



Universidad del
Rosario

Genomics and *in vitro* characterization of Actinobacteria from tropical environments reveal their bioactive potential

Genómica y caracterización *in vitro* de Actinobacterias de ambientes tropicales revela su potencial bioactivo

María Nathalia Vargas Flórez

**Universidad del Rosario
Facultad de Ciencias Naturales
Bogotá, Colombia
Agosto-2022**

Genomics and *in vitro* characterization of Actinobacteria from tropical environments reveals their bioactive potential

Genómica y caracterización *in vitro* de Actinobacterias de ambientes tropicales revela su potencial bioactivo

María Nathalia Vargas Flórez

Trabajo de grado presentado como requisito para obtener el título de:

(Magister en Ciencias Naturales)

Director

María Mercedes Zambrano PhD

Codirector

Adriana Corrales PhD

Tutor

María Mercedes Zambrano PhD

**Facultad en Ciencias Naturales
Maestría en Ciencias Naturales
Universidad del Rosario
Bogotá, Colombia
Agosto -2022-**

Abstract

Actinobacteria from pristine tropical environments such as the Colombian paramos could have novel adaptations in biosynthetic gene clusters (BGCs), resistance genes and bioactive compounds. In this study, the functional potential of seven Actinobacteria from the National Natural Park (NNP) of Nevados and the NNP of Chingaza was characterized under *in vitro* laboratory conditions and the genomic level. 16 isolates from these environments were tested for antimicrobial activity against clinically important bacteria such as *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae* and *Pseudomonas aeruginosa*. Seven out of these 16 isolates were confirmed as Actinobacteria through 16S rRNA gene sequencing. The supernatants from some of these Actinobacterial isolates, such as *Arthrobacter*, *Streptomyces*, *Subtercola*, *Amycolatopsis* and *Rhodococcus* inhibited the growth of *E. coli* by 47%, *K. pneumoniae* by 68%, *S. aureus* by 15% and *P. aeruginosa* by 28%. The genomes of these seven Actinobacteria were sequenced through Nanopore (MinION) and five of these genomes were also sequenced through Illumina (Novaseq 6000). After long-read *de novo* assembly (for Nanopore) and hybrid assemblies (for Nanopore and Illumina), we found BGCs coding for various biosynthetic pathways such as terpenes, Ribosomally synthesized and post-translationally modified peptides (RiPP-like) products, Nonribosomal peptides (NRPS), Polyketide synthases (PKS), ectoine, siderophores and oligosaccharides. We found coding potential for 33 bioactive compounds with antibiotic, antioxidant, antitumor, chelating and osmoregulatory functions. Moreover, we found 2557 resistance genes associated with antibiotic expulsion, antibiotic target alteration and replacement, antibiotic inactivation, and reduced antibiotic permeability. The *in vitro* antimicrobial activity of these isolates, along with the presence of BGCs and antibiotic resistance genes, shows the genomic diversity and functional capacity associated with the production of bioactive metabolites by Actinobacteria from the paramos. This work lays the groundwork to characterize new antimicrobial compounds and functional potential in environmental microorganisms.

Key Words

Antibacterial activity, natural products, Actinobacteria, genome sequencing.

Resumen

Las Actinobacterias de ambientes tropicales no intervenidos provenientes de los ecosistemas de páramo de Colombia, pueden ser novedosos en sus adaptaciones microbianas por la presencia de BGCs (clústers de genes biosintéticos), genes de resistencia y moléculas bioactivas. En este estudio se caracterizó el potencial funcional de siete Actinobacterias provenientes del Parque Nacional Natural (PNN) de los Nevados y PNN Chingaza en condiciones de laboratorio y a nivel genómico. 16 aislamientos fueron evaluados mediante pruebas de actividad antimicrobiana contra bacterias de importancia clínica como *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae* y *Pseudomonas aeruginosa*. De estos, se seleccionaron siete del filum Actinobacteria según clasificación por el gen 16S rRNA. Algunos de los sobrenadantes de estas Actinobacterias, pertenecientes a los géneros *Arthrobacter* sp., *Streptomyces*, *Subtercola* sp, *Amycolatopsis* sp. y *Rhodococcus* sp., inhibieron el crecimiento de *E. coli* en un 47%, *K. pneumoniae* en un 68%, *S. aureus* 15% y *P. aeruginosa* 28%. Se secuenciaron siete genomas mediante MinION (NanoPore) y cinco mediante Illumina (Novaseq 6000). Tras ensamblaje, *de novo* de secuencias largas (ONT) e híbridos (ONT+Illumina), más anotación funcional, se encontraron BGCs de diversas rutas biosintéticas como terpenos, RiPP-like, NRPS, PKS, ectoína, sideróforos y oligosacáridos. Estos BGCs se asociaron a 33 posibles compuestos bioactivos con funciones antibióticas, antioxidantes, antitumorales, quelantes y de osmorregulación. Se encontraron 2557 genes de resistencia asociados a mecanismos de resistencia por expulsión del antibiótico, alteración y reemplazo del diana del antibiótico, inactivación del antibiótico y reducción de la permeabilidad del antibiótico. Tanto la actividad antimicrobiana *in vitro* de estos aislamientos, como la presencia de BGCs y de genes de resistencia a antibióticos indican diversidad genómica y capacidad funcional asociada a producción de metabolitos bioactivos en Actinobacterias de páramo. Este trabajo sienta las bases para caracterizar nuevos compuestos antimicrobianos y potencial funcional en microorganismos ambientales.

Palabra Clave

Antibacterial, productos naturales, Actinobacteria

AGRADECIMIENTOS

- Agradecimientos a las Directoras, Dra. Patricia del Portillo y Dra. María Mercedes Zambrano y demás miembros de la Corporación Corpogen por abrirme las puertas del laboratorio y guiarme para realizar la experimentación de mi trabajo de grado.
- A la Universidad del Rosario por el apoyo financiero y a mis maestros del programa de maestría en Ciencias Naturales quienes siempre me brindaron las herramientas para mejorar como profesional.
- A la Dra. María Mercedes Zambrano, por abrirme las puertas en su grupo de investigación y por la confianza depositada en el desarrollo de este trabajo que me ha permitido crecer profesional y personalmente.
- A la Dra. Adriana Corrales por su guía y por el apoyo financiero en la secuenciación Illumina.
- A Fernando Vargas Zúñiga por brindarme su apoyo financiero y motivarme durante este proceso.
- A la Dra. Luz Nelly Díaz y la Dra. Luz Karime Afanador por su guía durante el desarrollo de las pruebas de actividad y experimentos durante el trabajo en el laboratorio
- A Christian Romero por su apoyo durante el montaje de librerías NanoPore.
- A Jeisson Alejandro Triana por el apoyo brindado con los protocolos de bioinformática.
- A la Dra. Maritsa Cruz, el Dr. Juan Manuel Anzola y el Dr. Joseph Petrone por la guía y el apoyo en protocolos bioinformáticos de ensamblaje
- Este trabajo de grado está dedicado a Rosaura Flórez Corredor, mi madre y mejor amiga, quien me ha brindado su amor, dedicación, paciencia y confianza incondicional a lo largo de mi vida académica.
- Proyecto financiado por MinCiencias (contrato No. 004-2020, proyecto No. 6396100270500)

INTRODUCCIÓN

El filo Actinobacteria comprende un extenso grupo de bacterias no móviles, Gram-positivas, con contenido G+C mayor a 60% en sus genomas (Barka *et al.*, 2016). Este filo en su mayoría son bacterias de vida libre que se encuentran en ambientes terrestres, marinos y de agua dulce (Barka *et al.*, 2016). Son particularmente abundantes en el suelo, en el que participan en procesos de descomposición de materia orgánica, ayudan al crecimiento de las plantas y enriquecen los ecosistemas del suelo (Barka *et al.*, 2016). Dentro de este filo se incluyen varios géneros que habitan suelos de distintos ambientes como *Streptomyces*, *Amycolatopsis*, *Arthrobacter*, *Subtercola* y *Rhodococcus* (Nouioui *et al.*, 2007). En este trabajo se aislaron y seleccionaron actinobacterias pertenecientes a los cinco géneros nombrados anteriormente.

El género *Amycolatopsis* comprende bacterias esporulantes que contienen genomas grandes de 5 Mb a 10,86 Mb en *Amycolatopsis halophila* y *A. balhimycina*, respectivamente. Además, contiene cromosomas circulares y no es común la presencia de plásmidos dentro del género (Kumari *et al.*, 2016). El género *Rhodococcus* son bacterias aerobias Gram-positivas, inmóviles y no esporulantes (Orro *et al.*, 2015). Su cromosoma es circular o lineal de aproximadamente 6 Mb y presentan plásmidos lineales grandes y circulares más pequeños (Orro *et al.*, 2015). Por su parte el género *Arthrobacter*, reconocido por promover el crecimiento en plantas e implicado en procesos de biorremediación posee un genoma con tamaños entre 3.24 – 5.28 Mb y algunos pueden poseer más de 3 plásmidos (Nakatsu *et al.*, 2013). El género *Subtercola* presenta un tamaño aproximado del genoma de 4 Mb, son bacterias extremófilas y su representante *Subtercola vilae* tiene un genoma con contenido GC mayor a 60% y no contiene plásmidos (Villalobos *et al.*, 2019). Por último, el género *Streptomyces* importante por la producción de una extensa variedad de metabolitos secundarios, contiene cromosomas lineales o circulares con un tamaño aproximado del genoma de 8 – 10 Mb y pueden poseer plásmidos, como es el caso de *S. rochei* en el que se reportó recientemente la presencia de más de 2 plásmidos lineales (Nindita *et al.*, 2019).

Varios miembros del filo Actinobacteria son reconocidos por la producción de una amplia gama de metabolitos secundarios, de los cuales alrededor de 14.000 son antibióticos de las clases betalactámicos (penicilinas), tetraciclinas (Clorotetraciclina), macrólidos (Eritromicina), aminoglucósidos, rifamicinas y glicopéptidos (Dávila *et al.*, 2020). El género *Streptomyces* produce cerca de 12.400 compuestos bioactivos, de los cuales 11.000 son antibióticos, siendo el género con mayor potencial biotecnológico y medicinal (Hamedi *et al.*, 2017). Además, las Actinobacterias tienen alto potencial en biorremediación y promoción del crecimiento de las plantas (Dávila *et al.*, 2020).

La actividad antimicrobiana de las Actinobacterias se atribuye a los metabolitos secundarios que producen. Un gran número de estos metabolitos, se producen a partir de genes cromosomales que se encuentran agrupados en clústers de genes biosintéticos (BGCs, por sus siglas en inglés) (Chen *et al.*, 2019). Estos BGCs están compuestos por un gran número de genes como aquellos para las rutas de síntesis de péptidos no ribosomales (NRPS) y policétidos sintasa (PKS). Estos clusters codifican una amplia gama de productos con aplicaciones en medicina e investigación, como los antibióticos, y son las más estudiadas en la actualidad (Chen *et al.*, 2019). Además de estos BGCs, los microorganismos también contienen genes para la síntesis de otras moléculas bioactivas, entre ellos los RiPPs que son péptidos ribosómicamente sintetizados y modificados post-traduccionalmente (Poorinmohammad *et al.*, 2019).

Los BGCs pueden estar silenciados, por lo cual se consideran también como crípticos en condiciones de laboratorio. Por tal razón ha sido un tema importante desarrollar nuevas aproximaciones para su activación. Entre las metodologías más conocidas se encuentran los sistemas de expresión heteróloga, la inserción de promotores constitutivamente activos mediante CRISPR-Cas9 y la selección de elicitores de alto rendimiento para la identificación de inductores de moléculas pequeñas (Mao *et al.*, 2018; Renn *et al.*, 2015). La mayoría de compuestos bioactivos provenientes de los comprende cuatro grupos químicos predominantes: terpenoides, alcaloides, policétidos y péptidos no ribosómicos (Poorinmohammad *et al.*, 2019).

Los BGC y los genes de resistencia a antibióticos (ARGs, por sus siglas en inglés) están presentes en varios ambientes naturales, incluyendo aquellos ambientes sin intervención humana considerados prístinos (Woo *et al.*, 2005). De hecho, la búsqueda de nuevos compuestos antimicrobianos mediante el minaje del genoma con herramientas como ARTS (Buscador de objetivos resistentes a los antibióticos ,por sus siglas en inglés), se realiza identificando genes de resistencia próximos a BGCs, teniendo en cuenta la premisa de que un productor de antibióticos debe ser a su vez resistente al mismo, por lo que los genes de resistencia pueden estar próximos a los BGCs en sus genomas (Mungan *et al.*, 2020). Por ejemplo, en el genoma de *Streptomyces* se han identificado ARGs que funcionan como mecanismo de protección frente a los metabolitos secundarios que ellos mismos generan (Jiang *et al.*, 2017). Algunos estudios han demostrado que los genes de resistencia en Actinobacterias son más ancestrales que los que se encuentran en bacterias patógenas (Woo *et al.*, 2005). También se ha descrito la posible transferencia horizontal de ARGs provenientes de Actinomicetos hacia Proteobacterias (Jiang *et al.*, 2017).

El presente estudio se realizó con el objetivo de conocer el potencial funcional de Actinobacterias provenientes de páramos de Colombia. Estos ecosistemas prístinos pueden albergar Actinobacterias, como las del género *Arthrobacter* o *Subtercola*, que son difíciles de cultivar y que han sido poco exploradas en su genoma pero que pueden ser novedosas en la producción de compuestos bioactivos desconocidos. En total se analizaron siete taxones de Actinobacterias funcionalmente y por secuenciación genómica. Se encontraron actividades antimicrobianas en ensayos *in vitro* y diversas rutas biosintéticas, como NRPS, PKS, RiPP-like, terpenos y sideróforos, posiblemente involucradas en síntesis de moléculas bioactivas. Se encontró que estos clústers se reportan 36 posibles compuestos bioactivos asociados a funciones antibióticas, antioxidantes, antitumorales, quelantes y de osmorregulación. También, se encontraron 2557 genes de resistencia asociados a expulsión del antibiótico, alteración del diana del antibiótico, reemplazo del diana del antibiótico, inactivación del antibiótico y reducción de la permeabilidad del antibiótico.

MATERIALES Y MÉTODOS

1. Activación y caracterización de cepas

Se realizó reactivación de 16 aislamientos almacenados a -80°C pertenecientes a la colección de CorpoGen. 10 aislamientos provenientes del PNN Chingaza (B1.2;C1.1;C3.20;C3.21;C3.25;G2.2;H2.1;L3.6; L3.8;N3.10) y seis aislamientos del PNN de los Nevados (PNA 1.2; PNC 2.0;PNC2.3; PNC 3.1;PNE 1.1;PNF 4.7). Las muestras fueron aisladas de ambientes ubicados aproximadamente a 4.000 m.s.n.m (Sierra *et al*, 2020).

La reactivación se hizo en medio sólido complejo (ISP2 “International Streptomyces Project-2 Medium (<http://actinobase.org/index.php/ISP2>)” + AIAG “Actinomycete Isolation Agar” (Difco) durante 7-15 días en temperaturas promedio de 20-28°C. Se hicieron descripciones macroscópicas de las colonias: color, superficie (lisa o rugosa), forma (circular, filamentosa, irregular o puntiforme)

Pruebas de actividad antimicrobiana.

Los 16 aislamientos se activaron en medio líquido complejo (ISP2 + AIAG) durante siete días en agitación constante a 120 rpm a 28°C. Después de los siete días se detuvo el crecimiento y se separó el sobrenadante del pellet mediante centrifugación. Se descartó el pellet, y el sobrenadante se guardó a -20°C hasta el montaje del ensayo. El mismo procedimiento se repitió tras 20 días de crecimiento.

En placas de 96 pozos se agregó el sobrenadante de cada cultivo, con sus respectivas réplicas. Se incluyeron controles, un control de crecimiento inoculado con la bacteria patógena y el control negativo (medio sin inocular). Se agregaron 20 µL de *E. coli* ATCC25022, crecido en caldo nutritivo y llevado al patrón de turbidez McFarland 1 y se hizo medición de absorbancia (595nm) de la placa en TECAN (Magellan). Después, la placa se tapó y se llevó a incubación en cámara húmeda (un recipiente con servilletas humedecidas con agua destilada), durante cuatro horas a 37°C y agitación a 120rpm. A las 4 horas se realizaron mediciones de absorbancia (595nm) en TECAN (Magellan). El mismo experimento fue realizado utilizando cepas patógenas de importancia médica: *Staphylococcus aureus* ATCC25923, *Klebsiella pneumoniae* ATCC1705 y *Pseudomonas aeruginosa* ATCC27853. Los ensayos de actividad antimicrobiana se hicieron según la metodología de Gozari *et al.* (2019) y Yong-pil *et al.* (2003), con modificaciones.

Los datos de absorbancia se transformaron a transmitancia mediante la fórmula $\%T = \text{antilog}(2 - \text{absorbancia})$ y se realizaron tablas de porcentaje de inhibición del crecimiento de las bacterias patógenas con cada uno de los sobrenadantes.

2. Identificación de cepas

Se realizó extracción de ADN de los 16 aislamientos utilizando el método de Ha *et al.* 2015 modificado. Se hizo lisis con 1 µL de buffer CTAB ((20 g/L) NaCl (1.4 M), Tris (0.1M), Na2EDTA (0.02 M)) por media hora, seguido de tratamiento con 10 uL de Proteinasa K (10mg/mL) durante media hora. Después, se hizo extracción con 1.5 ml de cloroformo-alcohol isoamílico (24: 1), se tomó la fase acuosa y se trató con 2 µL de RNasa (4 mg /ml) a 37°C por una hora. Luego de una segunda extracción con 1 mL de cloroformo-alcohol isoamílico (24: 1) y centrifugación por 10 minutos a 10.000 rpm, se obtuvo la fase acuosa y se precipitó el ADN con etanol al 100% durante una noche a -20°C. Después de centrifugar 15 min a 10.000 rpm, se hicieron tres lavados con etanol al 70%. El ADN se dejó secar por 20 minutos y se resuspendió en 50 µL de agua tipo I estéril.

Se verificó calidad e integridad del ADN mediante espectrofotometría con NanoDrop 2000 (ThermoFisher) y electroforesis en gel de agarosa al 1% en TBE 0.5X. El ADN se utilizó para clasificación taxonómica mediante amplificación del gen 16S rRNA con los iniciadores 27F-5'1AGAGTTTGGATCMTGGCTCAG'3 y 1492R-5'GGTTACCTTGTTACGACTT'3. Los fragmentos obtenidos de 1500 pb se visualizaron en gel al 1% de agarosa en TBE 0.5X.

Los fragmentos obtenidos del gen 16S rRNA se enviaron a secuenciar por Sanger a Macrogen, Corea. La secuenciación se realizó de forma bidireccional usando los iniciadores utilizados en la PCR. Las secuencias se editaron manualmente con PEARL (<https://www.gear-genomics.com/pearl/>) y se compararon en GenBank con la herramienta BLASTn (Altschul *et al.*, 1990). Se registraron las 2 secuencias con mayor similitud. Las secuencias editadas obtenidas en este trabajo y las secuencias reportadas en GenBank se usaron para realizar un cladograma a partir de matriz de distancia utilizando el método de Máxima Verosimilitud en MEGA (Kumar *et al.*, 2018). Se escogió el modelo de sustitución con valor BIC menor = K2+G+I, distribución gamma para uniformidad de las tasas evolutivas y modelo de sustitución Kimura 2 con Bootstrap (500 repeticiones) para estimar el soporte de los clados inferidos (Felsenstein, 1985; Kimura, 1980).

3. Extracción de ADN de alto peso molecular

La extracción de ADN de alto peso molecular se realizó con los métodos Somerville *et al.*, (1997) y Ha *et al.*, (2015), con las siguientes modificaciones. Los siete aislamientos se activaron en medio sólido complejo (ISP2 + AIA) y se dejaron crecer a 28°C, entre 5-7 días. Las cajas completas y crecidas con solo colonias axénicas se resuspendieron en 0.9 mL de buffer TE (10 mM Tris-HCl, 1 mM EDTA, pH 8) y se pasaron a tubos de 1.5mL. Se realizó lisis enzimática con 50 µL de lisozima (10 mg/ml) por 2 h a 37°C. Seguidamente, se añadió 2.5 mL de RNAsa A (4 mg/ml) 1 h 37°C para remover el RNA y tratamiento con 200 µL de NaCl al 5 M y 145 µL de una solución 2:1 (10% de SDS y Proteinasa K (10 mg/ml)) por 30 minutos a 55°C. Después, se añadió 75µL de una solución CTAB (20% CTAB, 1.4M NaCl) y se incubó 10 minutos a 65°C.

El lisado fue extraído con cloroformo: alcohol isoamílico (24: 1) y tras centrifugación a 13.000 rpm por 15 minutos se transfirió la fase acuosa cuidadosamente a otro tubo de 1.5mL. El ADN genómico fue incubado con 0.7 volúmenes de isopropanol durante dos días a -20°C y precipitado por centrifugación a 13.000 rpm durante 15 minutos. Después, el pellet fue lavado tres veces con 1 mL de etanol al 70%. Finalmente, el ADN se resuspendió en 50µL de agua tipo I, libre de nucleasas, estéril y se almacenó a -20°C.

La calidad e integridad del ADN fue verificada por espectrofotometría con NanoDrop 2000 y electroforesis en gel al 1% de agarosa con TBE 0.5X. Se utilizaron extracciones de ADN que cumplieran con una relación 260/280 entre 1,8-2,0 y peso molecular mayor a 10.000 pb, sin o con poco barrido visualizado en transiluminador UV.

4. Secuenciación de fragmentos largos con Nanopore (MinIon)

El ADN se cuantificó con Qubit (Invitrogen) con el kit de reactivos dsDNA High Sensitivity (Invitrogen). Debido a diferencias en concentración de cada ADN, todas las muestras se normalizaron a una concentración final de 1ug/50µL.

