



UNIVERSIDAD DE CHILE -FACULTAD DE CIENCIAS -ESCUELA DE PREGRADO

## **“Estudio metagenómico de la biodiversidad y potencial metabólico de hábitats terrestres de la antártica”**

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

**Mario Alejandro Fernández Hurtado**

Director del Seminario de Título:

Dr. Marcelo Baeza Cancino

Enero 2023  
Santiago - Chile



## INFORME DE APROBACIÓN SEMINARIO DE TITULO

Se informa a la Escuela de Pregrado de la Facultad de Ciencias, de la Universidad de Chile que el Seminario de Título, presentado por el Sr.

**Mario Alejandro Fernández Hurtado**

“Estudio metagenómico de la biodiversidad y potencial metabólico en hábitats terrestres de la antártica”

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

*Dr. Marcelo Baeza Cancino*  
**Director Seminario de Título**

---

### Comisión Revisora y Evaluadora

*Dr. Andrés Marcoleta*  
**Presidente**

---

*Dra. Jennifer Alcaíno*  
**Evaluadora**

---

Santiago de Chile, 30 de Enero del 2023

## **i. Agradecimientos**

A esta ciudad, que me ha acogido a su manera caótica y elegante.

A mis amigos, que sin ellos no hubiera podido.

A Margarita, con su paciencia y compañía.

A Nunca Logramos Nada, y su apoyo incondicional.

Agradezco a todos los que han formado parte de quien soy ahora.

## ii. Índice de contenidos

### Índice de contenidos

i. Agradecimientos .....	iii
ii. Índice de contenidos .....	iv
iii. Índice de tablas .....	v
iv. Índice de figuras.....	vi
v. Lista de abreviaturas.....	vii
RESUMEN.....	viii
ABSTRACT .....	ix
1. INTRODUCCIÓN.....	1
1.1 La metagenómica, su definición y aplicaciones .....	1
1.2 El impacto de las técnicas de secuenciación de nueva generación.....	1
1.3 Predicción de vías metabólicas y compuestos de interés .....	4
1.4 Metagenómica y la biodiversidad de ambientes extremos.....	5
1.5 La importancia del continente antártico .....	6
2. OBJETIVOS .....	8
2.1 Objetivo general: .....	8
2.2 Objetivos específicos: .....	8
3. MATERIALES Y MÉTODOS.....	9
3.1 Predicción de genes y asignación de funciones metabólicas .....	9
3.2 Identificación de módulos metabólicos .....	9
3.3 Análisis estadístico de datos metabólicos .....	10
3.4 Asignación de taxonomía .....	11
4. RESULTADOS .....	12
4.1 Disimilitud metabólica no se relaciona directamente con la distancia .....	12
4.2 Abundancia y representación de módulos metabólicos .....	13
4.3 Composición taxonómica de las localidades muestreadas .....	16
4.4 Diversidad taxonómica de bacterias.....	18
4.5 Diversidad taxonómica de fungi.....	20
5. DISCUSIÓN Y PROYECCIONES .....	24
6. CONCLUSIONES .....	27
REFERENCIAS .....	28

### iii. Índice de tablas

Tabla 1. Número de marcos de lectura alineados con la base de datos de KO. ....	9
Tabla 2. Número de marcos de lectura alineados con módulos completos de la base de datos KEGG MODULE. ....	10

#### iv. Índice de figuras

Figura 1. Mapas de calor derivados de comparaciones entre sitios. ....	12
Figura 2. Abundancias relativas de los módulos metabólicos en el total de los sitios muestreados. ....	14
Figura 3. Abundancia relativa de los doce módulos metabólicos menos representados en los 10 sitios de muestreo. ....	15
Figura 4. Abundancia relativa de filos bacterianos y fungi en cada localidad. ....	17
Figura 5. Diversidad bacteriana al nivel de familia. ....	19
Figura 6. Abundancia relativa de fungi a nivel de clase. ....	20
Figura 7. Diversidad de fungi al nivel de familia. ....	22

## v. **Lista de abreviaturas**

bp: Pares de bases

ITS: Espaciador transcrito interno

KEGG: Enciclopedia de Genes y Genomas de Kioto

KO: Base de datos de ortología de KEGG

LSU: Subunidad grande del ribosoma

NGS: Secuenciación de nueva generación

RPKM: Lecturas por millón de kilobases

### **Sitios de muestreo**

Decep: Isla Decepción

Dee: Isla Dee

KG1A – KG1B – KG2: Isla Rey Jorge

Lago: Isla Lagotellerie

Litch: Isla Litchfield

Nelson: Isla Nelson

Snow: Isla Nevada

UG: Glaciar Unión

## RESUMEN

La metagenómica es un término que describe tanto un campo de la biología como una serie de herramientas que permiten, entre otras cosas, el análisis de comunidades microbianas sin la necesidad de cultivar dichos microorganismos. El constante avance de la eficiencia y certeza de estas técnicas han convertido a la metagenómica en una poderosa herramienta para estudiar la biodiversidad microbiana de una zona en particular. En este estudio, se investigó el potencial metabólico y la diversidad de comunidades microbianas de diez muestras de suelo de distintas localidades de la antártica mediante un alcance metagenómico de secuenciación de escopeta. Se observó que la distancia geográfica no se correlaciona con el número de diferencias significativas en metabolismo entre las muestras; a pesar de que algunas localidades fueron agrupadas según lo esperado en base a sus distancias, la mayoría de los agrupamientos no muestran una relación entre distancia y similitud metabólica. Sin embargo, se encontró que la mayoría de los módulos metabólicos son compartidos entre localidades, no observándose ninguno que no estuviera en al menos dos localidades. Además, en el conjunto de diez muestras se observaron 17 filos, 177 familias y 297 géneros bacterianos y 5 filos, 53 familias y 72 géneros fúngicos. Tanto la diversidad como abundancia de fungi observada fue menor que la bacteriana en todas las localidades. Se presentaron diferencias notables en la diversidad presente en cada localidad a varios niveles taxonómicos, habiendo solo un puñado de localidades con diversidades similares. Adicionalmente, se encontró que la disimilitud en términos de diversidad no se correlaciona con las diferencias significativas en metabolismo entre cada localidad. Finalmente, este trabajo permitió dar un primer paso en cuanto a identificar qué organismos están presentes en estas localidades y cuáles funciones y procesos metabólicos ocurren a nivel de comunidad microbiológica en los suelos estudiados.

## **ABSTRACT**

Metagenomics is a concept that encompasses a field of biology as well as a set of tools that allow for, among other things, the analysis of microbial communities with culture-free approaches. The ongoing advances in both efficacy and accuracy of these techniques have allowed metagenomics to become a powerful tool in the study of microbial biodiversity of a given zone. In this study, the metabolic potential and diversity of microbial communities of ten soil samples of distinct Antarctic locales were analyzed with a shotgun sequencing metagenomic approach. It was determined that geographic distance does not correlate with the number of significant differences in metabolism between samples; despite some of the locales grouping as expected based on distance, most of them do not show a link between distance and metabolic similarity. Nonetheless, we found that most of the metabolic modules are indeed shared between all locales, with none being found in fewer than two locales. From ten soil samples we observed a total of 17 phyla, 177 families and 297 genera of bacteria and 5 phyla, 53 families and 72 genera of fungi. We found that fungi had both lower abundance and diversity metrics than bacteria in all the locales. Notable differences were observed in terms of diversity at multiple taxonomic levels, with just a handful of locales having comparable diversities. Additionally, we found that these differences in diversity between samples do not correlate with the dissimilarity in metabolism observed previously. Lastly, this study allows for a first step regarding the identification of both diversity and metabolic potential of the microbial communities present in the sampled soils with a culture-free approach.

## **1. INTRODUCCIÓN**

### **1.1 La metagenómica, su definición y aplicaciones**

La metagenómica es una rama de la biología molecular que engloba distintas técnicas y alcances, pero cuyo propósito principal se basa en el análisis comprensivo de todas las secuencias de DNA de una muestra ambiental (Marchesi & Ravel, 2015). El desarrollo de estas técnicas ha permitido dar pasos enormes en cuanto a la comprensión de comunidades microbianas completas, ya que la gran mayoría de los microorganismos no han sido (o derechamente no pueden ser) cultivados y aislados con las técnicas actuales de laboratorio (Liu y col., 2022). En algunos estudios se ha estimado que esta “materia oscura microbiana” representa incluso un 99% de los microorganismos del planeta (Solden y col., 2016). Las aproximaciones metagenómicas, que no requieren el aislamiento ni cultivo de microorganismos, han abierto las posibilidades en cuanto al descubrimiento de nuevos taxones, interacciones metabólicas e incluso compuestos novedosos de interés industrial (Bikel y col., 2015; Waschulin y col., 2022). Es más, estos descubrimientos gatillaron un cambio de paradigma en la microbiología, haciéndonos entender qué tan poco se conocía realmente acerca de los microorganismos y su importancia metabólica en distintos ambientes, incluyendo las interacciones entre microbiomas y sus huéspedes eucarióticos (Handelsman, 2004; Gilbert y col., 2018). Gracias a estos avances es que en las últimas décadas se han realizado cada vez más estudios con el fin de vislumbrar la biodiversidad de comunidades microbianas de distintos ambientes, en conjunto con los procesos metabólicos que realizan sus integrantes (Peterson y col., 2009; Gilbert y col., 2014; Sunagawa y col., 2015).

