



UNIVERSIDAD DE CHILE -FACULTAD DE CIENCIAS -ESCUELA DE PREGRADO

**“Estudio comparativo de genes que codifican enzimas
Polihidroxicanoato sintasas (PhaCs) en cepas antárticas del
género *Pseudomonas*”**

Seminario de Título entregado a la Universidad de Chile en cumplimiento parcial de los requisitos para optar al Título de Ingeniera en Biotecnología Molecular.

Amelia Eugenia Cox Fernandois

Director del Seminario de Título:

Dr. Andrés Marcoleta Caldera

Co-Director del Seminario de Título:

Dr. Ignacio Poblete

Septiembre 2023 Santiago - Chile



INFORME DE APROBACIÓN SEMINARIO DE TÍTULO

Se informa a la Escuela de Pregrado de la Facultad de Ciencias, de la Universidad de Chile que el Seminario de Título, presentado por la Srta. Amelia Eugenia Cox Fermandois.

“Estudio comparativo de genes que codifican enzimas Polihidroxiálcanoato sintasas (PhaCs) en cepas antárticas del género *Pseudomonas*”

Ha sido aprobado por la Comisión de Evaluación, en cumplimiento parcial de los requisitos para optar al Título de Ingeniera en Biotecnología Molecular.

Dr. Andrés Marcoleta Caldera:
Director Seminario de Título

Comisión Revisora y Evaluadora

Dra. Jennifer Alcaino:
Presidenta Comisión

Dr. Mario Tello:
Evaluador

Santiago de Chile, 20 de Septiembre 2023

BIOGRAFÍA

El 20 de marzo de 1997 mi mamá daba a luz a una niña que le preguntaría el por qué de todo por un largo tiempo. Hasta el año no hablaba ni hacía mucho, mi papá y mi mamá me llevaron al doctor pensando que quizás “tenía algo” como les gusta decir. Nunca pensaron que me transformaría en una persona con sobre dotación de mitocondrias, que hace un millón de cosas ni menos que me dedicaría al atletismo hasta hoy.

Viví 12 años en lo que un santiaguino probablemente llamaría provincia y/o pueblo, primero en San Esteban, después en San Felipe. Esos pueblos que me criaron con su naturaleza cautivante despertaron en mí la curiosidad. Jugaba a explorar, me escapaba con mi perro a un campo de parrones junto a mi casa, trepaba árboles y me creía Indiana Jones. Mi mamá nos avisaba cuando había eclipses y los veíamos desde el jardín. Tenía una bitácora donde escribía fragmentos del Larousse ilustrado y copiaba dibujos de plantas y animales que me llamaban la atención, no teníamos internet evidentemente. Cumpleaños número 9 y le pedí a mi papá un microscopio de juguete y en el número 10 un telescopio. A los 9 años, mi mamá me prestó un libro de Einstein formato cómic que explicaba la relatividad, recuerdo haber ido al colegio y preguntado al Mr. Marco de ciencias qué era el espacio-tiempo y me explicó tratándome como a una par. A los 10 años, mi tata Arturo me regaló un libro de James Lovelock que hablaba sobre Gaia y energía nuclear. Cuando veníamos a Santiago con mis hermanas a ver a mi papá, nos llevaba a museos, mi favorito: el MIM y el de Historia Natural, eran alucinantes, aún recuerdo la explicación de la repartición de la masa en la cama de clavos; también le gustaba partir sin rumbo con nosotras y decir: “¡Vamos a descubrir!”. Él me enseñó a observar la naturaleza.

Hoy, después de 26 órbitas completas al Sol, miro hacia atrás y me doy cuenta de que estos años en la Facultad de Ciencias en la Universidad de Chile fueron vivir el sueño que siempre tuve de niña: ser científica. Gracias familia y Gaia por inspirarme.

DEDICATORIA

A mis profesores de ciencias del colegio que me marcaron profundamente, los/as recuerdo a cada uno/a: Miss Claudia, Marco Orellana, Loreto Labbé, Soledad Rivas y Sebastián Figueroa.

Al LABVA, el futuro será biomaterial.

A Alexandra Elbakyan fundadora de Sci-Hub y a todas las mujeres que han pavimentado el camino para que podamos hacer ciencia.

A mis amigos, amigas y amigas, son de lo más lindo que tengo.

Y, con profundo amor, a mi familia, a toda, son un entramado de personas inspiradoras.

“Al ingresar al estrecho de Magallanes comenzó la fantasía.”

-Alice Guy (1873 - 1968) autora de la primera película de ficción

AGRADECIMIENTOS

En un país donde la educación de calidad es un privilegio, se hace menester sección como esta agradecer el acceso que tuve a ella y a quienes lo hicieron posible: mi familia. Agradezco también la responsabilidad que se desprende de lo mismo: trabajar por un país más justo. En cada paso pienso en la vocación que ha movido a mi familia en el anhelo de construir una sociedad mejor, y es ahí donde he encontrado la motivación cuando el escenario es complejo. En ese sentido, les agradezco por heredarme esa vocación, por enseñarme que mayores privilegios significan mayores responsabilidades y por inspirarme continuamente a ser mejor, pero sobre todo les agradezco por el amor que me han dado y porque nunca me ha faltado nada.

En el arduo camino de titularse, más aún con la pandemia, debo agradecer a mis profesores guías. Especialmente, a Andrés Marcoleta, por aventurarse a guiarme en un tema nuevo para el laboratorio y también, aquí sumo al profesor Francisco Chávez, por su apoyo sempiterno a mis proyectos y postulaciones que tantas puertas me han abierto. A quienes forman parte del BEM por su apoyo en todos los sentidos posibles, sobre todo a Camilo Berríos, Carlos Serrano y Marcelo Veloso, mis maestros Jedi, sin ustedes nada de esto hubiera sido posible.

Doy paso a un popurrí de agradecimientos necesarios: a la selección de atletismo de la U por ser mi familia todos estos años. Al equipo del EAC, por el apoyo mutuo. Al LABVA por enseñarme de biomateriales. A la XII edición del Programa para el Fortalecimiento de la Función Pública de la Fundación Botín, por inspirarme a cambiar el mundo desde mi vereda. Al equipo de CienciaF5 por creer. A José Ignacio Bernales y Claudio Moreira, amigos maravillosos. A Catalina Jara, Pedro Salinas, Diego Ticchione y Francisca Cymbron por hacer de mi paso por la universidad uno memorable. A todas/os/es mis amigas/os/ues por estar siempre. A la wela-wela por ser mi hogar estos últimos años. Finalmente, a Agustín, por aparecer en mi vida como una serendipia y darme el impulso final, como buen físico, que necesitaba para terminar este trabajo.

ÍNDICE DE CONTENIDOS

1. INTRODUCCIÓN.....	5
1.1 Chile: un país diverso, desconocido y con potencial biotecnológico.....	5
1.1.1 Generalidades geográficas de Chile.....	5
1.1.2 Antártica, territorio nacional poco explorado y con alto potencial biotecnológico.....	5
1.2 Género <i>Pseudomonas</i>: características generales y dificultades en su clasificación.....	7
1.2.1 Características generales del género <i>Pseudomonas</i>	7
1.2.2 Dificultades en la clasificación del género <i>Pseudomonas</i>	7
1.3 Polihidroxicarboxilatos: fuente de carbono bacteriana con potencial biotecnológico.....	9
1.3.1 Características generales de los polihidroxicarboxilatos.....	9
1.3.2 Propiedades y aplicaciones de los PHAs.....	9
1.3.3 Tipos de PHAs y rutas de síntesis.....	10
1.3.4 Microorganismos productores de PHA.....	14
1.3.5 MPC5 y MPC6: <i>Pseudomonas</i> antárticas con potencial biotecnológico.....	14
1.4 Objetivos.....	16
2. MATERIALES Y MÉTODOS.....	18
2.1 Construcción de una base de datos genómica de <i>Pseudomonas</i>	18
2.2 Cálculo de identidad promedio de nucleótidos entre genomas (ANI).....	19
2.3 Búsqueda y clasificación de genes codificantes para PhaCs.....	19
2.4 Extracción de las PhaCs identificadas.....	21
2.5 Agrupamiento de las PhaCs extraídas.....	21
2.6 Análisis de dominios conservados.....	22
2.7 Alineamiento múltiple de las PhaCs extraídas.....	22
2.8 Construcción de árbol filogenético de las PhaCs extraídas.....	23
2.9 Construcción de árbol filogenético de <i>Pseudomonas</i>	24
2.10 Comparación entre la presencia de PhaCs y las relaciones filogenéticas en las cepas de <i>Pseudomonas</i> analizadas.....	24
3. RESULTADOS.....	25
3.1 Construcción de una base de datos genómica de <i>Pseudomonas</i> provenientes de la Antártica y otras regiones geográficas.....	25
3.2 Diversidad y principales características de posibles genes codificantes de PhaCs en <i>Pseudomonas</i>	25
3.3 Principales características de las PhaCs identificadas en <i>Pseudomonas</i>	28
3.4 Relaciones filogenéticas entre las PhaCs de <i>Pseudomonas</i>	33
3.5 Relaciones filogenéticas entre <i>Pseudomonas</i> y asociaciones con PhaCs.....	35
4. DISCUSIÓN.....	39
5. CONCLUSIONES.....	50
6. BIBLIOGRAFÍA.....	53
7. ANEXO.....	57

RESUMEN

La llegada del siglo XXI se asocia, entre otras cosas, a la relevancia que cobra el cuidado del medio ambiente, el avance acelerado de la ciencia y la tecnología, y la necesidad de desarrollar materiales más amigables con el Planeta. En este sentido, los polihidroxicanoatos (PHAs) -biopolímeros producidos por microorganismos- se posicionan como buenos candidatos para reemplazar a los plásticos derivados del petróleo pues, a diferencia de estos, cumplen con parámetros sustentables relativos a su fuente de producción y biodegradabilidad. Así, los aportes que se puedan hacer en esta materia son relevantes para el desarrollo de nuevas y mejores materialidades.

Estudios previos han reportado el aislamiento desde el continente Antártico de cepas bacterianas del género *Pseudomonas*, que poseen capacidades únicas y ventajosas para la producción de PHAs. Sin embargo, no se ha explorado el repertorio de proteínas PHA sintasas (PhaCs), claves en este proceso, presente en un conjunto de *Pseudomonas* antárticas, puesto en el contexto de cepas aisladas desde otros ambientes. En este trabajo, se estudió la diversidad y distribución filogenética de genes de PhaCs en cepas de *Pseudomonas* antárticas, en comparación con cepas aisladas desde otros lugares del mundo. Para esto, se construyó una base de datos que contempla 855 genomas del género en cuestión, de los cuales 33 provienen de la Antártica. Estos, fueron estudiados con análisis bioinformáticos tales como alineamientos de secuencias, construcción de árboles filogenéticos, y análisis de dominios conservados, entre otros.

Las PhaCs se dividen en cuatro clases (I a IV), habiendo, normalmente, 2 PhaCs de clase II en el *cluster* génico para la producción de PHA en *Pseudomonas*. En este estudio, se identificaron 738 cepas de dicho género con presencia de 2 PhaCs de clase II. De estas 738, 6 poseen una tercera PhaC de esta clase y 1 posee una cuarta de la misma. Además, se identificaron 42 genomas que poseen PhaCs de clase I, donde 2 tienen dos PhaCs de dicha clase.

Se determinó que, normalmente, la ausencia o presencia de PhaCs y sus clases están vinculadas con las relaciones filogenéticas entre las cepas, aunque la presencia esporádica de algunas sugiere su adquisición por transferencia horizontal. Por otra parte, se concluyó que la mayoría de las *Pseudomonas* poseen la capacidad de producir PHAs asociado a la presencia de 2 PhaCs de clase II codificadas en su genoma, aunque en algunas *Pseudomonas* antárticas se encuentran características que llaman la atención en cuanto a la presencia de un número mayor de PhaCs de clase II o presencia de otras clases, destacando el caso de la cepa MPC6 con 5 PhaCs, incluyendo una perteneciente a una potencial nueva clase. Nuestros resultados confirman las peculiaridades y posibles ventajas de *Pseudomonas* antárticas para la producción de PHAs, siendo necesario estudios adicionales para develar la diversidad aún desconocida de enzimas que participan en la síntesis de estos biopolímeros con interés biotecnológico.

ABSTRACT

The arrival of the 21st century is associated, among other things, with the increasing relevance of environmental care, the accelerated advancement of science and technology, and the need to develop materials that are more environmentally friendly. In this regard, polyhydroxyalkanoates (PHAs) - biopolymers produced by microorganisms - emerge as good candidates to replace petroleum-derived plastics because, unlike them, they comply with sustainable parameters related to their source of production and biodegradability. Thus, contributions that can be made in this field are relevant for the development of new and better materials.

Previous studies have reported the isolation of bacterial strains from the Antarctic continent belonging to the *Pseudomonas* genus, which possess unique and advantageous capabilities for PHAs production. However, the repertoire of PHA synthase proteins (PhaCs), key enzymes in this process, present in a set of Antarctic *Pseudomonas* strains, in the context of strains isolated from other environments, has not been explored. In this work, the diversity and phylogenetic distribution of PhaC genes in antarctic *Pseudomonas* strains were studied, compared to strains isolated from other parts of the world. For this purpose, a database was constructed that includes 855 genomes of the relevant genus, of which 33 originate from Antarctica. These genomes were studied using bioinformatic analyses such as sequence alignments, construction of phylogenetic trees, and analysis of conserved domains, among others.

PhaCs are divided into four classes (I to IV), with typically 2 class II PhaCs in the gene cluster for PHA production in *Pseudomonas*. In this study, 738 strains of this genus were identified to have 2 class II PhaCs. Out of these 738 strains, 6 have a third PhaC of this class, and 1 has a fourth PhaC of the same class. Additionally, 42 genomes were found to have class I PhaCs, with 2 of them having two class I PhaCs.

It was determined that, normally, the absence or presence of PhaCs and their classes are linked to the phylogenetic relationships among the strains, although the sporadic presence of some suggests their acquisition through horizontal transfer. On the other hand, it was concluded that the majority of *Pseudomonas* strains have the capacity to produce PHAs associated with the presence of 2 class II PhaCs encoded in their genome. However, in some Antarctic *Pseudomonas* strains, unusual characteristics were found regarding the presence of a greater number of class II PhaCs or the presence of other classes, highlighting the case of the strain MPC6 with 5 PhaCs, including one belonging to a potential new class. Our results confirm the peculiarities and potential advantages of Antarctic *Pseudomonas* in PHA production, and further studies are needed to uncover the yet unknown diversity of enzymes involved in the synthesis of these biopolymers with biotechnological interest.

1. INTRODUCCIÓN

1.1 Chile: un país diverso, desconocido y con potencial biotecnológico

1.1.1 Generalidades geográficas de Chile

Chile es un país cuyas características geográficas lo hacen especialmente diverso. Este es tricontinental, estando constituido por tres zonas geográficas: Chile continental -paralelos 17°29'57 S y 56°32'12 S-; Chile insular y el Territorio Chileno Antártico -meridianos 53° O y 90° O- (Gobierno de Chile, s.f.). Esta tricontinentalidad, marcada por el hecho de abarcar una amplia latitud del continente americano -casi 40°- más una fracción de la Antártica, significa poseer los 5 grandes tipos macrobioclimáticos (determinados por Rivas-Martínez): Tropical, Mediterráneo, Templado, Boreal y Polar (Amigo y col., 2007). Al alero de esta diversidad climática, Chile presenta, a su vez, una diversidad ecosistémica importante que incluye desiertos, estando presente el más árido y antiguo del mundo -Atacama- (Azua-Bustos y col., 2012), cientos de volcanes (Lara y col., 2006), salares (Escudero y col., 2013) como el salar de Atacama en el Norte (Alonso & Risacher, 1996), la Selva Valdiviana en el Sur (Hajek, 1991), glaciares como el Pío XI en la región de Magallanes (Rivera y col., 1997), entre otros ecosistemas excepcionales. Esta pluralidad ecosistémica, al mismo tiempo, se traduce en una gran variedad de condiciones donde se gesta la vida en sus múltiples formas a lo largo del territorio.