Se prepararon librerías con las muestras normalizadas utilizando el kit de ligación SQK-LSK109 de Oxford Nanopore Technologies (ONT). La reparación y adenilación de extremos se hizo con los reactivos NEBNext ultra II End Repair/dA -Tailing Module (New England Biolabs) sin purificación con perlas antes del paso de ligación. Cada producto de ADN reparado se codificó con el kit NATIVE BARCOING KIT (ONT, NBD-EXP104) con los barcodes consecutivos del 1 al 7. La ligación de estos barcodes se hizo con los reactivos NEBNext ultra II Ligation Module (New England Biolabs). A continuación, las muestras se purificaron con perlas Agencourt AMPure XP (Beckman Coulter, EE. UU.) usando 80 µL de perlas para cada muestra y resuspendiendo en 25 µL de agua tipo I estéril. Las concentraciones resultantes de la purificación se midieron en Qubit para cada muestra con su respectivo barcode. Después se agrupó la misma concentración de ADN por cada barcode en un único pool para obtener una cantidad final de 1ug en 65 µL.

Los adaptadores de Nanopore se ligaron al pool usando los reactivos Quick Ligation Module (New England Biolabs). Después de esta ligación, hicimos la limpieza final con 80 µL de perlas AMPure XP (Beckman Coulter, EE. UU.) y el buffer de limpieza LFB (ONT). La librería final se resuspendió en 15 µL de buffer EB (ONT). La cuantificación final de la librería fue de 336 ng de ADN total en 14 µL. Esta concentración total fue sembrada en una celda R9.4.1 FLO-MIN106D nueva y se secuenció en un dispositivo MinION Mk1B (ONT) con el protocolo estándar del software MinKNOW (ONT, Reino Unido).

El llamado de bases se realizó con Guppy (<https://community.nanoporetech.com>)(Wick *et al.*, 2019). La calidad de las secuencias se verificó con pycoQC (Leger y Leonardi, 2019) (<https://github.com/aslide/pycoQC>). Finalmente, se aplicaron filtros de calidad mediante NanoFilt (De Coster

et al., 2018) (<https://github.com/wdecoester/nanofilt>). Los filtros se aplicaron con un valor umbral de 8. Es decir, secuencias con calidad por debajo de 8 se retiraron. También se quitaron 50 nt de cada uno de los extremos y las secuencias con longitudes inferiores a 500nt.

5. Secuenciación de fragmentos cortos por Illumina (NovaSeq)

Los ADNs se enviaron a secuenciar a NOVOGENE, Estados Unidos en la plataforma de secuenciación NovaSeq 6000. La secuenciación se hizo utilizando librerías paired-end. La calidad de las lecturas se revisó con fastQC (<https://qubeshub.org/resources/fastqc>) (Andrews, 2010) y MultiQC (Ewels *et al.*, 2016) (<https://github.com/ewels/MultiQC>). Las lecturas se limpiaron con trimmomatic (Bolger *et al.*, 2014) quitando los primeros 8 nucleótidos al extremo 3' y se removieron los adaptadores.

6. Ensamblajes

6.1 Ensamblaje *de novo* para secuencias largas (ONT).

Las lecturas de nanopore filtradas según calidad se ensamblaron *de novo* con Flye (Kolmogorov *et al.*, 2019) (<https://github.com/fenderglass/Flye>). Se revisó la calidad del ensamblaje con Quast (Gurevich *et al.*, 2013) (<https://github.com/ablab/quast>). Como se obtuvo más de un contig por cepa, se seleccionó el contig de mayor tamaño para los pasos subsecuentes.

Con el contig ensamblado de mayor tamaño, se realizaron cuatro rondas de pulido. Cada ronda de pulido se realizó con los programas Minimap2 (Heng *et al.*, 2018) (<https://github.com/lh3/minimap2>) para alinear el query contra las propias lecturas, seguido de Racon (Vaser *et al.*, 2017) (<https://github.com/isovic/racon>) para corregir los errores de ensamblaje. Por último, se hizo un pulido y consenso final con Medaka (Medaka, 2022) (<https://github.com/nanoporetech/medaka>). Se evaluó el porcentaje de contaminación de los genomas ensamblados y el porcentaje de genoma completado respecto a los genomas más similares de la base de datos con el programa CheckM (Parks *et al.*, 2014) (https://kbase.us/applist/apps/kb_Msuite/run_checkM_lineage_wf/release?gclid=EA1aIQobChMI0vrN4I62-QIVAgalCR0-twD_EAAYASAAEgLVffD_BwE).

6.2 Ensamblajes Híbridos

Los ensamblajes híbridos se realizaron con Unicycler (Wick *et al.*, 2017) (<https://github.com/rrwick/Unicycler>), que permite ensamblar lecturas cortas provenientes de Illumina y lecturas largas de NanoPore en una única secuencia consenso. Unicycler realiza ensamblajes internos con programas como Spades (Bankevich *et al.*, 2012) (<https://github.com/ablab/spades/releases>) utilizado en ensamblaje de secuencias cortas de Illumina. Además, corrige errores asociados a inserciones y deleciones con Racon (Vaser *et al.*, 2017) (<https://github.com/isovic/racon>) y Pilon (Walker *et al.*, 2014) (<https://github.com/broadinstitute/pilon>).

7. Asignación taxonómica de los genomas

La asignación taxonómica de los ensamblajes de Nanopore y los híbridos se realizó en primer lugar con el programa JSpecies WS (Richter *et al.*, 2016) (<http://jspecies.ribohost.com/jspeciesws/>) que permite calcular el grado de similitud entre el genoma de interés y una base de datos actualizada constantemente llamada GenomesDB (<http://jspecies.ribohost.com/jspeciesws/>) de genomas en borrador y completos. Primero se realizaron comparaciones mediante TCS (Tetra correlation search) que se basa en frecuencias de tetranucleótidos y coeficientes de correlación. Después se compararon los 10 genomas más cercanos obtenidos por TCS con el genoma query y se calculó el valor ANI (average nucleotide identity). Finalmente, se seleccionó el genoma más cercano según el valor ANI directamente en Jspecies.

Se utilizó una segunda base de datos para asignación taxonómica que corresponde a MiGA (Rodriguez *et al.*, 2018) (<http://microbial-genomes.org/>). MiGA contiene sólo genomas procariotas completos en la base de datos NCBI prok y además realiza comparaciones mediante el cálculo del ANI entre el query y la base de datos.

8. Anotación de genomas

La anotación de los genomas se realizó con Prokka (Seemann, 2014), con el que se determinó el número de regiones codificantes (CDS) para cada genoma, RNAs ribosómicos rRNA, RNAs de transferencia (tRNA) y RNAs mensajeros de transferencia (tmRNA). Después se realizó la anotación funcional de los genomas con Kegg (Minoru y Susumu, 2000) con el cual se determinaron genes implicados en rutas metabólicas de carbohidratos, proteínas, aminoácidos, nucleótidos y lípidos, entre otros.

Se realizó búsqueda de genes de resistencia en los genomas con la base de datos de CARD (Alcock *et al.*, 2020) <https://card.mcmaster.ca>. En la búsqueda de CARD se escogió el identificador de genes de resistencia (RGI) para predecir resistomas a partir de datos de nucleótidos mediante la búsqueda en una base de datos curada para la vigilancia de resistencias a antibióticos. En CARD se seleccionaron parámetros de búsqueda en modo perfecto, estricto y relajado. El algoritmo de búsqueda Perfecto detecta coincidencias perfectas con las secuencias de referencia seleccionadas y mutaciones en CARD. El algoritmo Estricto detecta variantes previamente desconocidas de genes AMR conocidos y el algoritmo Relajado funciona fuera de los puntos de corte del modelo de detección para proporcionar detección de amenazas nuevas y emergentes y homólogos más distantes de genes AMR (Alcock *et al.* 2020).

Finalmente, se realizó predicción de posibles moléculas que están siendo sintetizadas por los genes NRPS y PKS, entre otros, mediante antiSMASH 6.0 (Blin *et al.*, 2021) y se seleccionaron parámetros de búsqueda en antiSMASH en modo relajado.

RESULTADOS

1. Activación y caracterización de cepas

Inicialmente se realizó reactivación en medio sólido complejo (ISP2+AIAG) de 16 aislamientos del cepario de Corpogen provenientes de muestras de líquenes y algas de páramos (Sierra *et al.*, 2020). 10 aislamientos provenían de PNN Chingaza (B1.2, C1.1, C3.20, C3.21, C3.25, G2.2, H2.1, L3.6, L3.8, N3.10) y seis aislamientos del PNN de los Nevados (PNA 1.2, PNC 2.0, PNC2.3, PNC3.1, PNE 1.1, PNF 4.7). Estos ensayos mostraron diferencias en la forma de crecimiento, color, superficie, borde, temperatura óptima de crecimiento y tiempos de crecimiento que se enumeran en la **Tabla 1**.

Tabla 1. Características de crecimiento de aislamientos reactivados en medio (ISP2+AIAG)

Muestra	Código	Morfología macroscópica	Tiempo de crecimiento	Temperatura óptima
1	B 1.2	Color blanco opaco, textura cremosas, crecimiento puntiforme.	3 días	28 °C
2	C 1.1	Color blanco, textura seca, crecimiento irregular, olor a tierra húmeda.	3-5 días	28°C
3	C 3.20	Colonia blanca, textura seca, forma acojinada y ondulada, crecimiento irregular.	6-7 días	25-28°C
4	C 3.21	Colonia naranja, textura seca, forma acojinada, crecimiento irregular.	7 días	28 °C
5	C 3.25	Color curuba o rosado claro, textura muy cremosa, convexo y crece irregular.	2 días	20-28°C
6	G 2.2	Color curuba o rosado claro, textura muy cremosa, convexo y crece irregular.	2 días	20-28°C
7	H 2.1	Color fucsia, textura cremosa, crecimiento puntiforme.	6-10 días	28°C
8	L 3.6	Color amarillo pálido, textura cremosa, crecimiento puntiforme	7 días	28°C
9	L 3.8	Color blanco opaco, textura cremosa, crecimiento puntiforme	3 días	28°C
10	N 3.10	rosa a amarillo con la luz	2 días	20-28°C
11	PNA 1.2	Color amarillo claro, textura cremosa, borde plano, forma de crecimiento circular.	4 días	28°C
12	PNC 2.0	Color blanco, textura seca, borde convexo, forma irregular de crecimiento.	2 días	28°C
13	PNC 2.3	Color blanco, textura cremosa, crecimiento puntiforme.	2 días	28°C
14	PNC 3.1	Color blanco, textura seca, borde convexo, forma irregular de crecimiento	2 días	28°C
15	PNE 1.1	Color Amarillo claro, textura cremosas, crecimiento irregular.	4 días	28°C
16	PNF 4.7	Color Amarillo, textura cremosas, crecimiento irregular.	6 - 7 días	28°C

2. Pruebas de actividad antimicrobiana.

Para identificar si algunos de estos microorganismos tenían actividad antibacteriana, se hicieron pruebas en dos tiempos de crecimiento. La primera prueba se realizó con sobrenadantes de los 16 aislamientos a siete días de crecimiento contra *E. coli*. Debido a que ocho de los 16 sobrenadantes mostraron porcentajes de inhibición menores al 33% en este ensayo, se decidió hacer ensayos también a 20 días de crecimiento, con el propósito de evaluar si a mayor tiempo de crecimiento había inhibición del crecimiento en bacterias blanco. A los 20 días se hicieron experimentos sólo con ocho sobrenadantes que habían mostrado actividades menores a 33% de inhibición contra *E. coli*, *K. pneumoniae*, *S. aureus* y *P. aeruginosa*.

Los resultados a siete días de crecimiento fueron los siguientes: B1.2, PNC 2.3 y N 3.10 presentaron porcentajes de inhibición mayores a 80% contra *E. coli*. El resto de los sobrenadantes tuvieron porcentajes de inhibición menores al 33% contra *E. coli*. Contra las bacterias patógenas *K. pneumoniae* y *S. aureus* sólo B1.2, PNC 2.3, N 3.10, PNC 2.0 y PNE 1.1 mostraron porcentaje de inhibición del crecimiento mayor al 20%, el resto presentó porcentajes de inhibición menores al 20% (**Tabla 2**).

Los resultados a 20 días de crecimiento fueron los siguientes: solo los sobrenadantes C 1.1 y L 3.8 presentaron porcentajes de inhibición del crecimiento de *E. coli* de 47% y

31%, respectivamente. El resto de los sobrenadantes presentaron porcentajes de inhibición menores a 20% contra *E. coli*. Los sobrenadantes de PNF 4.7 y PNA 1.2 inhibieron en un 28% el crecimiento de *P. aeruginosa* y 14% a *K. pneumoniae*. Con el resto de los sobrenadantes no hubo inhibición del crecimiento.

En general, de los 16 aislamientos (x 7 días de crecimiento) sólo ocho aislamientos presentaron un porcentaje de inhibición mayor al 20% en alguna de las 4 bacterias patógenas. Los ocho aislamientos restantes presentaron porcentajes de inhibición menores al 20%, aun cuando se ensayaron a 20 días de crecimiento (**Tabla 2**).

Tabla 2 Inhibición del crecimiento de bacterias expuestas a sobrenadantes de aislamientos de Chingaza y Nevados.

No.	código	% inhibición <i>E. coli</i>				% inhibición de <i>K. pneumoniae</i>				% inhibición de <i>S. aureus</i>				% inhibición <i>P. aeruginosa</i>	
		7 días	DE	20 días	DE	7 días	DE	20 días	DE	7 días	DE	20 días	DE	20 días	DE
1	B 1.2	82	2	-	-	100	0.5	-	-	80	3.4	-	-	-	-
2	C 1.1	0	0.4	47	1.5	-	-	0	5	-	-	0	0.9	0	0.1
3	C 3.20	8.8	1.9	0	0.9	-	-	0	1.4	-	-	0	3.8	0	1.6
4	C 3.21	1.8	2.1	0	7.5	-	-	2	6.8	-	-	15	4.2	0	7.5
5	C 3.25	0	0.2	0	5	2.6	1.8	0	1.4	0	7.5	0	3.1	0	4
6	G 2.2	3.9	1.4	-	-	-	-	-	-	-	-	-	-	-	-
7	H 2.1	0	0.4	-	-	0	3.7	-	-	0	3.5	-	-	-	-
8	L 3.6	7	0.2	-	-	3	2.1	-	-	0	7.5	-	-	-	-
9	L 3.8	33	1.7	31	1.21	0	1.5	0	5	0	2.8	0	1.9	0	3.5
10	N 3.10	96	0.1	-	-	57	4	-	-	26	14	-	-	-	-
11	PNA 1.2	0	0.1	10.8	2.3	-	-	14	1.6	-	-	0	3.5	0	3
12	PNC 2.0	26	3.9	-	-	20	0.7	-	-	74	2.3	-	-	-	-
13	PNC 2.3	96.7	1	-	-	100	1.2	-	-	91	0.7	-	-	-	-
14	PNC 3.1	16	4.5	-	-	-	-	-	-	-	-	-	-	-	-
15	PNE 1.1	20	0.7	14.4	0.7	68	3.4	5	3.5	0	0.4	0	2.8	0	4.8
16	PNF 4.7	0	1.1	0	-	-	-	8	7	-	-	0	4.6	28	0.9
17	CONTROL-	0	1.9	0	1.3	0	1.4	0	2	0	1.7	0	2.3	0	3.1
Tot al	Total, muestras	16		8		9		8		9		8		8	

Se ensayaron sobrenadantes después de crecimiento por 7 y 20 días y se indica el % de inhibición de crecimiento. En negrilla se resaltan los aislamientos con inhibición mayor al 20% contra alguna de las cepas patógenas ensayadas. (Control -) control sin inocular para cada ensayo; (-) Experimentos no realizados; DE, Desviación estándar. Dato de 7 días para PNA1.2 contra *E. coli* se tomaron de experimentos hechos por Nelly Díaz y Valeria Gómez, 2021.

3. Identificación y selección de cepas.

Se hizo la identificación taxonómica de las cepas mediante análisis del gen 16S rRNA. Se observó que 9 de 16 cepas pertenecían al filo Actinobacteria, las cuales estaban distribuidas en los siguientes géneros: *Amycolatopsis* (C 3.20), *Streptomyces* (C 3.21 y C 1.1), *Rhodococcus* (C 3.25, G2.2, N 3.10), *Arthrobacter* (PNA 1.2 y PNE 1.1) y *Subtercola* (PNF 4.7). De estas 9 Actinobacterias identificadas se seleccionaron siete para la secuenciación genómica y para representar cada género. Estas cepas están resaltadas en rojo en la **Figura 1**, que representa los 5 clados identificados para las siete Actinobacterias aisladas de muestras de los PNNs Nevados y Chingaza. Las morfologías macroscópicas de los 7 aislamientos seleccionados se presentan en la **Figura 2**, donde se ven las morfologías diferentes de cada uno de los géneros.

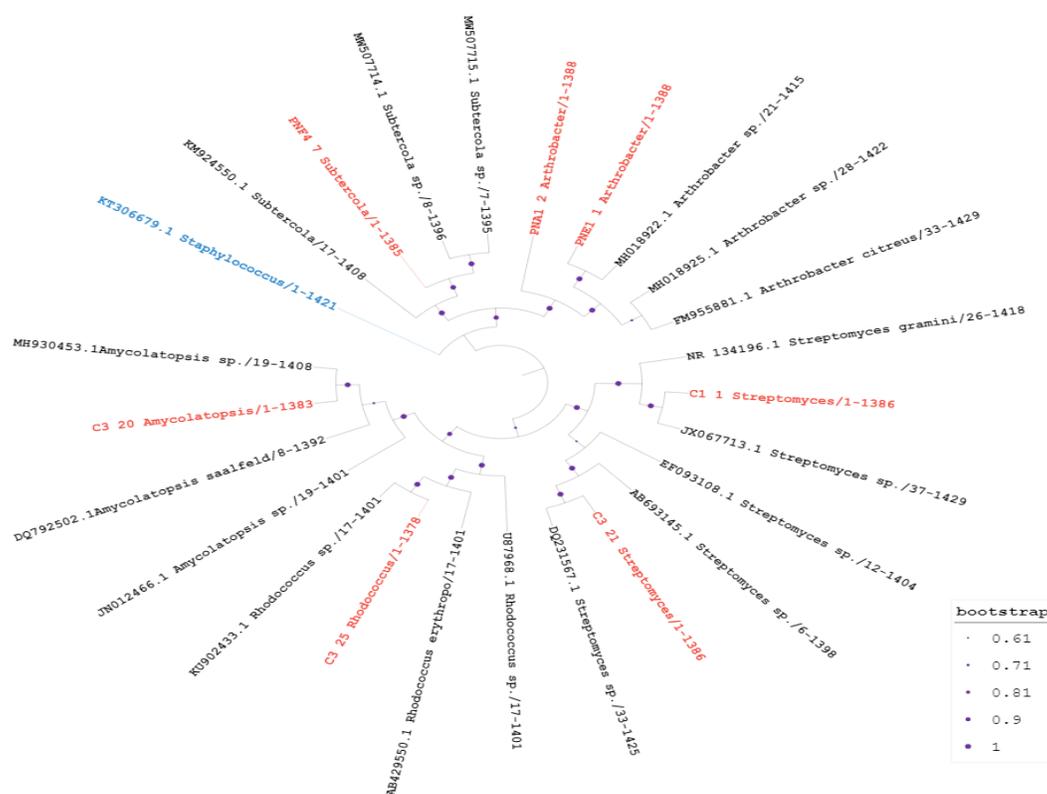


Figura 1. Cladograma de Actinobacterias a partir de secuencias de gen 16S rRNA. Máxima verosimilitud con el modelo BIC= K2+G+I y Bootstrap de 500 réplicas. Bootstrap mayor a 60% señalado en círculos morados. En rojo se señalan Actinobacterias para secuenciación genómica: PNF 4.7, PNA 1.2, PNE 1.1, C 1.1, C 3.21, C 3.25, C 3.20. En negro se señala secuencias más cercanas. En azul se señala el outgroup.

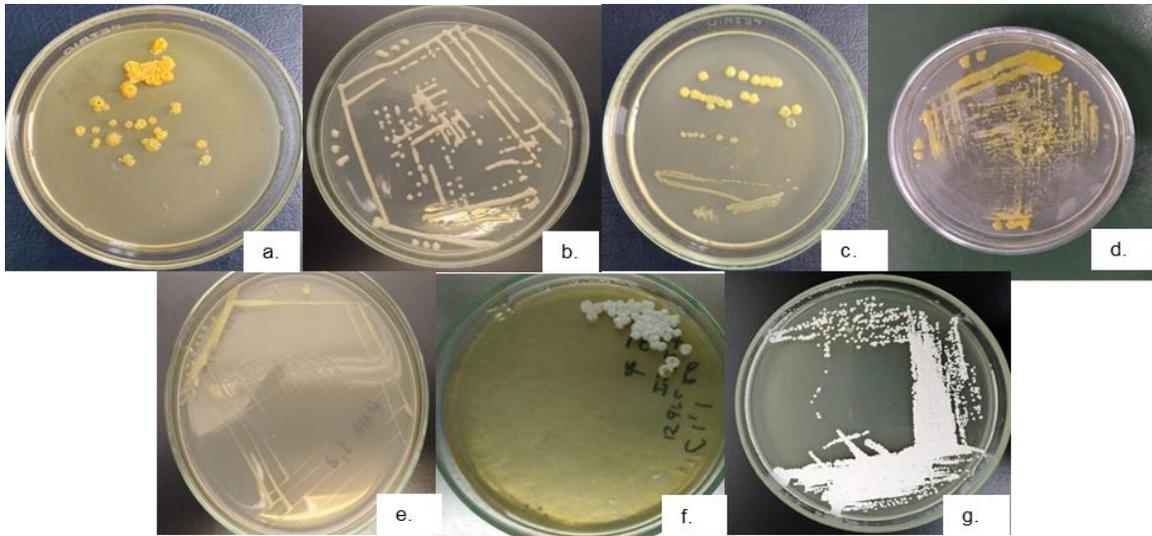


Figura 2 Morfología de Actinobacterias seleccionadas para secuenciación genómica. **a.** (C 3.21) *Streptomyces*. **b.** (C3.25) *Rhodococcus*. **c.** (PNE 1.1) *Arthrobacter*, **d.** (PNF 4.7) *Subtercola*, **e.** (PNA 1.2) *Arthrobacter*, **f.** (C 1.1) *Streptomyces*. **g.** (C 3.20) *Amycolatopsis*.