### **1.2 El impacto de las técnicas de secuenciación de nueva generación**

Para comprender el explosivo desarrollo que ha tenido la metagenómica y las disciplinas relacionadas (metatranscriptómica, metabolómica, etc.) es necesario mencionar la importancia de las técnicas o plataformas de secuenciación de nueva generación (NGS). A pesar de que existe una gran variedad de plataformas de secuenciación con protocolos totalmente distintos, las NGS son definidas por su capacidad de generar volúmenes enormes de datos en una sola reacción, siendo un método relativamente económico y de alta eficiencia (Metzker, 2010), diferenciándolas entonces de la secuenciación con método de Sanger, ahora conocida como secuenciación de primera generación. Si bien la metagenómica, entendida como el estudio del DNA de una comunidad microbiana completa, surgió incluso antes del desarrollo de las NGS, ha sido el desarrollo de estas el que ha permitido en gran parte la abundancia de conocimiento y estudios actuales (Escobar-Zepeda y col., 2015). Algo que hay que tener presente es saber diferenciar estas técnicas o plataformas de secuenciación respecto del alcance a utilizar. En este caso hay dos alcances bastante populares y eficientes, la secuenciación con amplicones (amplicon sequencing) y la secuenciación de escopeta (shotgun sequencing). En el primer alcance se seleccionan regiones de interés para amplificar previo a la secuenciación, específicamente mediante el uso de partidores universales que son capaces de amplificar dichas regiones. Se han propuesto en la literatura un número de conceptos distintos para encapsular mejor el objetivo y mira de este alcance, tales como “metaperfilado” o “metataxonómica”, ya que no se secuencia todo el material genético de una comunidad, sino que solo los genes ubicados en las regiones de interés que, en la gran mayoría de estudios, corresponden a distintas regiones del gen de rRNA 16S en bacterias y del espaciador transcrito interno (ITS) o el gen de la subunidad grande ribosomal (LSU) en eucariontes (Escobar-Zepeda y col., 2015; Durazzi y col., 2021). Por otro lado, la secuenciación de escopeta contempla la secuenciación de todo el DNA de

la muestra. El proceso consiste en fragmentar el DNA en segmentos pequeños (100-300 bp), los cuales son secuenciados generando lecturas o “reads”. Estos reads son ensamblados mediante algoritmos computacionales de acuerdo con distintos parámetros en secuencias contiguas o “contigs”. Estos contigs pueden ser agrupados para reconstruir el genoma de un organismo (“binning”) o incluso pueden ser directamente comparados con una o más bases de datos para identificar los genes que codifican y su posible función. De esta forma también es posible delinear las distintas especies presentes mediante la identificación de los distintos marcadores moleculares encontrados en estas colecciones de genes. Dichos marcadores (gen rRNA 16S; ITS; LSU) se han posicionado como tales debido a sus grados de conservación intra- e inter-especie, es decir, poseen regiones muy conservadas que permiten identificar taxones filogenéticamente cercanos, a la vez que poseen regiones hipervariables que permiten diferenciar entre especies particulares (Schoch y col., 2012; Sunagawa y col., 2013; Johnson y col., 2019; Cortés-López y col., 2020). Finalmente, independiente del alcance que se utilice, en ambos se pueden utilizar distintas técnicas de NGS disponibles en la actualidad, aunque un desafío existente es que a pesar de que se han demostrado consensos entre plataformas, se desconoce el total de diferencias que podrían generarse por el uso de estas distintas técnicas en cuanto a la cantidad y calidad de secuencias obtenidas (Luo y col., 2012; Allali y col., 2017). De todas maneras, al existir un grado de consenso entre las distintas plataformas, y al no haber un método específico que se considere como “estándar dorado”, a lo que se le suele dar importancia en estos estudios es nada más el alcance utilizado (Zhang y col., 2021). Finalmente, ningún alcance supera objetivamente a otro, habiendo estudios que demuestran que un alcance entrega mejores resultados que el otro y viceversa en investigaciones de biodiversidad microbiana (Ranjan y col., 2016; Tessler y col., 2017).

### **1.3 Predicción de vías metabólicas y compuestos de interés**

Como se ha establecido, con las distintas herramientas de NGS y metagenómica es posible identificar la composición taxonómica de una muestra ambiental, sin embargo, otra gran aplicación de la metagenómica consiste en predecir genes y sus posibles funciones dentro de la comunidad microbiana. Por ejemplo, mediante una secuenciación por escopeta y el uso de algoritmos como BLAST o MG-RAST y bases de datos como la NCBI-nr o Uniprot es posible alinear las distintas secuencias encontradas a genes conocidos o putativos. Luego, mediante otras búsquedas en bases de datos como la de Ortología de la Enciclopedia de Genes y Genomas de Kyoto (KO; [www.genome.jp/kegg/ko.html](http://www.genome.jp/kegg/ko.html)) se pueden identificar los procesos en los que están involucrados estos genes y de tal manera reconstruir las vías metabólicas y reacciones, tanto ubicuas como más específicas, presentes en las comunidades estudiadas (Raggi y col., 2020). Además, estas secuencias pueden ser agrupadas en módulos, esto es, unidades funcionales de genes en vías metabólicas, incluyendo complejos moleculares capaces de realizar procesos metabólicos completos ([www.genome.jp/kegg/module.html](http://www.genome.jp/kegg/module.html)). Con estos alcances es posible analizar las diferencias entre los procesos que ocurren en distintas localidades o incluso en localidades similares que han sido afectadas por eventos como contaminación con sustancias foráneas o, en el caso de la microbiota, las condiciones a las que estuvo expuesto el huésped (Zheng y col., 2015; Feng y col., 2018). Finalmente, una disciplina que se ha beneficiado especialmente de estas técnicas es la metagenómica funcional, facilitando el descubrimiento de nuevos compuestos biológicos de interés humano (Mirete y col., 2016). La creciente facilidad de acceso a herramientas de NGS y consecuente aumento de información respecto a genes funcionales, especialmente de

organismos no cultivables previamente desconocidos, han permitido expandir las fronteras del campo de la bioprospección y metagenómica funcional (Trindade y col., 2015)

#### **1.4 Metagenómica y la biodiversidad de ambientes extremos**

Un ambiente es considerado como extremo cuando presenta ciertas condiciones que dificultan la supervivencia de la mayoría de los organismos. Condiciones de temperaturas en los extremos o altas concentraciones de ciertos compuestos, tales como azufre, sal, dióxido de carbono, entre otros, permiten que solo un grupo muy particular de organismos, los cuales han desarrollado adaptaciones para vivir en estas condiciones, puedan sobrevivir y desarrollar comunidades. Históricamente se consideraba que estos ambientes carecían de vida, incluso microscópica, ya que ni siquiera esta podría soportar ciertas condiciones ambientales (Morita, 1999). Sin embargo, con el progresivo descubrimiento de extremófilos desde la década de los 50, el interés de estudiar estas comunidades no ha hecho más que aumentar. De hecho, los avances y nuevas técnicas en NGS delineadas previamente han sido de especial utilidad en el estudio de estos ambientes, permitiendo abrir una ventana a la diversidad de especies presente en estos biomas, junto a las interacciones y estructuras de las comunidades que los habitan (Cowan y col., 2015). Dentro de la gama de ambientes extremos, los ambientes fríos conforman la mayoría de la biósfera del planeta, concentrando gran parte de su biomasa. Debido a las adaptaciones que poseen estos organismos, representan un recurso valioso en la búsqueda de nuevas especies psicrófilas y compuestos con potenciales usos biotecnológicos (Aliyu y col., 2017; Dumorné y col., 2017). Urge investigar los aspectos de biodiversidad y potencial metabólico de estas zonas debido al especial riesgo que corren debido al calentamiento

global, en especial las regiones polares, que han visto los mayores aumentos de temperatura proporcionalmente en el planeta (Zucconi & Buzzini, 2021).

### **1.5 La importancia del continente antártico**

El continente antártico por un largo tiempo ha llamado la atención de la comunidad científica debido a que sus condiciones ambientales y geográficas lo han convertido en uno de los lugares más inhóspitos y aislados del planeta, donde además dichas condiciones propician la selección de cepas y especies de microorganismos capaces de realizar procesos metabólicos poco comunes con gran potencial biotecnológico (Doytchinov & Dimov, 2022). Estas últimas décadas han visto una creciente cantidad de estudios que buscan describir la biodiversidad microbiana presente en sus distintos ambientes, tanto para fines biotecnológicos (e.g. bioprospección) como para el estudio de la resistencia microbiana, entre otros (Lopatina y col., 2016; Varga y col., 2021; Marcoleta y col., 2022). Si bien la mayoría del continente está cubierto por una permanente capa de hielo y nieve, las regiones libres de hielo, que representan menos del 2% del área total del continente, abarcan una variada gama de ambientes tanto marinos como terrestres que concentran la gran mayoría de la biodiversidad del continente (Cowan & Tow, 2004; Lee y col., 2017). Debido a esta diversidad de factores abióticos y bióticos que presentan los distintos ambientes antárticos, los suelos de estas localidades presentan composiciones microbianas igualmente variadas y novedosas (Lambrechts y col., 2019). El riesgo que corren estos microbiomas debido a presiones como el cambio climático y la creciente actividad humana dada tanto por el aumento en investigaciones como en turismo señalan más que nunca la importancia de analizar estas comunidades y cómo podrían verse alteradas por dichos factores de cambio (Brooks y col., 2019; Wong y col., 2021).

