



UNIVERSIDAD NACIONAL DEL ALTIPLANO

FACULTAD DE CIENCIAS BIOLÓGICAS

ESCUELA PROFESIONAL DE BIOLOGÍA



**PERFIL TAXONÓMICO DE MICROBIOMAS BACTERIANOS
CON LA SECUENCIACIÓN SHOTGUN PRESENTES EN LA
CAPTACIÓN CHIMU Y BAHÍA INTERNA DEL LAGO TITICACA,
PUNO, PERÚ, 2023**

TESIS

PRESENTADA POR:

Bach. ABEL VIDAL MAYTA MENDIZABAL

PARA OPTAR EL TÍTULO PROFESIONAL DE:

**LICENCIADO EN BIOLOGÍA: MICROBIOLOGÍA Y
LABORATORIO CLÍNICO**

PUNO – PERU

2024



NOMBRE DEL TRABAJO

PERFIL TAXONÓMICO DE MICROBIOMAS BACTERIANOS CON LA SECUENCIACIÓN SHOTGUN PRESENTES EN LA CAPTACIÓN

AUTOR

ABEL VIDAL MAYTA MENDIZABAL

RECUENTO DE PALABRAS

20221 Words

RECUENTO DE CARACTERES

128186 Characters

RECUENTO DE PÁGINAS

127 Pages

TAMAÑO DEL ARCHIVO

5.8MB

FECHA DE ENTREGA

Jul 31, 2024 4:06 PM GMT-5

FECHA DEL INFORME

Jul 31, 2024 4:08 PM GMT-5

● **6% de similitud general**

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 6% Base de datos de Internet
- Base de datos de Crossref
- 4% Base de datos de trabajos entregados
- 1% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

● **Excluir del Reporte de Similitud**

- Material bibliográfico
- Material citado
- Bloques de texto excluidos manualmente
- Material citado
- Coincidencia baja (menos de 8 palabras)





DEDICATORIA

A mi querida madre: Esperanza Mendizabal Sarmiento, mi padre Justino Mayta Lerma, por no rendirse conmigo, gracias a ellos estoy en donde estoy y este es el regalo que los dedico y ofrezco, por todo el sacrificio que hicieron por mí y por mis hermanos para que podamos salir adelante.

También al amor de mi vida Carol Keitlin Colquehuanca Fernandez, quien me acompaño en todo momento y dándome el apoyo emocional que necesitaba, teniéndome paciencia y exigiéndome a la vez. Siempre será la mejor del mundo.

Abel Vidal Mayta Mendizabal



AGRADECIMIENTOS

*Al a doctora María Trinidad Romero Torrez
por la confianza, el esfuerzo de darme nuevas
oportunidades y caminos, por seguir
enseñándome después de muchos años y
espero seguir aprendiendo más de ella, por
exigirme y dándome a entender que todo lo
puedo hacer, siempre será la mejor.*

*Al lic. Ángel Francisco Soliz Gozar, ya más mi amigo
quien no dudo en enseñarme todo lo que sabe y
mostrarme más caminos, más objetivos los cuales llegar,
brindándome la confianza y su corazón te agradezco
hermano de todo corazón, y este es el resultado la cual
te dedico.*

*De manera muy especial a mi amiga mi loca
Evelin Quispe Calsin, quien, como ella para
decirme sin miedo al éxito, teniéndome la
paciencia, comprensión y la amabilidad para
superar mis momentos en los que me quería rendir.
Agradecido con Dios por haberla puesto en mi
camino.*

Abel Vidal Mayta Mendizabal



ÍNDICE GENERAL

| | Pag. |
|--|-------------|
| DEDICATORIA | |
| AGRADECIMIENTOS | |
| ÍNDICE GENERAL | |
| ÍNDICE DE FIGURAS | |
| ÍNDICE DE TABLAS | |
| ÍNDICE DE ACRÓNIMOS | |
| RESUMEN | 13 |
| ABSTRACT..... | 14 |
| CAPITULO I | |
| INTRODUCCION | |
| 1.1 OBJETIVO GENERAL | 17 |
| 1.2 OBJETIVOS ESPECIFICOS | 17 |
| CAPITULO II | |
| REVISIÓN DE LITERATURA | |
| 2.1 ANTECEDENTES..... | 18 |
| 2.1.1 Metagenomas de microbiomas bacterianos..... | 18 |
| 2.1.2 Metagenoma bacteriana en lagos | 19 |
| 2.1.3 Herramientas bioinformáticas para metagenomica | 20 |
| 2.2 MARCO TEÓRICO | 21 |
| 2.2.1. Lago Titicaca | 21 |
| 2.2.2. Microbiomas bacterianos | 25 |
| 2.2.3 Metagenoma..... | 28 |
| 2.2.4 Secuenciación Shotgun | 37 |