4. Obtención de ADN de alto peso molecular

A los 7 microorganismos clasificados en cada género por identificación del gen 16S rRNA se les realizó extracción de ADN de alto peso molecular. Las muestras de ADN se verificaron en su calidad con mediciones de absorbancia. Las calidades de absorbancia A260/280 promedio fueron de 1.88 que indica pureza óptima para secuenciación genómica en las siete Actinobacterias. Los valores de absorbancia A260/230 relacionado a la presencia de contaminantes, mostraron en promedio un valor 1.76 para las 7 muestras (**Tabla 3**). Este valor es óptimo para secuenciación genómica, solo disminuye la calidad A260/230 en las muestras C 3.25, con un valor de 1.03, y PNF 4.7 con un valor de 1.61 (**Tabla 3**). Estos valores pueden indicar presencia de contaminantes asociados a compuestos fenólicos o proteínas.

La concentración de ADN medida en Qbit fue mayor a 20,0 ng/uL para todas las muestras, lo cual es suficiente para normalizar a 1ug/50uL en el montaje de librerías de NanoPore (**Tabla 3**). Por último, la visualización del ADN mediante electroforesis muestra un peso molecular aproximado de 10.000pb (**Figura 3**). Estos resultados indican que el ADN de las siete Actinobacterias estaba con buena calidad y concentración para secuenciación.

Siete de las muestras (C 1.1, C3.20, C3.21, C3.25, PNA1.2, PNE 1.1 y PNF4.7) se secuenciaron por NanoPore y cinco de los mismos ADN (C3.20; C3.21; C3.25; PNE 1.1 y PNF4.7) se secuenciaron paralelamente por Illumina. Se hizo esta selección para

representar los cinco géneros *Amycolatopsis* (C 3.20), *Streptomyces* (C 3.21), *Rhodococcus* (C 3.25), *Arthrobacter* (PNE 1.1) y *Subtercola* (PNF 4.7) con ambas plataformas de secuenciación.

Tabla 3. Mediciones de calidad en NanoDrop y concentración de ADN genómico en Qbit

Muestra	código	16 S Tax	260/280	260/230	Método de Extracción	Concentración en Qbit (ng/uL)
1	C 1.1	<i>Streptomyces</i>	1,87	1,83	Lisozima	40
2	C 3.20	<i>Amycolatopsis</i>	1,88	1,99	Lisozima	114
3	C 3.21	<i>Streptomyces</i>	2	1,97	Lisozima	70
4	C 3.25	<i>Rhodococcus</i>	1,81	1,03	Lisozima	106
5	PNA 1.2	<i>Arthrobacter</i>	1,93	2,01	Lisozima	118
6	PNE 1.1	<i>Arthrobacter</i>	1,97	1,86	Lisozima	104
7	PNF 4.7	<i>Subtercola</i>	1,8	1,61	Lisozima	23,2
Promedio			1,89	1,76		82,17

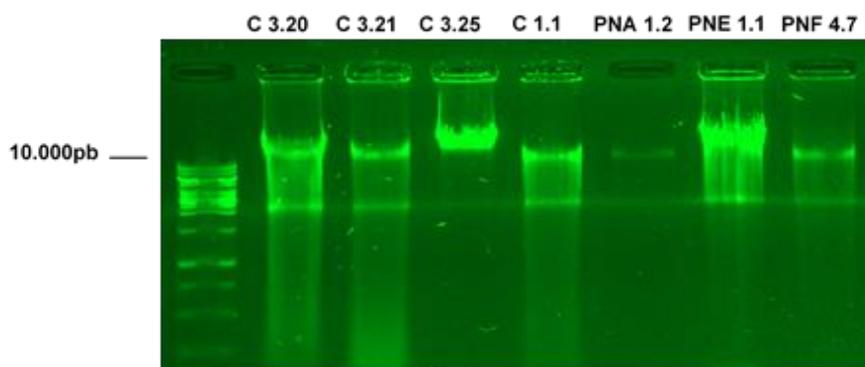


Figura 3. Electroforesis de ADN genómico de Actinobacterias en gel de agarosa 1% TBE 0.5X.

5. Secuenciación NanoPore

Se secuenciaron 7 muestras con la tecnología NanoPore y de estas secuencias se obtuvieron gráficos de calidad con PicoQC (Leger y Leonardi, 2019). El valor de la mediana de calidad PHRED fue de 10.21 (**Anexo 1**). La media de la calidad fue de 10 y la profundidad promedio de secuenciación fue de 26.47% con variaciones en cada una de las muestras (**Tabla 4 y Anexo 2**). Adicionalmente, la longitud promedio de las

lecturas varió de 549 pb en C3.21 a 4000 pb en PNA 1.2. Se esperaba un promedio de longitud de lectura de 1000 pb según lo descrito por Rajwani y colaboradores en 2021 (**Tabla 4**).

Adicional a las gráficas de calidad obtenidas en PicoQC (Leger y Leonardi, 2019), el mismo programa generó la gráfica de lecturas totales en la secuenciación (**Figura 4**). Del total de las lecturas (481.607) generadas en la secuenciación, el 37.3% no se clasificó. Las muestras PNE 1.1, C3.25 y C3.21 presentaron porcentajes de lecturas muy bajos con solamente 2.8%, 3.9% y 4.5% de representación sobre el total de las lecturas obtenidas en la secuenciación (**Figura 4**). En general la secuenciación de Nanopore arrojó resultados que variaron bastante entre muestras, respecto al número de lecturas, longitud media de las lecturas y profundidad. Las gráficas completas de calidad de secuenciación y longitud de lecturas se encuentran en el **Anexo 3**.

Tabla 4 Resumen de calidad y profundidad en la secuenciación NanoPore

No.	Código	16 S Tax	# de lecturas	longitud media pb	QC	Profundidad
1	C 1.1	<i>Streptomyces</i>	126.941	641	9,91	29,67
2	C 3.20	<i>Amycolatopsis</i>	41.035	859	10,1	11,4
3	C 3.21	<i>Streptomyces</i>	21.626	549	9,98	5,9
4	C 3.25	<i>Rhodococcus</i>	18.820	1993	10,34	23,76
5	PNA 1.2	<i>Arthrobacter</i>	39.523	4000	10,47	49,92
6	PNE 1.1	<i>Arthrobacter</i>	13.502	3579	10,44	18,58
7	PNF 4.7	<i>Subtercola</i>	40.698	2309	9,93	46,06
	No clasificadas	No clasificadas	179.462	1490	9,79	-
	Promedio		60.201	1927	10,12	26,47

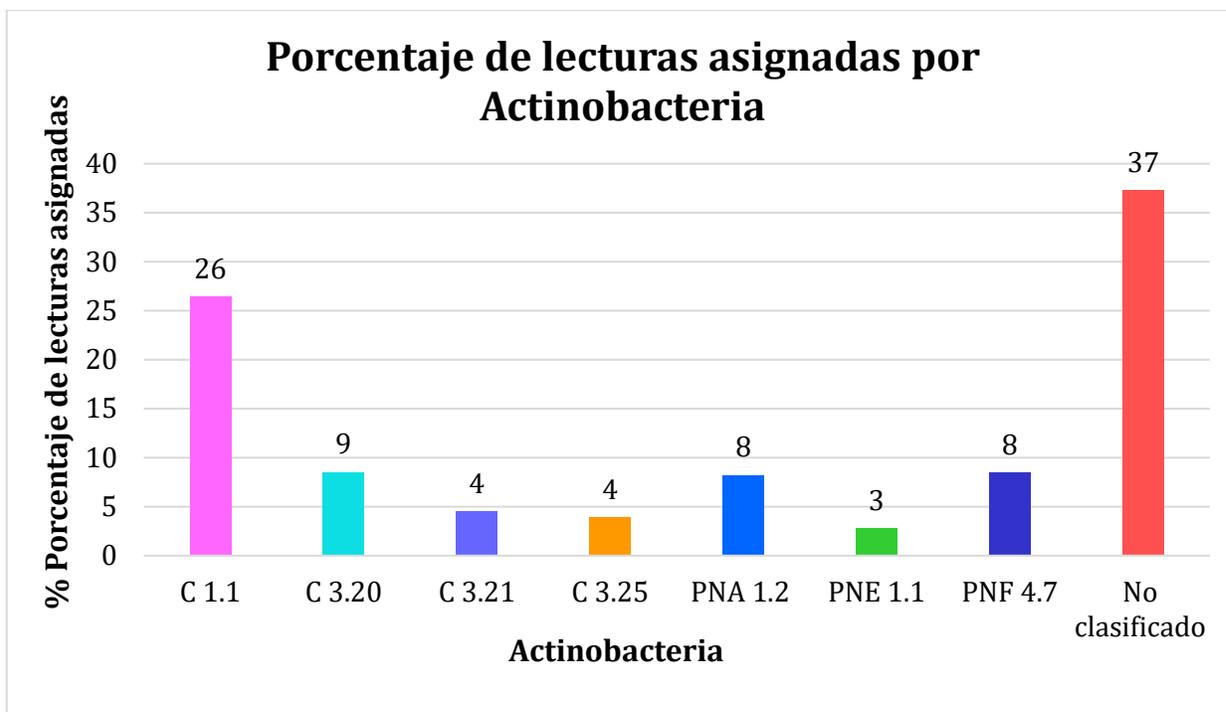


Figura 4. Porcentaje de lecturas asignadas en el set de datos de cada una de las Actinobacterias secuenciadas por Nanopore.

6. Secuenciación Illumina

Se seleccionaron cinco ADNs pertenecientes a los géneros *Amycolatopsis*, *Streptomyces*, *Rhodococcus*, *Arthrobacter* y *Subtercola* los cuales se secuenciaron en la plataforma de Illumina. En esta secuenciación se obtuvo un puntaje de calidad de 36 para todas las muestras (**Anexo 4**). El número de lecturas promedio fue de 4.000.000 (**Tabla 5**) y, a diferencia de la secuenciación por Nanopore, la longitud promedio de las lecturas fue corta, de 150 pb. (**Tabla 6**).

La tasa de error en lecturas Illumina es menor al 1%, en cambio en las lecturas Nanopore tienen una tasa de error de 10.5% (Delahaye y Jacques, 2021). La tasa de error en nanopore pudo interferir en el promedio de calidad de secuenciación.

Tabla 5. Resumen calidad de secuencias de Actinobacterias por Illumina

No.	código	16S Tax	Paired end	# de lecturas	longitud media (pb)	QC	# de secuencias (pb)
2	C 3.20	<i>Amycolatopsis</i>	1	5.505.366	150	36	8.200.000
			2	5.577.108	150	36	8.200.000

3	C 3.21	<i>Streptomyces</i>	1	4.654.615	150	36	8.600.000
			2	4.588.144	150	36	8.600.000
4	C 3.25	<i>Rhodococcus</i>	1	4.011.319	150	36	6.100.000
			2	4.180.667	150	36	6.100.000
6	PNE 1.1	<i>Arthrobacter</i>	1	3.522.868	150	36	5.900.000
			2	3.565.476	150	36	5.900.000
7	PNF 4.7	<i>Subtercola</i>	1	4.222.533	150	36	7.800.000
			2	4.215.316	150	36	7.800.000
Promedio				4.404.341	150	36,0 0	7.320.000

Tabla 6. Comparación secuenciación Nanopore e Illumina

Características	Secuenciación (ONT)	Illumina (NovaSeq)
Longitud de las lecturas	549 pb – 4000 pb	150 pb
Número de lecturas	13.502 – 179.462	4.404.341
Calidad promedio	10.12	36
Profundidad promedio	26.47	50

7. Ensamblaje de lecturas largas (ONT) y ensamblajes híbridos.

A partir de los resultados de secuenciación se realizaron los ensamblajes de lecturas largas con Flye (ONT), los ensamblajes de lecturas cortas con Spades (Illumina) y los ensamblajes híbridos de lecturas largas y cortas con Unicycler (ONT+Illumina). De los siete genomas ensamblados con Flye (ONT) tres generaron un único contig (PNA 1.2; PNE 1.1 y PNF 4.7) y cuatro más de un contig (C 1.1 C 3.20, C 3.21 y C 3.25) (**Tabla 7**). La presencia de varios contigs en estos genomas puede deberse a varios factores como errores de ensamblaje por causa de secuencias repetitivas, las cuales confunden los algoritmos. En géneros como *Streptomyces* la presencia de cromosomas lineales y plásmidos que contienen largas regiones repetitivas pueden alcanzar más de 1 Mb, son difíciles de resolver con las tecnologías actuales de secuenciación (Gómez *et al.*, 2016).

En los ensamblajes híbridos de los *Streptomyces* C3.20 y C3.21 y el *Rhodococcus* C 3.25 hubo una disminución considerable en el número de contigs finales respecto a los ensamblajes Flye (ONT). Específicamente, en C 3.25 hubo 63 contigs con Flye (ONT) y cuatro contigs con Unicycler (ONT + Illumina) y se presentó un aumento del valor N50. El valor N50 se define como la longitud del contig más corto que cubre por lo menos el 50% del genoma (Gurevich *et al.*, 2013) **(Tabla 8)**. Adicionalmente, el tamaño final de los ensamblajes presentó modificaciones. En el *Streptomyces* C 3.21 con Flye (ONT) se obtuvo un tamaño total de 734.381 pb. En cambio, el tamaño total del ensamblaje híbrido fue de 8.429.113 pb, este último tamaño es más cercano al tamaño de genoma reportado para *Streptomyces*, el cual debe ubicarse en el rango de 8Mb-10Mb (Nindita *et al.*, 2019).

Debido a que los valores de calidad de los ensamblajes híbridos presentaron N50 y longitud del genoma ensamblado mayor a comparación de los genomas con Flye, para los *Streptomyces* C3.20 y C3.21, y el *Rhodococcus* C 3.25 se seleccionaron únicamente los ensamblajes híbridos para continuar la anotación **(Tabla 8)**. A diferencia de los anteriores, los ensamblajes de *Arthrobacter* PNE 1.1 y del *Subtercola* PNF 4.7 provenientes de Flye presentaron N50 más altos y con un solo contig ensamblado a comparación de los híbridos, por lo cual se seleccionó el ensamblaje de Flye para continuar la anotación. Finalmente, debido a que el *Streptomyces* C 1.1 y el *Arthrobacter* PNA 1.2 solo se secuenciaron con NanoPore, se continuó la anotación con el ensamblaje de Flye para ambos genomas. No se realizó anotación de los ensamblajes de Spades debido a que en los cinco genomas este ensamblaje presentó un mayor número de contigs y un N50 menor a comparación de los ensamblajes de flye (ONT) y los híbridos (ONT + Illumina) **(Tabla 8)**.

Tabla 7. Ensamblaje *de novo* con secuencias ONT y calidad del ensamblaje en Quast

N o.	código	Especie 16S	Método de ensamblaje	# de contigs después del ensamblaje	longitud del contig más grande	Longitud del Ensamblaje	Calidad del ensamblaje Quast N50 (pb)	% GC
1	C 1.1	<i>Streptomyces</i>	Flye	21	4.552.434pb	10.110.931	1.098.813	70,1
2	C 3.20	<i>Amycolatopsis</i>	Flye	167	249.927 pb	10.707.162	88.504	68,0
3	C 3.21	<i>Streptomyces</i>	Flye	19	130.883 pb	734.381	74.260	69
4	C 3.25	<i>Rhodococcus</i>	Flye	63	6.349.031pb	9.913.200	6.349.031	62,3
5	PNA 1.2	<i>Arthrobacter</i>	Flye	1	3.971.880pb	3.971.880	3.971.880	65,2

6	PNE 1.1	<i>Arthrobacter</i>	Flye	1	3.854.995 pb	3.854.995	3.854.995	64,6
7	PNF 4.7	<i>Subtercola</i>	Flye	1	3.838.599pb	3.838.599	3.838.599	68,4

Tabla 8. Tabla comparativa de los ensamblajes de lecturas largas (ONT), cortas (illumina) y ensamblajes híbridos (ONT+illumina).

	Código	TAX 16S	QUAST					
			Ensamblaje	longitud Contig mas largo (pb)	No.De contigs ensamblados	%GC del ensamblaje	Tamaño Ensamblaje total (pb)	N 50 Quast
2	C.3.20	<i>Amycolatopsis</i>	Flye	249.927	167	68.0	10.707.162	88.504
			Spades	274.711	194	68.17	9.108.594	100.002
			Unicycler	762.162	81	68.13	9.125.515	195.946
	Código	TAX 16S	Ensamblaje	longitud Contig mas largo (pb)	No.De contigs ensamblados	%GC del ensamblaje	Tamaño Ensamblaje total (pb)	N 50 Quast
3	C.3.21	<i>Streptomyces</i>	Flye	130.883	19	69	734.381	74.260
			Spades	73.108	1.059	70.67	8.514.079	13.862
			Unicycler	381.950	127	70.58	8.429.113	134.384
	Código	TAX 16S	Ensamblaje	longitud Contig mas largo (pb)	No.De contigs ensamblados	%GC del ensamblaje	Tamaño Ensamblaje total (pb)	N 50 Quast
4	C3.25	<i>Rhodococcus</i>	Flye	6.349.031	63	62.3	9.913.200	6.349.031
			Spades	969.722	54	62.3	6.875.795	406.202
			Unicycler	6.911.887	4	62.27	6.911.887	6.353.612
	Código	TAX 16S	Ensamblaje	longitud Contig mas largo (pb)	No.De contigs ensamblados	%GC del ensamblaje	Tamaño Ensamblaje total (pb)	N 50 Quast
6	PNE 1.1	<i>Arthrobacter</i>	Flye	3.854.995	1	64.6	3.854.995	3.854.995
			Spades	520.062	26	64.6	3.823.103	299.434
			Unicycler	2.811.441	3	64.5	3.847.578	2.811.441
	Código	TAX 16S	Ensamblaje	longitud Contig mas largo (pb)	No.De contigs ensamblados	%GC del ensamblaje	Tamaño Ensamblaje total (pb)	N 50 Quast
7	PNF 4.7	<i>Subtercola</i>	Flye	3.838.599	1	68.4	3.838.599	3.838.599
			Spades	727.448	48	68.29	3.858.140	248.374
			Unicycler	2.433.627	2	68.45	3.839.936	2.433.627

Después de realizar la revisión de calidad de los ensamblajes con Quast (Gurevish *et al.*, 2013), se evaluaron los porcentajes de contaminación y de genoma completado con el programa Check M (Parks *et al.*, 2015). En los ensamblajes de lecturas Illumina

realizados con Spades y en los ensamblajes híbridos realizados con Unicycler, el porcentaje de contaminación fue de 1 % mientras que el porcentaje de genoma completado fue del 99% (**Figura 5**). En contraste con Flye (ONT) la contaminación promedio de los ensamblajes fue de 9% mientras que el porcentaje de genoma completado fue del 72%. En conclusión, los ensamblajes Spades (Illumina) y Unicycler (ONT + Illumina) mostraron un porcentaje mayor de genoma completado y una disminución de la contaminación (**Figura 5**). La contaminación en los ensamblajes de Flye puede deberse a la presencia de contigs con bajo contenido GC que se puede observar en las gráficas del **Anexo 5** obtenidas con Quast. En Actinobacterias se espera un contenido GC mayor a 55% (Doroghazi *et al.*, 2013). La presencia de contigs con contenidos inferiores a 55% en Quast indica también contaminación (Gurevich *et al.*, 2013) (**Anexo 5**).

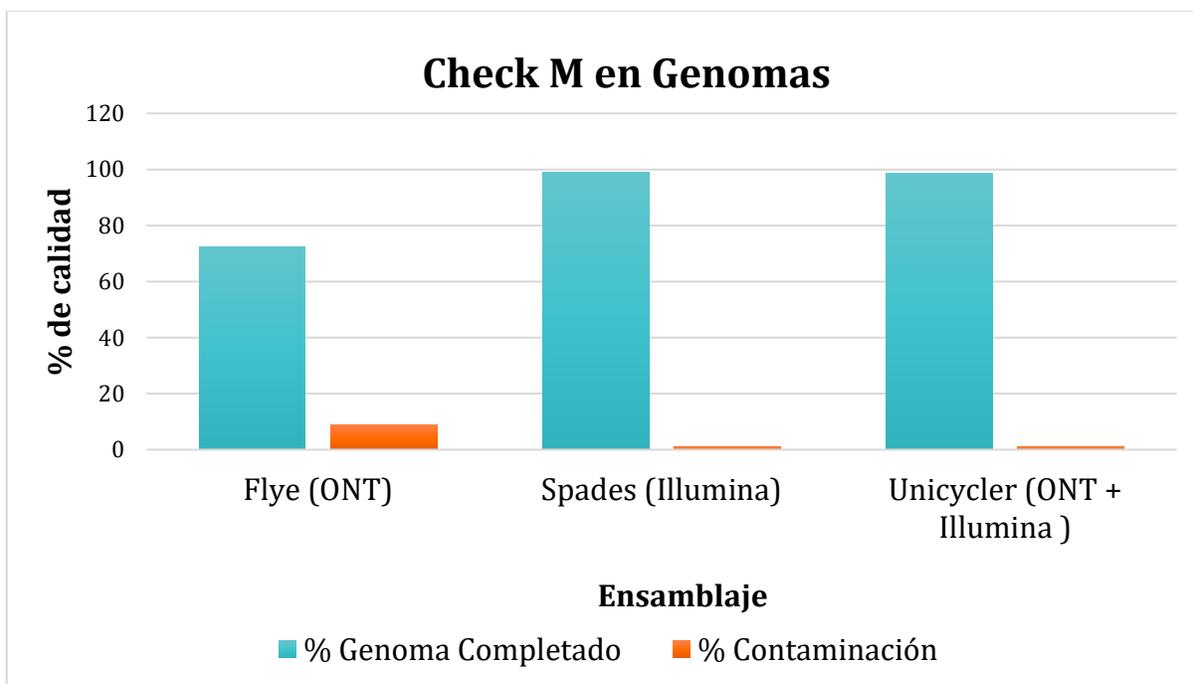


Figura 5. Porcentaje de genoma completado y porcentaje de contaminación de los genomas según metodología de ensamblaje

8. Asignación taxonómica de los genomas

Los siete genomas obtenidos a partir de los ensamblajes se evaluaron en sus taxonomías con JspeciesWS (Ritcher *et al.*, 2016). Los siete genomas tuvieron la misma asignación taxonómica a nivel de género respecto a la identificación inicial con el gen 16S rRNA (**Figura 1 y Tabla 9**). Además, los genomas híbridos de *Rhodococcus* C 3.25 y *Streptomyces* C3.21 presentaron asignación taxonómica hasta especie con ANI de

98.2% y 95.75%, respectivamente. Los otros genomas no se asignaron hasta especie porque presentaron identidades menores al 89%.