Para profundizar el conocimiento de la diversidad microbiana de este continente, en este estudio se analizaron datos metagenómicos obtenidos de muestras de suelos de cinco islas del archipiélago Shetland del sur, dos islas de la península antártica y el glaciar Unión con un alcance de secuenciación de escopeta. Se identificaron los procesos metabólicos presentes en las distintas localidades muestreadas además de la composición taxonómica de las comunidades microbianas de estos suelos.

## **2. OBJETIVOS**

### **2.1 Objetivo general:**

Determinar la diversidad de comunidades microbianas y potencial metabólico en suelos de distintas localidades antárticas.

### **2.2 Objetivos específicos:**

1. Determinar las diferencias en vías metabólicas presentes en suelos de distintas localidades antárticas
2. Analizar los distintos niveles de representación de las distintas vías y procesos metabólicos en cada localidad.
3. Examinar la composición taxonómica de las comunidades microbianas presentes en los suelos de las localidades estudiadas.

### 3. MATERIALES Y MÉTODOS

#### 3.1 Predicción de genes y asignación de funciones metabólicas

Los datos metagenómicos utilizados en este estudio provienen del material genético de diez muestras de suelo de localidades antárticas, el cual pasó por un tratamiento previo al desarrollo de este seminario de título. El ADN de las muestras fue aislado y posteriormente secuenciado siguiendo un alcance de secuenciación de escopeta. Se predijeron los marcos de lectura de los contigs ensamblados y se alinearon mediante BLAST contra una base de datos local curada previamente. De tal manera se identificó la posible proteína codificada por cada marco de lectura y se obtuvo un valor de RPKM en base al total de lecturas obtenidas en la muestra y el largo de cada secuencia.

Se compararon las proteínas predichas en BLAST contra la base de datos de KO para determinar la vía metabólica a la que pertenecían. Con esto se alinearon un total de 416.937 marcos de lectura distintos que fueron utilizados para los análisis de este estudio (Tabla 1).

**Tabla 1. Número de marcos de lectura alineados con la base de datos de KO.**

Sitio de Muestreo	N° de ORFs
Decep	36.827
Dee	22.925
KG1A	38.879
KG1B	39.449
KG2	58.918
Lago	38.504
Litch	54.529
Nelson	37.285
Snow	42.734
UG	46.887
Total de ORFs	416.937

#### 3.2 Identificación de módulos metabólicos

Además de la base de datos KO, previo a este estudio se alinearon los marcos de lectura predichos con la base de datos de módulos de KEGG (KEGG MODULE), utilizándose en este estudio solamente los módulos que se encontraron en su totalidad, es decir, si es que con el alineamiento se encontraron todos los genes que componen la unidad funcional. Con dicho alineamiento terminaron considerados un total de 175.723 marcos de lectura que fueron utilizados para análisis posteriores (Tabla 2).

**Tabla 2. Número de marcos de lectura alineados con módulos completos de la base de datos KEGG MODULE.**

Sitio de Muestreo	N° de ORFs
Decep	14.835
Dee	13.506
KG1A	15.853
KG1B	15.826
KG2	24.367
Lago	16.023
Litch	22.310
Nelson	16.411
Snow	17.673
UG	18.919
Total de ORFs	175.723

### 3.3 Análisis estadístico de datos metabólicos

Con los valores de RPKM de las secuencias asociadas a las distintas vías metabólicas, se calcularon las diferencias significativas entre la presencia de las vías metabólicas encontradas en cada localidad muestreada. Se realizó un análisis de varianza (ANOVA) entre los valores de RPKM de las lecturas, agrupadas según la vía metabólica a la que pertenecían; posteriormente se realizó una prueba de método de Tukey para determinar cuáles localidades específicas presentaban diferencias significativas entre ellas. Paralelamente se utilizó el programa Google Earth para cuantificar las distancias

geográficas entre las localidades muestreadas. Los análisis posteriores de estos datos se realizaron en el programa RStudio.

### **3.4 Asignación de taxonomía**

Previo a este estudio también se realizó un alineamiento para identificar las secuencias que codificaran ciertos marcadores moleculares, incluyendo el gen de rRNA 16S, LSU, ITS, entre otros. Los resultados entregados por el alineamiento fueron importados y trabajados en este estudio mediante el programa MS Excel, donde se filtraron según criterios de largo de secuencia (> 100 bp) y porcentaje de identidad (>80% para fungi y rRNA 16S y >90% para LSU de procariontes). En el caso de eucariontes, se omitieron los alineamientos que no correspondieran con fungi (3,28% del total de alineamientos de eucariontes). Luego, se ordenaron los datos asegurando su correcta clasificación en los distintos niveles taxonómicos. Finalmente, se realizaron análisis de índices de diversidad beta mediante el paquete vegan del programa RStudio.

## 4. RESULTADOS

### 4.1 Disimilitud metabólica no se relaciona directamente con la distancia

Para determinar qué localidades presentan un mayor grado de cercanía en cuanto a los genes putativos identificados en distintas vías metabólicas (que llamaremos similitud o disimilitud metabólica) las unas con otras, se tomaron los resultados del análisis estadístico de las diferencias significativas entre las localidades y se utilizó el programa RStudio para construir un mapa de calor denotando la disimilitud entre ellas (Figura 1a). De forma paralela, y con el fin de determinar si existe una relación entre esta disimilitud y la distancia geográfica entre cada localidad muestreada, se construyó otro mapa de calor en base a estas distancias (Figura 1b).

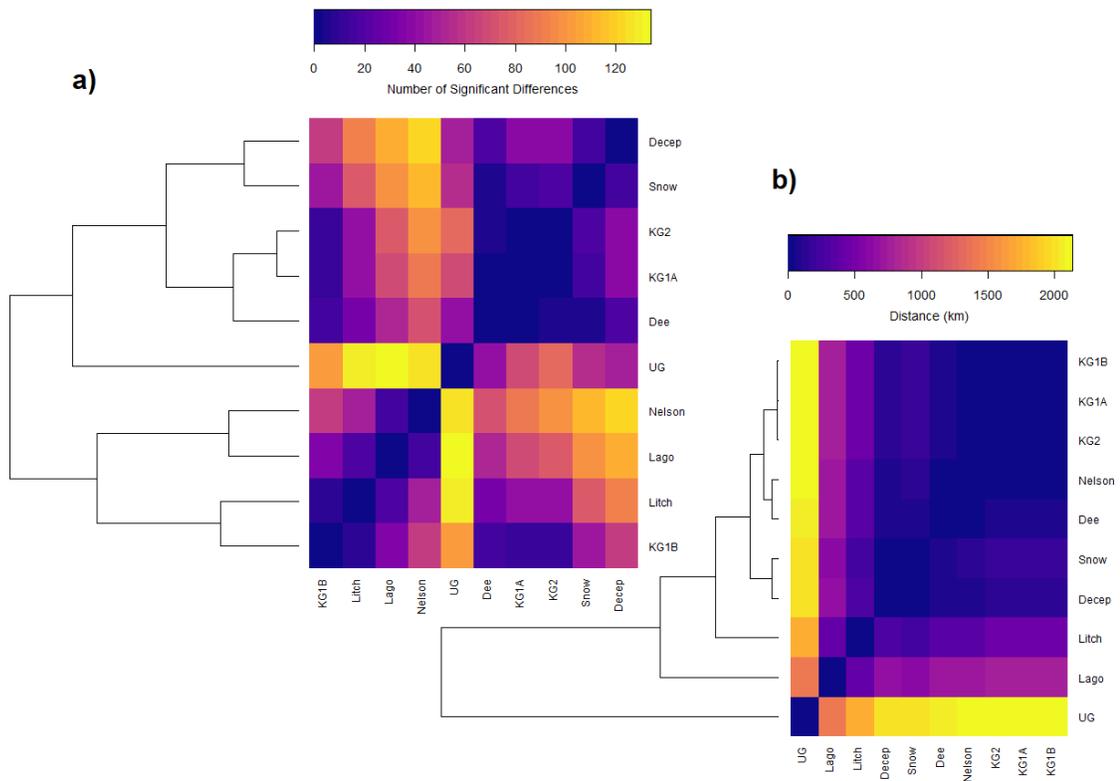


Figura 1. Mapas de calor derivados de comparaciones entre sitios. a) Ilustra el número de