| | | |
|-------|--|----|
| 2.2.5 | Perfil Taxonómico | 41 |
| 2.2.6 | Herramientas computacionales para el análisis metagenómico | 42 |

CAPITULO III

MATERIALES Y METODOS

| | | |
|------------|---|-----------|
| 3.1 | TIPO DE ESTUDIO | 51 |
| 3.2 | LUGAR DE ESTUDIO..... | 51 |
| 3.3 | PUNTO DE MUESTREO | 51 |
| 3.4 | RECOLECCIÓN Y PREPARACIÓN DE MUESTRAS..... | 54 |
| 3.4.1 | Recolección de muestra..... | 54 |
| 3.4.2 | Filtración de las muestras para la concentración de microorganismos ... | 54 |
| 3.5 | EXTRACCIÓN DE ADN BACTERIANO | 56 |
| 3.6 | CALIDAD E INTEGRIDAD DEL ADN BACTERIANO..... | 58 |
| 3.6.1 | Calidad/cantidad del ADN | 58 |
| 3.6.2 | Integridad de ADN | 59 |
| 3.7 | ANÁLISIS METAGENÓMICO POR SECUENCIACIÓN SHOTGUN | 60 |
| 3.8 | CALIDAD DE LA DATA SECUENCIADA | 64 |
| 3.9 | ANÁLISIS BIOINFORMÁTICO..... | 65 |
| 3.9.2 | Uso de Metaphlan4 para el perfil taxonómico por medio de árboles filogenéticos y el mapa de calor | 69 |

CAPITULO IV

RESULTADOS Y DISCUSIÓN

| | | |
|------------|---|-----------|
| 4.1 | EXTRACCIÓN DE ADN CALIDAD, CONCENTRACIÓN E INTEGRIDAD..... | 81 |
| 4.1.2 | Análisis de secuenciamiento Shotgun..... | 85 |



| | | |
|-------------|--|------------|
| 4.2 | ÁRBOL FILOGENÉTICO DE BACTERIAS PRESENTES EN LA BAHÍA INTERNA (BI)..... | 86 |
| 4.3 | ÁRBOL FILOGENÉTICO DE BACTERIAS PRESENTES DE LAS MUESTRAS DE CAPTACIÓN CHIMÚ (CH) | 89 |
| 4.4 | COMPARACIÓN DE ABUNDANCIA DE BACTERIAS PRESENTES EN LA BAHÍA INTERNA (BI) Y LA CAPTACIÓN CHIMÚ (CH) POR MEDIO DE MAPA DE CALOR..... | 91 |
| V. | CONCLUSIONES | 94 |
| VI. | RECOMENDACIONES | 95 |
| VII. | REFERENCIAS BIBLIOGRAFICAS..... | 96 |
| | ANEXOS..... | 102 |

AREA: Ciencias Biomédicas

SUBLINEA DE INVESTIGACIÓN: Diagnostico y Epidemiologia

FECHA DE SUSTENCTACIÓN: 31/05/2024



ÍNDICE DE FIGURAS

| | Pá g. |
|--|--------------|
| Figura 1 Cuenca del lago Titicaca (Stanish & Levine, 2011). | 22 |
| Figura 2 Diagrama de flujo de un proyecto metagenómico típico (Gao et al., 2021) | 31 |
| Figura 3 Representación del gel con lectura de barrido electroforético (Iranzo, 2022) | 37 |
| Figura 4 MetaPhlan4 integra secuencias de referencia de genomas aislados y ensamblados en Metagenomas para el perfil taxonómico del metagenoma. (Blanco et al., 2023) | 46 |
| Figura 5 Puntos de muestreo en la captación chimú. | 52 |
| Figura 6 Puntos de muestreo de la bahía interna | 53 |
| Figura 7 Diagrama de flujo de trabajo metagenómico. Fuente: Propia | 80 |
| Figura 8 Integridad de ADN de las muestras filtradas de la Bahía Interna y la Captación Chimú. | 84 |
| Figura 9 Árbol filogeético del metagenoma bacteriano aislado de la Bahía Interna del lago Titicaca..... | 88 |
| Figura 10 Árbol filogenético del metagenoma bacteriano aislado de la Captación Chimu del lago Titicaca..... | 90 |
| Figura 11 Relaciones genómicas y abundancia en el metagenoma bacteriano aislado de la Bahía Interior y Captación Chimu del lago Titicaca. | 93 |