El genoma de *Rhodococcus* C 3.25 se evaluó también con MiGA (Rodriguez *et al.*, 2018) porque contiene una base de datos curada. *Rhodococcus* C 3.25 en MiGA presentó una identidad de 98.47% con el genoma de *Rhodococcus qingshengii* (GCA_015034605.1), esta identidad fue mayor respecto a la asignación obtenida con JSpeciesWS (**Tabla 9**).

La cobertura de los 7 genomas ensamblados respecto al genoma más similar fue mayor al 85% (**Tabla 9**). Los genomas de C 3.20 *Streptomyces*, C 3.25 *Rhodococcus*, PNA 1.2 *Arthrobacter* y PNE 1.1 *Arthrobacter*, mostraron una cobertura mayor al 100% respecto al genoma de referencia. La cobertura mayor al 100% se debe a que los genomas ensamblados presentan mayor tamaño en comparación con los genomas de referencia (**Tabla 9**). Para los genomas ensamblados de *Amycolatopsis* C 3.20, los *Arthrobacter* (PNA 1.2 y PNE 1.1) y el *Rhodococcus* C 3.25 se obtuvo el mismo contenido GC respecto a los genomas de referencia. Además, los genomas de los *Streptomyces* (C 1.1 y C 3.21) presentaron el mayor contenido GC con valores por encima del 70%. Por otro lado, el número de contigs en los genomas ensamblados fue menor respecto al número de contigs en los genomas de referencia para *Streptomyces* C 1.1, *Subtercola* PNF 4.7 y los dos *Arthrobacter* PNE 1.1 y PNA 1.2. En cambio, aumentó el número de contigs en comparación con los genomas de referencia solo en C 3.20 *Amycolatopsis* y C 3.21 *Streptomyces*.

Tabla 9. Comparación de genomas ensamblados vs. genomas más similares con JSpeciesWS.

No.	Ensamblajes Flye y Unicycler							JSpecies WS							
	Código	TAX16S	Ensamblaje	longitud Contig mas largo (pb)	No. De contigs ensamblados	%GC del ensamblaje	Tamaño Ensamblaje total (pb)	No. Accesoión	Especie más similar	No. De Contigs genoma, mas similar	%GC del genoma más similar	Longitud genoma más similar (pb)	TCS	ANI	% Cobertura
1	C.1.1	<i>Streptomyces</i>	F	4.552.434	19	70,1	10.110.931	GCA_000367365.1	<i>Streptomyces prunicolor NBRC 13075</i>	202	69,73	11.760.649	99,9	89,2	85,97
2	C.3.20	<i>Amycolatopsis</i>	U	249.927	167	68	10.707.162	GCA_001558125.2	<i>Amycolatopsis regifaucium GY080</i>	62	68,5	8.283.019	97,93	75,62	110
3	C.3.21	<i>Streptomyces</i>	U	381.950	127	70,58	8.429.113	GCA_000818175.1	<i>Streptomyces sp. ACH</i>	27	67,54	9.005.794	99,79	95,75	94
4	C.3.25	<i>Rhodococcus</i>	U	6.349.031	4	62,3	6.911.887	GCA_000174835.1	<i>Rhodococcus erythropolis SK121</i>	124	62,3	6.785.398	99,9	98,2	101,86
5	PNA.1.2	<i>Arthrobacter</i>	F	3.971.780	1	65,2	3.971.880	GCA_009192735.1	<i>Arthrobacter gandavensis JCM 13316</i>	12	65,33	3.722.081	99,61	88,36	106,71
6	PNE.1.1	<i>Arthrobacter</i>	F	3.854.995	1	64,6	3.854.995	GCA_009192735.1	<i>Arthrobacter gandavensis JCM 13316</i>	12	65,33	3.722.081	98,07	81,14	103,57
7	PNF.4.7	<i>Subtercola</i>	F	3.838.599	1	68,4	3.838.599	GCA_006716115.1	<i>Subtercola boreus (high GC Gram+)</i>	2	67,98	4.214.389	98,25	80,09	91,08

(F) ensamblaje (ONT) con Flye. **(U)** ensamblaje (ONT+Illumina) con Unicycler. **TCS** (Tetra correlation search)

9. Anotación de los genomas

Después de la asignación taxonómica se realizó la anotación de los 7 genomas. Primero se identificaron las regiones codificantes (CDS), los RNA de transferencia (tRNA), los RNA mensajeros de transferencia (tmRNA) y los RNA ribosomales (rRNA). Se obtuvo un mayor número de CDS en el genoma de *Amycolatopsis* (C 3.20), respecto al resto de los genomas (**Tabla 10**). Luego se realizó la comparación del número CDS entre los genomas de referencia más similares obtenidos por JSpecies y los genomas de las Actinobacterias de este estudio, ensamblados con las tres metodologías (**Figura 6**). Los genomas de *Rhodococcus* C 3.25, *Amycolatopsis* C 3.20 y *Streptomyces* C 1.1 generados por Nanopore y ensamblados con Flye presentaron un mayor número de CDS que los genomas ensamblados con Unicycler, Spades y los respectivos genomas de referencia (**Figura 6**). Esto puede deberse a errores del ensamblaje propios de la tecnología de Nanopore ya que esta tecnología presenta un porcentaje de error mayor al 10% a comparación de Illumina cuya tasa de error es menor al 1% (Delahaye y Jacques, 2021). Además, estos ensamblajes mostraron en Quast contigs con contenidos GC variables que indican contaminación (**Anexo 5**) (Gurevich *et al.*, 2013).

Luego de anotar los CDS se realizó la búsqueda de genes de resistencia, clústeres de genes biosintéticos y se hizo anotación con Kegg de los genes implicados en distintas rutas metabólicas para los 7 genomas seleccionados con Quast y Check M (**Tabla 9**).

Tabla 10. Anotación general de genomas de Actinobacterias con Prokka.

No.	Ensamblaje	código	Especie 16S	CDS	tRNA	tmRNA	rRNA
1	F	C 1.1	<i>Streptomyces</i>	12.838	90	1	18
2	U	C 3.20	<i>Amycolatopsis</i>	8.417	82	1	3
3	U	C 3.21	<i>Streptomyces</i>	7.291	88	1	7
4	U	C 3.25	<i>Rhodococcus</i>	6.361	55	1	0
5	F	PNA 1.2	<i>Arthrobacter</i>	3.688	57	2	0
6	F	PNE 1.1	<i>Arthrobacter</i>	3.987	53	2	18
7	F	PNF 4.7	<i>Subtercola</i>	3.940	51	1	6

(F) Flye (U) Unicycler (CDS) regiones codificantes (tRNA) RNA de transferencia, (tmRNA) RNA mensajero de transferencia, (rRNA) RNA ribosomal

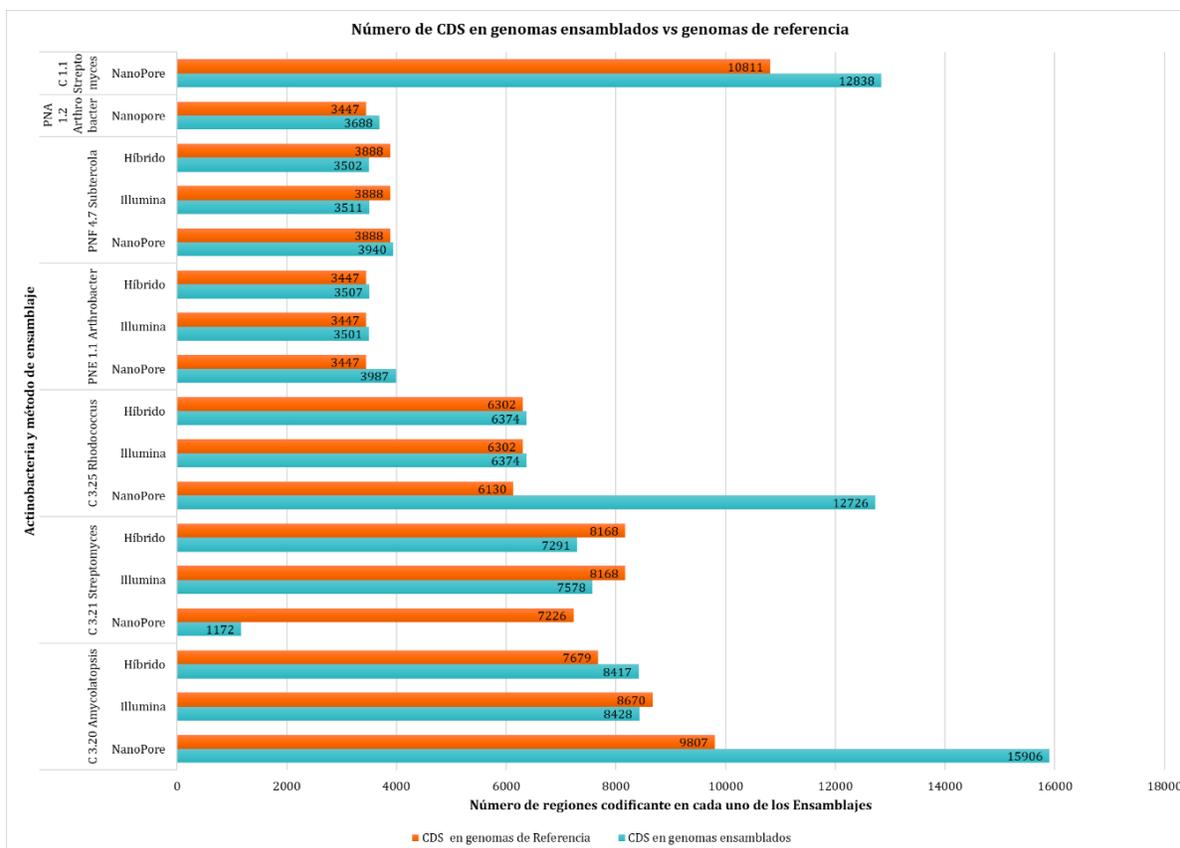


Figura 6. Número de regiones codificantes (CDS) en genomas de referencia y genomas ensamblados con las tres metodologías de Flye (ONT), Spades (Illumina) e Híbrido (ONT + Illumina).

Se realizó la búsqueda de genes de resistencia en los genomas de Actinobacterias con el fin de conocer posibles mecanismos de resistencia presentes en este grupo de Actinobacterias. La búsqueda de genes de resistencia con la base de datos de CARD (Alcock *et al.*, 2020) arrojó resultados con el algoritmo de búsqueda en modo relajado, el cual es un modelo de detección de homólogos distantes de genes de resistencia con identidades menores a 60%. Solamente en *Rhodococcus* C 3.25 se encontraron coincidencias con genes de resistencia en modo de búsqueda estricto, lo que indicó la presencia de genes de resistencia conocidos con identidades mayores al 90%.

El principal mecanismo de resistencia en las siete Actinobacterias fue el mecanismo de expulsión del antibiótico (Antibiotic efflux) con un total de 1777 genes (**Figura 7 y Tabla 11**), lo cual indica que las Actinobacterias pueden estar secretando sustancias tóxicas y antibióticos hacia el exterior de sus células mediante bombas de eflujo (Webber y Piddock, 2003) (**Tabla 11**).

En segundo lugar, se encontraron 433 genes implicados en la alteración o modificación de la diana del antibiótico (Antibiotic target alteration) (**Figura 7 y Tabla 11**). En estos casos alterar el blanco de un antibiótico puede disminuir la afinidad del antibiótico por su diana (Miller *et al.*, 2014).

El tercer mecanismo de resistencia encontrado, con un total de 205 genes en los 7 genomas, fue el de inactivación del antibiótico (**Tabla 11**). Los genes de resistencia asociados a la inactivación del antibiótico pueden deberse a procesos como degradación de antibióticos de forma química por hidroxilación o por la transferencia de grupos acetilo, fosforilo o adenilo (Reygaert, 2018).

El mecanismo de reducción a la permeabilidad del antibiótico con menor número de genes identificados en los 7 genomas (13 genes) (**Tabla 11**). Este mecanismo ha sido descrito sobre todo en bacterias Gram-negativas como *E. coli* y *K. pneumoniae*. La membrana externa de bacterias Gram-negativas contiene diferentes adaptaciones que les permiten resistir el ingreso de antibióticos al interior celular, como las modificaciones conformacionales en las porinas (Delcour, 2009). Es probable, por lo tanto, que bacterias Gram-positivas como las Actinobacterias de este estudio no tengan tantos genes de resistencia involucrados en este mecanismo debido a que no tienen la membrana exterior característica de bacterias Gram-negativas.

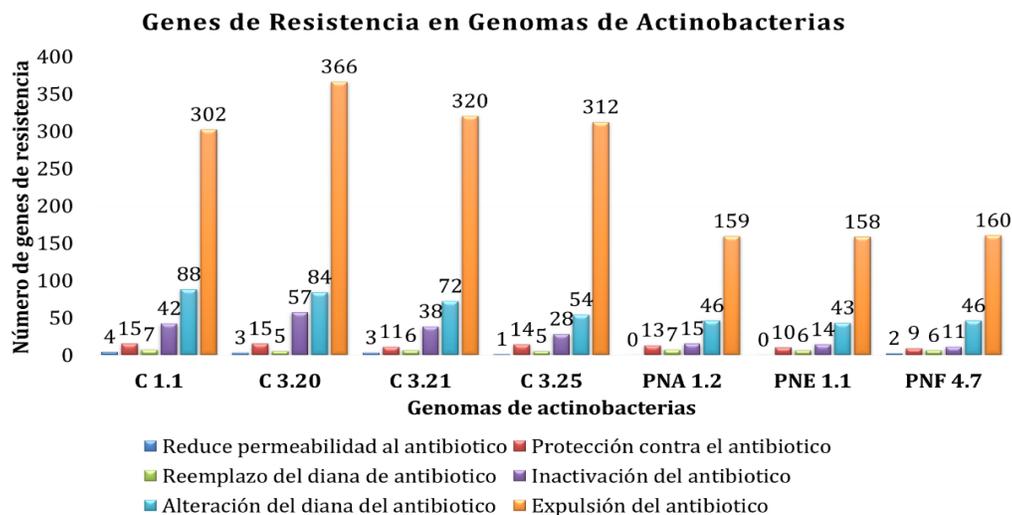


Figura 7. Genes de resistencia según el mecanismo de resistencia en genomas de Actinobacterias. **C 1.1** *Streptomyces*, **C 3.20** *Amycolatopsis*, **C 3.21** *Streptomyces*, **C 3.25** *Rhodococcus*, **PNA 1.2** *Arthrobacter*, **PNE 1.1** *Arthrobacter*, **PNA 4.7** *Subtercola*.

Tabla 11 Genes de resistencia en cada Actinobacteria.

N o.	código	Especie 16S	número de Hits	P. B	R.P. A	P.C. A	R.D. A	I.A	A.D. A	E.A	Total
1	C 1.1	<i>Streptomyces</i>	458	Relajado	4	15	7	42	88	302	458
2	C 3.20	<i>Amycolatopsis</i>	399	Relajado	3	15	5	57	84	366	530
3	C 3.21	<i>Streptomyces</i>	450	Relajado	3	11	6	38	72	320	450
4	C 3.25	<i>Rhodococcus</i>	414	Estricto y Relajado	1	14	5	28	54	312	414
5	PNA 1.2	<i>Arthrobacter</i>	240	Relajado	0	13	7	15	46	159	240
6	PNE 1.1	<i>Arthrobacter</i>	231	Relajado	0	10	6	14	43	158	231
7	PNF 4.7	<i>Subtercola</i>	234	Relajado	2	9	6	11	46	160	234
Total					13	87	42	205	433	1777	2557

Orden de las siglas de izquierda a derecha. **P.B** parámetros de búsqueda, **R.P.A** reduce permeabilidad del antibiótico (reduced permeability to antibiotic), **P.C.A** protección contra el antibiótico (antibiotic target protection), **R.D.A** reemplazo del diana del antibiótico (antibiotic target replacement) **I.A** inactivación del antibiótico (antibiotic inactivation), **A.D.A** alteración del diana del antibiótico (antibiotic target alteration), **E.A** expulsión del antibiótico (antibiotic efflux).

Adicionalmente solamente en el genoma híbrido (ONT + Illumina) de C 3.25 *Rhodococcus* se identificaron tres genes con identidad mayor al 60% y porcentaje de longitud mayor al 98%, con el parámetro de búsqueda estricto. Los genes identificados fueron: *iri*, *rbpA* y *folC* con mutación para resistencia a ácido paraaminosalicílico (**Tabla 12**). *iri* y *rbpA* están relacionados a la resistencia al antibiótico rifampicina que se usa en el tratamiento de la tuberculosis.

Tabla 12 Genes de resistencia en C 3.25 *Rhodococcus* con mayor porcentaje de identidad encontrados en CAD.

Código y Especie	Hits estrictos	Criterio de detección	AMR familia de genes	Nombre del gen	clase de medicamento	Mecanismo de resistencia	% de identidad	% longitud
C 3.25 <i>Rhodococcus</i>	1	Homología de proteínas	Rifampin Monooxygenase	<i>Iri</i>	Antibiotico Rifampicina	Inactivación del Antibiótico	95.4	99.79
	1	Homología de proteínas	RbpA bacteria-RNA polimerasa binding protein	<i>RbpA</i>	Antibiotico Rifampicina	protección contra el antibiótico	86.49	98.25
	1	Homología de proteínas	<i>Mycobacterium tuberculosis</i> folC con mutación para resistencia a ácido paraaminosalicílico	<i>modo de proteína</i>	Acido para-aminoasilico	Alteración contra el antibiótico	67.77	98.36

				varia nte				
--	--	--	--	--------------	--	--	--	--

La búsqueda de BGCs con AntiSMASH (Blin *et al.*, 2021) arrojó los siguientes resultados. En primer lugar, de un total de 97 BGCs encontrados en los 7 genomas solo 24 presentaron una similitud mayor al 50% (**Figura 7**). Estos BGCs se escogieron puesto que contenían genes con un % de identidad mayor comparados con los BGCs de la base de datos de AntiSMASH. Sin embargo, BGCs con similitud menor al 50% pueden ser novedosos respecto a los que están indexados en la base de datos, dado que estas Actinobacterias de páramo pueden contener nueva información que no ha sido caracterizada a nivel genómico (**Tabla 13**).

Entre los BGCs con similitud mayor al 50%, se encontraron las siguientes rutas biosintéticas: 4 de terpenos, 3 son RiPP-like, 6 para NRPS, 4 PKS, 3 para ectoína, 3 sideróforos y 1 de oligosacáridos (**Figura 8 y Tabla 13**). El genoma con mayor número de BGCs fue el C 1.1 *Streptomyces*, en el que se encontraron BGCs de las rutas para terpenos, RiPP-like, sideróforos, NRPS, PKS y ectoína, seguido de *Amycolatopsis* (C 3.20) y *Rhodococcus* (C 3.25) (**Figura 8**). En los dos *Arthrobacter* (PNA 1.2 y PNE 1.1) se encontraron sólo clústeres de las rutas de los sideróforos y en *Subtercola* PNF 4.7 solo se encontraron terpenos y un oligosacárido (**Figura 8**).

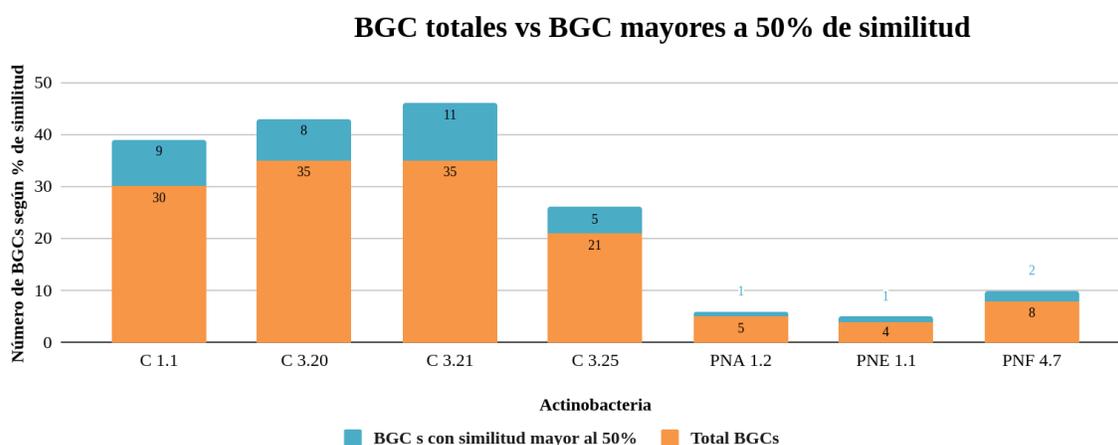


Figura 7. BGCs totales y BGCs con porcentaje de similitud mayor al 50%. *C 1.1 Streptomyces*, *C 3.20 Amycolatopsis*, *C 3.21 Streptomyces*, *C 3.25 Rhodococcus*, *PNA 1.2 Arthrobacter*, *PNE 1.1 Arthrobacter*, *PNA 4.7 Subtercola*. Eje Y. número de BGC de acuerdo al porcentaje de similitud.