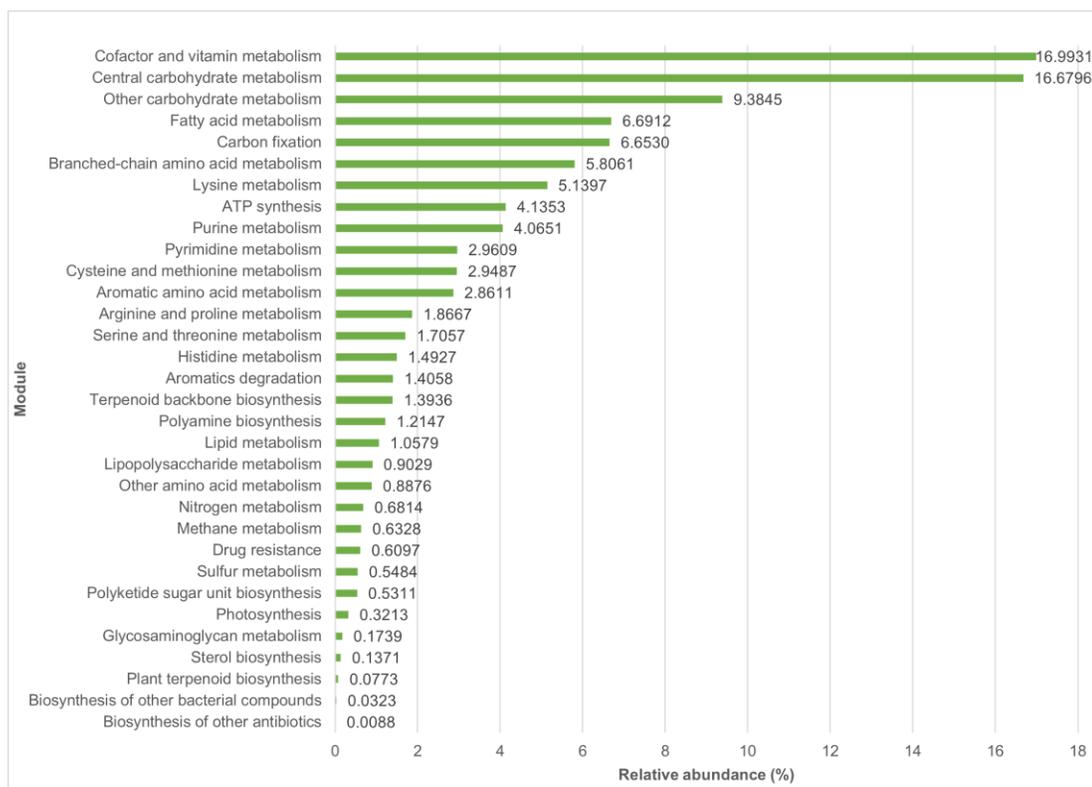
diferencias significativas en metabolismo entre sitios. **b)** Ilustra la distancia en kilómetros entre las localidades muestreadas. Los dendrogramas indican los niveles de cercanía o similitud entre localidades según los criterios estipulados previamente.

Los datos obtenidos no muestran una correlación directa entre la distancia y similitud metabólica; a pesar de estar ubicadas más cerca del glaciar unión (UG, la localidad más alejada de todas las demás) las islas Lagotellerie (Lago) y Litchfield (Litch) muestran uno de los niveles más altos de disimilitud con UG. Igualmente, a pesar de estar en relativa cercanía, las muestras de las islas Rey Jorge (KG1A; KG1B; KG2), Dee y Nelson se encuentran en distintas ramas del dendrograma, estando KG1B agrupada con Litchfield y Nelson agrupada con Lagotellerie. Finalmente, las islas Decepción (Decep) y Nevada (Snow) sí se encuentran agrupadas de acuerdo con lo esperado según sus distancias geográficas, a diferencia de las demás muestras.

#### **4.2 Abundancia y representación de módulos metabólicos**

Para determinar qué proceso metabólico particular se encuentra más o menos representado en su totalidad en cierta localidad, se utilizaron los datos correspondientes a los módulos metabólicos completos identificados en cada muestra. Estos módulos se encuentran agrupados en distintas categorías según el tipo de proceso que llevan a cabo, esto es, si corresponden a procesos de biosíntesis de aminoácidos, resistencia a antibióticos, procesos fotosintéticos, etc. Además, cada módulo identificado por la base de datos de KEGG es subdividido en tres categorías de distinta especificidad, con el primer nivel clasificando a grandes rasgos el proceso y el tercer nivel siendo generalmente la reacción específica en la que participa el gen identificado. Para tener una cantidad trabajable de datos sin perder el grado necesario de precisión, se utilizó la segunda categoría de clasificación, con lo que se tuvo un total de treinta y dos módulos.

Luego, con los valores de RPKM asociados a cada secuencia, se calculó la abundancia relativa de cada módulo en las distintas localidades. Para evitar que los valores absolutos de RPKM crearan una sobrerrepresentación de ciertas localidades, se calcularon las abundancias relativas de los módulos de cada localidad por separado. Luego, con el promedio de estos porcentajes (abundancia relativa promedio) se pudieron discernir los procesos más abundantes en el conjunto de todas las localidades (Figura 2).

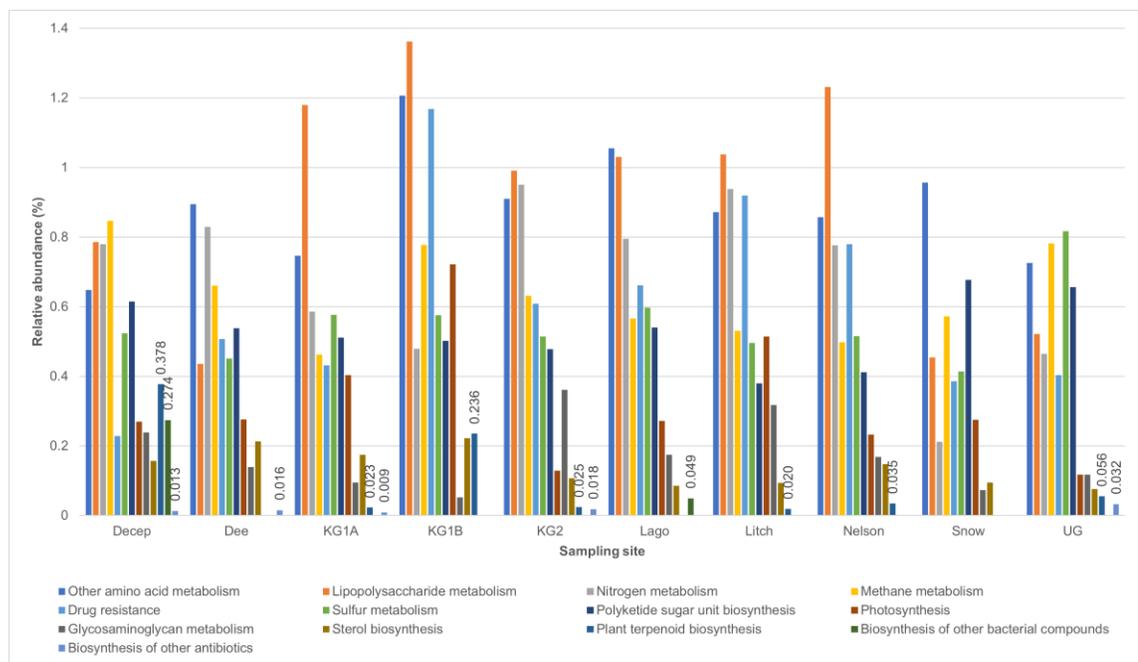


**Figura 2. Abundancias relativas de los módulos metabólicos en el total de los sitios muestreados.** Las abundancias relativas de cada localidad fueron calculadas por separado y el promedio entre ellas utilizado para la elaboración de este gráfico. Los valores de abundancia se indican al costado del módulo respectivo.

De los treinta y dos procesos identificados, los cinco más representados concentran más del 50% de la abundancia relativa en todas las localidades. Cuatro de estos coinciden con procesos bastante ubicuos, como el metabolismo de carbohidratos y ácidos grasos,

mientras que el quinto proceso corresponde a la fijación de carbono, principalmente asociada a procesos fotosintéticos que no son ubicuos para los organismos heterótrofos. Inversamente, las tres categorías menos representadas: biosíntesis de otros antibióticos (*i.e.* kanosamina); biosíntesis de otros compuestos bacterianos (*i.e.* estafiloferrina A y B) y biosíntesis de terpenoides de plantas (*i.e.* ácido abscísico y betacaroteno) solo se encontraron en una fracción de las localidades, siendo Decep la única localidad en donde se encontraron los tres; inversamente, en Snow no se encontró ninguno de estos módulos en su totalidad. A pesar de esto, no se observó ningún módulo único (*i.e.* que se encuentre en una sola muestra), estando todos los módulos presentes en al menos dos localidades.

Para visualizar de mejor manera los módulos menos representados, se seleccionaron los 12 módulos con menor abundancia relativa, los cuales concentraban un poco más del 5% de la abundancia total en todas las localidades (Figura 3).