ÍNDICE DE TABLAS

| | Pág. |
|--|-------------|
| Tabla 1 Listado de plantas de tratamiento. | 24 |
| Tabla 2 Límites legales permitidos en aguas utilizadas para el consumo humano según normativa peruana DS N°031-2010-SA (DIGESA, 2011)..... | 24 |
| Tabla 3 Análisis microbiológico de aguas para el consumo humano..... | 24 |
| Tabla 4 Análisis de metales pesados de aguas para el consumo humano..... | 25 |
| Tabla 5 Comparación de tiempo de ejecución y memoria "Consumo de MetaPhlAn 3, MetaPhlAn2, mOTUs2 y Kraken+Bracken en los 5 metagenomas de HMP ". | 32 |
| Tabla 6 Valores indicativos de pureza y muestras de ADN previamente filtradas. .. | 34 |
| Tabla 7 Criterios de validez y puntuación asignada según la integridad de ADN observada en gel de agarosa..... | 36 |
| Tabla 8 Puntos de muestreo para la evaluación de análisis metagenómico..... | 53 |
| Tabla 9 Valores de pureza de ADN genómico. | 59 |
| Tabla 10 Recursos bioinformáticos empleados para el análisis metagenómico shotgun..... | 65 |
| Tabla 11 Lectura de valores que resaltan la calidad de DNA de la muestra de Bahía interna | 81 |
| Tabla 12 Lectura de valores que resaltan la calidad de DNA de la muestra de Captación chimu | 82 |
| Tabla 13 Lectura de valores que resaltan la concentración de DNA de la muestra de Bahía Interna..... | 83 |
| Tabla 14 Lectura de valores que resaltan la concentración de DNA de la muestra de Captación chimu | 83 |



| | | |
|-----------------|--|----|
| Tabla 15 | Resumen de datos de la secuenciación genómica de la BI. | 85 |
| Tabla 16 | Resumen de datos de la secuenciación genómica de la Captación chimu. ... | 86 |



ÍNDICE DE ACRÓNIMOS

SI: Sistema operativo.

Pb: Pares de bases

NBCI: Centro Nacional para la Información Biotecnológica (National Center for Biotechnology Information)

BI: Bahía Interna.

CH: Captación Chimú.

NGS: Secuenciación de próxima generación.

CLI: Interfaz de línea de comando.

SO: Sistema operativo.

LPS: Áreas menos pobladas.

HPS: Áreas altamente pobladas.

ARG: Genes resistentes a antibióticos.

MRG: Genes de resistencia a metales.

BE: Buffer de elución.

ADN: Ácido desoxirribonucleico.

ARN: Ácido ribonucleico.

MAG: Genomas ensamblados en Metagenomas.

ETAP: Estación de tratamiento de aguas residuales.

EDAR: Estaciones depuradoras de aguas residuales.



UFC: Unidades formadoras de colonias.

NTU: Unidades nefelométricas de turbidez.

HMP: Proyecto de microbioma humano.