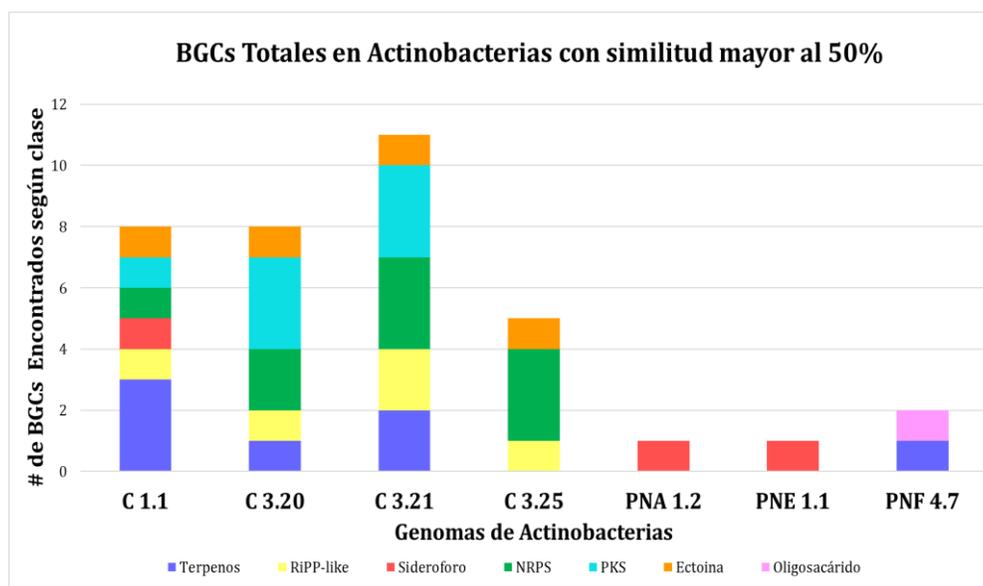


Figura 8. BGCs en genomas de Actinobacterias. C 1.1 *Streptomyces* , C 3.20 *Amycolatopsis* , C 3.21 *Streptomyces* , C 3.25 *Rhodococcus* , PNA 1.2 *Arthrobacter* , PNE 1.1 *Arthrobacter* , PNA 4.7 *Subtercola*.

Tabla 13. BGCs totales encontrados en los 7 genomas de Actinobacterias.

N o.	código	Especie 16S	Total BGCs	similitud mayor al 50%	Terpenos	RiPP-like	NRPS	PKS	Ectoína	Sideróforo
1	C 1.1	<i>Streptomyces</i>	30	8	3	1	1	1	1	1
2	C 3.20	<i>Amycolatopsis</i>	35	6	1	1	0	3	1	0
3	C 3.21	<i>Streptomyces</i>	35	11	2	2	0	3	3	1
4	C 3.25	<i>Rhodococcus</i>	21	5	0	1	3	0	1	0
5	PNA 1.2	<i>Arthrobacter</i>	5	1	0	0	0	0	0	1
6	PNE 1.1	<i>Arthrobacter</i>	4	1	0	0	0	0	0	1
7	PNF 4.7	<i>Subtercola</i>	8	1	1	0	0	0	0	0
Total			138	33	7	5	4	7	6	4

Después de la búsqueda de los BGCs en estos 7 genomas se trabajó con los 33 BGCs con mayor similitud (**Tabla 13**) para identificar sus posibles funciones. Los BGCs

encontrados y los compuestos bioactivos reportados para cada uno de los BGCs se presentan en el **Anexo 6**. En general los BGCs encontrados fueron muy similares a BGCs reportados y descritos en AntiSMASH provenientes de genomas de Actinobacterias de los géneros *Streptomyces*, *Rhodococcus*, *Amycolatopsis* y *Saccharopolyspora*. Los BGCs de genomas de diferentes *Streptomyces* fueron los de mayor prevalencia en los genomas ensamblados en este estudio (**Anexo 6**).

Los compuestos bioactivos asociados a cada BGC se identificaron mediante MIBiG (Minimum Information about a Biosynthetic Gene cluster) y PubMed. Las funciones principales de cada uno de los 24 compuestos y sus estructuras químicas se presentan en el **Anexo 7**. En *Subtercola* PNF 4.7 y *Rhodococcus* C 3.25 se encontró un ácido graso de cadena ramificada involucrado en la biosíntesis de daptomicina. La daptomicina es un antibiótico producido naturalmente por *Streptomyces roseosporus*, con función bactericida en Gram positivas de importancia clínica como *S. aureus* y *Enterococcus spp.* (Araos *et al.*, 2012) (**Anexos 6 y 7**). En *Rhodococcus* C 3.25 se encontró un clúster biosintético de tipo NRPS involucrado en la biosíntesis de marformicina A, marformicina B, marformicina C, marformicina D, marformicina E y marformicina F (**Anexo 6 y 7**). Las marfomicinas A-F son producidas por *Streptomyces drozdowiczii*, y se han descrito con actividad antiinfecciosa selectiva contra *Micrococcus luteus* (Zhou *et al.*, 2014). En los **Anexos 6 y 7** se muestran 25 compuestos bioactivos con funciones antibióticas, antioxidantes, antitumorales, quelantes y de osmorregulación.

Finalmente se hizo la anotación funcional con Kegg (Minoru y Sususmu, 2000) del genoma de C3.25 *Rhodococcus*, cuyo ensamblaje híbrido dio la mejor calidad según Quast (Gurevich *et al.*, 2013) (**Tabla 7**). En esta anotación se muestran las regiones codificantes distribuidas a lo largo del genoma completo, así como genes asociados con funciones metabólicas importantes como el metabolismo energético, metabolismo de proteínas y de policétidos y terpenoides (**Figura 9**).

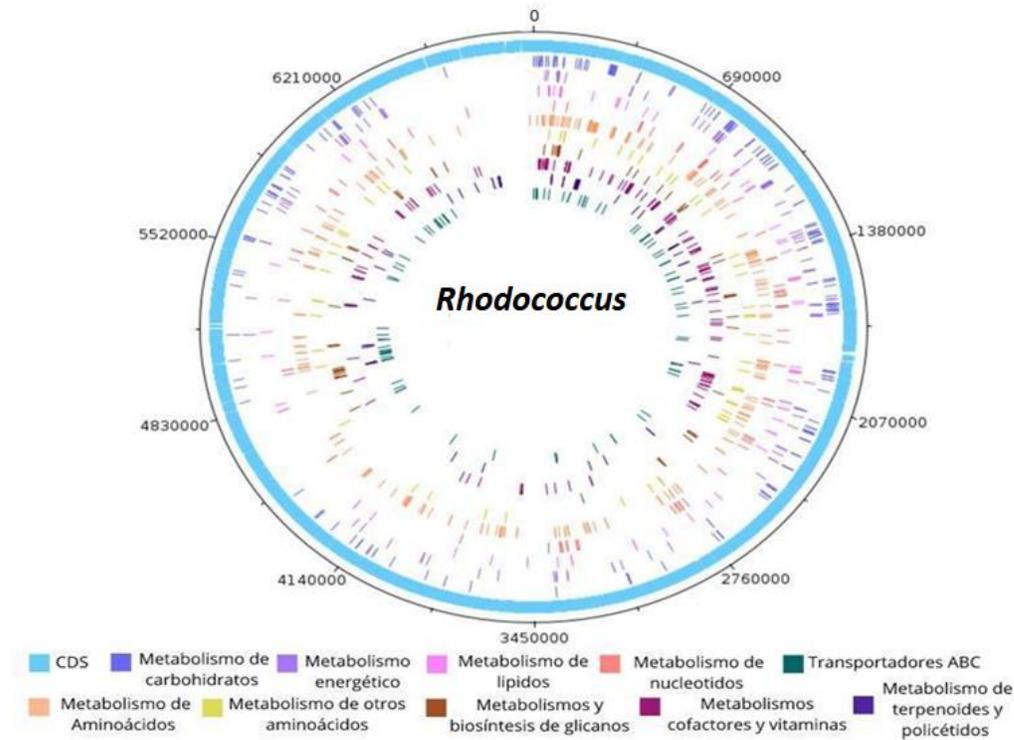


Figura 9. Genoma híbrido de *Rhodococcus* ensamblado a partir de (Illumina+ONT). Asignación taxonómica hasta género.

La anotación de los siete genomas y de las pruebas de actividad antimicrobiana contra bacterias de importancia clínica se resumen en la **Tabla 14**, que muestran las actividades antimicrobianas de las Actinobacterias secuenciadas y el contenido de BGCs. En general en *Arthrobacter* (PNA 1.2 y PNE 1.1) con % de inhibición promedio de 22.6 % (ISP2+AIAG) se encontraron cuatro BGCs de los cuales solo un BGC presentó un porcentaje de similitud mayor al 50% respecto a la base de datos (Tabla 14). Ese BGC en las dos *Arthrobacter* corresponde a un sideróforo que es un compuesto producido en bacterias con actividad quelante de hierro y que ha exhibido actividad antitumoral (Yu *et al.*, 2012). Sin embargo, en este trabajo no se realizaron ensayos de actividad antitumoral o actividad quelante de las Actinobacterias.

En las Actinobacterias de los géneros *Streptomyces* (C 1.1 y C 3.21), *Amycolatopsis* (C 3.20) y *Rhodococcus* (C 3.25) se presentó un % de inhibición promedio de 11.75% (ISP2+AIAG) y a su vez presentaron más de 30 BGCs en sus genomas (**Tabla 14**). Los valores bajos de la actividad antimicrobiana y el contenido alto de BGCs en sus genomas puede deberse a la presencia de BGCs crípticos que no se expresan en las condiciones de laboratorio evaluadas (Mao *et al.*, 2018; Renn *et al.*, 2015). Adicionalmente, pueden tener actividades contra otras bacterias o hongos no evaluados en los ensayos de

actividad antimicrobiana de este estudio como los enfrentamientos en medio sólido y contra otras especies bacterianas..

Tabla 14. Contenido de BGCs y actividad antimicrobiana en los 7 genomas secuenciados del filo Actinobacteria.

No.	Ensamblaje	Especie 16S	código	% Inhibición	BGC s totales	BGC s con similitud mayor al 50%
1	F	<i>Streptomyces</i>	C 1.1	47	30	8
2	U	<i>Amycolatopsis</i>	C 3.20	0	35	8
3	U	<i>Streptomyces</i>	C 3.21	0	35	11
4	U	<i>Rhodococcus</i>	C 3.25	0	21	5
5	F	<i>Arthrobacter</i>	PNA 1.2	0	5	1
6	F	<i>Arthrobacter</i>	PNE 1.1	68	4	1
7	F	<i>Subtercola</i>	PNF 4.7	28	8	2
Promedio/ Total				20.4	138	36

DISCUSIÓN

En este trabajo se estudiaron microorganismos de páramos Colombianos con el fin de conocer su potencial antimicrobiano e identificar posibles BGCs, genes de resistencia y su diversidad genómica. Los ecosistemas de páramo como Chingaza y parque nacional de los nevados pueden albergar una alta diversidad de microorganismos que debido a presiones selectivas particulares de sus ambientes, se han adaptado y pueden estar produciendo metabolitos secundarios con posibles actividades antimicrobianas desconocidas (Sierra 2018; Sierra *et al.*, 2020). Por lo tanto, en este trabajo se activaron 16 aislamientos provenientes de ecosistemas de páramo pertenecientes a la colección de Actinobacterias de CorpoGen. De los 16 aislamientos iniciales solo 8 mostraron alguna actividad antimicrobiana en ensayos a 7 ó 20 días (**Anexo 8**). Del grupo de 16 microorganismos, 9 se identificaron como Actinobacterias por análisis del gen 16S rRNA y se seleccionaron 7 actinobacterias de 5 géneros diferentes por análisis del gen 16S rRNA. Los géneros fueron *Arthrobacter* (PNA 1.2, PNE 1.1), *Streptomyces* (C 1.1, C 3.21), *Subtercola* (PNF 4.7), *Amycolatopsis* (C 3.20) y *Rhodococcus* (C 3.25). Estas bacterias presentaron porcentajes de inhibición del crecimiento en un rango entre 0-68% contra las 4 bacterias patógenas ensayadas. Aunque solo 7 de 16 aislamientos pertenecían al filo Actinobacteria otros sobrenadantes pertenecientes a los

aislamientos B1.2 y PNC 2.3 inhibieron el crecimiento de *E. coli* y las patógenas en más de un 80% (**Tabla 2**).

Los resultados de porcentaje de inhibición variables están condicionados por varios factores, como el tiempo de crecimiento, el medio de cultivo, resistencia desconocida a los compuestos producidos por parte de las bacterias patógenas ensayadas, degradación del compuesto antimicrobiano, baja concentración del compuesto liberado, entre otros.

Según Charousová y colaboradores (2017), las Actinobacterias muestran buena actividad antimicrobiana contra bacterias Gram-positivas y actividades antimicrobianas moderadas contra Gram-negativas. La baja actividad antimicrobiana contra Gram-negativas puede deberse a las modificaciones en porinas o bombas de eflujo presentes en su membrana externa (Bolla *et al.*, 2011). Estos antecedentes pueden explicar algunas de las causas de la baja actividad antimicrobiana contra las bacterias Gram-negativas: *K. pneumoniae*, *P. aeruginosa* y *E. coli* evaluadas en el presente estudio (**Tabla 2**).

Además de la membrana externa en Gram-negativas, las condiciones de cultivo, como el uso de medio sólido o medio líquido, pueden conducir a la producción de distintos antibióticos activos (Charousová *et al.*, 2017). En este estudio solo se realizaron actividades antimicrobianas en medio líquido (ISP2+AIAG). En el trabajo realizado por Vargas y colaboradores (2022) (Datos publicados en póster) se ensayaron también los aislamientos *Arthrobacter* (PNA1.2) y *Streptomyces* (C3.21) en medio sólido ISP2+ISP4+AIAG por 35 días y se encontró una inhibición en el crecimiento de *E. coli* del 100% y 50%, respectivamente. Estos resultados sugieren mejor actividad antimicrobiana a 35 días y en medio sólido, por lo menos para estas dos cepas, en comparación con los ensayos realizados en este estudio en medio líquido. Otro motivo por el cual se pudo presentar la baja actividad antimicrobiana en los ensayos de laboratorio puede ser la presencia de BGCs silenciosos o crípticos que no se activaron o se expresaron débilmente en las condiciones de laboratorio estudiadas (Mao *et al.*, 2018).

El análisis de los genomas identificó varios clusters de genes biosintéticos, con diferencias en cada Actinobacteria. Estos BGCs se identificaron como genes involucrados en síntesis de terpenos, RiPP-like, NRPS, PKS, ectoína y sideróforos. Solo en los *Streptomyces* C1.1 y C 3.21 se encontró un terpeno denominado geosmina, el cual es responsable del olor característico a tierra húmeda en los suelos. Cabe destacar que durante las experimentaciones en laboratorio esta Actinobacteria presentaba dicho olor (Jiang *et al.*, 2007) (**Tabla 1**). En PNA 1.2 y PNE 1.1 *Arthrobacter* se encontró un

clúster involucrado en la biosíntesis de “desferrioxamine E” el cual es un compuesto que tiene un papel como sideróforo, metabolito bacteriano, metabolito marino y agente antineoplásico (Yamanaka *et al.*, 2005) **(Anexo 7)**. En el caso del *Subtercola* PNF 4.7 se identificó un terpeno de tipo carotenoide, el cual también fue identificado en el estudio reciente del genoma de *Subtercola endophyticus* (Jiang *et al.*, 2022) **(Anexo 7)**.

En los genomas de *Arthrobacter* (PNA 1.2 y PNE 1.1) y *Subtercola* (PNF 4.7) se identificaron en total solo 5, 4 y 8 BGCs, respectivamente. El bajo número de BGCs en los *Arthrobacter* (PNA 1.2 y PNE 1.1) y el *Subtercola* (PNF 4.7) puede estar asociado a diferentes factores como los algoritmos de detección de BGCs por parte de AntiSMASH **(Tabla 14)**. En AntiSMASH la detección de BGCs en los genomas se hace mediante la identificación de las enzimas involucradas en la biosíntesis de metabolitos secundarios conocidos y descritos por investigadores. El algoritmo con base en HMMER (Modelos ocultos de Markov) utiliza matrices de puntuación probabilísticas que determinan la identidad entre perfiles específicos de las enzimas conservadas en bases de datos como PFAM y en los genomas de interés (Blin *et al.*, 2019). En consecuencia, la identificación de BGCs con AntiSMASH depende de la base de datos y de si los genes involucrados en la biosíntesis de los metabolitos buscados son conocidos y codifican enzimas conocidas e identificadas previamente. La presencia de enzimas desconocidas y no relacionadas no se puede detectar con AntiSMASH (Blin *et al.*, 2019). Por tal razón, en los genomas de *Arthrobacter* y *Subtercola* la búsqueda de BGCs podría complementarse con programas como ClusterFinder, ARTS ó EvoMining que pueden detectar BGCs bajo diferentes reglas de detección utilizadas en AntiSMASH (Blin *et al.*, 2019). Adicional al modelo de detección, la identificación de solo 8 BGCs en el genoma de *Subtercola sp.*, de los cuales solo 2 tuvieron porcentaje de similitud mayor al 50% puede deberse a que el género ha sido poco caracterizado a nivel genómico. En la base de datos de genomas en borrador y completos del genbank (www.ncbi.nlm.nih.gov/genbank/) están cargados solo 28 genomas correspondientes a 4 taxones del género: *S. frigoramans*, *S. boreus*, *S. lobariae* y *S. vilae*.

En los dos *Arthrobacter sp.* PNA 1.2 y PNE 1.1 se encontraron 5 y 4 BGCs en total y solo un BGCs en ambos tuvo una similitud mayor al 50% **(Anexo 7)**. Estos dos actinomicetos mostraron porcentajes de inhibición de crecimiento de 14% (x20 días) y 68% (x 7 días) contra *K. pneumoniae*, respectivamente **(Tabla 14)**. En contraste, en experimentos de Vargas y colaboradores (2022) el *Arthrobacter sp.* PNA 1.2 (x 35 días) inhibió en un 100% el crecimiento de *E. coli*. La baja cantidad de BGCs conocidos encontrados en ambos genomas (similitud mayor 50%), y la actividad antimicrobianas alta en el *Arthrobacter sp.* PNA 1.2 pueden indicar la presencia de clusters nuevos no indexados en AntiSMASH (Tabla 8). Adicionalmente, ensayos en diferentes condiciones de experimentación pudieron interferir con los porcentajes de inhibición obtenidos. En el

trabajo de Munaganti y colaboradores en *Arthrobacter kerguelensis* encontraron porcentajes de inhibición de hasta 30% contra *S. aureus* y *E. coli* en medio con agar ISP2 , a un pH 7, temperatura de 30°C y después de 5 días de incubación (Munaganti *et al.*, 2016).

En el genoma de *Amycolatopsis* (C 3.20) se encontraron BGCs de tipo PKS asociados a la producción de “arimetamycin A, arimetamycin B y arimetamycin C”. La arimetamycin A es una antraciclina identificada por primera vez mediante el estudio metagenómico de suelo (Kang y Brady, 2013). Este compuesto tiene funciones anticancerígenas (Kang y Brady, 2013). En este trabajo, *Amycolatopsis* (C 3.20) no inhibió el crecimiento de ninguna de las cuatro bacterias patógenas evaluadas en este estudio. Sin embargo, se ha reportado actividad antimicrobiana con halos de inhibición de 20 mm contra *E. coli* en un *Amycolatopsis* aislada en Tailandia y ensayado después de 11 días de incubación a 30°C, en medio líquido y cuyos compuestos se extrajeron con Etil-acetato y se ensayaron por difusión en disco contra *E. coli* (Sripairoj *et al.*, 2013). Por tal razón, sería interesante evaluar diferentes metodologías y condiciones de cultivo para ensayar la actividad antimicrobiana de *Amycolatopsis* C 3.20 evaluado en este estudio.

La búsqueda preliminar de genes de resistencia en los siete genomas ensamblados resultó en un total de 2557 genes de resistencia asociados a diversos mecanismos de resistencia (**Figura 7**). Encontrar genes de resistencia en Actinobacterias de páramo (lugares poco intervenidos) es interesante debido a que estos ecosistemas no están cercanos a fuentes de resistencia como los hospitales. La presencia de genes de resistencia en Actinobacterias (no patógenas) se ha descrito como un mecanismo de autoprotección frente a los mismos compuestos antimicrobianos que producen o también como posibles moléculas señalizadoras implicados en la comunicación interbacteriana ambiental (Berglund, 2015; Jian *et al.*, 2017). Por ejemplo, se ha descrito que géneros productores de una amplia gama de antibióticos como *Streptomyces* (secuenciado en este estudio) contienen genes de resistencia como mecanismo de autoprotección frente a los antibióticos que ellos mismos producen (Jian *et al.*, 2017).

Específicamente, sólo en el genoma de *Rhodococcus* se encontró (mediante búsqueda estricta) alta identidad con los genes *iri* y *rbpA* asociados a la resistencia a Rifampicina (**Tabla 12**). Se ha descrito que en las Actinobacterias patógenas *Rhodococcus equi*, *R. jostii* y *R. opacus* también hay bombas de eflujo que comparten 55% de identidad con otras bombas de eflujo de *M. tuberculosis* (De Carvalho *et al.*, 2014).

Los ensayos preliminares de actividad antimicrobiana mostraron bajos porcentajes de inhibición del crecimiento de bacterias patógenas. Sin embargo, el minaje y caracterización del genoma de 6 Actinobacterias reveló su potencial bioactivo, por la de

detección de BGCS implicados en la producción de 24 compuestos con actividades ya reportados y asociados a funciones antibióticas, antioxidantes, antitumorales y quelantes. Estos resultados, aportan al conocimiento de las comunidades microbianas de suelos de páramos y nevados de Colombia que han sido poco estudiados. Además, sienta las bases para la exploración de genes de resistencia en bacterias de ambientes prístinos y su relación con genes de resistencia en ambientes intervenidos.