Es pues la territorialidad de Chile una privilegiada si se trata de estudiar organismos de diferentes condiciones; más aún si se habla de condiciones extremas. Siendo esto una gran oportunidad para estudiar lo desconocido, dónde quizás se podrían encontrar las soluciones a problemas del presente y futuro a las que aún no se ha llegado.

1.1.2 Antártica, territorio nacional poco explorado y con alto potencial biotecnológico

Una de las zonas menos exploradas del país y del mundo corresponde al continente Antártico (Orellana-Saez y col., 2019). Este se caracteriza por ser uno de los ambientes terrestres

más fríos, secos y químicamente extremos para ser habitado por seres vivos (Orellana-Saez y col., 2019). La vegetación no es muy diversa, de hecho, no hay árboles, y sus temperaturas pueden alcanzar los -40°C e incluso ser inferiores durante los meses de invierno (Tindall, 2004). Por esta y otras razones, se le considera un ambiente extremo. Sin embargo, se ha visto que, si bien estas condiciones pueden ser hostiles para los eucariotas, pueden albergar un amplio rango de procariotas (Tindall, 2004). En este sentido, es importante considerar que la capa superior de la tundra polar alterna entre un estado descongelado y congelado según la estación, lo que significa un ambiente desafiante al que el metabolismo de los microorganismos debe adaptarse para funcionar adecuadamente en un rango extenso de temperaturas (Orellana-Saez y col., 2019).

En línea con lo anterior, es que hay dos tipos de microorganismos en particular que han llamado la atención: psicrófilos y psicrotolerantes, puesto que poseen mecanismos moleculares únicos que les permiten sobrevivir en condiciones ambientales bastante duras, además de su potencial para sintetizar químicos y proteínas que podrían ser utilizables en la industria (Orellana-Saez y col., 2019). Los psicrófilos se definen por su capacidad de propagarse por debajo del punto de congelación del agua y que no pueden replicarse por sobre los 20°C . Por su parte, los psicrotolerantes pueden crecer en un rango que va desde los -5°C hasta los 40°C (Orellana-Saez y col., 2019).

Es a partir de la premisa de querer indagar lo desconocido, y aprovechando la ubicación privilegiada de Chile, que surge el interés por estudiar la Antártica y sus zonas prístinas, específicamente uno de los más desconocidos de sus componentes: sus microorganismos y su potencial biotecnológico. En concordancia con ello, nuestro grupo de investigación, realizó dos expediciones a la Antártica durante los años 2017 y 2019, en el marco de las 53^a y 55^a Expediciones Científicas Antárticas Chilenas, con el fin de muestrear su suelo en pos de caracterizar su diversidad microbiana y los genes de resistencia a antibióticos que podrían tener, además de la prospección de nuevos antibióticos, considerando tanto zonas intervenidas como no

intervenidas, y estudiar sus microorganismos (Marcoleta y col., 2022). Se encontraron distintos tipos de estos, entre ellos bacterias, distinguiéndose varias del género *Pseudomonas* (Marcoleta y col., 2022).

1.2 Género *Pseudomonas*: características generales y dificultades en su clasificación

1.2.1 Características generales del género *Pseudomonas*

Pseudomonas no solamente es el género de las bacterias gram negativo con el mayor número de especies reconocidas (Lalucat, 2020), sino que también dentro de todos los géneros bacterianos descritos a la fecha es, probablemente, el más diverso del Planeta (Spiers y col., 2000). Prueba de ello, es que *Pseudomonas* se encuentra presente en todos los principales ambientes naturales, es decir, terrestre, agua dulce y marino (Spiers y col., 2000) (Palleroni, 1993). Sumado a esto, se tiene que forman asociaciones con plantas y animales (Spiers y col., 2000). Una distribución tan amplia como esta sugiere que hay un grado, tanto genético como fisiológico, extraordinario de adaptación (Spiers y col., 2000). Una característica particular del género en cuestión es su impresionante capacidad de degradar un amplio rango de sustratos, el que incluye compuestos aromáticos, derivados halogenados y residuos orgánicos recalcitrantes (Spiers y col., 2000) y que rara vez son degradados por otros grupos microbianos (Palleroni, 1993). En línea con esto, es que, debido a su resistencia a compuestos tóxicos y su versatilidad metabólica, *Pseudomonas* se plantea como uno de los géneros bacterianos más relevantes para la biorremediación y para la producción industrial de compuestos de interés como lo son los polihidroxicanoatos (PHAs) (Pacheco y col., 2019).

1.2.2 Dificultades en la clasificación del género *Pseudomonas*

Un género tan diverso como este presenta ciertas complejidades en cuanto a su clasificación, siendo esta una problemática que se arrastra desde el siglo pasado y que aún no encuentra soluciones certeras. Durante el siglo XX, se realizaron varios intentos por desarrollar un sistema confiable de clasificación de las especies de *Pseudomonas*; mas estos fueron fallidos

(Palleroni, 1993). Esta situación, como lo analizan los autores Jorge Lalucat, Magdalena Mulet, Margarita Gomila y Elena García-Valdés, da un giro importante con la llegada del siglo XXI y la introducción de la genómica que cambia drásticamente la taxonomía bacteriana (Lalucat y col., 2020). La filogenómica posee métodos que permiten identificar especies e inferir la filogenia de rangos taxonómicos más altos incluso al nivel de subespecies. Así, el conocimiento de todo el genoma de cepas bacterianas, junto con el análisis de sus genes, se han instaurado como métodos con mayor valor en comparación con las pruebas fenotípicas a la hora de realizar clasificación e identificación bacteriana (Lalucat y col., 2020). Se tiene que métodos como comparaciones digitales de genoma completo usando identidades de nucleótidos promedio, ANI por sus siglas en inglés, es uno de los estándares de oro para la circunscripción de especies. Por otro lado, con el objetivo de clasificar a las bacterias en base a sus relaciones evolutivas, es posible realizar análisis filogenéticos estudiando las secuencias de nucleótidos de ciertos genes. Un ejemplo de esto es el procedimiento de tipificación de secuencias multilocus, MLST por sus siglas en inglés. Se ve cómo el desarrollo de herramientas computacionales ha permitido brindar clasificaciones más confiables, desplazando incluso métodos experimentales de larga trayectoria como la hibridación ADN-ADN, sin embargo, aún no existe una forma que permita establecerlas de manera 100% certera. A esto se debe agregar el factor de que es probable que las nuevas tecnologías lleven a un aumento importante en el número de especies, lo que podría llevar a dividir el género *Pseudomonas* en otros diferentes o incluso subgéneros (Lalucat y col., 2020).

Si bien se ha avanzado en cuanto a metodologías de clasificación taxonómica más confiables, los límites aún no están del todo definidos. Esto significa un desafío a la hora de hacer estudios comparativos que apuntan a asociar ciertas características al género *Pseudomonas* y/o a cepas dentro de este, como pretende justamente el presente trabajo.

1.3. Polihidroxicanoatos: fuente de carbono bacteriana con potencial biotecnológico

1.3.1. Características generales de los polihidroxicanoatos

Se mencionaba en la sección 1.2.1 que *Pseudomonas* sería un género relevante, entre otras cosas, para la producción de PHAs. Los PHAs son biopolímeros, específicamente poliésteres formados por varios hidroxialcanoatos, los cuales son sintetizados por arqueas y bacterias, tanto gram negativo como positivo. Estos actúan como reserva de carbono y energía y se depositan en el citoplasma como inclusiones insolubles en agua cuando hay plena disponibilidad de fuentes de carbono y otros nutrientes que limitan el crecimiento. No obstante, cuando esta situación cambia y hay escasez de dichas fuentes, los PHAs son movilizados por PHA depolimerasas intracelulares (Reddy y col., 2003) (Rehm, 2003).

Los PHAs pueden llegar a acumularse intracelularmente a niveles altísimos llegando al 90% del peso seco de la célula en condiciones de estrés por nutrientes (Reddy y col., 2003). Los gránulos de PHA tienen un diámetro que va desde los 0,2 a los 0,5 mm, estos son amorfos y forman un núcleo hidrofóbico rodeado de proteínas unidas al gránulo de PHA (Mozejko-Ciesielska y col., 2019). Además de servir como fuente de reserva de energía, se ha comprobado que los gránulos de PHAs son factores clave en cuanto a la resistencia contra el congelamiento -como crioprotectores-, presiones altas y compuestos tóxicos, además, ayudan a mantener la integridad celular ante desbalances osmóticos (Pacheco y col., 2019; Mozejko-Ciesielska y col., 2019).

1.3.2 Propiedades y aplicaciones de los PHAs

El interés científico por los PHAs se ha intensificado debido a sus propiedades y aplicaciones biotecnológicas. Dentro de sus propiedades se encuentra que son biodegradables, biocompatibles, no tóxicos, tienen un alto grado de polimerización, son altamente cristalinos, ópticamente activos, insolubles en agua, entre otras. Cabe destacar que son termoplásticos, es decir, se tornan flexibles a altas temperaturas y se solidifican al enfriarse; se pueden producir a

partir de fuentes renovables y pueden variar desde ser un material transparente y elástico a uno opaco, sólido y con alta resistencia. Teniendo todo esto en consideración, es casi intuitivo pensar en los PHAs como una alternativa sustentable a los plásticos petroquímicos, sabiendo que estos no dependen de combustibles fósiles para su producción. De hecho, se considera que los PHAs son muy competitivos con el polipropileno. Dentro de las aplicaciones que tienen los PHAs, se encuentran algunas como micro polvo para cosméticos, suturas o películas para uso médico, andamios para la adhesión y proliferación celular, bioimplantes, sistema de *delivery* de drogas, como plastificante, películas para *packaging*, para la fabricación de artículos desechables, como precursor para químicos finos y producción de biocombustible, entre muchas otras (Reddy y col., 2003; Mozejko-Ciesielska y col., 2019; Tan y col., 2020).

1.3.3 Tipos de PHAs y rutas de síntesis

Las investigadoras Justyna Mozejko-Ciesielska, Karolina Szacherska y Paulina Marciniak describen los distintos tipos de PHAs, sus rutas de síntesis y la clasificación de las enzimas involucradas en un artículo que corresponde a una revisión en torno a las *Pseudomonas* y la síntesis de estos polímeros (Mozejko-Ciesielska y col., 2019) y es en este trabajo en el que se basa gran parte de la presente subsección. Los PHAs son poliésteres compuestos por monómeros formados por una cadena principal de carbono con un grupo hidroxilo y uno carboxilo, cada uno unido a un átomo de carbono diferente. Estos monómeros se unen entre sí por un enlace éster que se forma al combinar el grupo carboxilo del monómero con el grupo hidroxilo del grupo adyacente (Figura 1). Sin embargo, la composición estructural de los PHAs varía según la cepa bacteriana y la fuente de carbono utilizada. Hasta el año 2019, se habían descubierto al menos 150 monómeros diferentes con estructuras diversas, además de grupos funcionales en la cadena monomérica, pudiendo clasificarse estos en dos grandes grupos según el número de átomos de carbono en las unidades monoméricas: a) PHAs de cadena corta o scl-PHAs por sus siglas en

inglés y b) PHAs de cadena media o mcl-PHAs por sus siglas en inglés. En el caso de a) la cantidad de carbonos por unidad monomérica va desde los 3 a 5 átomos y en el de b) de 6 a 14.

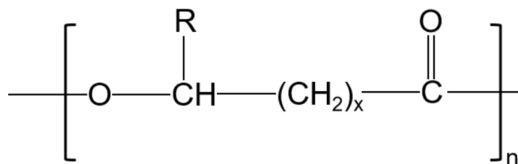


Figura 1. Representación del monómero de los PHAs donde **R** corresponde a grupos alquilo de cadena lateral, estos pueden ir desde metilo (C1) hasta tridecilo (C13); **n** a la cantidad de veces que se repite el monómero y **x** el número de grupos metileno.

Tomado de Wei y col. (2022).

Dentro de los scl-PHAs, el más común en cuanto a síntesis es el poli-3-hidroxiбутirato o [P(3HB)], el cual es un polímero cristalino de estructura rígida, frágil, con un punto de fusión alto y una temperatura de transición vítrea baja. Por otro lado, en el caso de los mcl-PHAs, se encuentra el poli-3-hidroxi octanoato [P(3HO)] y el poli-3-hidroxi hexanoato [P(3HHx)], los que destacan por su alta elasticidad con baja cristalinidad y resistencia a la tracción, con punto de fusión alto. Es por todo esto que los mcl son preferidos antes que los scl; cabe mencionar que estos mcl son principalmente sintetizados por especies de *Pseudomonas*. Dentro de este género bacteriano, existen dos rutas metabólicas para obtener los precursores para la síntesis de mcl-PHAs, donde cada una tiene un sustrato diferente para generar los precursores del 3-hidroxiacil (3HA) que son utilizados para la síntesis de los mcl-PHAs. Por un lado, fuentes de carbono alifático como ácidos grasos, los que son degradados a través de la β -oxidación; por otro lado, pueden utilizar fuentes de carbono como glucosa, gluconato o etanol para sintetizar monómero de mcl-PHA mediante la vía de síntesis de ácidos grasos *de novo*.

Dentro de todas las enzimas involucradas en la síntesis de los PHAs, cobran especial relevancia las PHA sintasas (PhaC). Esta enzima está a cargo de la actividad de polimerización, controla los pesos moleculares y la composición monomérica del PHA (Tan y col., 2020). Las PhaCs se clasifican en 4 grupos considerando la especificidad de sustrato y las subunidades que las componen: las clases I y II poseen una sola subunidad PhaC que tiene una masa molecular

que varía entre los 61 y 73 kDa. Por su parte, la clase III posee dos subunidades diferentes: PhaC y PhaE, ambas con una masa molecular de unos 40 kDa; y, finalmente, la clase IV también posee dos subunidades diferentes: PhaC y PhaR. En cuanto a los monómeros que usa, las clases I, III y IV tienen la característica común de usar monómeros de cadena corta, mientras que la clase II usa de cadena mediana y es justamente esta última clase la que se encuentra presente en *Pseudomonas*.

Se ha descrito que *Pseudomonas* tiene un *cluster* de genes para la síntesis y acumulación de PHA conservado y está compuesto por dos operones: *phaC1ZC2D* y *phaFI* que codifican 6 proteínas claves: 2 polimerasas (PhaC1 y PhaC2) responsables de la síntesis de mcl-PHAs, 1 depolimerasa (PhaZ), 2 fasinas (PhaI y PhaF) encargadas de la estabilización y regulación de los gránulos de PHA y 2 proteínas regulatorias (PhaD y PhaG). Los genes de las proteínas PhaC1, PhaC2, PhaZ y PhaD se transcriben en la misma dirección, por su parte, los genes de PhaF y PhaI se transcriben en la dirección contraria. A continuación, se presenta un esquema representativo del *cluster* génico mencionado y de la estructura granular del PHA (Figura 2).