PE: Tipo de lectura emparejada (paired-end)

RAM: Memoria de acceso aleatorio (Random Access Memory)

PCR: Reacción en cadena de la polimerasa

MB: Megabyte

TB: Terabyte

GB: Gigabyte

uSGB: Contenedores genómicos a nivel de especies aun por caracterizar (desconocidas)

kSGB: Contenedores genómicos a nivel de especies existentes (conocidas)

SE: Tipo de lectura simple (single end)

NLC: Número de lecturas por carrera

MSNM: Metros sobre el nivel del mar

SED: Stream editor



RESUMEN

El Lago Titicaca, ubicado a una altura de 3,825 metros sobre el nivel del mar y reconocido como el lago navegable más extenso (8000 km²) y alto del mundo; es un ecosistema de agua dulce expuesto a contaminación permanente, por lo tanto, es de vital importancia conocer la presencia y diversidad de las comunidades bacterianas que mantienen el equilibrio ecológico, y que a la fecha aún no han sido estudiadas. En tal razón, se planteó como objetivo de trabajo determinar el perfil taxonómico de microbiomas bacterianos mediante la secuenciación Shotgun en la captación Chimú y Bahía Interna del Lago Titicaca. La metodología empleada fue de tipo observacional y descriptivo. Se recolectaron muestras de agua, con cinco repeticiones cada una. Posteriormente, se procedió a la extracción y evaluación de la cantidad, calidad e integridad del DNA genómico bacteriano de cada una de las muestras obtenidas y enviadas a -80°C a BTSconsultores para la construcción de librerías de los genomas completos por la secuenciación Shotgun. Los resultados muestran que en la Bahía Interna al clado más representativo son las Alphaproteobacterias, con la especie *Alphaproteobacterium SCGC AAA028 D10*. En la captación Chimú, el clado con mayor representatividad son las Actinomicetas, con la especie *Candidatus nanopelagicus abundans*. En conclusión, se obtuvo el árbol filogenético de la Bahía Interna y la captación Chimú resaltando la presencia de *Alpha proteobacterium SCGC AAA028 D10*, como especie representativa en los dos árboles filogenéticos; y según el mapa de calor, *Alpha proteobacterium SCGC AAA028 D10* presenta la abundancia relativa más alta para ambas zonas de estudio.

Palabras clave: Árbol filogenético, Mapa de calor, Metagenoma, Shotgun, Taxonomía.



ABSTRACT

Lake Titicaca, located at a height of 3825 meters above sea level and recognized as the largest (8000 km²) and highest navigable lake in the world; It is a freshwater ecosystem exposed to permanent contamination; therefore, it is vitally important to know the presence and diversity of the bacterial communities that maintain ecological balance, and which to date have not yet been studied. For this reason, the work objective was set to determine the taxonomic profile of bacterial microbiomes through Shotgun sequencing in the Chimu catchment and Internal Bay of Lake Titicaca. The methodology used was observational and descriptive. Water samples were collected, with five repetitions each. Subsequently, the quantity, quality and integrity of the bacterial genomic DNA was extracted and evaluated from each of the samples obtained and sent at -80°C to BTSconsultores for the construction of complete genome libraries by Shotgun sequencing. The results show that in the Inner Bay the most representative clade is the Alphaproteobacteria, with the species *Alphaproteobacterium SCGC AAA028 D10*. In the Chimú catchment, the most representative clade is the Actinomycetas, with the species *Candidatus nanopelagicus abundant*. In conclusion, the phylogenetic tree of the Internal Bay and the Chimu catchment was obtained, highlighting the presence of *Alpha proteobacterium SCGC AAA028 D10*, as a representative species in the two phylogenetic trees; and according to the heat map, *Alpha proteobacterium SCGC AAA028 D10* presents the highest relative abundance for both study areas.

Keywords: Phylogenetic tree, Heat map, Metagenome, Shotgun, Taxonomy.



CAPITULO I

INTRODUCCION

El lago Titicaca cuenta con una antigüedad de 3 millones de años aproximadamente, la cual es catalogada como el lago navegable más alto del mundo ya que se encuentra a 3.800 metros sobre el nivel del mar (msnm) y se extiende por más de 50.000 km², el paisaje del lago Titicaca ofrece una rica variedad de nichos ecológicos explotables con diversas poblaciones microbianas (Stanish & Levine, 2011).

La investigación se refiere a un trabajo genómico bacteriológico, el cual se basa en el estudio de poblaciones microbianas, esto debido a dificultades operativas y la información limitada obtenida de microorganismos individuales. El paso que se sigue para determinar la identidad de un microorganismo es la “identificación”. Un microorganismo se identifica correctamente cuando se encuentra que la descripción de la especie es idéntica a las características observadas en ese microorganismo. Una unidad taxonómica comúnmente utilizada es la especie. Esta clase sistemática se define como una colección de poblaciones clonales con alta similitud fenotípica que son distintas de otras poblaciones clonales. Una propiedad de los microorganismos es el gran aumento del tamaño de la población en un corto período de tiempo, lo que determina los cambios genéticos que pueden establecer una continuidad adquirida. Esto dificulta la distinción de especies basándose en un pequeño número de características (Usyk et al., 2023).