CONCLUSIONES

1. Se caracterizó el potencial funcional de Actinobacterias de páramo de Colombia mediante ensayos *in vitro* e identificación de BGCs en sus genomas.
2. 7 Actinobacterias inhibieron el crecimiento de bacterias patógenas en un rango de 0 a 68%, variación que se debe a factores de experimentación como medio de cultivo, tiempo de crecimiento, entre otros.
3. La secuenciación de 7 genomas de Actinobacterias por Nanopore permitió ensamblajes con cobertura mayor al 85%, en comparación con genomas de referencia de especies cercanamente emparentadas.
4. Los ensamblajes híbridos en géneros como *Streptomyces* y *Amycolatopsis* mostraron mayor % de genoma completado y poca contaminación respecto a los ensamblajes de Flye de ellos mismos por factores de error en la secuenciación con Nanopore.
5. En los 7 genomas se encontraron 138 clústers biosintéticos de las rutas NRPS, PKS, RiPP-like, terpenos y sideróforos, de los cuales 33 presentaron similitud mayor a 50%.
6. Se encontraron compuestos ya reportados asociados a funciones antibióticas, antioxidantes, antitumorales, quelantes y de osmorregulación.

RECOMENDACIONES FUTURAS

- Continuar con los ensayos de actividad antimicrobiana bajo diferentes condiciones de medio de cultivo y contra diferentes patógenos como hongos o levaduras.
- Continuar con la búsqueda complementaria de BGCs en otros programas como ClusterFinder, ARTS ó EvoMining con el propósito de encontrar BGCs interesantes no encontrados con AntiSMASH.
- Realizar comparación de los genes de resistencia encontrados en este estudio en bases de datos como ARTS para determinar proximidad o cercanía a BGCs desconocidos.

CONTRIBUCIONES ESPECÍFICAS DEL ESTUDIANTE A LA INVESTIGACIÓN

1. La estudiante contribuyó con sus resultados de 16S rRNA a la identificación taxonómica de 16 aislamientos almacenados en el cepario de CopoGen.
2. La estudiante caracterizó mediante pruebas de actividad antimicrobiana 16 aislamientos provenientes del cepario de CopoGen.
3. La estudiante realizó estandarización completa de un protocolo casero para extracción de ADN de alto peso molecular y con concentración mayor a 20ng/uL a partir de Actinobacterias crecidas en medio sólido.
4. La estudiante apoyó la estandarización del protocolo para secuenciación de genoma completo en bacterias con el equipo MinIon (Nanopore).
5. La estudiante desarrolló un protocolo óptimo para el ensamblaje *de Novo* de genomas bacterianos completos a partir de lecturas largas provenientes de Nanopore, así como el pulido de esta.
6. La estudiante contribuyó a la estandarización de un nuevo protocolo para el ensamblaje híbrido de secuencias largas y cortas (ONT + Illumina) en genomas completos de Actinobacterias de páramo.
7. La estudiante analizó los resultados genómicos a partir de los ensamblajes *de novo*, híbrido y de lecturas largas en los 7 genomas de este estudio.
8. La estudiante realizó e interpretó los resultados de anotación genómica con AntiSMASH, CARD y Kegg.

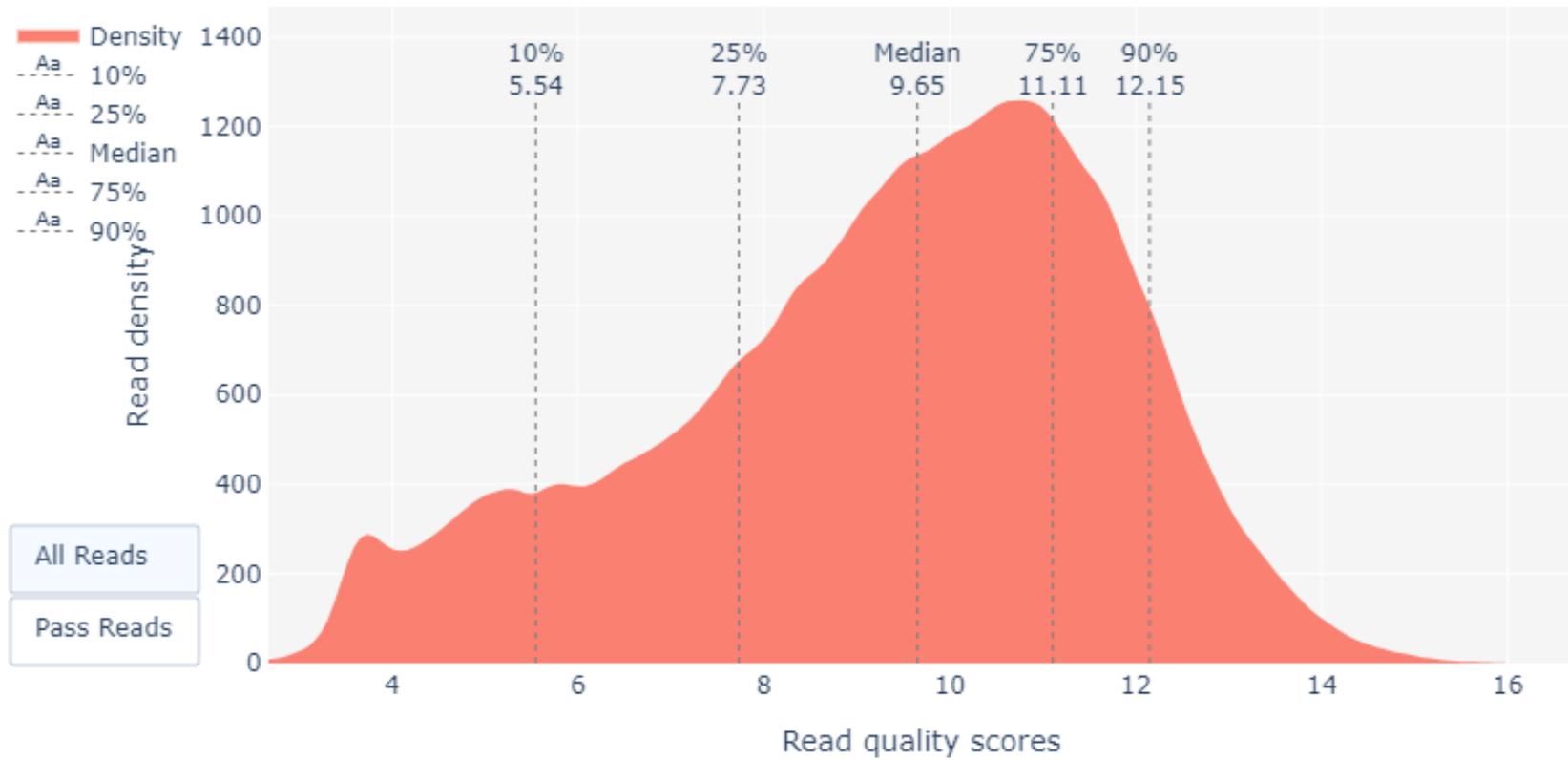
AGRADECIMIENTOS

- Agradecimientos a las Directoras, Dra. Patricia del Portillo y Dra. María Mercedes Zambrano y demás miembros de la Corporación Corpogen por abrirme las puertas del laboratorio y guiarme para realizar la experimentación de mi trabajo de grado.
- A la Universidad del Rosario por el apoyo financiero y a mis maestros del programa de maestría en Ciencias Naturales quienes siempre me brindaron las herramientas para mejorar como profesional.
- A la Dra. María Mercedes Zambrano, por abrirme las puertas en su grupo de investigación y por la confianza depositada en el desarrollo de este trabajo que me ha permitido crecer profesional y personalmente.
- A la Dra. Adriana Corrales por su guía y por el apoyo financiero en la secuenciación Illumina.
- A Fernando Vargas Zúñiga por brindarme su apoyo financiero y motivarme durante este proceso.
- A la Dra. Luz Nelly Díaz y la Dra. Luz Karime Afanador por su guía durante el desarrollo de las pruebas de actividad y experimentos durante el trabajo en el laboratorio
- A Christian Romero por su apoyo durante el montaje de librerías NanoPore.
- A Jeisson Alejandro Triana por el apoyo brindado con los protocolos de bioinformática.
- A la Dra. Maritsa Cruz, el Dr. Juan Manuel Anzola y el Dr. Joseph Petrone por la guía y el apoyo en protocolos bioinformáticos de ensamblaje
- Este trabajo de grado está dedicado a Rosaura Flórez Corredor, mi madre y mejor amiga, quien me ha brindado su amor, dedicación, paciencia y confianza incondicional a lo largo de mi vida académica.

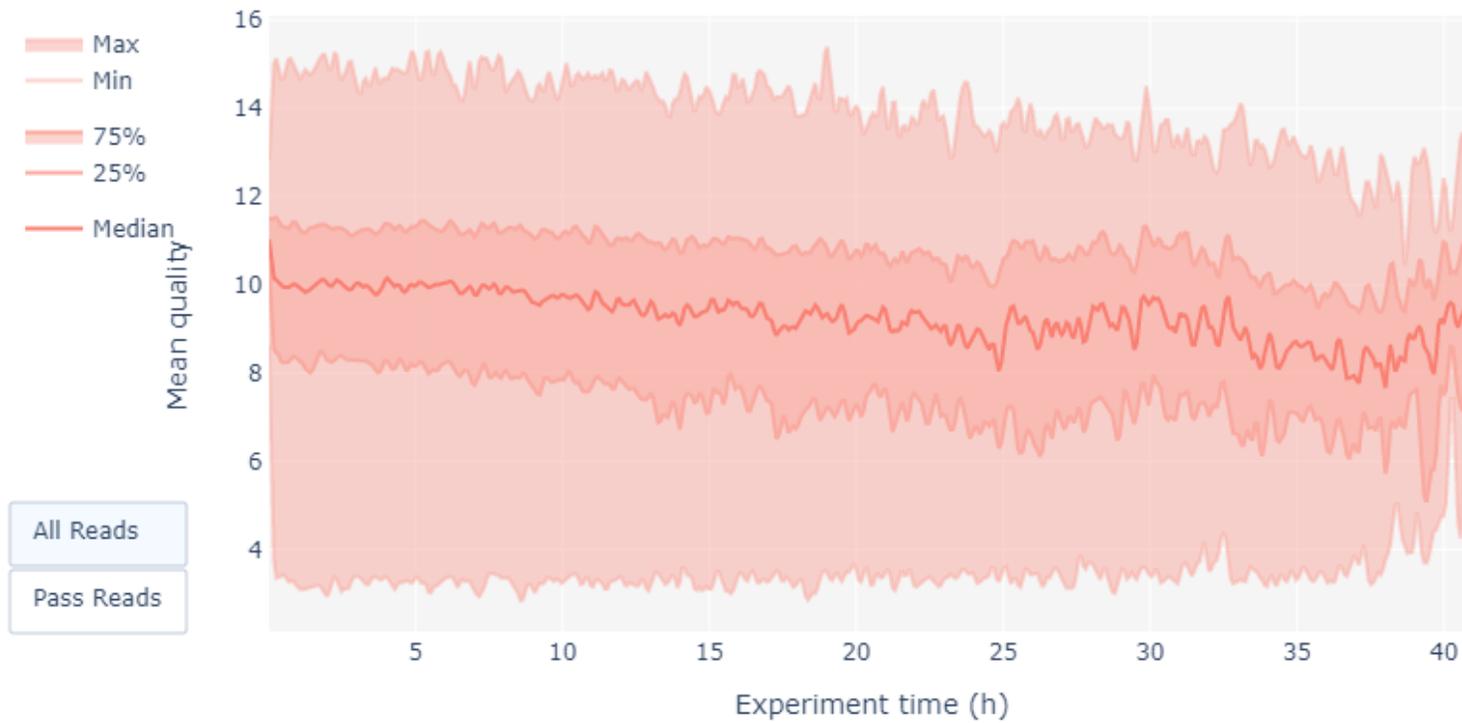
- Proyecto financiado por MinCiencias (contrato No. 004-2020, proyecto No. 6396100270500).

ANEXOS

Anexo 1. valor de mediana calidad de las lecturas PHRED. Eje X puntaje de calidad de la lectura. Eje Y densidad de lecturas (o número de lecturas)

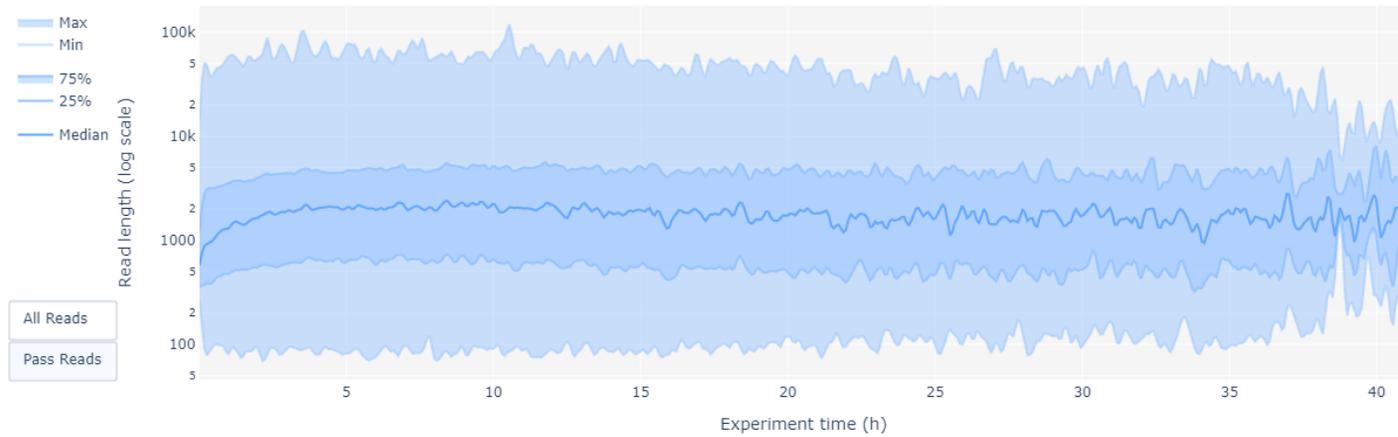


Anexo 2. Calidad promedio de las lecturas secuenciadas por NanoPore. Eje X tiempo de corrido de la secuenciación. Eje Y calidad media de las lecturas (8-14)

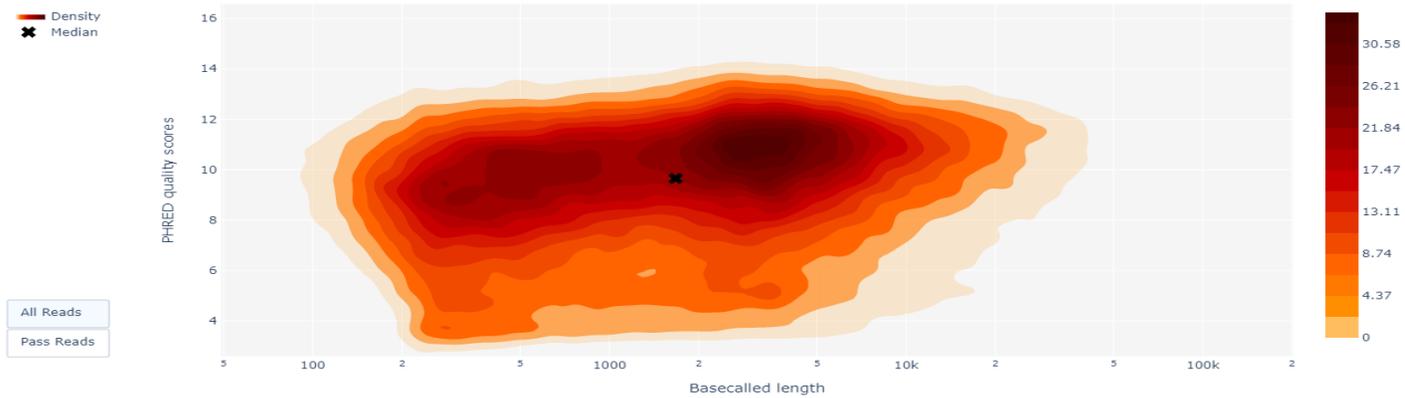


Anexo 3 Graficas adicionales PicoQC |

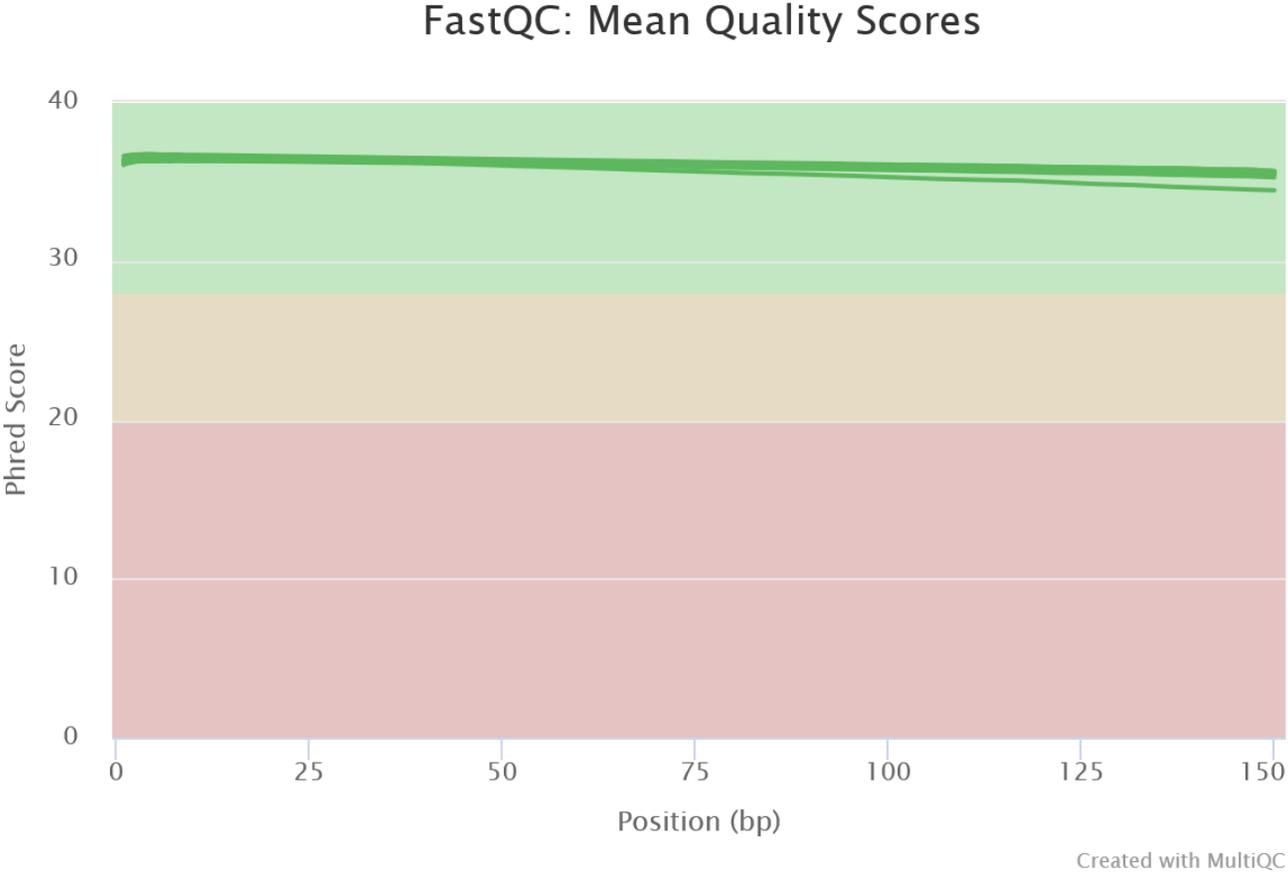
1. Longitud de las lecturas a través del tiempo de secuenciación.



2. Longitud de las lecturas vs calidad de las lecturas.

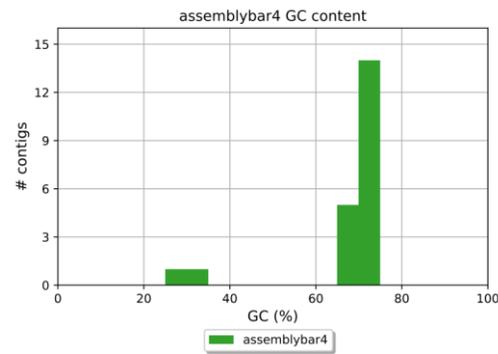


Anexo 4. Calidad media de las lecturas provenientes de illumina y grafica obtenida con fastQC y MultiQC.

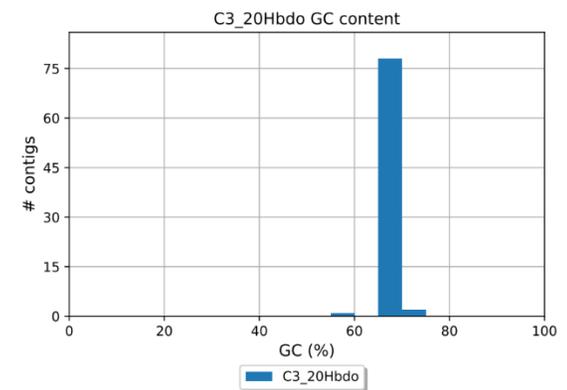
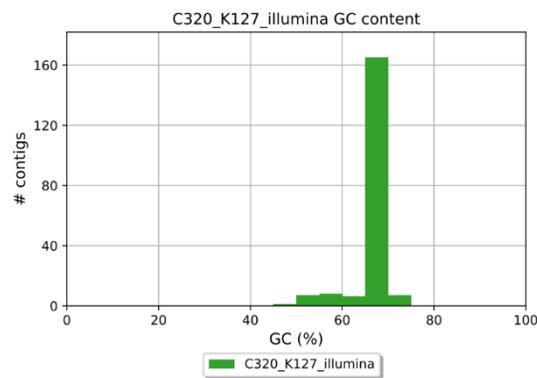
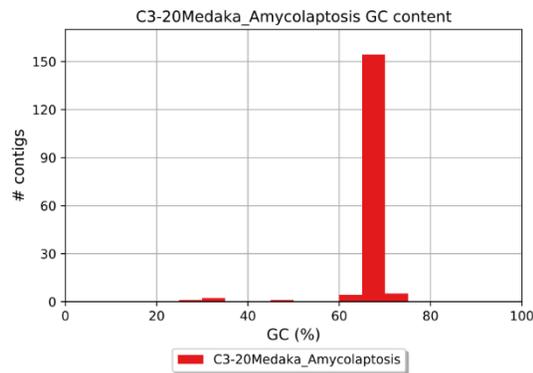


Anexo 5. Graficas GC comparativas con las distintas metodologías de ensamblaje realizadas con Quast.