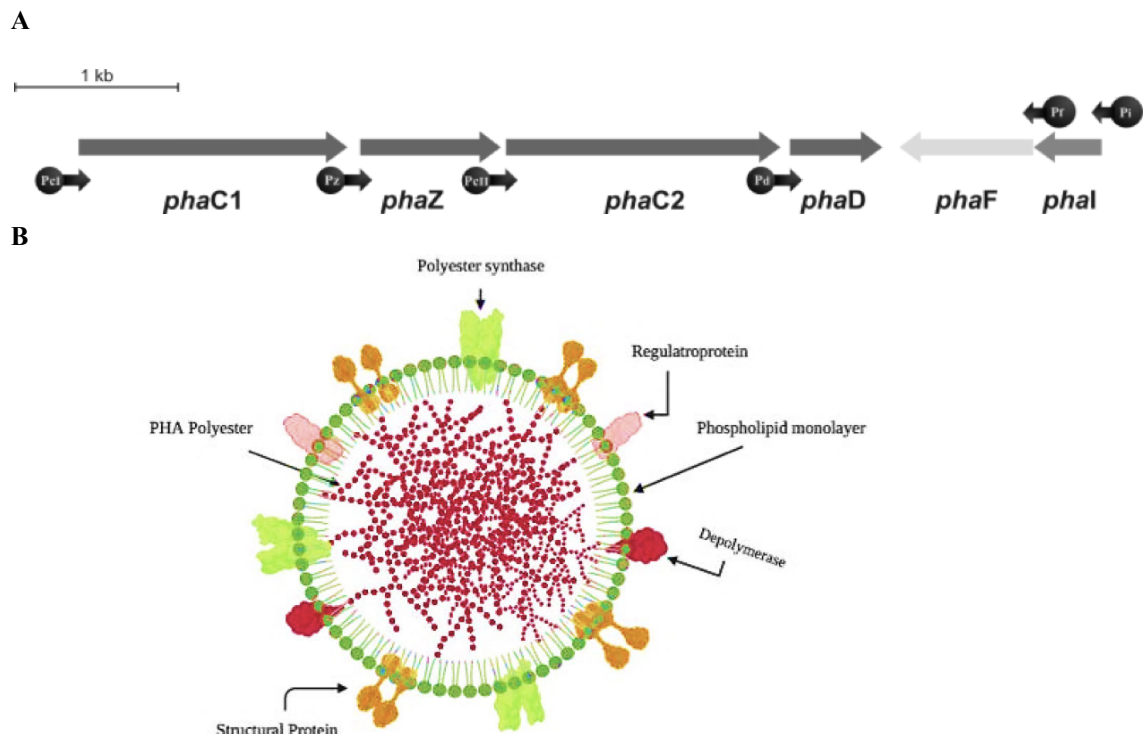


Figura 2. (A) Representación del locus *pha* (genes *phaC1ZC2DFI*) donde los genes *phaC1*, *phaZ*,

PhaCs presentes en la naturaleza que puedan cumplir con las características mencionadas anteriormente. En ese sentido, evaluar la presencia de PhaCs en géneros bacterianos donde hay una preponderancia en cuanto a la producción de PHA, podría ser un primer paso en esa senda.

1.3.4 Microorganismos productores de PHA

Existe un extenso rango de microorganismos capaces de acumular PHA los que, a su vez, han sido aislados de diversos hábitats incluyendo ambientes extremos (fuentes termales, lagos salados y regiones polares). Hasta el año 2020, se habían descrito 167 géneros microbianos productores de PHA de los cuales 150 corresponden a bacterias y 17 a arqueas. Sumado a esto, se han propuesto una gran cantidad de posibles PhaCs de bacterias no cultivables (Tan y col., 2020).

Pseudomonas es uno de los 167 géneros bacterianos productores de PHAs, siendo este y *Janthinobacterium* los mayores productores dentro de los microorganismos aislados desde suelo antártico (Pacheco y col., 2019). *Pseudomonas* ha ido atrayendo interés como género productor de PHAs, lo que se evidencia en la cantidad de estudios recientes que han elegido como candidatos productores de estos polímeros a bacterias del género en cuestión. Esto se debe a características que se han mencionado en subsecciones anteriores como su versatilidad metabólica y por el amplio rango de fuentes de carbono que puede utilizar (Mozejko-Ciesielska y col., 2019).

1.3.5 MPC5 y MPC6: *Pseudomonas* antárticas con potencial biotecnológico

Se mencionaba anteriormente -en la subsección 1.1.2- que nuestro equipo de investigación aisló bacterias del género *Pseudomonas* desde muestras de suelo del continente blanco. Sumado a ello, se comentaba en la subsección 1.2.1 el potencial biotecnológico de este género. En línea con ambos elementos, se tiene que dentro de este grupo de *Pseudomonas* aisladas se encuentran las cepas MPC5 y MPC6, las que destacan por su potencial biotecnológico relacionado con la producción de biopolímeros (Pacheco y col., 2019).

Como describen los investigadores Nicolas Pacheco y colaboradores (Pacheco y col.,

2019), MPC5 y MPC6 son bacterias extremófilas psicrotolerantes que comparten no solamente su origen territorial -Antártica-, sino que también una tasa de crecimiento específica similar en condiciones de 4°C (0,18 1/h) o 30°C (0,2 1/h) en glicerol y la capacidad de producir altas cantidades de PHAs a partir de dicho sustrato en el rango de temperatura señalado. Esto es algo llamativo, puesto que nunca se había descrito este atributo en bacterias del género *Pseudomonas*.

Se estudió el tipo de PHA producido por cada una de las cepas en cuestión en diferentes condiciones. En el caso de MPC5, se encontró que a 30°C sintetiza un mcl-PHA donde su cadena polimérica está compuesta mayormente por los monómeros 3-hidroxicanoato, 3-hidroxi-octanoato y 3-hidroxi-dodecanoato, mientras que a 4°C tiene un 31,7% del monómero 3-hidroxi-5-cis-dodecanoato. Es interesante observar que, si bien a diferentes temperaturas cambia la composición del PHA, la concentración de este en ambas condiciones es casi idéntica: 0,58 (g/L) a 30°C y 0,54 (g/L) a 4°C. En el caso de MPC6, más del 80% de su composición corresponde al monómero 3-hidroxi-butirato y el resto a monómeros de cadena mediana. La concentración de PHA para MPC6 corresponde a 1,28 (g/L) a 30°C y a 0,94 (g/L) a 4°C. A continuación, se presenta el detalle de la composición de los PHAs, su concentración y la biomasa en ambas cepas a diferentes temperaturas (Tabla 1).

Tabla 1. Características de los PHAs producidos por MPC6 y MPC5 en diferentes condiciones. Se muestra la composición de los PHAs, concentración y biomasa en MPC5 y MPC6 cultivadas en matraces con glicerol (30 g/L) a 4°C o 30°C durante 72 h en agitación. Tomado de Pacheco y col. (2019).

Strain	Temp. (°C)	CDM (g/L)	PHA ^b (%)	PHA (g/L)	Monomer composition (%) ^c						
					C4	C6	C8	C10	C12	C12:1	C14
MPC5	4	2.75 ± 0.1	12.5	0.54 ± 0.0	N.D	3.1	11.5	37.7	13.1	31.7	2.8
	30	3.03 ± 0.1	19.2	0.58 ± 0.1	N.D	4.2	23.1	42.7	20.9	8.6	N.D
MPC6	4	3.54 ± 0.0	26.6	0.94 ± 0.1	82.1	2.1	2.7	9.7	3.5	3.7	N.D
	30	3.76 ± 0.1	33.4	1.28 ± 0.1	89.5	1.8	3.3	4.4	1.1	N.D	N.D

^aThe data reflect mean values and deviation from three replicates

^bPHA content relative to cell dry mass (CDM)

^cC4: 3-hydroxybutyrate, C6: 3-hydroxyhexanoate, C8: 3-hydroxyoctanoate, C10: 3-hydroxydecanoate, C12: 3-hydroxydodecanoate, C12:1: 3-hydroxy-5-cis-dodecanoate, C14: 3-hydroxytetradecanoate

Cabe mencionar, que en el estudio llevado a cabo previamente por nuestro equipo de investigación, donde se caracterizaron las cepas MPC5 y MPC6 en cuanto a la producción de PHAs, también se evaluó la producción de PHAs de otras 23 cepas aisladas desde Isla Decepción en Antártica. No obstante, el foco se ha puesto específicamente en MPC5 y MPC6, debido a que son interesantes en cuanto a aplicaciones biotecnológicas, puesto que fueron las que mostraron una mayor acumulación del biopolímero en cuestión. Es pertinente mencionar que se encontró que MPC6 posee una PhaC -PhaC3- que no pudo ser clasificada, siendo esta una variante novedosa de este tipo de enzimas. Es entonces en esto y en la alta producción de PHA en rangos de temperatura amplios, donde radica el potencial biotecnológico de estas *Pseudomonas* antárticas: producción de PHAs nuevos utilizando sustratos de bajo costo, como el glicerol, sin la necesidad del control de temperatura con la participación de un set único de PhaCs.

Sumado a lo anterior, caracterizaciones adicionales de la cepa MPC6 realizadas por nuestro grupo de investigación (Orellana-Saez y col., 2019), han resaltado aún más sus propiedades únicas e interesantes. Primero, esta sería una nueva especie dentro del género *Pseudomonas*, proponiéndose el nombre *P. frigusceleri*, debido a su capacidad de proliferar y producir compuestos valiosos en condiciones cercanas a los 0°C. Segundo, tiene un potencial de resistencia antibiótico y patogénico reducido. Tercero, posee atributos que permitirían aplicaciones biotecnológicas importantes como biorremediación a bajas temperaturas de arsénico, cadmio, alcanos y compuestos cloro- y nitroaromáticos, y la síntesis de biopolímeros como PHAs y alginato.

1.4 Objetivos

Considerando los antecedentes descritos anteriormente, en el presente trabajo de Seminario de Título se proponen los siguientes objetivos:

Objetivo general: Estudiar la diversidad y distribución filogenética de genes de Polihidroxicanoato sintasas (PhaCs) en *Pseudomonas* antárticas.

Objetivos específicos:

1. Identificar y clasificar genes de PhaCs presentes en genomas secuenciados de *Pseudomonas* provenientes de Antártica y otras regiones del Planeta.
2. Evaluar la distribución de las PhaCs identificadas en distintos linajes de *Pseudomonas*, buscando posibles asociaciones entre tipos de PhaCs, lugar de origen, y filogrupos característicos de este género.

Estos dos objetivos fueron abordados utilizando diversas herramientas bioinformáticas e información experimental otorgada por nuestro equipo de investigación.

2. MATERIALES Y MÉTODOS

2.1 Construcción de una base de datos genómica de *Pseudomonas*

Con el fin de poder llevar a cabo el estudio comparativo entre cepas de *Pseudomonas* procedentes de la Antártica y otras del mismo género de diferentes partes del mundo, un primer paso fue enriquecer la base de datos de genomas de *Pseudomonas* de nuestro grupo de investigación, en cuanto a presencia de cepas antárticas del género en cuestión. Para ello, se realizó una búsqueda bibliográfica de *Pseudomonas* de dicho continente que hubiesen sido caracterizadas en términos de su genoma, esto se hizo con el buscador *web* de artículos académicos *Google Scholar* utilizando palabras claves como *Pseudomonas*, *antarctic* y *genome*, y considerando publicaciones desde el año 2020 hasta el 2021, puesto que la base de datos fue construida previamente con información actualizada hasta el año 2019. A partir de las nuevas cepas encontradas en esta búsqueda, se obtuvieron los códigos de acceso en la base de datos de nucleótidos GeneBank que permitieron la descarga de los genomas desde los repositorios del Centro Nacional para la Información Biotecnológica (NCBI), que es parte de los Institutos Nacionales de la Salud (NIH) de Estados Unidos (Sayers y col., 2021). El nivel de ensamblaje considerado para toda la base de datos corresponde a cromosoma o genoma completo, excepto en el caso de algunas de las *Pseudomonas* antárticas, donde también se incluyeron a nivel de *contig* y *scaffold*. Esto con el fin de incluir el mayor número de cepas antárticas posibles. En base a esto y sumado a los genomas de *Pseudomonas* ya presentes previamente en la base de datos construida por nuestro grupo de investigación, se llegó a un total de 859 cepas, inicialmente clasificadas como *Pseudomonas*, de las cuales 33 provienen de la Antártica. A cada genoma se le asignó un número o ID (1 a 859).

Luego, se realizó la asignación taxonómica (género y especie) a dichos genomas utilizando el programa GTDB-Tk, por sus siglas en inglés *Genome Taxonomy Database Toolkit* (Chaumeil y col., 2020) que asigna clasificaciones taxonómicas basado en la Taxonomía de la

Base de Datos de Genomas, GTDB por sus siglas en inglés *Genome Taxonomy Database* (Parks y col., 2018). Esto, con la finalidad de tener un criterio homogéneo para clasificar e identificar estos grupos en el árbol, corroborando su correcto posicionamiento. A partir de los resultados obtenidos, se construyó una tabla en la que se encuentran tanto el genoma como la predicción de la especie (Apéndice Digital 1). Se eliminaron aquellos genomas que no pudieron ser clasificados o que no pertenecen al género *Pseudomonas* según GTDB-Tk.

Cabe mencionar que la base de datos de *Pseudomonas* construida no considera plásmidos, los que se filtraron revisando los archivos uno a uno de manera manual.

2.2 Cálculo de identidad promedio de nucleótidos entre genomas (ANI)

Buscando identificar grupos de cepas de una misma especie, se calculó la identidad promedio de nucleótidos entre genomas (ANI), por sus siglas en inglés *Average Nucleotide Identity*, para todos los genomas de la base de datos. Para ello, se utilizó el programa fastANI (Jain y col., 2018) en su versión 1.32, en la modalidad *todos contra todos (many to many)*, dado que son múltiples genomas de consulta y de referencia, donde lo que se busca es comparar cada genoma con respecto a la colección completa. El *input* del programa fueron los genomas en formato .fna y el *output* corresponde a un archivo .txt que contiene el genoma de consulta, el de referencia, el valor ANI y el recuento de mapeos de fragmentos bidireccionales y el total de fragmentos de consulta, respectivamente. Cabe mencionar que, si el valor ANI para un par de genomas es menor a 76%, no se muestra dentro de los resultados.

2.3 Búsqueda y clasificación de genes codificantes para PhaCs

Para buscar genes que codifican PhaCs en los genomas de *Pseudomonas* se usaron como referencia las PhaCs utilizadas para caracterizar a MPC6 (Orellana-Saez y col., 2019) y una de clase IV, específicamente la subunidad PhaC perteneciente a *Bacillus megaterium* (Tan y col., 2020), teniéndose 12 PhaCs que cubren todas las clases descritas hasta ahora (I a IV). Se consideró una mayor cantidad de referencias de la clase II, puesto que es la que suele estar

mayoritariamente presente en *Pseudomonas*. Además, se usó como referencia la proteína PhaC3 identificada previamente en la cepa MPC6, la cual no se ha podido clasificar, pues queda en otra rama en el árbol (Orellana-Saez y col., 2019) y que para efectos de este trabajo hemos denominado clase X. A partir de los códigos de estas, se pudo acceder a sus secuencias aminoacídicas en formato fasta en la base de datos de NCBI. En el caso de la PhaC3, su secuencia aminoacídica se extrajo a partir del genoma secuenciado de MPC6 utilizando SnapGene versión 6.2.1 (SnapGene, 2022). Los códigos de acceso de cada una de las PhaCs utilizadas, la clase y bacteria de la que provienen se detallan en el Anexo 1. Con dichas secuencias se confeccionó un archivo multifasta que incorpora las 12 PhaCs de referencia. Utilizando este y los genomas de la base de datos de *Pseudomonas*, se hizo la búsqueda de PhaCs con BLASTX, por sus siglas en inglés *Basic Local Alignment Search Tool eXtended* (Gish & States, 1993), utilizando un script de Python para su versión 3.9.13 (Anexo Digital 1).