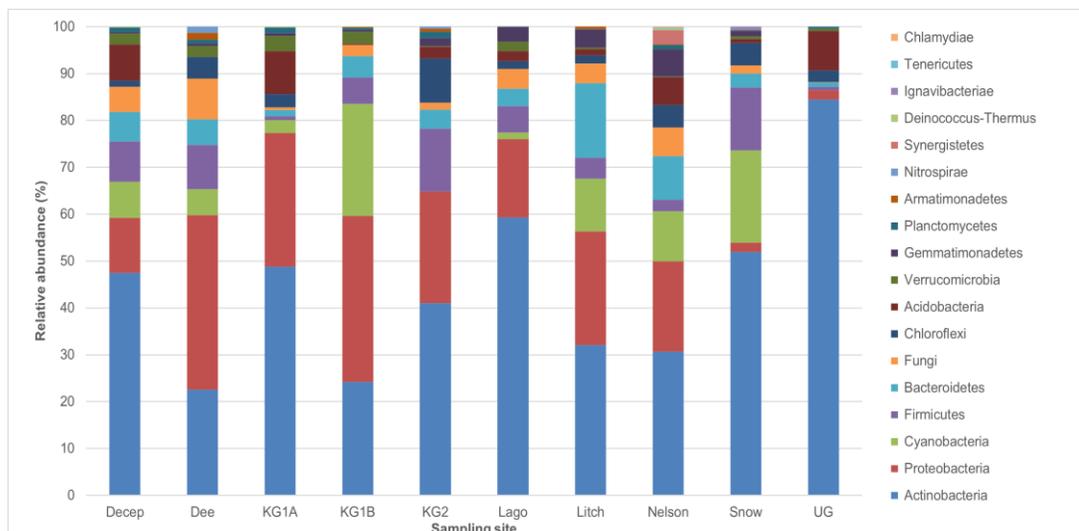
**Figura 3. Abundancia relativa de los doce módulos metabólicos menos representados en los 10 sitios de muestreo.** Los tres módulos menos abundantes solo se encontraron en una

fracción de las localidades (abundancia indicada con etiquetas donde corresponda). En conjunto estos doce módulos concentran aproximadamente el 5% de la abundancia relativa promedio entre todas las localidades.

Se pudo notar que de los tres módulos que no se encuentran en todas las localidades, solamente la biosíntesis de otros antibióticos presentó consistentemente la menor abundancia en todas las localidades en las que se encontró completo, a diferencia de los otros dos módulos, que en muestras como la de la isla Decepción, presentaron abundancias mayores incluso que las de otros módulos como la biosíntesis de esteroides. Por otro lado, se ve que las abundancias relativas de procesos como la biosíntesis de azúcares policétidos se mantienen relativamente constantes entre las localidades, mientras que otros como el metabolismo de lipopolisacáridos y metabolismo de nitrógeno fluctúan considerablemente.

#### **4.3 Composición taxonómica de las localidades muestreadas**

Para identificar la composición taxonómica de las localidades muestreadas, se realizaron alineamientos mediante BLAST con distintos marcadores moleculares. Del total de estos alineamientos, ya filtrados según criterios de largo de secuencia y porcentaje de identidad, se obtuvo que un total de 873 contigs alinearon a algún taxón: 755 a bacterias y 118 a fungi. De igual manera que con los módulos, se calculó la abundancia relativa según los valores de RPKM de cada taxón y localidad (Figura 4).

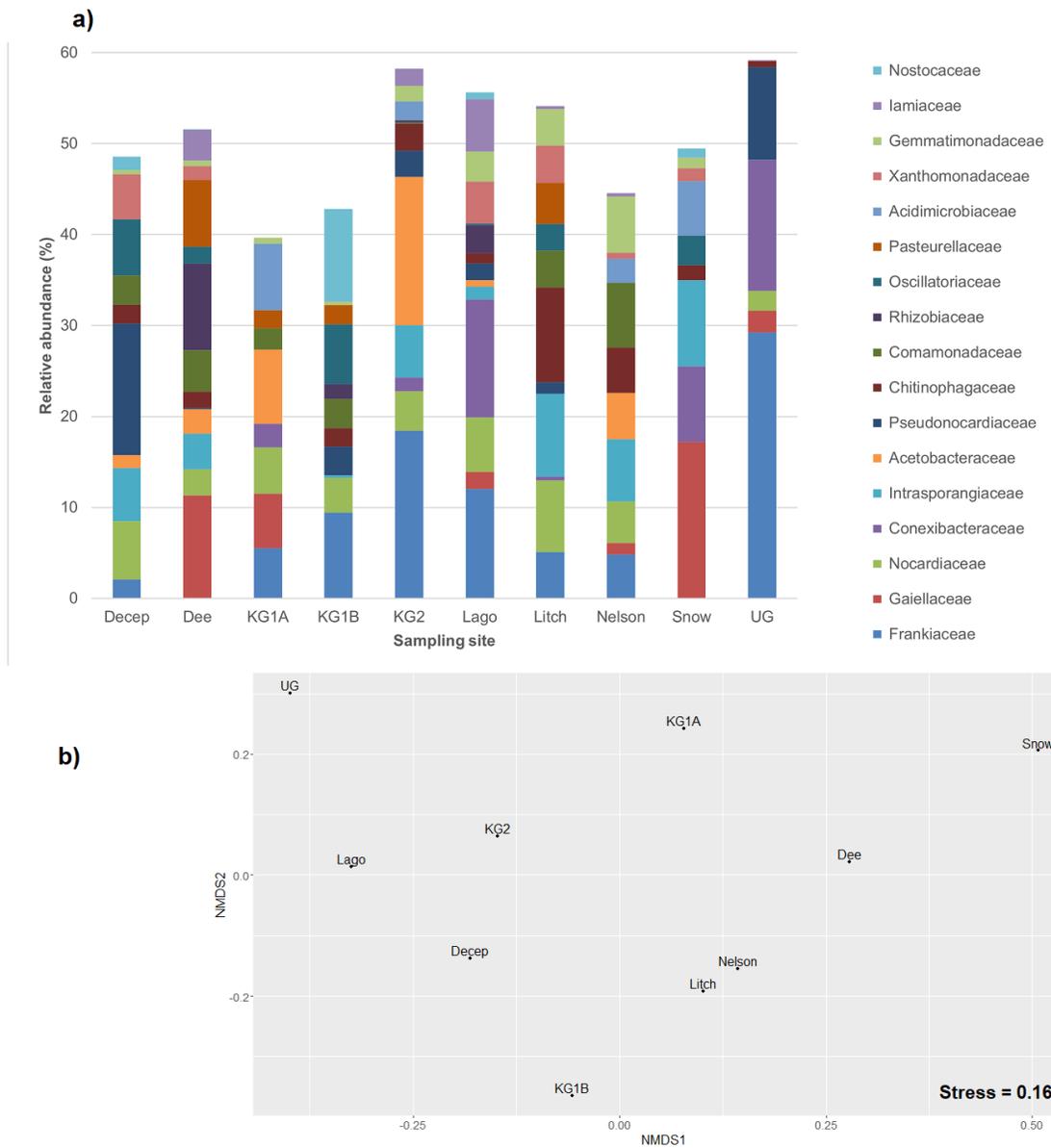


**Figura 4. Abundancia relativa de filios bacterianos y fungi en cada localidad.** Los valores de abundancia relativa fueron calculados en base a los RPKM de cada taxón y el RPKM total de cada localidad.

Se encontró que, si bien los filios más representados fueron Actinobacteria y Proteobacteria, estos se observaron en proporciones bastante distintas según cada localidad. En sitios como Dee y KG1B Proteobacteria es el filo más abundante mientras que en Snow Cyanobacteria es el segundo filo más abundante por sobre Proteobacteria. De los diecisiete filios bacterianos identificados, un total de seis fueron encontrados en todas las localidades: los dos más abundantes, Actinobacteria y Proteobacteria junto a Firmicutes, Bacteroidetes, Verrucomicrobia y Planctomycetes. Inversamente, se encontraron cuatro filios únicos en un total de dos sitios, siendo Nelson el sitio donde se encontró la mayor cantidad: Synergistetes, Tenericutes y Chlamydiae. A la vez solo en la isla Snow se registró la presencia de Ignavibacteriae. Finalmente, se encontraron fungi en todas las localidades, sin embargo, en total estos alineamientos presentaron abundancias relativas en el rango de 0,01% a 8,6% (UG y Dee respectivamente) del total de secuencias, por lo que para los análisis más detallados se decidió trabajar con estos distintos dominios por separado.

#### **4.4 Diversidad taxonómica de bacterias**

Se encontraron un total de 177 familias bacterianas, con las 17 más representadas concentrando aproximadamente el 50% de la abundancia relativa en promedio (Figura 5a). A pesar de esto, 100 de las 177 familias se encontraron con más de un 1% de abundancia relativa en al menos una de las localidades. En el caso de las familias más abundantes de cada localidad, estas no siempre pertenecieron a los filos más abundantes de su respectiva muestra y solo en una localidad superaron el 20% de abundancia relativa (*Frankiaceae* en UG); en 6 localidades la familia más abundante perteneció a Actinobacteria; en 2 localidades a Proteobacteria; en una localidad a Cyanobacteria y en una localidad a Gemmatimonadetes.