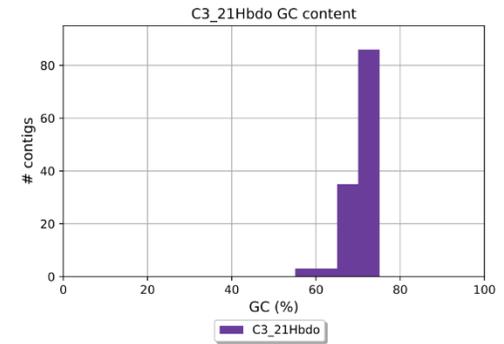
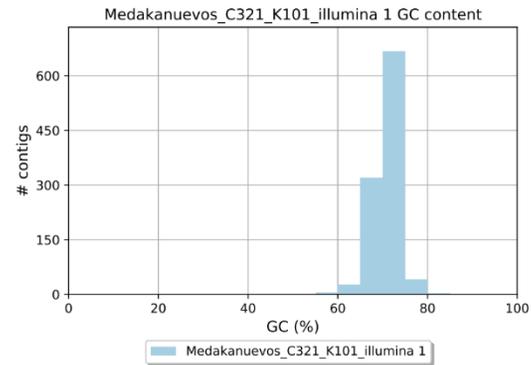
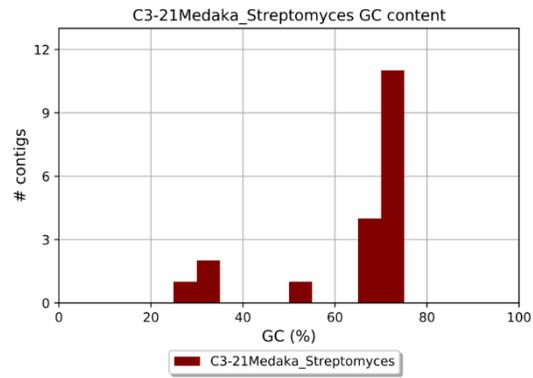
1. Contenido GC en el genoma de C 1.1 *Streptomyces* solo un ensamblaje (ONT).



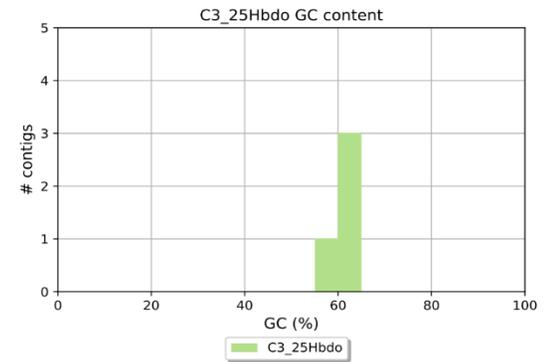
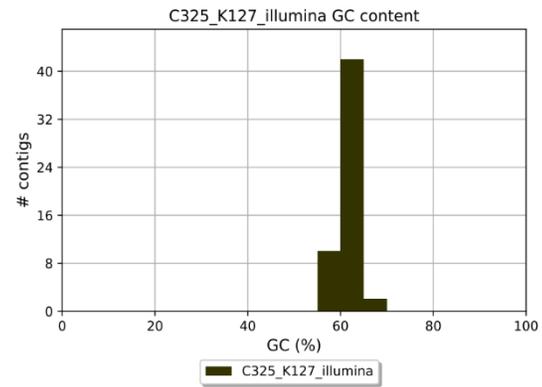
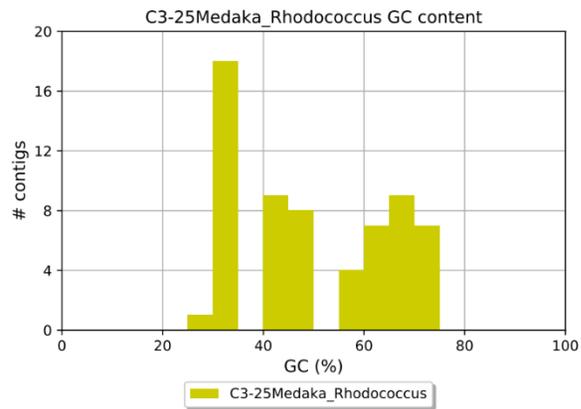
2. Contenido GC en los genomas de C 3.20 *Amycolatopsis* comparación de los tres ensamblajes.



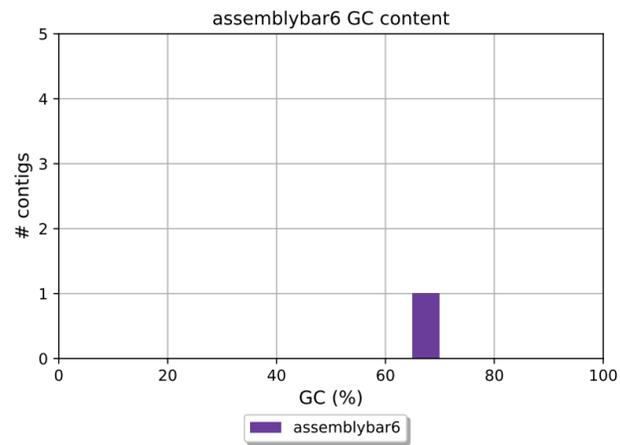
3. Contenido GC en los genomas de C 3.21 *Streptomyces* comparación de los tres ensamblajes.



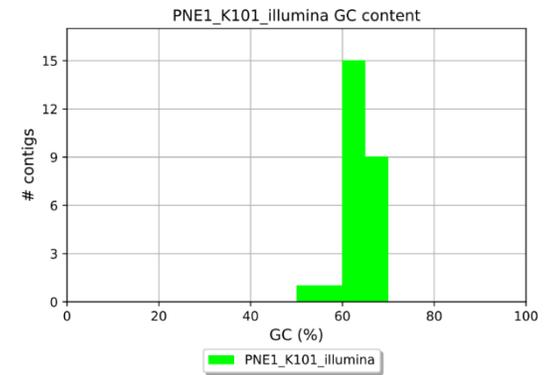
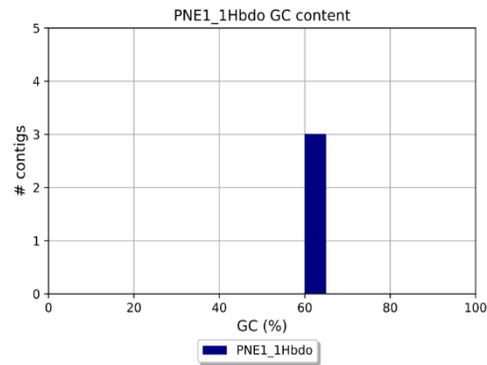
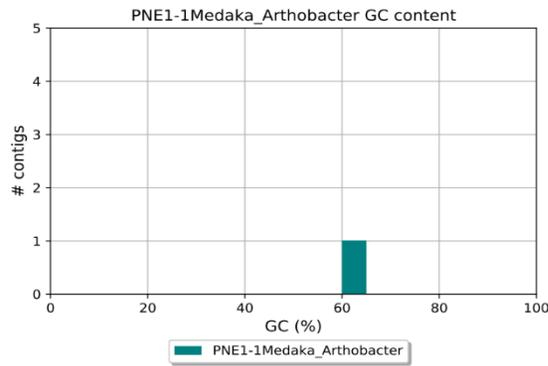
4. Contenido GC en el genoma de C 3.25 *Rhodococcus* comparación de los tres ensamblajes.



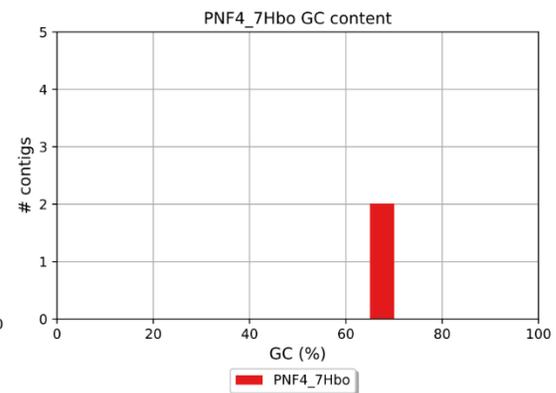
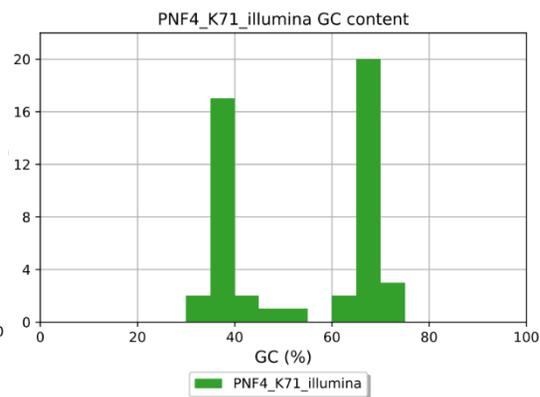
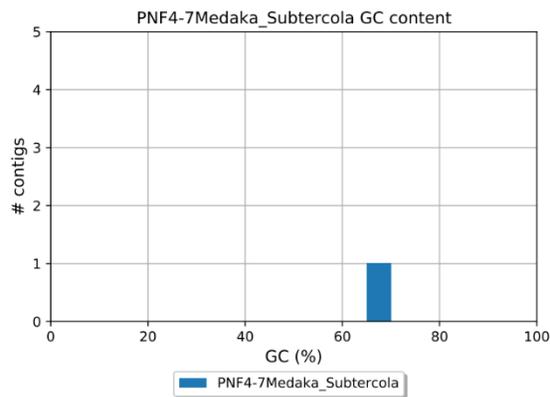
5. Contenido GC en el genoma de PNA 1.2 *Arthrobacter* comparación de los tres ensamblajes.



6. Contenido GC en el genoma de PNE 1.1 *Arthrobacter* comparación de los tres ensamblajes.



7. Contenido GC en el genoma de PNF 4.7 *Subtercola* comparación de los tres ensamblajes.



Anexo 6. BGCs en genomas de Actinobacterias más similares respecto a la base de datos de antiSMASH y compuestos reportados para cada BGC.

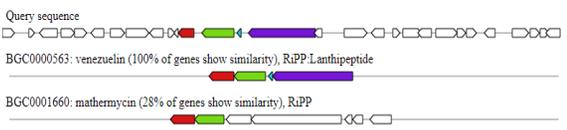
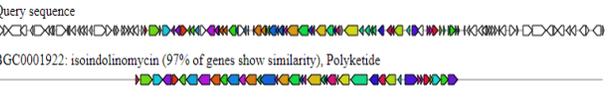
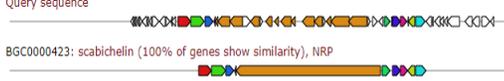
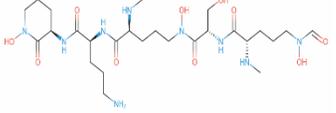
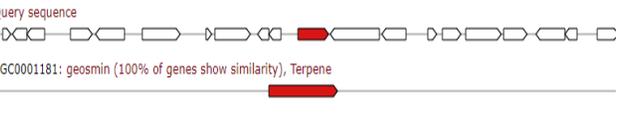
No. De clusters	Genoma	Clúster	Tamaño	% similitud	Nombre BGC más similar	No. De Accesoión MIBiG	Especie BGCs más similar	Compuest o químico
1	C 1.1 <i>Streptomyces</i>	RiPP Lanthipepti do-clase 4	22.400 nt	100%	Venezuelin	BGC000056 3	<i>Streptomyces venezuelae</i> ATCC 10712	Sin estructura
2	C 1.1 <i>Streptomyces</i>	PKS tipo 2	72.489 nt	97%	isoindolino mycin	BGC000192 2	<i>Streptomyces</i> sp.	Sin estructura actual
3	C 1.1 <i>Streptomyces</i>	NRPS	30.020 nt	100%	Scabichelin	BGC000042 3	<i>Streptomyces scabiei</i>	Scabichelin
4	C 1.1 <i>Streptomyces</i>	Terpeno	2.181 nt	100%	Geosmin	BGC000118 1	<i>Streptomyces coelicolor</i>	Geosmina
5	C 1.1 <i>Streptomyces</i>	Ectoina	3.366 nt	100%	Ectoine	BGC000085 3	<i>Streptomyces anulatus</i>	Ectoina

6	C 1.1 <i>Streptomyces</i>	RiPP Lanthipeptide	14.886 nt	85%	informatipeptin	BGC0000518	<i>Streptomyces viridochromogenes</i>	Sin estructura
7	C 3.20 <i>Amycolatopsis</i>	RiPP Clase 3 Lanthipeptide	21.740 nt	100%	Ery-9 / Ery-6 / Ery-8 / Ery-7 / Ery-5 / Ery-4 / Ery-3	BGC0000513	<i>Saccharopolyspora erythraea</i>	7 compuestos lantibioticos
8	C 3.20 <i>Amycolatopsis</i>	NRPS	33.663 nt	70%	Albachelin	BGC0001211	<i>Amycolatopsis alba DSM 44262</i>	Sin estructura actual
9	C 3.20 <i>Amycolatopsis</i>	PKS tipo 2	72.906 nt	67 %	arimetamicin B	BGC0000199	<i>uncultured bacterium</i>	arimetamicin A arimetamicin C arimetamicin B
10	C 3.20 <i>Amycolatopsis</i>	PKS modular tipo 1	53.336 nt	63%	iso-migrastatin migrastatin dorrigin A dorrigin B 13-epi-dorrigin A	BGC0000177	<i>Streptomyces platensis subsp. rosaceus</i>	Isomigrastatin migrastatin dorrigin A dorrigin B 13-epi-dorrigin A
11	C 3.20 <i>Amycolatopsis</i>	NRPS	41.443 nt	56%	amychelin	BGC0000300	<i>Streptomyces sp. AA4</i>	amychelin

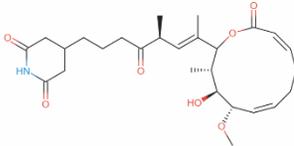
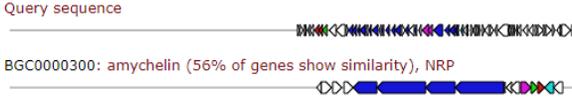
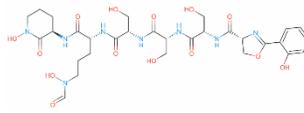
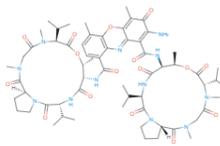
12	C 3.21 <i>Streptomyces</i>	RiPP Clase 3 Lanthipeptide	22,751 nt	100%	AmfS	BGC000049 6	<i>Streptomyces griseus</i>	AmfS
13	C 3.21 <i>Streptomyces</i>	NRPS	67,977 nt	89%	actinomycin D	BGC000029 6	<i>Streptomyces anulatus</i>	actinomycin D
14	C 3.21 <i>Streptomyces</i>	PKS	81,565 nt	63%	Nystatin A	BGC000011 5	<i>Streptomyces noursei</i>	Nystatina A
15	C 3.21 <i>Streptomyces</i>	PKS	41,053 nt	100%	alkylresorcinol	BGC000028 2	<i>Streptomyces griseus</i>	alkylresorcinol
16	C 3.21 <i>Streptomyces</i>	NRPS	63,507 nt	100%	isorenieratene	BGC000066 4	<i>Streptomyces griseus</i>	isorenieratene
17	C 3.21 <i>Streptomyces</i>	PKS	33,307 nt	75%	flaviolin	BGC000090 2	<i>Saccharopolyspora erythraea</i>	flaviolin
18	C 3.25 <i>Rhodococcus</i>	NRPS	22,529 nt	100%	heterobactin A / heterobactin S2	BGC000037 1	<i>Rhodococcus erythropolis PR4</i>	heterobactin A heterobactin S2
19	C 3.25 <i>Rhodococcus</i>	RiPP-like	5,762 nt	75%	branched-chain fatty acids biosynthetic gene	BGC000153 4	<i>Streptomyces filamentosus</i>	branched-chain fatty acids

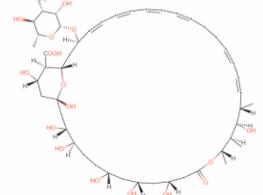
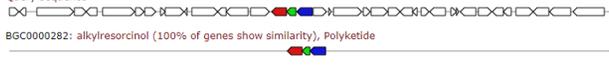
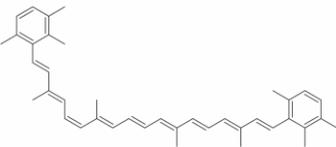
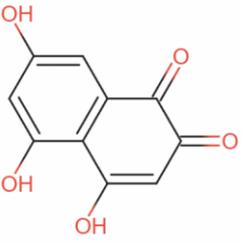
20	C 3.25 <i>Rhodococcus</i>	NRPS	47,529 nt	54%	marformycin A biosynthetic gene cluster	BGC0001214	<i>Streptomyces drozdowiczii</i>	marformycin A a marformycin F
21	C 3.25 <i>Rhodococcus Sp.</i>	NRPS	24,145 nt	57%	erythrochelin	BGC0000349	<i>Saccharopolyspora erythraea</i>	erythrochelin
22	PNA 1.2 <i>Arthrobacter</i>	Siderophore	5,420 nt	100%	desferrioxamine E	BGC0001478	<i>Streptomyces sp. ID38640</i>	desferrioxamine E
23	PNE 1.1 <i>Arthrobacter</i>	Siderophore	5,420 nt	100%	desferrioxamine E	BGC0001478	<i>Streptomyces sp. ID38640</i>	desferrioxamine E
24	PNF 4.7 <i>Subtercola</i>	oligosaccharide	4,391 nt	100%	branched-chain fatty acids gene	BGC0001535	<i>Streptomyces filamentosus</i>	branched-chain fatty acids
25	PNF 4.7 <i>Subtercola</i>	Terpeno	6,312 nt	50%	carotenoide	BGC0000644	<i>Dietzia sp. CQ4</i>	carotenoid

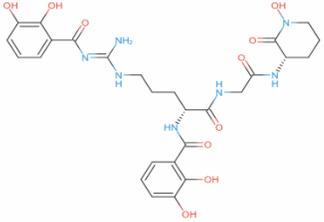
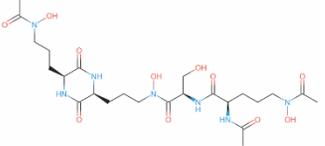
Anexo 7. Compuestos reportados para BGCs encontrados en antiSMASH y MiBIG.

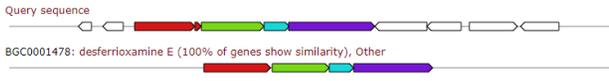
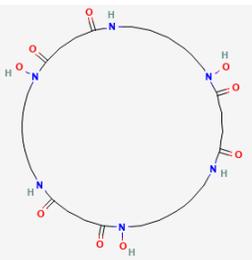
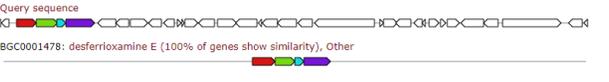
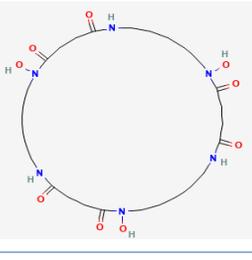
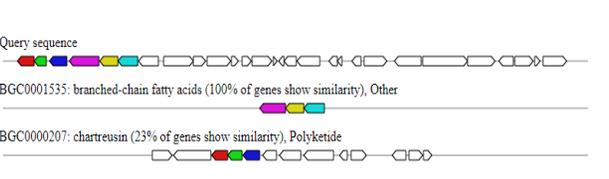
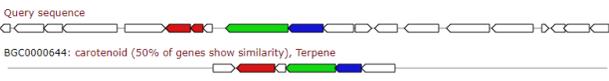
No	Genoma	Clúster	Compuesto químico	Función del compuesto	Comparación de BGCs más similares. AntiSMASH	Estructura química
1	C 1.1 <i>Streptomyces</i> sp.	RiPP Lanthipeptido-clase 4	Venezuelin	Implicado en la regulación de "Lantibiotic synthetases"	<p>Query sequence</p>  <p>BGC0000563: venezuelin (100% of genes show similarity), RiPP Lanthipeptide</p> <p>BGC0001660: mathermycin (28% of genes show similarity), RiPP</p>	No está disponible
2	C 1.1 <i>Streptomyces</i> sp.	PKS tipo 2	Isoindolinomycin	Policetido bioactivo tetra-cíclico que esta involucrado en la producción de Isoindolinone	<p>Query sequence</p>  <p>BGC0001922: isoindolinomycin (97% of genes show similarity), Polyketide</p>	No está disponible
3	C 1.1 <i>Streptomyces</i> sp.	NRPS y Sideróforo	Scabichelin	Su función esta asociada a los sideróforos que secuestran el hierro y proporcionan este nutriente metálico esencial a la bacteria.	<p>Query sequence</p>  <p>BGC0000423: scabichelin (100% of genes show similarity), NRP</p>	
4	C 1.1 <i>Streptomyces</i> sp.	Terpeno	Geosmina	Responsable del olor a tierra húmeda en suelos	<p>Query sequence</p>  <p>BGC0001181: geosmin (100% of genes show similarity), Terpene</p>	