Se construyeron gráficos de frecuencias de *percentage of identical matches* para los resultados arrojados por BLASTX para cada una de las 13 PhaCs, a partir de los cuales se definieron intervalos que corresponderían a probables verdaderos positivos. Estos fueron corroborados extrayendo las secuencias de los genomas, a partir de sus archivos .fna con la herramienta SnapGene, y que corresponden a la zona de *match* con la PhaC de referencia en los límites inferiores y superiores de los intervalos de porcentaje de identidad en cuestión, usando la herramienta BLASTp de NCBI para verificar que tuviesen similitud con otras PhaCs e incluyesen los dominios característicos de estas proteínas. También, se compararon los resultados de los límites de los intervalos con los resultados para distintas PhaCs buscando sus coordenadas en el cromosoma, con el fin de evaluar si los porcentajes de identidad de estos límites eran mayores o menores para las mismas coordenadas con otras PhaCs. Dada la amplia diferencia entre los porcentajes de identidad entre las 7 PhaCs de clase II, lo que se evaluó con un alineamiento con Clustal Omega (Sievers y col., 2011), se eligieron 5 que cubrieran la mayor

diversidad posible.

Se identificaron aquellos *match* que cumplían las condiciones de verdaderos positivos utilizando filtros y funciones de Excel. En los casos en que se encontró más de una PhaC de una misma clase en un genoma, estas se nombraron con una letra mayúscula (A, B, C, etc.) según la región del cromosoma y el contexto genómico en la que típicamente se encontraban.

Para corroborar la efectividad de la estrategia de identificación de PhaCs, se utilizó como control positivo la cepa MPC6, cuyas PhaCs ya fueron identificadas (Orellana-Saez y col., 2019).

2.4 Extracción de las PhaCs identificadas

A partir de las zonas de los diferentes genomas identificadas como genes codificantes de PhaCs, se obtuvieron las secuencias de aminoácidos para realizar análisis a nivel de proteína. Para obtener la secuencia de las proteínas, se utilizaron dos *scripts* de Python (Anexo Digital 1 y Anexo Digital 2): el primero, realiza un BLASTX considerando todas las PhaCs de referencias utilizadas contra la base de datos de *Pseudomonas* y entrega dentro del *output* la secuencia de la proteína con la que la PhaC de referencia hace *match*. Luego, el segundo, utiliza como *input* dicha información y las coordenadas que se establecieron como verdaderos positivos para los *matches* de las distintas PhaCs con los diferentes genomas en base a la metodología descrita anteriormente. Con esta información, el algoritmo busca dichas coordenadas en las secuencias proteicas que arrojó el primer *script* y extrae la proteína completa. Así, independientemente de si las coordenadas de los verdaderos positivos abarcan la totalidad de la proteína que está codificada en la zona del *match*, se extrae la proteína completa evitando obtener solo un fragmento de ella.

2.5 Agrupamiento de las PhaCs extraídas

Con el objetivo de observar la diversidad de PhaCs y así definir si luego el alineamiento se realizaba con representantes de cada grupo o con todas las proteínas de cada clase, se hizo un agrupamiento por clase -I y II- de las proteínas extraídas en el paso anterior ejecutando el programa *Cluster Database at High Identity with Tolerance* o CD-HIT (Li & Godzik, 2006) (Fu y

col., 2012) en su versión 4.8.1. Este agrupa proteínas considerando cierto porcentaje de identidad y establece una proteína dentro del *cluster* como la representativa del mismo. Se estableció como parámetro para el agrupamiento un 60% de identidad.

2.6 Análisis de dominios conservados

A partir de los resultados del paso anterior, se procedió a realizar un análisis de los dominios conservados de cada una de las proteínas representantes de cada *cluster* arrojado por CD-HIT. En el caso del *cluster* único de la clase I, la representante fue una PhaC del subgrupo IA del genoma -en base al ID asignado- 8; en el caso del *cluster* 0 de las PhaCs de clase II la proteína representante es una del genoma 560 clase IIB; para el *cluster* 1 una del genoma 5 clase IIB y, finalmente, para el *cluster* 2 una del genoma 740 clase IIA. También, se hizo este análisis para la PhaC3 clase X de MPC6.

Estos análisis se realizaron usando dichas secuencias aminoacídicas como *input* para ejecutar la herramienta BLASTp para cada una de ellas. Se utilizó este programa en su versión *web* disponible en la página de NCBI y que arroja dentro de los resultados un análisis de los dominios conservados. De todos los dominios arrojados para cada representante, se seleccionó aquel que se encontró presente en todas ellas, con el fin de establecer una comparación del mismo entre las distintas PhaCs de las clases I y II en el paso que viene a continuación.

2.7 Alineamiento múltiple de las PhaCs extraídas

Con el fin de analizar a nivel de aminoácidos las PhaCs extraídas en el paso 2.4 de esta sección, se realizaron alineamientos múltiples de las mismas por clase, es decir, se realizó un alineamiento considerando todas las PhaCs de clase I y otro considerando todas las PhaCs de clase II. Esto se hizo utilizando el programa de alineamiento múltiple MAFFT, por sus siglas en inglés *Multiple Sequence Alignment Software*, en su versión 7 (Kato & Standley 2013). Se usó esta herramienta en su formato *web* incorporado en NGPhylogeny.fr (Lemoine y col., 2019), esta última corresponde a una plataforma que integra varias herramientas bioinformáticas para llevar a

cabo análisis filogenéticos y que como *input* utiliza un archivo fasta con las secuencias aminoacídicas de interés y entrega como *output* distintos archivos, entre ellos un alineamiento en formato fasta y un árbol en formato Newick. Cabe mencionar que para realizar el alineamiento de las PhaCs de clase I, se incorporó dentro de las secuencias *input* la PhaC3 de clase X de MPC6, puesto que el análisis de dominios conservados mostró mayor cercanía de esta con la clase I de PhaCs.

Utilizando la información de los dominios conservados obtenida en el paso 2.6 y que contempla la posición de los aminoácidos que los conforman, se anotó el alineamiento generado para cada una de las clases y se destacó el dominio conservado seleccionado, comparando entre las distintas PhaCs. Para observar estos alineamientos se utilizó la herramienta Jalview en su versión 2.11.2.6 (Waterhouse y col., 2009) y que sirve justamente para este fin. Esta herramienta también se utilizó para colorear los aminoácidos de cada proteína según el porcentaje de identidad, además de para obtener información relevante como la conservación, la ocupación, la calidad y el consenso. Toda esta información de los alineamientos se incorporó en las imágenes generadas por Jalview.

2.8 Construcción de árbol filogenético de las PhaCs extraídas

Buscando establecer las relaciones filogenéticas entre las PhaCs extraídas, es que se procedió a construir dos árboles filogenéticos a partir de los alineamientos generados en el paso anterior, es decir, un árbol por clase. En el caso del árbol de la clase I, también se incorporó la PhaC3 clase X de MPC6. Para determinar estos árboles, se utilizó PhyML en su versión 3.0 (Guindon y col., 2010) en su formato *web* incorporado en NGPhylogeny.fr. Luego, se utilizó el programa iTOL, por sus siglas en inglés *Interactive Tree Of Life* (Letunic & Bork, 2007) para visualizar los árboles obtenidos y destacar en ellos el tipo de PhaC: A, B, C, D o X, según corresponda.

2.9 Construcción de árbol filogenético de *Pseudomonas*

A fin de establecer relaciones filogenéticas entre las distintas *Pseudomonas* de nuestra base de datos, es que se utilizó la herramienta bioinformática MASHTREE en su versión 1.2.0 (Katz y col., 2019) para confeccionar un árbol filogenético con los genomas de las mismas considerando el universo total de la base de datos. Luego, fueron eliminados 13 genomas del árbol, dado que se alejaban considerablemente del resto (probablemente por problemas a nivel de ensamblaje), y se ejecutó nuevamente el programa. Sumado a esto, se construyó un árbol adicional con el programa MASHTREE, el cual contempla solamente las *Pseudomonas* antárticas de la base de datos. Cabe mencionar que para generar los dos árboles nombrados se ejecutó MASHTREE configurando el parámetro kmerlegth en 21.

2.10 Comparación entre la presencia de PhaCs y las relaciones filogenéticas en las cepas de *Pseudomonas* analizadas

Con el fin de mostrar de manera clara la presencia de PhaCs y las relaciones filogenéticas entre las *Pseudomonas* de este estudio, se complementaron los dos árboles filogenéticos creados en el paso anterior con mapas de presencia/ausencia de genes utilizando la herramienta iTOL. Así, en el árbol que contempla todas las *Pseudomonas* de nuestra base de datos (excepto las que fueron eliminadas), se buscó plasmar la presencia o ausencia de las PhaCs y sus clases en los diferentes linajes con una simbología para ello y el bootstrap. Además, se incorporó un código de colores para mostrar las 13 especies más abundantes y destacar a las cepas MPC5, y MPC6. En el caso del árbol que solo contempla *Pseudomonas* antárticas, la herramienta iTOL se utilizó con el fin de mejorar su visualización y destacar elementos importantes como el bootstrap y la presencia de PhaCs de distintas clases. La información para asignar las especies en cada uno de los 2 árboles se obtuvo del análisis hecho con GTDB-Tk realizado en el paso 2.1.

3. RESULTADOS

3.1 Construcción de una base de datos genómica de *Pseudomonas* provenientes de la Antártica y otras regiones geográficas

La base de datos genómica de *Pseudomonas* en su versión actualizada contempla un total de 859 cepas de las cuales 33 son antárticas, incorpora información relativa a la asignación de especies realizada con GTDB-Tk, el código de acceso GenBank, la especie, la cepa, un número de identificación o ID asignado de manera arbitraria (que van del 1 al 859) y si procede o no de la Antártica (Apéndice Digital 2). Cabe mencionar que la clasificación con GTDB-Tk arrojó entre sus resultados (Apéndice Digital 1) que el genoma 510 no pertenecería al género *Pseudomonas*, siendo clasificado como *Entomomonas*, y los genomas 196, 517 y 722 no pudieron ser clasificados. Estos genomas no fueron considerados en los análisis filogenéticos posteriores de las cepas. De esta manera, nuestra base de datos incluye 855 cepas del género *Pseudomonas* contemplado un total de 186 especies, siendo la más abundante *P. aeruginosa* (331) seguida por *P. chlororaphis* (53) y *P. fluorescens* (23) (Apéndice Digital 2).

Con el objetivo de complementar el análisis hecho con GTDB-Tk, se calcularon los valores de ANI de todas las parejas de genomas presentes en nuestra base de datos. Esto permite confirmar si dos genomas pertenecen a la misma especie (ANI \geq 95%) (Apéndice Digital 3). Estos resultados se utilizaron principalmente para compararlos con algunos de los resultados arrojados por GTDB-Tk, como fue en el caso de MPC5 y MPC6, donde el valor ANI fue menor a 90%.

3.2 Diversidad y principales características de posibles genes codificantes de PhaCs en *Pseudomonas*

Utilizando BLASTX y las diferentes PhaCs de referencia sobre nuestro set de genomas de *Pseudomonas*, se identificaron genes que codifican PhaCs (Apéndice Digital 4). Dado que los porcentajes de identidad entre PhaCs de una misma clase pueden ser bajos, menores a 60%, fue

necesario establecer intervalos de porcentajes de identidad que pudiesen considerarse verdaderos positivos. Para ello se realizó una serie de análisis, según cada PhaC y sus resultados, los que se detallan a continuación.

Para las PhaCs de clase I, III o IV, se realizó el mismo análisis. Primero, en base a gráficos de frecuencia, se identificaron intervalos de porcentajes de identidad para los *matches* con la PhaC de referencia correspondiente. Segundo, se evaluaron los límites superiores e inferiores de estos, realizando una inspección visual de los genomas correspondientes a estos límites con la herramienta SnapGene. Así, se buscaron las coordenadas de los *match* de los genomas en cuestión con la PhaC de referencia y se extrajeron sus secuencias aminoacídicas en las coordenadas del *match*. Tercero, se hizo un BLASTp con la herramienta de NCBI, buscando coincidencias con PhaCs de la clase respectiva. En paralelo, se buscaron las coordenadas de los *matches* en cuestión en los resultados de BLASTX para las otras PhaCs de referencia utilizadas, con el fin de evaluar si el porcentaje de identidad y cobertura eran mayores, evitando tener resultados redundantes. En base a esto se establecieron los intervalos correspondientes a posibles verdaderos positivos.

Así, se obtuvo que para la PhaC de clase I el intervalo a considerar, en cuanto a resultados que indiquen posiblemente la presencia de PhaCs de clase I, va desde 50% hacia arriba. En el caso de las PhaCs de clase III o IV, los análisis permitieron descartar su presencia en las cepas de *Pseudomonas* de nuestra base de datos, dado que los porcentajes de identidad de los *matches* en ambos casos fueron menores a 38%. Por otro lado, en el caso de la PhaC3 de clase X de MPC6, los resultados arrojados por BLASTX mostraron una única coincidencia con un porcentaje de identidad y cobertura indicativos de un resultado positivo y es, justamente, con el genoma de MPC6. De esta manera, se concluye que no hay genomas de nuestra base de datos que codifiquen una enzima de tipo PhaC3, como aquella identificada en *Pseudomonas* MPC6.

En el caso de las PhaCs de referencia de clase II, uno de los primeros pasos para acotar

la búsqueda de PhaCs en la base de datos de *Pseudomonas*, fue realizar un alineamiento con Clustal Omega considerando las 7 PhaCs de clase II utilizadas para el BLASTX. Esto se hizo para evaluar la similitud entre ellas con el fin de definir si se requería utilizar todas las PhaCs de clase II para los análisis de los resultados de BLASTX o si se podía prescindir de algunas. Los resultados arrojaron hasta sobre un 80% de identidad entre algunas de las PhaCs y, también, bajo 60% (Tabla 2).

Tabla 2. Porcentajes de identidad alineamiento PhaCs de clase II

PhaC clase II	AAM63407.1	ACK57560.1	BAA36200.1	AAM63409.1	WP_003115652.1	ACK57562.1	BAA36202.1	Porcentaje de identidad
AAM63407.1	100.00	81.75	81.93	53.49	55.04	53.13	52.77	
ACK57560.1	81.75	100.00	89.80	52.95	56.83	55.81	55.10	
BAA36200.1	81.93	89.80	100.00	52.77	55.04	53.67	53.31	
AAM63409.1	53.49	52.95	52.77	100.00	68.94	72.86	73.04	
WP_003115652.1	55.04	56.83	55.04	68.94	100.00	72.89	73.97	
ACK57562.1	53.13	55.81	53.67	72.86	72.89	100.00	85.54	
BAA36202.1	52.77	55.10	53.31	73.04	73.97	85.54	100.00	

Así, se pudo simplificar el análisis de los resultados de BLASTX, al acotar el espectro seleccionando 5 referencias de las 7 totales para PhaCs de clase II, y que capturan la mayor parte de la diversidad observada: ACK57560.1; ACK57562.1; AAM63409.1; AAM63407.1 y WP_003115652.1. Dado que esta selección aborda una amplia diversidad de PhaCs de clase II, se pusieron parámetros más exigentes para determinar verdaderos positivos que coincidan con estas proteínas: un 60% de identidad y un 80% de cobertura.