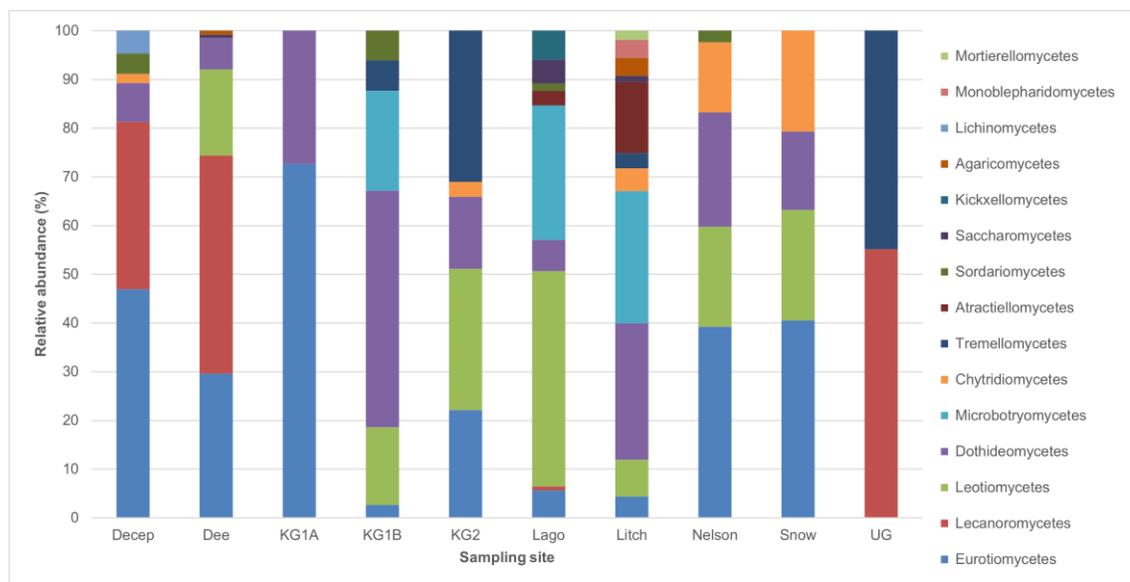
**Figura 5. Diversidad bacteriana al nivel de familia. a)** Abundancia relativa de las 17 familias más representadas en promedio. **b)** Escalamiento multidimensional no métrico basado en el índice de disimilitud de Bray-Curtis entre sitios al nivel de familia. Los tamaños de las letras en b) es muy pequeño

Finalmente, se calculó el índice de disimilitud de Bray-Curtis entre localidades en base a las abundancias relativas de las familias bacterianas y se realizó un análisis de escalamiento multidimensional no métrico (NMDS) (Figura 5b). De esta manera se

aprecia que existe un considerable grado de disimilitud entre todos los sitios, quizás a excepción de Litch y Nelson. Además, Snow parece ser la localidad con la mayor disimilitud a todos los otros sitios en el contexto de diversidad bacteriana.

#### 4.5 Diversidad taxonómica de fungi

De las 118 secuencias alineadas con fungi, 92 (77,97%) correspondieron a Ascomycota, 16 a Basidiomycota (13,56%) y 8 a Chytridiomycota (6,78%) con las dos secuencias restantes alineando a organismos dentro de las divisiones Mucoromycota y Zoopagomycota respectivamente (0,85% cada uno). Al analizar con más detalle estos alineamientos se observa que la distribución de los distintos niveles taxonómicos inferiores es significativamente desigual entre los sitios de muestreo, no habiendo una tendencia clara en cuanto a la proporción de cada clase (Figura 6).

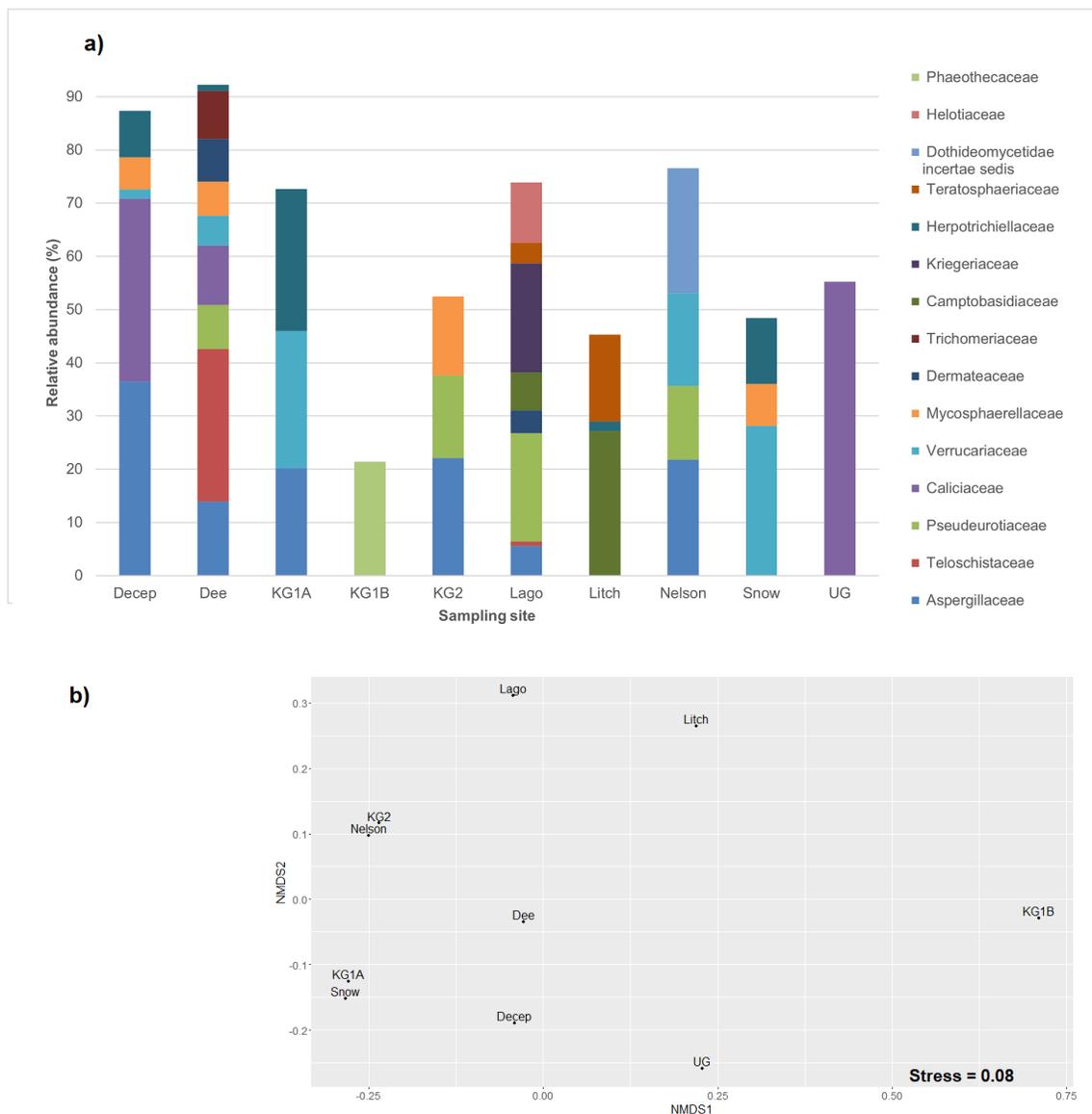


**Figura 6. Abundancia relativa de fungi a nivel de clase.** Los valores de abundancia relativa fueron calculados en base a los RPKM de cada taxón y el RPKM total de cada localidad.

Se encontraron un total de 15 clases distintas de fungi, siendo Eurotiomycetes la más abundante en general. Sin embargo, ninguna de las clases se observó en común en

todas las muestras, con dos clases (Eurotiomycetes y Dothideomycetes) estando presentes en todas las localidades excepto UG. De las quince clases, siete pertenecen a Ascomycota, cuatro a Basidiomycota, dos a Chytridiomycota y una a Mucoromycota y Zoopagomycota respectivamente. A pesar de la distribución desigual, las cuatro clases más abundantes pertenecen a Ascomycota; además, la clase más abundante de cada localidad siempre fue una de estas cuatro, las cuales son Eurotiomycetes; Lecanoromycetes; Leotiomycetes y Dothideomycetes. A pesar de lo anterior, entre las diez muestras no se presenta ninguna clase compartida en su totalidad, con Eurotiomycetes y Dothideomycetes presentes en nueve de las diez muestras. La localidad que concentra más clases únicas es la isla Litchfield, con dos clases ausentes en todas las demás comunidades, siendo Decepción y Lagotellerie las únicas otras localidades que contienen una clase única para su muestra.

A nivel de familia, se encontró que del total de 53 familias observadas solo una no poseía una abundancia relativa mayor a 1% en al menos una localidad (*Powellomycetaceae*, de Chytridiomycota). Además, si bien las 15 familias con mayor abundancia en promedio concentraron más del 60% de la abundancia relativa promedio, existió un alto sesgo, ya que en localidades como Dee estas familias concentraban casi el 90% de la abundancia total mientras que en KG1B solo una familia de estas 15 se veía representada (Figura 7a).