La secuenciación metagenómica Shotgun es una técnica de laboratorio que determina la secuencia del ADN genómico de un organismo. El método consiste en descomponer el genoma en una colección de pequeños fragmentos de ADN dispuestos individualmente. El método más eficaz para secuenciar largos tramos de ADN es la secuenciación Shotgun. Para ello, el ADN a analizar se corta aleatoriamente en varios



fragmentos más pequeños y cada fragmento se secuencia individualmente. La secuencia resultante se lee a partir de diferentes fragmentos y luego se analiza mediante un programa informático en busca de fragmentos de secuencia que sean idénticos entre sí en diferentes lecturas. Cuando se identificaron regiones idénticas que se superponían entre sí, se unieron las secuencias de las dos lecturas. Este proceso informático se repite una y otra vez hasta que, con el tiempo, se obtiene a partir del ADN la secuencia completa de ese fragmento. La fragmentación aleatoria inicial y la lectura del ADN llevaron a que este método se denominara "secuenciación de Shotgun" (Wang & Bao, 2022).

Aunque existe información limitada sobre el conocimiento molecular del microbioma, incluidos los microorganismos no cultivados del lago Titicaca. Como resultado, la gran mayoría de los recursos genómicos de la región siguen estando subexplotados. Siendo que no se encontró estudios que describan la composición de las comunidades microbianas se decidió abordar este problema, para lo cual se aplicó la metagenómica para tener una mejor comprensión de los microorganismos en el medio ambiente. Para esto se aplicó, análisis de secuenciamiento Shotgun, además de herramientas bioinformáticas, sistemas de información (SI) y del sistema operativo (SO) Linux/Unix que emplea recursos de la terminal con una interfaz Línea-comando (CLI), para crear y manipular la data del secuenciamiento (Zubairi & Weber, 2018). Por lo cual, en este trabajo se realizó el perfil taxonómico del metagenoma bacteriano presente en la BI y CH del lago Titicaca.



1.1 OBJETIVO GENERAL

- Determinar el perfil taxonómico de los microbiomas bacterianos presentes en el Lago Titicaca.

1.2 OBJETIVOS ESPECIFICOS

- Obtener el árbol filogenético de las bacterias presentes en la Bahía Interna.
- Obtener el árbol filogenético de las bacterias presentes en la Captación Chimú.
- Comparar la abundancia de las bacterias presentes tanto en la Bahía Interna como en la Captación Chimú por medio de un mapa de calor.



CAPITULO II

REVISIÓN DE LITERATURA

2.1 ANTECEDENTES

2.1.1 Metagenomas de microbiomas bacterianos

Como bien sabemos de la existencia de una variedad de ecosistemas complejos que albergan una gran cantidad de biodiversidad procariota, según Xu et al., (2014) en su investigación, denominada la caracterización de metagenómica microbiana de suelos, utilizó 33 metagenomas de diversos sitios de suelo e integraron herramientas computacionales para explorar el desarrollo y la funcionalidad del sistema, características y comunidades microbianas del suelo. La composición microbiana y el potencial metabólico del suelo se dilucidaron exhaustivamente a nivel metagenómico. Se descubrió una serie de biomarcadores metagenómicos que contenían 46 taxones y 33 módulos metabólicos eran significativamente diferentes y podían usarse como indicadores para diferenciar comunidades del suelo.

Mediante métodos computacionales es posible estudiar diferentes microbiomas, que son los conjuntos de genes que posee un grupo específico de microorganismos, como los de la microbiota humana y diversas muestras de suelo, agua y sedimentos, así lo demuestra Santiago, (2018) que tomo un gran número de muestras de agua de sitios afectados por derrames de petróleo en el Golfo de México, con el fin de conocer la comunidad procariota, donde aprovecharon dos técnicas metagenómicas similares: secuenciación de amplicones de la región V4 del gen 16S del ARNr, y secuenciación Shotgun. Los resultados muestran los filos preponderantes: Proteobacteria, Thaumarchaeota, Planctomycetes, Acidobacteria



y Chloroflexi. El estudio permitió inferir los procesos biogeoquímicos activados por los microorganismos para proporcionar el planteamiento de estudios dirigidos para la obtención de productos de interés biotecnológicos.