Por lo tanto, en base a los intervalos y criterios definidos a lo largo de esta subsección 3.2, se identificaron genes codificantes de PhaCs de diferentes clases: 41 para PhaCs de la clase IA, 2 clase IB, 774 clase IIA, 756 clase IIB, 5 clase IIC, 3 clase IID y 1 clase X (MPC6). En ninguno de los genomas se encontraron genes codificantes de PhaCs de las clases III y IV, por lo que se descarta su presencia en los mismos. Los resultados correspondientes a las coordenadas de los genes codificantes de PhaCs encontrados, la clase y sus respectivos genomas se encuentran en el Apéndice Digital 5.

El control positivo, la cepa MPC6, permitió corroborar que la búsqueda se hizo con criterios acertados, puesto que se encontraron todos los genes codificantes para PhaCs que posee MPC6 y que son 5: 1 de clase I, 3 de clase II y la PhaC3 de clase X.

3.3 Principales características de las PhaCs identificadas en *Pseudomonas*

Con el fin de analizar las PhaCs identificadas en los genomas de la base de datos de *Pseudomonas*, se extrajeron las secuencias aminoacídicas que se reconocieron como PhaCs en base a los análisis mencionados anteriormente. Luego, para evaluar la diversidad presente de estos elementos y simplificar el alineamiento a realizar a continuación, es que las proteínas extraídas se agruparon con CD-HIT, buscando de esta forma obtener una representante de cada grupo. El resultado arrojó un solo grupo para las proteínas extraídas que fueron clasificadas como clase I y tres grupos para aquellas clasificadas como clase II (Apéndice Digital 6). A continuación, se presenta en la tabla 3 un resumen de los grupos obtenidos y la secuencia representante de cada uno de ellos considerando clase, ID del genoma y coordenadas.

Tabla 3. Secuencia representante de cada grupo en base a CD-HIT

	PhaCs clase I	PhaCs clase II		
Grupo	0	0	1	2
Clase	IA	IIB	IIB	IIA
ID Genoma	8	560	5	740
Coordenadas	787.653 - 789.257	3.713.939 - 3.715.708	428.935 - 430.617	2.364.192 - 2.365.871

Posteriormente, y dado que el resultado obtenido con CD-HIT no mostró una diversidad muy grande (un grupo en el caso de las clase I, y 3 grupos en el caso de las clase II), se decidió no usar exclusivamente las secuencias representativas para los alineamientos, sino todas las proteínas extraídas y realizar dos alineamientos múltiples: uno con las de clase I más la PhaC3, dado que el análisis de dominios conservados indicó una mayor cercanía de esta última con la clase I, y otro con las de clase II.

Para complementar el análisis anterior, se evaluaron los dominios conservados de las

representantes de cada grupo arrojado por CD-HIT y de la PhaC3 de la cepa MPC6, usando BLASTp. Se obtuvo que, en todos los casos, el dominio presente en cada una de ellas correspondió a phaC_N (pfam07167) y que representa la región N-terminal. Además, en todos los casos las coordenadas de la zona, en cuanto a la posición de los aminoácidos, es similar. En el caso de la representante de la clase I, el dominio conservado va desde el aminoácido 49 al 216; en el de la PhaC3 desde el 69 al 239; en el de la clase II grupo 0 desde el 75 al 243, en el grupo 1 desde el 76 al 244 y en el del grupo 2 desde el 75 al 243.

En los alineamientos generados para la clase I (Apéndice Digital 7) y la clase II (Apéndice Digital 9) se destacó el dominio conservado que se seleccionó en las representantes de cada grupo. En el caso de la clase I más la PhaC3, en la Figura 4, se observa una fracción del alineamiento, en la que se destaca el dominio conservado que va desde el aminoácido 49 al 216. En la zona abarcada por la franja color menta se observan partes con un mayor grado de conservación que otras, como muestra el color morado intenso de los aminoácidos en posiciones con alto grado de conservación. Se tiene que al inicio de la zona en cuestión -hacia N-terminal- hay un menor grado de conservación que hacia C-terminal.

Es posible notar que en más de una coordenada se ha generado un espacio o *gap*, el que ha sido reemplazado por un guión. Se observa que el grado de conservación para la PhaC3 de MPC6 (genoma 545) es bastante bajo.

En el caso de la clase II, se muestra en la Figura 5 una región del alineamiento donde se destaca la zona correspondiente al dominio conservado seleccionado y, ya que se tiene que no hay una variación considerable de la posición del aminoácido de inicio y de término del dominio conservado para cada una de las distintas representantes de los *cluster* de la clase II, se consideró este desde el aminoácido 75 al 244. Cabe mencionar que en la Figura 5 se muestran solo algunas de las PhaCs de clase II utilizadas para el alineamiento, mas se muestra la representante del *cluster* 1 y que corresponde a una PhaC de clase IIB del genoma 5. Se observa que el dominio

marcado con una franja color menta, contempla bastantes aminoácidos conservados, lo que se refleja tanto en el color morado intenso de los aminoácidos coloreados como en el color amarillo pálido y altura de las barras que indican el nivel de conservación, siendo más conservada la zona final (hacia el C terminal). Empero, también se observan zonas que no presentan tanta conservación.

Las figuras de los alineamientos completos para las PhaCs de clase I y para las PhaCs de clase II, se encuentran en Apéndice Digital 8 y 10, respectivamente.

3.4 Relaciones filogenéticas entre las PhaCs de *Pseudomonas*

Buscando comprender las relaciones filogenéticas entre las PhaCs presentes, a partir de los alineamientos de las PhaCs de cada clase, se infirieron dos árboles filogenéticos de máxima verosimilitud ejecutando PhyML (integrado en la plataforma NGPhylogeny.fr), uno para las PhaCs de clase I más la PhaC3 de MPC6 (Figura 6) y otro para las PhaCs de clase II (Figura 7).

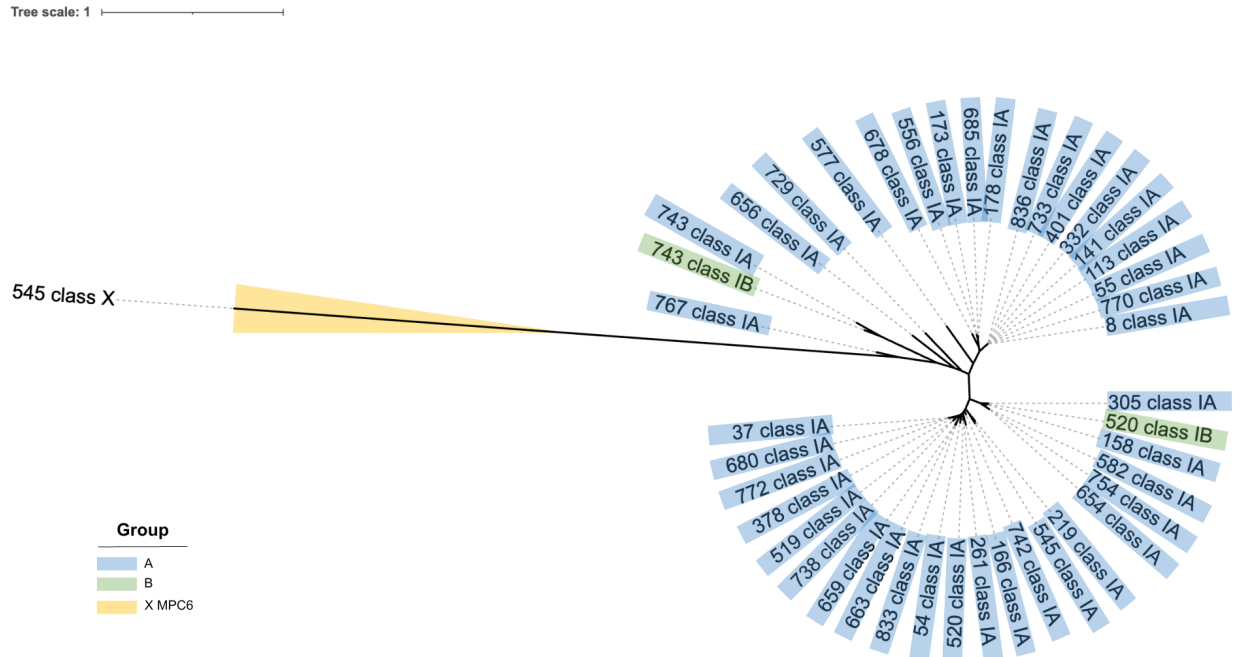


Figura 6. Árbol filogenético de las PhaCs de clase I extraídas desde las diferentes *Pseudomonas* de nuestra base de datos agrupadas como A o B según corresponda. Se incluye PhaC3 clase X de MPC6 y que corresponde al genoma 545.

Se observa que la PhaC3 de MPC6 queda bastante lejos del resto de PhaCs en el árbol (Figura 6). Lo que indica que, si bien el análisis de dominios conservados muestra una cercanía mayor a la clase I, aún así es bastante lejana a las PhaCs de dicha clase encontradas en las *Pseudomonas* de nuestra base de datos. Por otra parte, se observó una alta variabilidad entre enzimas de la clase I, distinguiéndose al menos 3 grandes clados que agrupan proteínas que comparten una mayor similitud. Se requieren más estudios para indagar sobre las posibles diferencias en términos de actividad que poseen las distintas variantes encontradas. Además, las

dos PhaC IB identificadas formaron clados mixtos con PhaC IA, indicando que, si bien son proteínas codificadas en distintas regiones del genoma, poseen una alta similitud de secuencia.

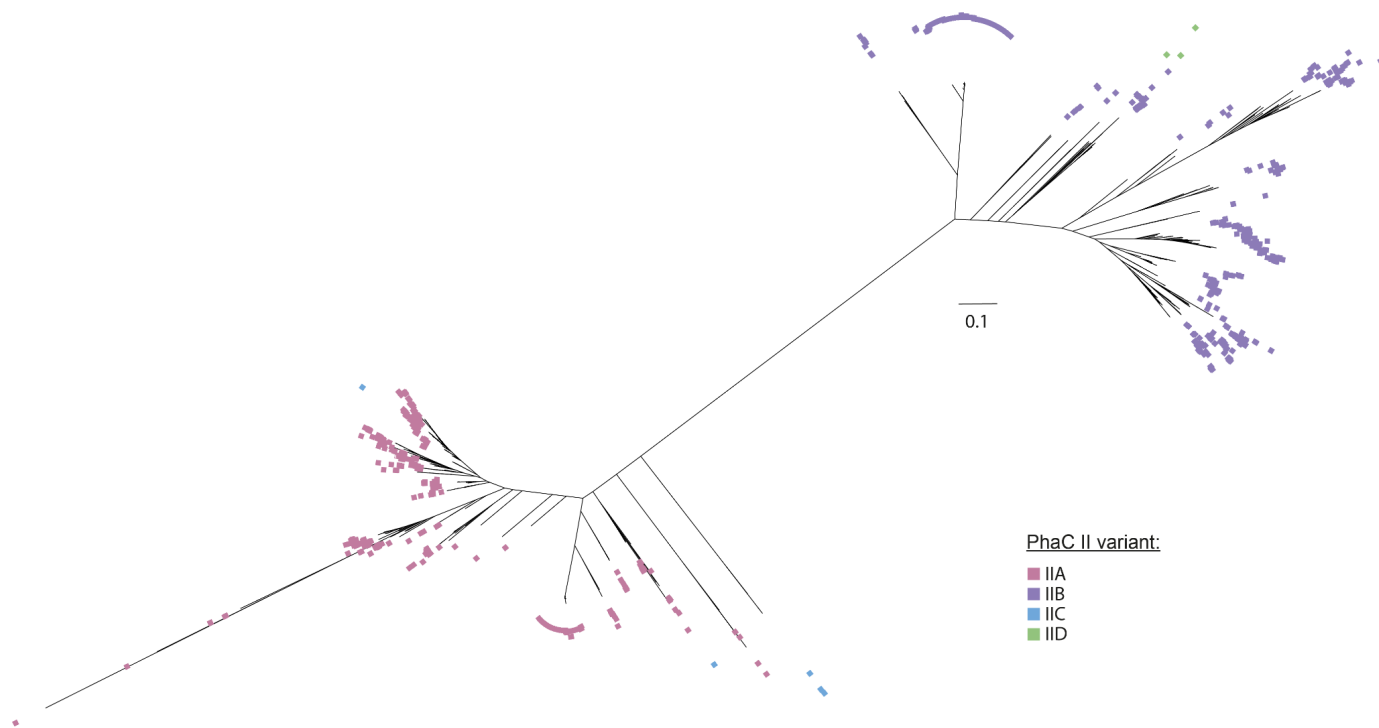


Figura 7. Árbol filogenético de las PhaCs de clase II extraídas desde las diferentes *Pseudomonas* de nuestra base de datos agrupadas como A, B, C o D, según corresponda.

En la Figura 7, se observa que al interior de la clase II existe una diversidad importante de PhaCs, formando parte de al menos 5 grupos bastante definidos. Teniendo en cuenta los resultados obtenidos con Clustal Omega (Tabla 2) al alinear las PhaCs de referencia de clase II para realizar la búsqueda en los genomas, era de esperar que se obtuviera una diversidad considerable al establecer el árbol filogenético en cuestión. Esto, dado que entre las mismas PhaCs de referencia, a pesar de ser de la misma clase, los porcentajes de identidad entre ellas son variados (desde 52,77% hasta 89,80%). En el árbol es posible observar dos grandes grupos, que corresponden a las dos polimerasas clase II que componen el *cluster* canónico (IIA y IIB). Las variantes poco frecuentes que aparecen en otras regiones del genoma, IIC y IID, corresponden a enzimas bastante cercanas a variantes IIA y IIB, respectivamente. Esto indica que no se trataría de enzimas adicionales con propiedades funcionales muy diferentes a las canónicas, sino más

bien copias adicionales de estas enzimas, que por razones desconocidas se encuentran en otros contextos genómicos. Por otra parte, es posible observar una gran variabilidad entre enzimas que pertenecen a cada uno de los dos grandes grupos mencionados anteriormente. Se requieren más estudios para indagar respecto a cómo esta variabilidad afecta las propiedades funcionales de dichas enzimas.

3.5 Relaciones filogenéticas entre *Pseudomonas* y asociaciones con PhaCs

Con el fin de establecer las relaciones filogenéticas entre las *Pseudomonas* de nuestra base de datos, se construyeron 2 árboles filogenéticos en base a la secuencia genómica de dichas cepas usando la herramienta MASHTREE. Dicha inferencia filogenética fue complementada con el resultado de la asignación de especies obtenida con GTDB-Tk (Apéndice Digital 2).

El primer árbol, incorpora el universo casi total de *Pseudomonas* de la base de datos (Figura 8), puesto que solo hay presencia de 842 *Pseudomonas*. Esto, porque, sumado a la eliminación de los genomas 196, 510, 517 y 722, fue necesario excluir 13 genomas más, debido a que el primer árbol arrojado por MASHTREE presentaba genomas que mostraban una distancia anormalmente grande con respecto al resto de cepas, lo que probablemente se debe a problemas en el ensamblaje de dichos genomas. Los genomas eliminados fueron, en base al ID asignado en nuestra base de datos: 005, 017, 018, 019, 022, 272, 285, 333, 543, 577, 759, 786 y 790. Luego de su eliminación, se ejecutó nuevamente MASHTREE obteniéndose así el árbol definitivo. A este árbol, se le agregó, en base a la identificación y clasificación de PhaCs realizada para cada una de las cepas, la información correspondiente a ausencia o presencia de PhaCs para cada una de las clases y el bootstrap.

Cabe mencionar que los genomas que fueron eliminados, solamente lo fueron para el análisis filogenético entre las distintas *Pseudomonas*, pues estos sí se consideraron para la identificación y extracción de de PhaCs y los análisis relacionados a ello.

