propia.

\* Consensus sequence.

\*\* Self-Targeting: Start: 8536769, End: 8536796.