5	C 1.1 <i>Streptomyces sp.</i>	Ectoína	Ectoína	La ectoína es un derivado de aminoácido natural producido por bacterias que viven en condiciones ambientales extremas y actúa como osmorregulador.	<p>Query sequence</p> <p>BGC0000853: ectoine (100% of genes show similarity), Other</p>	
6	C 1.1 <i>Streptomyces sp.</i>	RiPP Lanthipeptid ide Clase 3	informatip eptin	Proteasa que actúa como regulador de la esporulación	<p>Query sequence</p> <p>BGC0000518: informatipeptin (85% of genes show similarity), RiPP:Lanthipeptide</p>	No registra
7	C 3.20 <i>Amycolatopsis sp.</i>	RiPP Clase 3 Lanthipept ide	7 compuestos Antibióticos	Antibiótico de tipo Lantibióticos Erythreapeptin, Avermipeptin y Griseopeptin	<p>Query sequence</p> <p>BGC0000513: Ery-9 / Ery-6 / Ery-8 / Ery-7 / Ery-5 / Ery-4 / Ery-3 (100% of genes show similarity), RiPP:Lanthipeptide</p>	<p>Ery-9 C₁₁₉H₁₇₃N₃₅O₃₅S₂ Ery-6 C₁₀₄H₁₅₆N₃₂O₂₉S₂ Ery-8 C₁₁₀H₁₆₄N₃₄O₃₃S₂ Ery-7 C₁₀₈H₁₆₁N₃₃O₃₂S₂ Ery-5 C₉₉H₁₄₉N₃₁O₂₈S₂ Ery-4 C₉₆H₁₄₆N₃₀O₂₇S₂ Ery-3 C₉₀H₁₃₉N₂₇O₂₅S₂</p>
8	C 3.20 <i>Amycolatopsis sp.</i>	NRPS y Sideróforo	Albachelin	Su función está asociada a los sideróforos que sequestran el hierro y proporcionan este nutriente metálico esencial a la bacteria	<p>Query sequence</p> <p>BGC0001211: albachelin (70% of genes show similarity), NRP</p>	No disponible
9	C 3.20 <i>Amycolatopsis sp.</i>	PKS tipo 2	arimetamycin A arimetamycin C arimetamycin B	Antibióticos, pertenecientes a la clase de compuestos orgánicos conocidos como antraciclina	<p>Query sequence</p> <p>BGC0000199: arimetamycin B / arimetamycin C / arimetamycin A (67% of genes show similarity), Polyketide:Type II + Sac</p>	<p>Arimetamycin B</p> <p>arimetamycin C C₂₈H₃₀O₁₂ arimetamycin A</p>

						$C_{39}H_{52}N_2O_{13}$
10	C 3.20 <i>Amycolatopsis sp.</i>	PKS modular tipo 1	Isomigrastatin migrastatin dorrigocin A dorrigocin B 13-epi-dorrigocin A	Antibióticos dentro de los Macrólidos. La isomigrastatina son potentes inhibidores de la migración de células tumorales.	<p>Query sequence</p>  <p>BGC0000177: iso-migrastatin / migrastatin / dorrigocin A / dorrigocin B / 13-epi-dorrigocin A (63% of genes show similarity);</p>	<p>Iso-migrastatin</p>  <p>Migrastatin = $C_{27}H_{39}N_1O_7$ dorrigocin A = $C_{27}H_{41}N_1O_8$ dorrigocin B = $C_{27}H_{41}N_1O_8$ 13-epi-dorrigocin A = $C_{27}H_{41}N_1O_8$</p>
11	C 3.20 <i>Amycolatopsis sp.</i>	NRPS y Sideróforo	amychelin	Su función esta asociada a los sideróforos que secuestran el hierro y proporcionan este nutriente metálico esencial a la bacteria	<p>Query sequence</p>  <p>BGC0000300: amychelin (56% of genes show similarity), NRP</p>	
12	C 3.21 <i>Streptomyces</i>	RiPP-like	AmfS	El grupo de genes amf se identificó previamente como un regulador del inicio de la formación de micelio aéreo en <i>Streptomyces griseus</i> .	<p>Query sequence</p>  <p>BGC0000496: AmfS (100% of genes show similarity), RiPP:Lanthipeptide</p>	No registra
13	C 3.21 <i>Streptomyces</i>	NRPS	actinomycin D	Actinomycin D – Antibiótico, antineoplásico y citotóxico	<p>Query sequence</p>  <p>NZ_CP026652 (8339216-8408852): Streptomyces dengpaensis strain XZHG99 chromos... (38% of genes show similarity), N</p>	

14	C 3.21 <i>Streptomyces</i>	PKS	Nistatina	El antifúngico polieno macrólido nistatina producido por <i>Streptomyces noursei</i> ATCC 11455 es un importante agente antifúngico.	<p>Query sequence</p>  <p>NZ_JT1Y01000001 (6423575-6583345): <i>Streptomyces</i> sp. Ach 505 scaffold00001, wh... (37% of genes show similarity), T1f</p>	
15	C 3.21 <i>Streptomyces</i>	PKS	alkylresorcinol	Función desconocida	<p>Query sequence</p>  <p>BGC0000282: alkylresorcinol (100% of genes show similarity), Polyketide</p>	No registra
16	C 3.21 <i>Streptomyces</i>	NRPS	isorenieratene	Carotenoide	<p>Query sequence</p>  <p>BGC0000664: isorenieratene (85% of genes show similarity), Terpene</p>	
17	C 3.21 <i>Streptomyces</i>	PKS	flaviolin	Flaviolin función asociada a la biosíntesis de melanina	<p>Query sequence</p>  <p>NZ_JT1Y01000001 (6683460-6724600): <i>Streptomyces</i> sp. Ach 505 scaffold00001, wh... (78% of genes show similarity), T3f</p>	
18	C 3.25 <i>Rhodococcus</i>	NRPS	heterobactin A / heterobactin S2	Su función esta asociada a los sideróforos que secuestran el hierro y proporcionan este nutriente metálico esencial a la bacteria	<p>Query sequence</p>  <p>BGC0000371: heterobactin A / heterobactin S2 (100% of genes show similarity), NRP</p>	heterobactin A

						<p>heterobactin S2= C28H35N7O12S1</p> 
19	C 3.25 <i>Rhodococcus</i>	RiPP-like	branched-chain fatty acids	Compuesto involucrado en la biosíntesis de ácidos grasos que regulan la La daptomicina que es un antibiótico lipopeptídico cíclico	<p>Query sequence</p>  <p>BGC0001534: branched-chain fatty acids (75% of genes show similarity), Other</p> 	No disponible
20	C 3.25 <i>Rhodococcus</i>	NRPS	marformycin in A marformycin in B marformycin in C marformycin in D marformycin in E marformycin in F	actividad antiinfecciosa contra otras bacterias	<p>Query sequence</p>  <p>BGC0001214: marformycin A / marformycin B / marformycin C / marformycin D / marformycin E / marformycin F (54% of .</p> 	<p>marformycin A= C₄₃H₆₈N₈O₁₀ marformycin B= C₄₄H₇₀N₈O₁₀ marformycin C = C₄₃H₆₈N₈O₁₁ marformycin D= C₄₄H₇₀N₈O₁₁ marformycin E= C₄₂H₆₆N₈O₁₁ marformycin F= C₄₂H₆₆N₈O₁₀</p>
21	C 3.25 <i>Rhodococcus</i>	NRPS	erythrochelin	Su función esta asociada a los sideróforos que secuestran el hierro y proporcionan este nutriente metálico esencial a la bacteria	<p>Query sequence</p>  <p>BGC0000349: erythrochelin (57% of genes show similarity), NRP</p> 	

22	PNA 1.2 <i>Arthrobacter</i>	Siderophore	desferrioxamine E	La desferrioxamina E es un sideróforo del ácido hidroxámico cíclico producido por varias especies bacterianas que exhibe actividad antitumoral	<p>Query sequence</p>  <p>BGC0001478: desferrioxamine E (100% of genes show similarity), Other</p>	
23	PNE 1.1 <i>Arthrobacter</i>	Siderophore	desferrioxamine E	La desferrioxamina E es un sideróforo del ácido hidroxámico cíclico producido por varias especies bacterianas que exhibe actividad antitumoral	<p>Query sequence</p>  <p>BGC0001478: desferrioxamine E (100% of genes show similarity), Other</p>	
24	PNF 4.7 <i>Subtercola</i>	Oligosaccharide	"branched-chain fatty acids" Ácidos grasos de cadena ramificada	Involucrado en la biosíntesis de daptomicina que es un antibiótico lipopeptídico cíclico.	<p>Query sequence</p>  <p>BGC0001535: branched-chain fatty acids (100% of genes show similarity), Other</p> <p>BGC0000207: chartreusin (23% of genes show similarity), Polyketide</p>	No registra
25	PNF 4.7 <i>Subtercola</i>	Terpeno	Carotenoid e	Asociado a la coloración de la bacteria. también pueden ser antioxidantes	<p>Query sequence</p>  <p>BGC0000644: carotenoid (50% of genes show similarity), Terpene</p>	

Resultados obtenidos de AntiSMASH, NPAtlas, MIBiG Minimum Information about a Biosynthetic Gene cluster y PUB med. gov

REFERENCIAS BIBLIOGRÁFICAS

1. Alcock B, Raphenya A, Lau T, Tsang K, Bouchard M, Edalatmand A, *et al.*, (2020). CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 8:517-525.
2. Altschul S.F, Gish W, Miller W, Myers E. y Lipman D. (1990) Basic local alignment search tool. *J Mol Biol.* 215: 403-410.
3. Andrews S. (2010).FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2015), "FastQC," <https://qubeshub.org/resources/fastqc>
4. Araos R, García P, Chanqueo L y Labarca J. (2012). Daptomicina: características farmacológicas y aporte en el tratamiento de infecciones por cóceas gram positivas. *R chilena de infectología.* 29: 127-131.
5. Bankevich A, Nurk S, Antipov D, *et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.*19:455-477.
6. Barka E, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Klenk H-P, Clément C, Ouhdouch Y,van Wezel G. (2016)Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbio. Mol. Biol. Reviews.* 80:1-43.
7. Berglund B. (2015). Environmental dissemination of antibiotic resistance genes and correlation to anthropogenic contamination with antibiotics. *Infect. ecology & epidemiol.* 5: 28563-28564.
8. Blin K, Kim HU, Medema MH, Weber T. (2019). Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Brief Bioinform.* 19: 1103-1113.
9. Blin K, Shaw S, Kloosterman A, Charlop-Powers Z, van Weezel G, Medema MH y Weber T. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49:29-35.
10. Bolger A, Lohse M y Usadel B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics,* btu170.

11. Bolla J, Alibert-Franco S, Handzlik J, Chevalier J, Mahamoud A, Boyer G, Pagès J, *et al.* (2011). Strategies for bypassing the membrane barrier in multidrug resistant Gram-negative bacteria. *FEBS letters*. 585: 1682-1690.
12. Charousová I, Medo J, Halenárová E, Javoreková S. (2017). Antimicrobial and enzymatic activity of actinomycetes isolated from soils of coastal islands. *J Adv Pharm Technol Res*. 8:46-51.
13. Chen R, Wong H, Burns BP. (2019). New Approaches to Detect Biosynthetic Gene Clusters in the Environment. *Medicines*. 6: 9-11.
14. Culp E, Yim G, Waglechner N, Wang W, Pawlowski A, Wright G. (2019) Hidden antibiotics in actinomycetes can be identified by inactivation of gene clusters for common antibiotics. *Nat. Biotechnol.* 37: 1149–1154.
15. Dávila J, Hoskisson P, Paterlini P, Romero C, Álvarez A. (2020). Whole genome sequence of the multi-resistant plant growth-promoting bacteria *Streptomyces* sp. Z38 with potential application in agroindustry and bio-nanotechnology. *Genomics*. 112: 4684–4689.
16. De Carvalho C, Costa S, Fernandes P, Couto I y Viveiros M. (2014). Membrane transport systems and the biodegradation potential and pathogenicity of genus *Rhodococcus*. *Front. in physiology*. 5: 133.
17. De Coster W, D’Hert S, Schultz D.T, Cruts M y Van Broeckhoven C. (2018). NanoPack: visualizing and processing long-read sequencing data, *Bioinformatics*. 34: 2666–2669.
18. Delahaye C, Nicolas J. (2021). Sequencing DNA with nanopores: Troubles and biases. *PLoS ONE*. 16: 257521.
19. Delcour A. (2009). Outer membrane permeability and antibiotic resistance. *Biochimica et biophysica acta*. 1794: 808-16.
20. Doroghazi J, Metcalf W. (2013). Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics*. 14: 609-611.
21. Ewels P, Magnusson M, Lundin S, Käller M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report, *Bioinformatics*. 32: 3047–3048.

22. Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*. 39: 783-791.
23. Gozari M, Zaheri A, Jahromi S., Gozari M, Karimzadeh R. (2019). Screening and characterization of marine actinomycetes from the northern Oman Sea sediments for cytotoxic and antimicrobial activity. *Int. J. Microbiol.* 22: 521-530.
24. Gomez-Escribano J, Alt S, Bibb M. (2016). Next Generation Sequencing of Actinobacteria for the Discovery of Novel Natural Products. *Mar Drugs*. 14:78.
25. Gurevich A, Saveliev V, Vyahhi N y Tesler G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics*. 29: 1072–1075.
26. Ha L, Vanlerberghe L, Toan T, Dewettinck K y Messens K. (2015). Comparative evaluation of six extraction methods for DNA quantification and PCR detection in cocoa and cocoa-derived products. *Food Biotechnol.* 29: 1-19.
27. Hamed J, Poorinmohammad N. (2017). The cellular structure of Actinobacteria. En: Wink J, Mohammadipanah F, Hamed J. *Biology and biotechnology of Actinobacteria*. Springer. 5-28.
28. Heng L. (2018). Minimap2: pairwise alignment for nucleotide sequences, *Bioinformatics*. 34: 3094–3100.
29. Jiang J, He X, Cane D. (2007). Biosynthesis of the earthy odorant geosmin by a bifunctional *Streptomyces coelicolor* enzyme. *Nat Chem Biol.* 3: 711-5.
30. Jiang L, Peng Y, Seo J, Jeon D, Jo M, Lee J, Lee J, *et al.* (2022). *Subtercola endophyticus* sp. nov., a cold-adapted bacterium isolated from *Abies koreana*. Pre-print. Acceso: 21-04-22.
31. Jiang X, Hashim M, Charusanti P, Munck C, Blin K, Tong Y, Weber T, Sommer M, Lee Y. (2017). Dissemination of antibiotic resistance genes from antibiotic producers to pathogens. *Nat. commu.* 8: 15784.
32. Kang H y Brady S. (2013). Arimetamycin A: improving clinically relevant families of natural products through sequence-guided screening of soil metagenomes. *Angewandte Chemie International Edition*, 52: 11063-11067.

33. Kimura M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120.
34. Kolmogorov M, Yuan J, Lin Y y Pevzner P.(2019). Assembly of Long Error-Prone Reads Using Repeat Graphs. *Nature Biotechnology.* 37: 540-546.
35. Kumari S, Stecher G, Li M, Knyaz C y Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* 35: 1547-1549.
36. Kumari R, Singh P, Lal R. Genetics and Genomics of the Genus *Amycolatopsis*. (2016). *Indian J Microbiol.* 56: 233-46.
37. Leger A. y Leonardi T. (2019). pycoQC, interactive quality control for Oxford Nanopore Sequencing. *J. Open Source Softw.* 4: 1236-12340.
38. Mao D, Okada B, Wu Y, Xu F y Seyedsayamdost M. (2018). Recent advances in activating silent biosynthetic gene clusters in bacteria. *Current opinion in microbiology*, 45, 156–163.
39. Mao D, Okada B. K, Wu Y, Xu F y Seyedsayamdost M. R. (2018). Recent advances in activating silent biosynthetic gene clusters in bacteria. *Curr. Opin. Microbiol.* 45:156–163.
40. medaka: Sequence correction provided by ONT Research. <https://github.com/nanoporetech/medaka>, Accessed 24 Apl 2022. [Google Scholar]
41. Miller W, Munita J y Arias C. (2014). Mechanisms of antibiotic resistance in enterococci. *Expert Rev Anti Infect Ther.* 12: 1221-1236.
42. Minoru K y Susumu G. (2000). Kegg: Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids Res.* 28: 27–30.
43. Mungan M, Alanjary M, Blin K, Weber T, Medema MH, Ziemert N. (2020). ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. *Nucleic Acids Res.* 48:546-552.

44. Munaganti R, Muvva V, Konda S, Naragani K, Mangamuri U, Dorigondla K, Akkewar D. (2016). Antimicrobial profile of *Arthrobacter kerguelensis* VL-RK_09 isolated from Mango orchards. *Braz J Microbiol.* 47:1030-1038.
45. Nakatsu C, Barabote, R., Thompson, S. *et al.* (2013) Complete genome sequence of *Arthrobacter* sp. strain FB24. *Stand in Genomic Sci.* 9: 106–116
46. Nindita Y, Cao Z, Fauzi A, Teshima A, Misaki Y, Muslimin R, *et al.* (2019). The genome sequence of *Streptomyces rochei*, which carries a linear chromosome and three characteristic linear plasmids. *Scien. reports.* 9: 1-11.
47. Nouioui I, Carro L, García-López M, Meier-Kolthoff JP, Woyke T, Kyrpides NC, Pukall R, Klenk H-P, Goodfellow M and Göker M. (2018). Genome-Based Taxonomic classification of the Fílum Actinobacteria. *Front. Microbiol.* 9:2007.
48. Orro A, Cappelletti M, D'Ursi P, Milanese L, Di Canito A, Zampolli J, *et al.* (2015) Genome and Phenotype Microarray Analyses of *Rhodococcus* sp. BCP1 and *Rhodococcus opacus* R7: Genetic Determinants and Metabolic Abilities with Environmental Relevance. *PLoS One.* 10: e0139467.
49. Parks D, Imelfort M, Skennerton C, Hugenholtz P y Tyson G. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research,* 25: 1043-1055.
50. Poorinmohammad N, Bagheban-Shemirani R y Hamed J. (2019). Genome mining for ribosomally synthesised and post-translationally modified peptides (RiPPs) reveals undiscovered bioactive potentials of actinobacteria. *Antonie van Leeuwenhoek.* 112: 1477–1499.
51. Rajwani R, Ohlemacher S, Zhao G, Liu H y Bewley C. (2021). Genome-Guided Discovery of Natural Products through Multiplexed Low-Coverage Whole-Genome Sequencing of Soil Actinomycetes on Oxford Nanopore Flongle. *Msystems.* 6: 1020-21.
52. Reen FJ, Romano S, Dobson AD, O'Gara F. (2015). The Sound of Silence: Activating Silent Biosynthetic Gene Clusters in Marine Microorganisms. *Mar Drugs.* 13: 4754-4783.
53. Reygaert W.(2018). An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS microbiology,* 4: 482–501.

54. Richter M, Rosselló-Móra R, Glöckner F, Peplies J. (2016). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics*. 32: 929–931.
55. Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. (2018) The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res*. 46:282-288.
56. Seemann T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 30:2068-2069.
57. Sierra A. (2018). Análisis de la diversidad y el potencial bioactivo de bacterias asociadas a líquenes de páramo en el parque nacional natural Chingaza (Tesis de pregrado no publicada) Universidad industrial de Santander.
58. Sierra M, Danko D, Sandoval T, Pishchany G, Moncada B, Kolter R y Zambrano M M. (2020). The microbiomes of seven lichen genera reveal host specificity, a reduced core community and potential as source of antimicrobials. *Front. Microbiol*. 11: 398.
59. Sriparoj P, Suwanborirux K y Tanasupawat S. (2013). Characterization and antimicrobial activity of *Amycolatopsis* strains isolated from Thai soils. *J. of Applied Pharmaceutical Science*, 3: 011-016.
60. Somerville W, Thibert L, Schwartzman K, Behr M. (1997). Extraction of *Mycobacterium tuberculosis* DNA: a question of Containment. *J. Clin. Microbiol*. 43: 2996-2997.
61. Vargas M-N, Triana J, Diaz-Puentes N, Gómez V, Romero C, Villabona N, Cruz M, Corrales A, Chaib M, Del Portillo P, Zambrano M. (2022). Potencial funcional y genómico en Actinobacterias de ecosistemas tropicales. Poster presentado en: VI congreso colombiano de biología computacional y bioinformática: 28 de marzo al 1 abril de 2022. Cartagena. Colombia.
62. Vaser R, Sović I, Nagarajan N y Šikić M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome research*, 27: 737–746.
63. Villalobos A, Wiese J, Imhoff J, Dorador C, Keller A, Hentschel U. (2019). Systematic Affiliation and Genome Analysis of *Subtercola vilae* DB165T with Particular Emphasis on Cold Adaptation of an Isolate from a High-Altitude Cold Volcano Lake. *Microorganisms*. 23: 7-107.

64. Walker B, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, *et al.* (2014) Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS ONE 9: e112963.
65. Webber M y Piddock J. (2003). La importancia de las bombas de expulsión en la resistencia bacteriana a los antibióticos. Revista de quimioterapia antimicrobiana. 51: 9-11.
66. Wick R, Judd L, Gorrie C, Holt K. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 13: e1005595.
67. Wick R, Judd L. y Holt E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 20: 1-10.
68. Woo P, Lau S, Huang Y, Yuen K-Y. (2005). Genomic evidence for antibiotic resistance genes of actinomycetes as origins of antibiotic resistance genes in pathogenic bacteria simply because actinomycetes are more ancestral than pathogenic bacteria. Med Hyp. 67: 1297-1304.
69. Yu Y, Gutierrez E, Kovacevic Z, Saletta F, Obeidy P, Suryo Rahmanto Y, Richardson D. (2012). Iron chelators for the treatment of cancer. Curr Med Chem. 19: 2689-702.
70. Yamanaka K, Oikawa H, Ogawa HO, Hosono K, Shinmachi F, Takano H, Sakuda S, Beppu T, Ueda K. (2005). Desferrioxamine E produced by *Streptomyces griseus* stimulates growth and development of *Streptomyces tanashiensis*. Microbiology (Reading). 151: 2899-2905.
71. Yong-pil K, Hiroshi T, Kousuke I, Takashi F, Atsuko M, Yoko T y Satoshi O.(2003). Akanawaenes, Novel antifungal antibiotics produced by *Streptomyces sp.* J. Antibiot. 56:5218.
72. Zhou X, Huang H, Li J, Song Y, Jiang R, Liu J, Ju J, *et al.* (2014). New anti-infective cycloheptadepsipeptide congeners and absolute stereochemistry from the deep sea-derived *Streptomyces drozdowiczii*. Tetrahedron. 70: 7795-7801.