De acuerdo con lo esperado, se observó la presencia ubicua en casi todo el árbol de las PhaCs IIA y IIB (Figura 8), las cuales corresponden a las dos de clase II que conforman el *cluster* génico canónico para la producción de PHA en *Pseudomonas*. Una observación interesante, es que ciertos genomas poseen además PhaCs de otras clases, lo que no es común en *Pseudomonas*; en la literatura se menciona que, de hecho, poseen un *cluster* bien conservado, que incluye específicamente dos PhaCs de clase II (Mozejko-Ciesielska y col., 2019). Por otra parte, se identificaron algunas cepas que carecen totalmente de genes que codifiquen PhaCs. Se observó que en general, la presencia o ausencia de PhaCs es un patrón relativo a la especie, por ejemplo, se observa que ninguna cepa de *Pseudomonas syringae* ni de *Pseudomonas stutzeri* posee PhaCs, siendo algo común a la especie. En el caso de *Pseudomonas aeruginosa* se observa que casi todas las cepas poseen ambas PhaCs de clase II y solamente 6 de ellas poseen, adicionalmente, una PhaC de clase IA. Lamentablemente, la gran mayoría de los genomas secuenciados de *Pseudomonas* presente en bases de datos corresponde a *P. aeruginosa*, dejando a otras especies bastante menos representadas. Esto impide un análisis más exhaustivo de la presencia de determinadas PhaCs al interior de ciertos linajes.

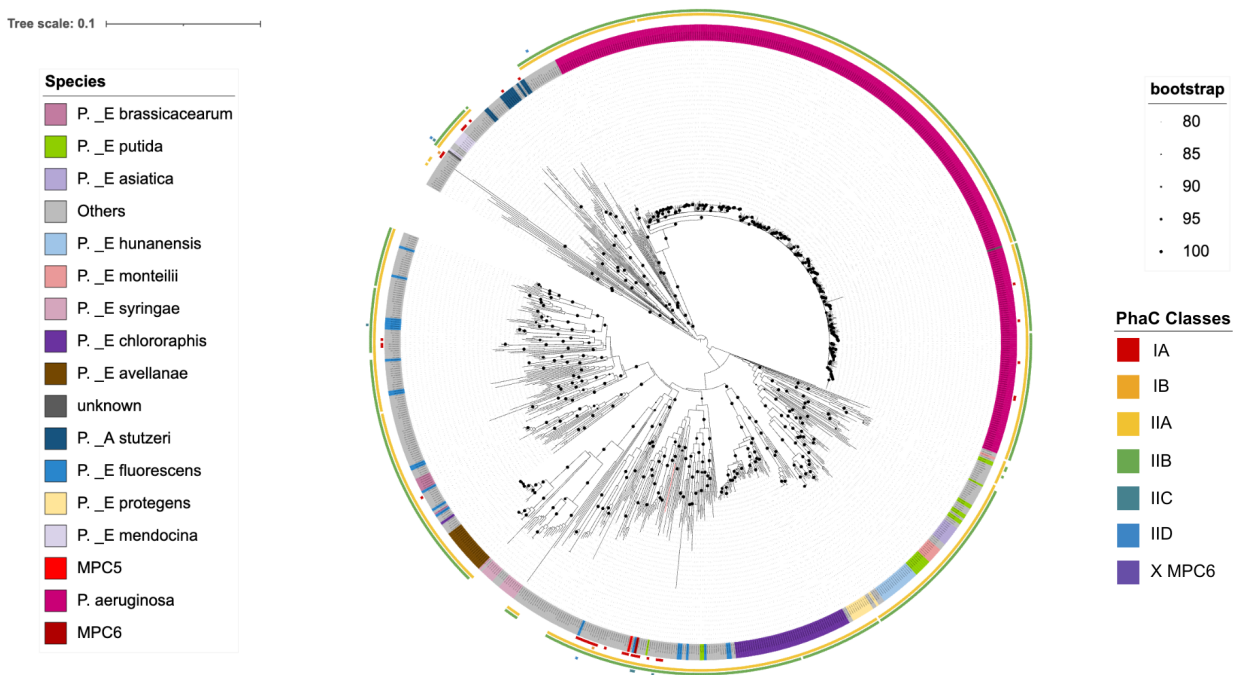


Figura 8. Árbol filogenético que incorpora 842 *Pseudomonas* de nuestra base de datos, se muestran las especies más abundantes, la presencia de PhaCs de diferentes clases y el bootstrap.

Para mostrar más claramente si existen patrones en el contenido de PhaCs característicos de cepas antárticas, se construyó un segundo árbol, que incluye solo 33 genomas de *Pseudomonas* originarias de dicho continente (Figura 9). En él, se muestra la presencia o ausencia de las distintas clases de PhaCs, y se destacan las cepas MPC5 y MPC6 en colores azul cerúleo y azul rey, respectivamente. Llama la atención la alta diversidad de especies capturada en este set relativamente pequeño de genomas de *Pseudomonas* antárticas, lo que anticipa una gran cantidad de especies por conocer de este género a nivel genómico en el continente blanco. 29 de las 33 cepas poseen las PhaCs IIA y IIB, es decir, aproximadamente el 87% de las cepas antárticas de *Pseudomonas*. Sumado a ello, 8 de estas cepas presentan además otra PhaC, ya sea de clase I, II o X. Entre estas, destacan el aislado *Pseudomonas* sp. ACM7 (sp004a36015, genoma 520), el cual posee dos PhaCs de clase I, la cepa MPC5 y otras de distintas especies que poseen una PhaC de clase I, y *P. fluorescens* (genoma 858) que posee una tercera PhaC de clase II. Por otra parte, llama la atención la ausencia de PhaCs en *P. versuta* y *P. paraversuta*, considerando que los PHAs podrían tener un rol en la adaptación al ambiente extremo.

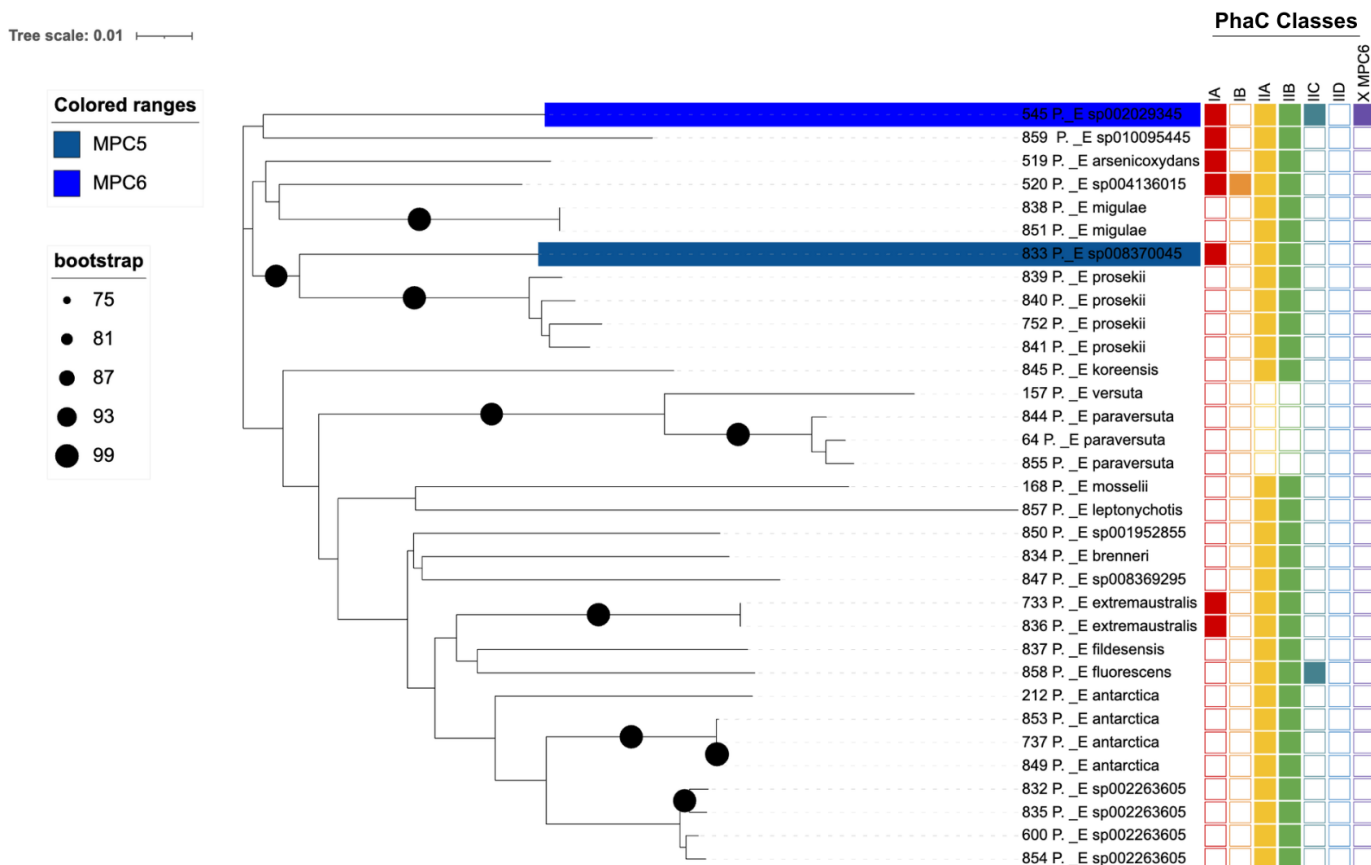


Figura 9. Árbol filogenético de *Pseudomonas* antárticas y la correspondiente presencia de PhaCs de diferentes clases.

En síntesis, los resultados presentados anteriormente indican que el continente Antártico alberga una gran diversidad de cepas y especies del género *Pseudomonas*, probablemente muchas de ellas aún por descubrir, parte de las cuales poseen una dotación de enzimas PhaC bastante peculiar, y con ello, un potencial por explorar para la síntesis de nuevas variedades de PHAs. Por otra parte, se constató que el género *Pseudomonas* alberga una gran diversidad de enzimas PhaCs de distintas clases, cuyas peculiaridades en términos funcionales se desconocen. Se requieren más estudios para conocer la diversidad de PhaCs de *Pseudomonas* en Antártica y otros ambientes del Planeta.

4. DISCUSIÓN

Las PhaCs son enzimas fundamentales en la síntesis de PHA que catalizan la unión de los respectivos monómeros para formar polímeros de este tipo de sustancias (Neoh y col., 2022). Por ello, la predicción de estos genes en genomas es crucial para evaluar el potencial de síntesis de PHAs en diferentes representantes bacterianos y descubrir nuevas PhaCs ventajosas para su uso industrial; por ejemplo, que permitan abaratar costos de producción o mejorar las propiedades mecánicas de los polímeros. En este sentido, el presente estudio evaluó la presencia de PhaCs en 855 cepas del género *Pseudomonas*, indagando sobre el potencial de este género para producir PHAs, e identificar un grupo de PhaCs no descritas anteriormente.

Es importante destacar que un estudio de estas características no había sido realizado previamente. Hasta ahora, se sabía que: a) el género *Pseudomonas*, junto a *Janthinobacterium*, es uno de los mayores productores de PHA dentro de los microorganismos aislados desde suelo antártico (Pacheco y col., 2019); b) el cluster génico para la producción de PHA en *Pseudomonas* es conservado y cuenta con la presencia de dos PhaCs de clase II (Mozejko-Ciesielska y col., 2019); c) no todas las cepas de *Pseudomonas* tienen la capacidad de producir PHA; y d) hay un alto interés en este género por el potencial que posee para la producción de PHAs (Mozejko-Ciesielska y col., 2019). Sin embargo, no se había evaluado el potencial para producir PHAs de este género, en cuanto a la dotación de enzimas PhaCs claves en este proceso, en un grupo numeroso de aislados. Sumado a esto, tampoco se había realizado una caracterización enfocada en dichos términos y que considerara un alto número de cepas antárticas de *Pseudomonas*. Esto último no solamente es relevante por los aportes en cuanto a creación de conocimiento, sino que también por sus contribuciones al desarrollo de la ciencia antártica con miras hacia soluciones para la producción de materiales sustentables como el PHA.

Entonces, es a partir de este estudio y de la literatura revisada, que es posible determinar ciertas características del género *Pseudomonas* en cuanto a las capacidades de producción de

PHA, asociadas a la presencia de PhaCs, que no se habían descrito previamente, haciéndose un especial énfasis en cepas antárticas.

Se ha mencionado en varias secciones del presente trabajo que el *cluster* génico para la producción de PHAs en *Pseudomonas* contempla dos PhaCs de clase II (Mozejko-Ciesielska y col., 2019). Es importante tener en cuenta que no todas las cepas del género *Pseudomonas* tienen la capacidad de sintetizar PHA. Es de esperar que, si una cepa de este género sí la tiene, esta posea las dos PhaCs de clase II del *cluster* en cuestión. Justamente, en este estudio se encontró que, de las 855 cepas de *Pseudomonas* de nuestra base de datos, 738 poseen 2 PhaCs de clase II, lo que ya establece una tendencia. De estas 738 cepas, 6 poseen una tercera PhaC de clase II y 1 posee una cuarta PhaC de clase II. Dada lo bien conservado que es el *cluster* génico para la producción de PHA en *Pseudomonas*, el que contempla dos PhaCs de clase II (Mozejko-Ciesielska y col., 2019), llama la atención encontrar cepas que posean 3 o más PhaCs de clase II, entre ellas MPC6 que posee 3 y que fue descrita por nuestro grupo de investigación el año 2019. Más aún, se encontraron 42 genomas que poseen PhaCs de clase I, algo muy poco usual, si se considera lo anteriormente mencionado respecto del *cluster* génico para la producción de PHA. Aún más llamativo, es el hecho de que, de esas 42 *Pseudomonas*, dos tienen 2 PhaCs de clase I. Si bien otros estudios han reportado PhaCs de clase I en cepas de *Pseudomonas*, estos casos son poco usuales (Ayub y col., 2007). Cabe mencionar que, de estos casos poco frecuentes, algunos pertenecen a cepas antárticas del género en cuestión (Ayub y col., 2007; Tan y col., 2020). Esto a su vez resulta interesante, pues va en línea con el haber encontrado en este trabajo 7 cepas antárticas con PhaCs de clase I.

Dentro de las novedades encontradas, resulta interesante el caso de MPC6. Si bien ya se tenía conocimiento de que poseía 5 PhaCs y que una de ellas no pudo ser clasificada -PhaC3- (Orellana-Saez y col., 2019), no se había realizado un estudio más detallado de sus PhaCs, comparando con otras cepas Antárticas. Así, en este trabajo, a partir del análisis de dominios

conservados se encontró que la PhaC3, sería más cercana a las PhaCs de clase I, sin embargo, se observa en el árbol filogenético que contempla las PhaCs de clase I (Figura 6) que esta queda alejada del resto de PhaCs de clase I encontradas en nuestra base de datos de *Pseudomonas*, lo que podría indicar que podría pertenecer a una nueva clase de PhaCs. Esta PhaC3 es algo bastante excepcional, puesto que en los resultados arrojados por BLASTX no se encontraron coincidencias significativas con esta en ningún otro genoma de nuestro set además del de MPC6. No obstante, al realizar una búsqueda más amplia mediante BLASTp contra la base de datos NCBI no redundante, se identificaron cuatro cepas de *Pseudomonas* que codifican una proteína con sobre un 95% de identidad respecto a la PhaC de clase X de MPC6. Lo anterior indica que, si bien esta proteína es muy poco frecuente, no sería un artefacto solo presente en MPC6.