**Figura 7. Diversidad de fungi al nivel de familia. a)** Abundancia relativa de las 15 familias más representadas en promedio. **b)** Escalamiento multidimensional no métrico basado en el índice de disimilitud de Bray-Curtis entre sitios al nivel de familia.

Se observa nuevamente un alto grado de diferencias entre cuáles son las familias más abundantes e incluso representadas en absoluto entre las distintas localidades, con sitios como KG1B y UG teniendo solo una familia de las 15 más abundantes en general presentes en sus muestras. Para visualizar de mejor manera las diferencias entre

diversidad al nivel de familia, se realizó el mismo tratamiento que con las familias bacterianas y se realizó un análisis de escalamiento multidimensional no métrico (NMDS) (Figura 7b). Con esto se permite apreciar que a pesar de que existen notables diferencias en la abundancia entre las familias más representadas, al considerar el total de las familias observadas, sitios como KG1A y Snow, junto a KG2 y Nelson presentan similitudes considerables. Por otro lado, los demás sitios parecen presentar diferencias notables entre ellos, especialmente KG1B y, en menor medida, UG y Litch.

## 5. DISCUSIÓN Y PROYECCIONES

En este estudio se analizaron datos metagenómicos derivados de una secuenciación por escopeta de diez muestras de localidades antárticas. El alcance de secuenciación de escopeta utilizado permitió tener una visión tanto de la composición taxonómica como del “perfil metabólico” de las comunidades microbianas presentes en los suelos de estas localidades. Nuestros resultados indican que existe un alto grado de disimilitud tanto en metabolismo como en composición taxonómica entre las localidades estudiadas. Cinco de las localidades estudiadas se encuentran en relativa cercanía, perteneciendo al archipiélago Shetland del sur, y aun así no se encontró un grado significativo de similitud en los ámbitos estudiados. Este resultado no deja de estar dentro de lo esperado, ya que es conocido que ambientes de suelo con condiciones biogeoquímicas muy distintas pueden existir a tan solo centímetros de distancia y con comunidades microbianas muy distintas acorde a esta diferencia de condiciones ambientales (Fierer, 2017). Sin embargo, se ha observado que la composición bacteriana antártica sí suele agruparse en base a latitudes geográficas a gran escala (Chong y col., 2012), lo cual se ve reflejado en cierta medida en nuestros resultados, estando agrupadas las islas Shetland del sur principalmente entre ellas antes que con UG o las islas de la península antártica, tanto en metabolismo como en taxonomía.

En cuanto a la identificación de procariontes, si bien se realizó un alcance de secuenciación de escopeta, una vez se aplicaron los filtros correspondientes, solamente se encontraron lecturas codificantes de regiones de marcadores moleculares de bacterias. A pesar de que no podemos atribuir la ausencia de lecturas de arqueas a ningún factor en particular, está descrito que una de las falencias de la secuenciación de genoma completo es la dependencia que presenta con la información existente en las

bases de datos utilizadas, en donde no solo arqueas, sino que también organismos endémicos poco estudiados podrían no estar presentes (Doytchinov & Dimov, 2022). De todas maneras, en estudios previos se ha observado que la abundancia y diversidad de arqueas es considerablemente baja en suelos antárticos (Lambrechts y col., 2019; Marcoleta y col., 2022). Por otro lado, se ha descrito que los filos bacterianos que se observaron con mayor abundancia suelen encontrarse altamente representados en una variedad de suelos antárticos (Chong y col., 2012; Pearce y col., 2012; Wang y col., 2015). Al mismo tiempo, en todas las muestras existió una mayor representación de bacterias por sobre fungi, lo cual podría indicar que la biomasa de bacterias supera considerablemente a la de fungi a pesar del menor tamaño de las células procariontes (Schenk y col., 2019). Si bien la diversidad de Fungi no fue tan alta como la de bacterias, las proporciones de los filos encontrados (77.97% Ascomycota y 13.56% Basidiomycota) se condicen con varios trabajos realizados con alcances libres de cultivo en la antártica y otros suelos polares (Ferrari y col., 2015; Baeza y col., 2017) en donde Ascomycota y luego Basidiomycota comprenden la mayoría de los fungi identificados.

La mayoría de los estudios que describen la taxonomía de microbiomas antárticos en la actualidad han sido realizados en base a un alcance de secuenciación con amplicones, lo que presenta ciertas ventajas en cuanto a costos monetarios, acceso a conductos de trabajo preestablecidos y existencia de una mayor cantidad de trabajos similares con los cuales comparar y apoyarse (Ranjan y col., 2015). Sin embargo, si se superan las dificultades asociadas con realizar un estudio con secuenciación por escopeta, i.e. mayores costos de trabajo; necesidad de una mayor cantidad de DNA recolectado y por consecuencia un mejor trabajo de conservación de este, este alcance nos permite identificar una mayor cantidad de taxones poco representados en la muestra, además de entregar una mejor resolución que la obtenida por alcances de secuenciación de

amplicones (Durazzi y col., 2021; Doytchinov & Dimov, 2022). Gracias a este alcance es que en este trabajo fuimos capaces de identificar tanto la composición microbiana como ciertos procesos metabólicos presentes en estos suelos. Sin embargo, no es posible asegurar que las diferencias en metabolismo están dadas netamente por las especies identificadas en este trabajo, por lo que este trabajo debe ser considerado principalmente un primer paso en cuanto a la identificación del potencial metabólico y composición taxonómica de estos suelos. Esto abre una avenida de investigación en cuanto a la búsqueda de compuestos de interés presentes en estos sitios, por ejemplo, mediante un análisis más meticuloso de los genes funcionales encontrados en estas muestras. Paralelamente, se podría complementar y profundizar la identificación de la composición taxonómica mediante búsquedas en otras bases de datos que contengan una mayor cantidad de taxones no cultivados o endémicos del continente antártico y, de esta forma, observar qué otras especies podrían estar presentes en estas localidades.

## 6. CONCLUSIONES

En conclusión, el análisis de datos metagenómicos realizado en este trabajo permitió observar parte de la diversidad microbiana y potencial metabólico de estos ambientes antárticos. Los resultados obtenidos se condicen en gran parte con la literatura existente, con una mayor abundancia de filos bacterianos como Proteobacteria, Actinobacteria, Bacteroidetes y Firmicutes, además de una relativamente baja representación de fungi, con Ascomycota y Basidiomycota como principales taxones encontrados. Por otro lado, se encontró que la distancia no corresponde a un factor predominante en cuanto a la disimilitud metabólica y taxonómica de estas localidades. Adicionalmente, la mayoría de los módulos metabólicos se encontraron en todas las localidades muestreadas, con una notable abundancia de procesos de fijación de carbono. Si bien se requieren más análisis para determinar las relaciones que existen entre los individuos presentes y los procesos metabólicos que ocurren en estos suelos, esta investigación abre las puertas a la búsqueda de especies y compuestos de interés novedosos contenidos en estas muestras.

## REFERENCIAS

- **Aliyu, H., De Maayer, P., Sjöling, S., & Cowan, D. A. (2017).** Metagenomic Analysis of Low-Temperature Environments. *Psychrophiles: From Biodiversity to Biotechnology*, 389–421.
- **Allali, I., Arnold, J. W., Roach, J., Cadenas, M. B., Butz, N., Hassan, H. M., Koci, M., Ballou, A., Mendoza, M., Ali, R., & Azcarate-Peril, M. A. (2017).** A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. *BMC Microbiology*, 17(1).
- **Baeza, M., Barahona, S., Alcaíno, J., & Cifuentes, V. (2017).** Amplicon-Metagenomic Analysis of Fungi from Antarctic Terrestrial Habitats. *Frontiers in Microbiology*, 8.
- **Bikel, S., Valdez-Lara, A., Cornejo-Granados, F., Rico, K., Canizales-Quinteros, S., Soberón, X., Del Pozo-Yauner, L., & Ochoa-Leyva, A. (2015).** Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. *Computational and Structural Biotechnology Journal*, 13, 390–401.
- **Brooks, S. T., Jabour, J., van den Hoff, J., & Bergstrom, D. M. (2019).** Our footprint on Antarctica competes with nature for rare ice-free land. *Nature Sustainability*, 2(3), 185–190.
- **Chong, C., Pearce, D., Convey, P., Yew, W., & Tan, I. (2012).** Patterns in the distribution of soil bacterial 16S rRNA gene sequences from different regions of Antarctica. *Geoderma*, 181–182, 45–55.
- **Cortés-López, N. G., Ordóñez-Baquera, P. L., & Domínguez-Viveros, J. (2020).** Herramientas moleculares utilizadas para el análisis metagenómico. Revisión. *Revista Mexicana De Ciencias Pecuarias*, 11(4), 1150–1173.
- **Cowan, D., & Tow, L. (2004).** Endangered Antarctic Environments. *Annual Review of Microbiology*, 58(1), 649–690.
- **Cowan, D., Ramond, J. B., Makhalyane, T., & De Maayer, P. (2015).** Metagenomics of extreme environments. *Current Opinion in Microbiology*, 25, 97–102.
- **Doytchinov, V. V., & Dimov, S. G. (2022).** Microbial Community Composition of the Antarctic Ecosystems: Review of the Bacteria, Fungi, and Archaea Identified through an NGS-Based Metagenomics Approach. *Life*, 12(6), 916.
- **Dumorné, K., Cordova, D. C., Astorga-Elo, M., & Renganathan, P. (2017).** Extremozymes: A Potential Source for Industrial Applications. *Journal of Microbiology and Biotechnology*, 27(4), 649–659.