Se ha visto que los microorganismos productores de PHA tienen una ventaja asociada a la producción de este polímero ante condiciones de estrés abiótico como lo son el osmótico, las limitaciones de nutrientes, el frío, entre otras (Pacheco y col., 2019) (Mozejko-Ciesielska y col., 2019). Es por ello, que se podría pensar que en un ambiente antártico donde este tipo de condiciones son muchas veces la norma, la producción de PHA en cepas de *Pseudomonas* del territorio en cuestión sea algo común, siendo esto una ventaja adaptativa. Lo que no sería extraño, si se considera que el género *Pseudomonas* tiene un alto grado de adaptación genético y fisiológico (Spiers y col., 2000). En línea con ello, se podría pensar que una cuestión tan llamativa como que MPC6 posea 5 PhaCs podría deberse a las ventajas que otorga el producir PHA y, justamente, el tener más de una PhaC y de diferentes clases podría traer beneficios en ese sentido. Dado que las PhaCs de clase I utilizan monómeros de cadena corta y las de clase II de cadena mediana, podría ser que la presencia de PhaCs de clase I o cercanas, como es el caso de la PhaC3, responda a una adaptación que permite un uso más amplio de monómeros. De hecho, el PHA de MPC6 -de carácter novedoso- corresponde a un copolímero de cadena corta y mediana, es decir un scl-co-mcl-PHA, donde más del 80% de su composición corresponde al monómero

3-hidroxi-butirato (3HB) el que justamente corresponde a uno de cadena corta (Pacheco y col., 2019). Entonces, podría ser que esta PhaC de clase I y la PhaC3 estén involucradas en el uso de dicho monómero para la síntesis del scl-co-mcl-PHA.

De ser cierto que las PhaCs de clase I representan una ventaja adaptativa, y considerando que una de esta clase fue encontrada en una cepa de *Pseudomonas* antártica (Orellana y col., 2019), se podría pensar que esta adaptación podría ser común a este grupo como una respuesta al alto estrés abiótico que enfrentan los microorganismos en el continente blanco. En concordancia con esto, en el presente trabajo se encontró que, si bien no todas las *Pseudomonas* antárticas del estudio presentan PhaCs, por lo que la producción de PHAs no sería inherente a cepas antárticas de *Pseudomonas*, sí ocurre que varias de ellas tienen características que llaman la atención como poseer más de 2 PhaCs de clase II e incluso poseer PhaCs de clase I (Figura 9). Por ello, podría ser que la excepcionalidad de MPC6 de poseer una PhaC de clase I y poseer más de una PhaC de clase II, sea una característica de algunas *Pseudomonas* antárticas. Sin embargo, es importante notar que estas anomalías no son exclusivas de este grupo, puesto que de los 42 genomas que poseen PhaCs IA, 7 son antárticos; de los 2 que poseen PhaCs IB, 1 es antártico; de los 5 que poseen PhaCs IIC, 2 son antárticos y, finalmente, de los 3 que poseen una IID, ninguno es antártico. Cabe recordar que la clasificación correspondiente IA y IB, se hizo para distinguir entre diferentes *matches* en el genoma con la única PhaC de clase I de referencia. Y en el caso de las clases IIA, IIB, IIC y IID, para distinguir entre distintos tipos de regiones del genoma que alinean dentro de la misma categoría de PhaCs de clase II.

Resulta interesante notar que, dado que se encontró que MPC5 poseería una PhaC de clase I, podría esperarse que produjera PHA con monómeros de cadena corta como lo hace MPC6, sin embargo, el PHA de MPC5 fue caracterizado por nuestro grupo de investigación y no posee ese tipo de monómeros (Pacheco y col., 2019). Esto último arroja una serie de interrogantes: ¿Puede ser que el gen de la PhaC de clase I de MPC5 se expresa en ciertas

condiciones que no fueron las del estudio realizado por nuestro equipo y por eso su PHA no tiene monómeros de cadena corta? Para responder esta pregunta, habría que cultivar MPC5 en otras condiciones y evaluar si hay transcritos del gen y la composición de su PHA. Otra pregunta que se desprende es ¿Será que en MPC6 no es la PhaC de clase I la que está involucrada en el uso de monómeros de cadena corta, sino que en realidad sea su PhaC3? ¿Será que las PhaCs de clase II de MPC6 pueden usar monómeros de cadena corta? Se requieren estudios adicionales para responder estas preguntas.

Cabe mencionar que, dado que MPC5 y MPC6 pueden producir altas cantidades de PHA en amplios rangos de temperatura (Pacheco y col., 2019), se podría esperar que tuvieran ciertas similitudes en cuanto a las PhaCs que poseen. Por su parte, MPC5 posee 3 PhaCs siendo dos de ellas de clase II y una de clase I, mientras que MPC6 tiene 5, donde 3 pertenecen a PhaCs de clase II, una a la clase I y una PhaC3 cercana a esta clase. Por lo que solamente comparten el tener 2 PhaCs de clase II y la anomalía de poseer una de clase I. Cabe mencionar que ambas cepas fueron aisladas desde el mismo sitio en Isla Decepción, Antártica, sin embargo, pertenecen a especies diferentes según los resultados entregados por GTDB-Tk y ANI. Este último entregó un valor menor a 90% (Apéndice Digital 3). Esto podría explicar las diferencias en relación con la presencia de PhaCs en ambas.

A partir de los análisis que se han hecho en torno a la PhaC3 de MPC6, se trataría de una propiedad bastante particular de dicha cepa, puesto que ninguna otra cepa ni antártica ni de otros lugares del mundo presente en nuestra base de datos posee una PhaC similar. Podría ser que el PHA novedoso de MPC6 se deba, en parte, a esta PhaC3. Resulta interesante señalar que en un estudio donde se evaluaron las PhaCs de cuatro cepas aisladas de la Antártica, una cifra bastante menor a la del presente estudio, donde dos corresponden al género *Pseudomonas* y dos a *Janthinobacterium*, utilizando herramientas bioinformáticas como las de este trabajo, se encontraron dos rarezas: a) PhaCs de clase I en cepas de *Pseudomonas* y b) una PhaC que no

pudo ser clasificada en *Janthinobacterium*, debido a que queda en un clado separado de las clases I, II, III y IV, por lo que proponen una nueva clase, la V, basada en criterios filogenéticos (Tan y col., 2020). Por lo que podría ser que Antártica sea un reservorio de PhaCs novedosas y donde es posible encontrar de clase I en el género *Pseudomonas*.

Es importante destacar que, si bien varias de las cepas antárticas con presencia de PhaCs de este estudio tienen anomalías como el poseer más de dos PhaCs de clase II o poseer PhaCs de clase I, ninguna posee tantas PhaCs como MPC6, siendo esta una característica que llama la atención. Si bien la mayoría de las cepas antárticas fueron aisladas desde la península antártica, no habiendo así mucha variación territorial en su origen, sí son bastante diversas en cuanto a los ambientes desde los que fueron aisladas en dicho continente (muestras de agua, suelo, plantas, entre otros). Considerando esto y con relación a la presencia o ausencia de genes codificantes para PhaCs, resulta pertinente preguntarse si es que estas diferencias genéticas podrían deberse a que las condiciones abióticas en la Antártica no son uniformes, pues cambian según el ambiente. Así, podría ser que la diversidad en relación con la presencia de PhaCs se debiera a diferencias abióticas dependientes del ambiente, los que no son iguales para cada cepa. Es decir, la presencia de las PhaCs podría estar condicionada por los factores abióticos del lugar.

Actualmente, la clasificación de las PhaCs (I a IV) pasa por tres criterios propuestos por Rehm: a) especificidad de sustrato *in vivo*, b) subunidades que las componen (PhaC, PhaE or PhaR) y c) secuencia aminoacídica (Rehm, 2003). Sin embargo, el hecho de encontrar elementos fuera de lo común en cepas de *Pseudomonas* como PhaCs de clase I y una PhaC3 no clasificable en MPC6, dan pie para pensar que las clasificaciones actuales de las PhaCs podrían presentar problemas ante la ampliación de la diversidad conocida de este tipo de enzimas. Sumado a ello, ocurre que a la hora de usar herramientas bioinformáticas que permiten dilucidar cuestiones como relaciones filogenéticas, conservación de dominios y similitudes de secuencia, como se hace precisamente en este trabajo, se obtienen porcentajes de identidad muy amplios y variables entre

PhaCs de una misma clase (entre 50% y 80%) y PhaCs que quedan muy lejos en el árbol, no siendo posible su clasificación. Podría ser que la situación de las PhaCs clasificadas como de clase I y la de la PhaC3 de clase no determinada, en base a análisis bioinformáticos, cambiara si se evaluara con los dos primeros criterios mencionados al inicio de este párrafo (a y b), pudiendo pasar las PhaCs inicialmente clasificadas como de clase I a ser de clase II y la PhaC3 podría ser considerada dentro de algunas de las clases existentes. En línea con este cuestionamiento a la forma de clasificación actual de las PhaCs, cabe mencionar que otros investigadores han mencionado que, en general, las PhaCs presentan una gran variación en la similitud de secuencia aminoacídica desde 8 a 96% (Tan y col., 2020). De hecho, en el estudio previo en que se caracterizó MPC6, a su PhaC1 le fue asignada la clase II, teniendo esta un 53% de identidad con una PhaC de dicha clase y que fue usada como referencia para la clasificación a partir de alineamientos de secuencias y la construcción de un árbol filogenético (Orellana-Saez y col., 2019).

Considerando que las PhaCs de clase I y II no poseen una subunidad accesoria, a diferencia de las PhaCs de clase III y IV, podría ser que esto, sumado a que, quizás, la especificidad de sustrato por parte de estas PhaCs pueda lograrse con secuencias no muy conservadas –lo que explicaría los porcentajes de identidad tan variables entre PhaCs de una misma clase–, esté dando lugar a errores en las clasificaciones. Esto, pues se obtienen porcentajes de identidad muy bajos admitiendo un espectro mayor, haciéndose así poco claro el límite entre las clases I y II. Así, los análisis bioinformáticos resultan fútiles al intentar alinearse con el criterio de la especificidad de sustrato, mencionado en el párrafo anterior, establecido para la clasificación de PhaCs.

Más antecedentes que respaldan la problemática de la clasificación de las PhaCs, son que, si bien las PhaCs de clase I, III y IV sólo pueden incorporar monómeros de scl-PHA y las de clase II solo mcl-PHA, en un estudio (Neoh y col., 2022) se menciona que hay excepciones y se

nombran PhaCs muy similares a la clase I, pero que pueden incorporar mcl-PHA. En ese mismo trabajo también se mencionan la PhaC1 y la PhaC2 de *Pseudomonas sp. 61-3*, que tiene una alta similitud con PhaCs de clase II, mas puede incorporar monómeros de scl-PHA.

Es por todo aquello, que se plantea que aún falta incorporar de mejor manera la dimensión bioinformática en las clasificaciones de PhaCs. Es importante mencionar que el área de clasificación de las PhaCs se encuentra en desarrollo, lo que se evidencia en que, por ejemplo, hasta el año 2022 solamente existían dos PhaCs cuyos dominios catalíticos poseían una estructura resuelta, una de *C. necator* y otra de *Chromobacterium sp. USM2* (Neoh y col., 2022). Así, cuestionamientos apoyados en evidencia en cuanto a la clasificación de las PhaCs, como los planteados aquí, resultan pertinentes para el progreso de esta área.

En cuanto a los análisis de dominios conservados y alineamientos múltiples, se tiene que tanto para las PhaCs de clase I como para las de clase II, el dominio conservado presente en todas las representantes de cada uno de los *clusters* arrojados por CD-HIT corresponde al phaC_N (pfam07167) y que representa la región N-terminal (Figura 4 y Figura 5). Con relación a esto, hasta el año 2022 la función exacta del dominio en cuestión no había sido determinada, de hecho, se han realizado varios estudios por parte de distintos investigadores con el objetivo de resolver esta interrogante, sin embargo, esta sigue sin poder ser contestada (Neoh y col., 2022). Sumado a ello, su estructura no se ha podido obtener mediante cristalización; además, se plantea que este dominio varía en cuanto a su longitud entre diferentes PhaCs (Neoh y col., 2022).

Respecto de las relaciones filogenéticas entre las *Pseudomonas* de nuestra base de datos, llama la atención que en el árbol que incorpora a las 842 cepas de *Pseudomonas* (Figura 8), el grupo de *Pseudomonas fluorescens* se encuentra distribuido en distintos lugares de aquel, teniéndose grupos parafiléticos. Esto, plantea la pregunta de si es que la especie *fluorescens* estaría abarcando *Pseudomonas* de otras especies y, justamente, esto concuerda con lo que se ha planteado recientemente en la literatura: *Pseudomonas fluorescens* sería un complejo de especies,

siendo este el grupo más diverso dentro del género *Pseudomonas* (Garrido-Sanz y col., 2017). Si bien este fenómeno de tener a individuos de la misma especie repartidos en distintos lugares del árbol (Figura 8) se da con otras especies además de *fluorescens*, el fenómeno es mayor o más notorio en esta. En el caso de MPC5, se tiene que el análisis de GTDB-Tk la asigna como sp008370045 (Apéndice Digital 2), por lo que estaría pendiente de bautizar en cuanto a su clasificación como especie. Al compararla con MPC6, se observa que difieren lo suficiente en el árbol como para ser diferentes especies (Figura 8 y Figura 9), además, la asignación taxonómica entregada por GTDB-Tk arrojó especies diferentes para ambas, en el caso de MPC6 le asigna la especie sp002029345 (Apéndice Digital 2) y en el estudio realizado por nuestro equipo de investigación se propone una nueva especie: *frigusceleri* (Orellana y col., 2019). Sumado a ello, el porcentaje ANI entre ambas cepas es menor a 90%.

Otra cuestión que también viene a develar las dificultades en cuanto a las clasificaciones taxonómicas es el hecho de que nuestra base de datos inicialmente contemplara 859 cepas de *Pseudomonas*. Sin embargo, tuvieron que ser eliminadas 17 de ellas para los análisis filogenéticos de las cepas, debido a lo lejos que quedaban en el árbol -pudiendo no ser realmente bacterias de este género- o a que GTDB-Tk les asignó otro género o a que no les pudo asignar uno, quedando finalmente 842 cepas para la construcción del árbol filogenético de *Pseudomonas* (Figura 8) que, se plantea, serían del género en cuestión.

Al buscar asociaciones entre filogrupos del árbol que contempla a las 842 cepas (Figura 8) y la presencia de PhaCs y sus clases, es evidente notar que la mayoría de los genomas presentan 2 PhaCs de clase II, específicamente IIA y IIB, indicando la posible capacidad de producción de PHA. Esto va en línea con el antecedente que se tiene de que el *cluster* génico para la producción de PHA está bastante conservado en este género y, justamente, contempla 2 PhaCs de clase II. Dado que es un grupo minoritario el que no las posee, se evidencia la preponderancia de los individuos de este género en cuanto a capacidad de síntesis de PHA. Por otro lado, resulta

interesante notar que la mayoría de casos en los que se encontraron PhaCs de clase I, se encuentran acumulados a las 6:30 hrs. del árbol (especies *P. fluorescens*, *P. mucoides*, *P. mandelii*, *sp003151075*, *P. arsenicoxydans*, *P. sp004136015*, *P. frederiksbergensis*, *P. lini*, *P. sp012935755*, *P. sp008370045*, *P. sp002029345*, *P. silesiensis*, *P. sp010095445*, *P. sp002356535* y *P. umsongsensis*) (Figura 8), lo que podría indicar que entre individuos cercanos filogenéticamente se da esta característica, sin embargo, también hay otros casos de PhaCs de clase I en otros puntos del árbol. A partir de esto, se desprende que podría ocurrir que en algunos casos haya una relación entre la filogenia de las cepas y la presencia de anomalías como presencia de PhaCs de clase I, pero también puede ocurrir que en otros casos no sea así, y que la presencia esporádica de PhaCs de clase I en un filogrupo se pudiera deber a adquisición por transferencia horizontal. Se requieren estudios adicionales que examinen en detalle las regiones del genoma donde se encuentran codificadas estas proteínas para identificar posibles elementos genéticos móviles que puedan estar movilizandando estos genes.