- **Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D., & De Cesare, A. (2021).** Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Scientific Reports*, 11(1).
- **Escobar-Zepeda, A., Vera-Ponce de León, A., & Sanchez-Flores, A. (2015).** The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. *Frontiers in Genetics*, 6.
- **Feng, G., Xie, T., Wang, X., Bai, J., Tang, L., Zhao, H., Wei, W., Wang, M., & Zhao, Y. (2018).** Metagenomic analysis of microbial community and function involved in cd-contaminated soil. *BMC Microbiology*, 18(1).
- **Ferrari, B. C., Bissett, A., Snape, I., van Dorst, J., Palmer, A. S., Ji, M., Siciliano, S. D., Stark, J. S., Winsley, T., & Brown, M. V. (2015).** Geological connectivity drives microbial community structure and connectivity in polar, terrestrial ecosystems. *Environmental Microbiology*, 18(6), 1834–1849.
- **Fierer, N. (2017).** Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, 15(10), 579–590.
- **Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., & Knight, R. (2018).** Current understanding of the human microbiome. *Nature Medicine*, 24(4), 392–400.
- **Gilbert, J. A., Jansson, J. K., & Knight, R. (2014).** The Earth Microbiome project: successes and aspirations. *BMC Biology*, 12(1).
- **Handelsman, J. (2004).** Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, 68(4), 669–685.
- **Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., Leopold, S. R., Hanson, B. M., Agresta, H. O., Gerstein, M., Sodergren, E., & Weinstock, G. M. (2019).** Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications*, 10(1).
- **Lambrechts, S., Willems, A., & Tahon, G. (2019).** Uncovering the Uncultivated Majority in Antarctic Soils: Toward a Synergistic Approach. *Frontiers in Microbiology*, 10.
- **Lee, J. R., Raymond, B., Bracegirdle, T. J., Chadès, I., Fuller, R. A., Shaw, J. D., & Terauds, A. (2017).** Climate change drives expansion of Antarctic ice-free habitat. *Nature*, 547(7661), 49–54.
- **Liu, S., Moon, C. D., Zheng, N., Huws, S., Zhao, S., & Wang, J. (2022).** Opportunities and challenges of using metagenomic data to bring uncultured microbes into cultivation. *Microbiome*, 10(1).

- **Lopatina, A., Medvedeva, S., Shmakov, S., Logacheva, M. D., Krylenkov, V., & Severinov, K. (2016).** Metagenomic Analysis of Bacterial Communities of Antarctic Surface Snow. *Frontiers in Microbiology*, 7.
- **Luo, C., Tsementzi, D., Kyripides, N., Read, T., & Konstantinidis, K. T. (2012).** Direct Comparisons of Illumina vs. Roche 454 Sequencing Technologies on the Same Microbial Community DNA Sample. *PLoS ONE*, 7(2), e30087.
- **Marchesi, J. R., & Ravel, J. (2015).** The vocabulary of microbiome research: a proposal. *Microbiome*, 3(1).
- **Marcoleta, A. E., Arros, P., Varas, M. A., Costa, J., Rojas-Salgado, J., Berríos-Pastén, C., Tapia-Fuentes, S., Silva, D., Fierro, J., Canales, N., Chávez, F. P., Gaete, A., González, M., Allende, M. L., & Lagos, R. (2022).** The highly diverse Antarctic Peninsula soil microbiota as a source of novel resistance genes. *Science of the Total Environment*, 810, 152003.
- **Metzker, M. L. (2010).** Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11(1), 31–46.
- **Mirete, S., Morgante, V., & González-Pastor, J. E. (2016).** Functional metagenomics of extreme environments. *Current Opinion in Biotechnology*, 38, 143–149.
- **Morita, R. Y. (1999).** Extremes of Biodiversity. *BioScience*, 49(3), 245–248.
- **Pearce, D. A., Newsham, K. K., Thorne, M. A. S., Calvo-Bado, L., Krsek, M., Laskaris, P., Hodson, A., & Wellington, E. M. (2012).** Metagenomic Analysis of a Southern Maritime Antarctic Soil. *Frontiers in Microbiology*, 3.
- **Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., Bonazzi, V., McEwen, J. E., Wetterstrand, K. A., Deal, C., Baker, C. C., Di Francesco, V., Howcroft, T. K., Karp, R. W., Lunsford, R. D., Wellington, C. R., Belachew, T., Wright, M., Giblin, C., . . . Guyer, M. (2009).** The NIH Human Microbiome Project. *Genome Research*, 19(12), 2317–2323.
- **Raggi, L., García-Guevara, F., Godoy-Lozano, E. E., Martínez-Santana, A., Escobar-Zepeda, A., Gutierrez-Rios, R. M., Loza, A., Merino, E., Sanchez-Flores, A., Licea-Navarro, A., Pardo-Lopez, L., Segovia, L., & Juarez, K. (2020).** Metagenomic Profiling and Microbial Metabolic Potential of Perdido Fold Belt (NW) and Campeche Knolls (SE) in the Gulf of Mexico. *Frontiers in Microbiology*, 11.
- **Ranjan, R., Rani, A., Metwally, A., McGee, H. S., & Perkins, D. L. (2016).** Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and Biophysical Research Communications*, 469(4), 967–977.

- **Schenk, J., Geisen, S., Kleinboelting, N., & Traunspurger, W. (2019).** Metabarcoding data allow for reliable biomass estimates in the most abundant animals on earth. *Metabarcoding and Metagenomics*, 3.
- **Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W., Bolchacova, E., Voigt, K., Crous, P. W., Miller, A. N., Wingfield, M. J., Aime, M. C., An, K. D., Bai, F. Y., Barreto, R. W., Begerow, D., Bergeron, M. J., Blackwell, M., . . . Schindel, D. (2012).** Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. *Proceedings of the National Academy of Sciences*, 109(16), 6241–6246.
- **Solden, L., Lloyd, K., & Wrighton, K. (2016).** The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Current Opinion in Microbiology*, 31, 217–226.
- **Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d’Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., . . . Velayoudon, D. (2015).** Structure and function of the global ocean microbiome. *Science*, 348(6237).
- **Sunagawa, S., Mende, D. R., Zeller, G., Izquierdo-Carrasco, F., Berger, S. A., Kultima, J. R., Coelho, L. P., Arumugam, M., Tap, J., Nielsen, H. B., Rasmussen, S., Brunak, S., Pedersen, O., Guarner, F., de Vos, W. M., Wang, J., Li, J., Doré, J., Ehrlich, S. D., . . . Bork, P. (2013).** Metagenomic species profiling using universal phylogenetic marker genes. *Nature Methods*, 10(12), 1196–1199.
- **Tessler, M., Neumann, J. S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L. F. M., Segovia, B. T., Lansac-Toha, F. A., Lemke, M., DeSalle, R., Mason, C. E., & Brugler, M. R. (2017).** Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Scientific Reports*, 7(1).
- **Trindade, M., van Zyl, L. J., Navarro-Fernández, J., & Abd Elrazak, A. (2015).** Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. *Frontiers in Microbiology*, 6.
- **Varga, A., Csuka, P., Sonesouphap, O., Bánóczy, G., Toşa, M. I., Katona, G., Molnár, Z., Bencze, L. C., Poppe, L., & Paizs, C. (2021).** A novel phenylalanine ammonia-lyase from *Pseudozyma antarctica* for stereoselective biotransformations of unnatural amino acids. *Catalysis Today*, 366, 185–194.
- **Wang, N. F., Zhang, T., Zhang, F., Wang, E. T., He, J. F., Ding, H., Zhang, B. T., Liu, J., Ran, X. B., & Zang, J. Y. (2015).** Diversity and structure of soil bacterial communities

in the Fildes Region (maritime Antarctica) as revealed by 454 pyrosequencing. *Frontiers in Microbiology*, 6.

- **Waschulin, V., Borsetto, C., James, R., Newsham, K. K., Donadio, S., Corre, C., & Wellington, E. (2021).** Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. *The ISME Journal*, 16(1), 101–111.
- **Zhang, L., Chen, F., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y., Hao, H., Yi, W., Li, M., & Xie, Y. (2021).** Advances in Metagenomics and Its Application in Environmental Microorganisms. *Frontiers in Microbiology*, 12.
- **Zheng, W., Zhang, Z., Liu, C., Qiao, Y., Zhou, D., Qu, J., An, H., Xiong, M., Zhu, Z., & Zhao, X. (2015).** Metagenomic sequencing reveals altered metabolic pathways in the oral microbiota of sailors during a long sea voyage. *Scientific Reports*, 5(1).
- **Zucconi, L., & Buzzini, P. (2021).** Editorial: Microbial Communities of Polar and Alpine Soils. *Frontiers in Microbiology*, 12.