Lo anterior respondía a un panorama más general, pero al ir a un espectro más acotado como el de las *Pseudomonas* antárticas, se tiene que, en relación a la diversidad filogenética de estas, se observa (Figura 9) que se forman diferentes ramas, indicando la presencia de varias especies distintas. También, resulta interesante notar que aquellas que son más cercanas comparten similitudes asociadas a la presencia o ausencia de PhaCs, por ejemplo, todas las cepas antárticas que no poseen ninguna PhaC se encuentran cerca. Lo mismo ocurre con la mayoría de las cepas que poseen una o dos PhaCs de clase I, están casi todas cerca formando 3 grupos: uno de 4 cepas, otro de 2 y otro de 1 que es la MPC5. Se observa que las cepas de una misma especie poseen las mismas clases de PhaCs; todas las cepas de las especies *prosekii*, *migulae*, *antarctica* y *sp002263605* poseen PhaCs de clase IIA y IIB; todas las *extremaustralis* poseen una PhaC de clase IIA, IIB y IA. En el caso de aquellas cepas que no poseen PhaCs, resulta que todas son

paraversuta o *versuta*. Estos resultados plantean que hay una relación directa entre ausencia o presencia de PhaCs, la clase de estas, y las relaciones filogenéticas entre las *Pseudomonas*.

Finalmente, es importante dar cuenta de que son necesarios más estudios para poder confirmar que la capacidad de producción de PHA predicha a partir de la presencia de PhaCs existe, para lo que es necesario evaluar si es que están presentes los *cluster* génicos completos para la producción de PHA y no solamente las PhaCs. Además, se requiere hacer las pruebas experimentales que corroboren la producción de este tipo de compuestos, mas esto también tiene sus propias limitaciones, pues es de conocimiento común entre la comunidad científica que no todos los microorganismos son cultivables, sumando esto un grado de dificultad considerable. Respecto de esto último, justamente es en esos casos -en los que hay microorganismos no cultivables- que herramientas *in silico*, como las de la bioinformática, ofrecen la posibilidad de caracterizarlos y dar un acercamiento al que no se puede llegar experimentalmente.

5. CONCLUSIONES

El presente estudio permitió identificar y clasificar genes de PhaCs presentes en genomas secuenciados de *Pseudomonas* provenientes de Antártica y otras regiones del planeta, cumpliéndose así el objetivo 1. También, permitió evaluar la distribución de las PhaCs identificadas en distintos linajes de *Pseudomonas*, encontrándose asociaciones entre tipos de PhaCs, lugar de origen, y filogrupos característicos de este género, lográndose el objetivo 2. Así, se pudo cumplir el objetivo general de este trabajo: estudiar la diversidad y distribución filogenética de genes de Polihidroxicanoato sintasas (PhaCs) en *Pseudomonas* antárticas, estableciéndose comparaciones con cepas de *Pseudomonas* de todo el globo.

Este estudio permitió observar que la situación canónica en *Pseudomonas*, donde solo habrían PhaCs de clase II pareciera no ser tan así, pues la búsqueda realizada apunta a la presencia de PhaCs de clase I en varias de las cepas analizadas. Por otro lado, no se encontraron PhaCs de clase III ni IV, las cuales están presentes en organismos filogenéticamente distantes a *Pseudomonas*. También, se pudo establecer que la presencia de características que llaman la atención como PhaCs de clase I, o más de una PhaC de clase II no es restrictiva a un solo filogrupo, sino que estas ocurren en varios diferentes. En concordancia con esto, se encontró que habría una relación entre ausencia o presencia de PhaCs y sus clases y las relaciones filogenéticas entre las cepas de *Pseudomonas*. Sin embargo, los casos de presencia esporádica de PhaCs de clase I en el árbol sugieren que habría adquisición por transferencia horizontal en ellos.

Con relación a las cepas de *Pseudomonas* originarias de Antártica, se determinó que la capacidad productora de PHA no sería una característica propia de este grupo, puesto que varios de sus miembros no poseen ninguna PhaC. Sin embargo, sí se encontraron anomalías interesantes, como la presencia de más de dos PhaCs de clase II y PhaCs de clase I. Sumado a esto, se logró establecer que la presencia o ausencia de PhaCs y sus clases tienen relación con

los vínculos evolutivos de las cepas. Esto, se evidenció en que las cepas cercanas en el árbol de *Pseudomonas* antárticas comparten el tener las mismas clases y cantidades de PhaCs, a excepción de MPC6 que posee más PhaCs que cualquiera de las 855 *Pseudomonas* restantes de nuestra base datos. En relación con esto último, se comprobó que MPC6 sería una cepa absolutamente fuera de lo común al poseer 5 PhaCs y, además, entre ellas una novedosa que no pudo ser clasificada: la PhaC3.

En cuanto a la PhaC3, que fue encontrada en un estudio previo realizado por nuestro grupo de investigación y cuya clase no pudo ser determinada, los análisis realizados en el presente trabajo permitieron hacer avances en dicha materia, al encontrarse que de todas las clases de PhaCs existentes, esta sería más cercana a la I. Sin embargo, sigue siendo bastante lejana filogenéticamente al resto de la PhaCs de clase I de *Pseudomonas* como se pudo apreciar en el árbol que agrupa las PhaCs de clase I y esta PhaC3. Por ello, se requieren mayores estudios para determinar si esta corresponde a un nuevo tipo de PhaC, *ergo* sería una nueva clase.

Los resultados obtenidos a lo largo de los distintos análisis del presente estudio dieron cuenta de una problemática importante: la clasificación propuesta para las PhaCs requiere mejoras para poder establecer límites claros. Esto, puesto que los rangos de identidad de secuencias aminoacídicas entre PhaCs son muy amplios para enzimas de una misma clase, por lo que clasificar PhaCs basado en esto no sería certero ni suficiente, requiriendo complementar con análisis filogenéticos, que tampoco logran satisfacer completamente la clasificación, pues se obtienen varios grupos de PhaCs de una misma clase bastante lejanos entre sí. Sumado a ello, el criterio de especificidad de sustrato no es fácil de medir y podría no alinearse con el de identidad de secuencias. Así, este estudio aporta una gran cantidad de data que respalda la problemática en cuestión, contribuyendo al desarrollo del área de clasificación de las PhaCs.

Finalmente, este trabajo se posiciona como el primero en el que se evalúa el potencial de producción de PHA asociado a la presencia de PhaCs y las clases de las mismas en una muestra

tan grande de *Pseudomonas* y que considera cepas de todo el mundo -855 cepas-, recogiendo una cantidad importante del continente antártico -33 cepas-, estableciendo asociaciones entre relaciones filogenéticas y presencia o ausencia de PhaCs y sus clases. Además, debido a la magnitud de cepas evaluadas, este trabajo entrega un número importante de PhaCs que no habían sido identificadas, las que pueden ser estudiadas para su aplicación en la producción de PHAs, contribuyendo así al área de desarrollo de materiales más sustentables.

BIBLIOGRAFÍA

1. Alonso, H., & Risacher, F. (1996). Geoquímica del Salar de Atacama, parte 1: origen de los componentes y balance salino. *Andean Geology*, 23(2), 113-122.
2. Amigo, J., Izco, J., & Rodríguez-Gutián, M. A. (2007). Rasgos bioclimáticos del territorio templado de Chile. *Phytocoenologia*, 37(3), 739.
3. Ayub, N. D., Pettinari, M. J., Méndez, B. S., & Lopez, N. I. (2007). The polyhydroxyalkanoate genes of a stress resistant Antarctic *Pseudomonas* are situated within a genomic island. *Plasmid*, 58(3), 240-248.
4. Azua-Bustos, A., Urrejola, C., & Vicuña, R. (2012). Life at the dry edge: microorganisms of the Atacama Desert. *Febs Letters*, 586(18), 2939-2945.
5. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database.
6. Dotmatics. (2022). SnapGene (Versión 6.2.1). Disponible en: <https://www.snapgene.com>
7. Escudero, L. V., Bijman, J., Chong, G., Pueyo, J. J., & Demergasso, C. (2013). Geochemistry and microbiology in an acidic, high altitude (4,000 m) salt flat—high andes, Northern Chile. In *Advanced Materials Research* (Vol. 825, pp. 28-32). Trans Tech Publications Ltd.
8. Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150-3152.
9. Garrido-Sanz, D., Arrebola, E., Martínez-Granero, F., García-Méndez, S., Muriel, C., Blanco-Romero, E., ... & Redondo-Nieto, M. (2017). Classification of isolates from the *Pseudomonas fluorescens* complex into phylogenomic groups based in group-specific markers. *Frontiers in Microbiology*, 8, 413.
10. Gish, W., & States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nature genetics*, 3(3), 266-272.
11. Gobierno de Chile. (s.f.). Nuestro País. Gobierno de Chile. Recuperado el 2 de Diciembre de 2022, desde <https://www.gob.cl/nuestro-pais/>
12. Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New

- algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), 307-321.
13. Hajek, E. R. (1991). Medio ambiente en Chile. La situación ambiental en América Latina. CIEDLA. Buenos Aires, 237-294.
 14. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature communications*, 9(1), 5114.
 15. Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772-780.
 16. Katz, L. S., Griswold, T., Morrison, S. S., Caravas, J. A., Zhang, S., den Bakker, H. C., ... & Carleton, H. A. (2019). Mashtree: a rapid comparison of whole genome sequence files. *Journal of Open Source Software*, 4(44).
 17. Lalucat, J., Mulet, M., Gomila, M., & García-Valdés, E. (2020). Genomics in bacterial taxonomy: impact on the genus *Pseudomonas*. *Genes*, 11(2), 139.
 18. Lara, L. E., Clavero, J., Hinojosa, M., Huerta, S., Wall, R., & Moreno, H. (2006). NVEWS-CHILE: Sistema de Clasificación semicuantitativa de la vulnerabilidad volcánica. In *Congreso Geológico Chileno (Vol. 11, pp. 487-490)*.
 19. Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., & Gascuel, O. (2019). NGPhylogeny. fr: new generation phylogenetic services for non-specialists. *Nucleic acids research*, 47(W1), W260-W265.
 20. Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, 23(1), 127-128.
 21. Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658-1659.
 22. Marcoleta, A. E., Arros, P., Varas, M. A., Costa, J., Rojas-Salgado, J., Berríos-Pastén, C., ... & Lagos, R. (2022). The highly diverse Antarctic Peninsula soil microbiota as a source of novel resistance genes. *Science of The Total Environment*, 810, 152003.

23. Mozejko-Ciesielska, J., Szacherska, K., & Marciniak, P. (2019). *Pseudomonas* species as producers of eco-friendly polyhydroxyalkanoates. *Journal of Polymers and the Environment*, 27, 1151-1166
24. Muneer, F., Rasul, I., Azeem, F., Siddique, M. H., Zubair, M., & Nadeem, H. (2020). Microbial polyhydroxyalkanoates (PHAs): efficient replacement of synthetic polymers. *Journal of Polymers and the Environment*, 28, 2301-2323.
25. Neoh, S. Z., Chek, M. F., Tan, H. T., Linares-Pastén, J. A., Nandakumar, A., Hakoshima, T., & Sudesh, K. (2022). Polyhydroxyalkanoate synthase (PhaC): The key enzyme for biopolyester synthesis. *Current Research in Biotechnology*, 4, 87-101.
26. Orellana-Saez, M., Pacheco, N., Costa, J. I., Mendez, K. N., Miossec, M. J., Meneses, C., ... & Poblete-Castro, I. (2019). In-depth genomic and phenotypic characterization of the antarctic psychrotolerant strain *Pseudomonas* sp. MPC6 reveals unique metabolic features, plasticity, and biotechnological potential. *Frontiers in Microbiology*, 10, 1154.
27. Pacheco, N., Orellana-Saez, M., Pepczynska, M., Enrione, J., Bassas-Galia, M., Borrero-de Acuña, J. M., ... & Poblete-Castro, I. (2019). Exploiting the natural poly (3-hydroxyalkanoates) production capacity of Antarctic *Pseudomonas* strains: from unique phenotypes to novel biopolymers. *Journal of Industrial Microbiology and Biotechnology*, 46(8), 1139-1153.
28. Palleroni, N. J. (1993). *Pseudomonas* classification. *Antonie van Leeuwenhoek*, 64(3), 231-251.
29. Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarszewski, A., Chaumeil, P. A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature biotechnology*, 36(10), 996-1004.
30. Reddy, C. S. K., Ghai, R., & Kalia, V. (2003). Polyhydroxyalkanoates: an overview. *Bioresource technology*, 87(2), 137-146
31. Rehm, B. H. (2003). Polyester synthases: natural catalysts for plastics. *Biochemical Journal*, 376(1), 15-33.
32. Rivera, A., Aravena, J. C., & Casassa, G. (1997). Recent fluctuations of Glaciar Pío XI, Patagonia: discussion of a glacial surge hypothesis. *Mountain Research and Development*,

309-322.

33. Sandoval, Á., Arias-Barrau, E., Arcos, M., Naharro, G., Olivera, E. R., & Luengo, J. M. (2007). Genetic and ultrastructural analysis of different mutants of *Pseudomonas putida* affected in the poly-3-hydroxy-n-alkanoate gene cluster. *Environmental Microbiology*, 9(3), 737-751.
34. Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., ... & Sherry, S. T. (2021). Database resources of the national center for biotechnology information. *Nucleic acids research*, 49(D1), D10.
35. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), 539.
36. Spiers, A. J., Buckling, A., & Rainey, P. B. (2000). The causes of *Pseudomonas* diversity. *Microbiology*, 146(10), 2345-2350.
37. Tan, I. K. P., Foong, C. P., Tan, H. T., Lim, H., Zain, N. A. A., Tan, Y. C., ... & Sudesh, K. (2020). Polyhydroxyalkanoate (PHA) synthase genes and PHA-associated gene clusters in *Pseudomonas* spp. and *Janthinobacterium* spp. isolated from Antarctica. *Journal of biotechnology*, 313, 18-28.
38. Tindall, B. J. (2004). Prokaryotic diversity in the Antarctic: the tip of the iceberg. *Microbial ecology*, 47, 271-28.
39. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189-1191.
40. Wei, T., & Fang, Q. (2022). Regulating the monomer of polyhydroxyalkanoate from mixed microbial culture: with particular emphasis on substrate composition: A review. *Environmental Engineering Research*, 27(5).

ANEXO

Anexo 1. Códigos de acceso PhaCs de referencia (clases I a IV).

Bacteria	Código	Clase de PhaC
<i>Pseudomonas putida</i> KT2440	AAM63409.1; AAM63407.1	II
<i>Pseudomonas fluorescens</i>	ACK57560.1; ACK57562.1	II
<i>Pseudomonas sp. 61-3</i>	BAA36200.1; BAA36202.1	II
<i>Pseudomonas aeruginosa</i> PAO1	WP_003115652.1	II
<i>Haloferax mediterranei</i>	WP_004060696.1; WP_004056138.1	III
<i>Bacillus megaterium</i>	WP_013055939.1	III
<i>Cupriavidus necator</i>	WP_011615085.1	I
<i>Bacillus megaterium</i>	AAD05260	IV (subunidad PhaC)