2.1.2 Metagenoma bacteriana en lagos

Según Ahmad et al., (2021) que evaluaron la constitución y partición del microbiota en el lago Dal en Srinagar, India, descubrieron que las proteobacterias eran el filo bacteriano más cuantioso, encontrando 3097 especies en la muestra del área menos poblada durante el verano (LPS), mientras que la muestra de verano del área altamente poblada (HPS) fue la más diversa entre todas, como lo indica el análisis de diversidad taxonómica.

Du et al., (2022), investigaron la distribución y variedad de genes de resistencia a antibióticos (ARG) y genes de resistencia a metales (MRG) en el sedimento del lago Daihai, China mediante secuenciación y análisis metagenómico, descubriendo que los lugares de muestreo tenían comunidades bacterianas similares siendo las más resaltantes: Proteobacteria, Actinobacteria, Firmicutes y Bacteroidetes.

También Quillaguamán et al., (2021) analizaron el microbioma del lago Alalay Cochabamba, Bolivia, donde utilizaron la secuenciación metagenómica de Shotgun para examinar la diversidad bacteriana, descubriendo que las proteobacterias dominaban la abundancia variable de muestras de la cuenca y del sedimento (25%) donde un segmento significativo de la diversidad bacteriana estaba sin clasificarse. Resultados similares fueron obtenidos por Wei et al., (2023) quienes revelaron perfiles moleculares y de capacidad metabólica de las comunidades microbianas encontradas en el lago Bamucuo, Tibet, teniendo en



cuenta el análisis metagenómico Shotgun, mediante el cual se ensamblaron 75 genomas completos y 74 genomas sin caracterizar, de las cuales las proteobacterias y actinobacterias son los filos más dominantes y hay buenos resultados en función a ARG como también para MRG.

Es cierto que la variedad que distintas comunidades bacterianas dependen mucho de sus entornos ecológicos, así Wang & Bao, (2022) describen la variabilidad de microorganismos procarióticos en suelos del lago salado de Qarhan, Qunghai Tibet, para lo cual utilizaron secuenciamiento de tercera generación para determinar las arqueas y bacterias de sueño salino, alcalino, hallando que las Betaproteobacterias (Proteobacteria) y Halobacterias (Euryarchaeota) son las más dominantes en las áreas de suelos de la zona, comunes en ambientes de alta salinidad que al parecer son necesarios en el ciclo del nitrógeno y el carbono por medio de vías metabólicas singulares.

2.1.3 Herramientas bioinformáticas para metagenómica

En los últimos años, las comunidades microbianas se han convertido en un importante tema de investigación en múltiples disciplinas. Estas comunidades a menudo se examinan mediante secuenciación metagenómica rápida, una técnica que puede proporcionar información única sobre el contenido genómico de las comunidades microbianas. Al respecto, Lindgreen et al., (2016) nos mencionan que, en una evaluación de la precisión y velocidades de las herramientas de análisis metagenómicas para diversos ecosistemas, ya sean terrestres o acuáticos, como también de la piel y el intestino delgado, muestran claramente que las herramientas más utilizadas no son necesariamente las más precisas, y que las herramientas más precisas no son necesariamente las que consumen más tiempo



y que existe un gran grado de variabilidad entre las herramientas disponibles. Estos hallazgos son importantes porque las conclusiones de cualquier estudio metagenómico se verían afectadas por errores en la predicción de la composición de la comunidad y las capacidades funcionales.

Asimismo, Meslier et al., (2022) en su evaluación de plataformas de secuenciación de distintas generaciones, que abarcaron secuenciadores de 2G (Illumina HiSeq 300, MGI DNBSEQ-G400 y DNBSEQ-T7, ThermoFisher Ion GeneStudio S5 e Ion Proton P1) y 3G (Oxford Nanopore Technologies MinION R9 y Pacific Biosciences Sequel II) demostraron que los secuenciadores de la 3G tienen ventajas significativas sobre las plataformas 2G.

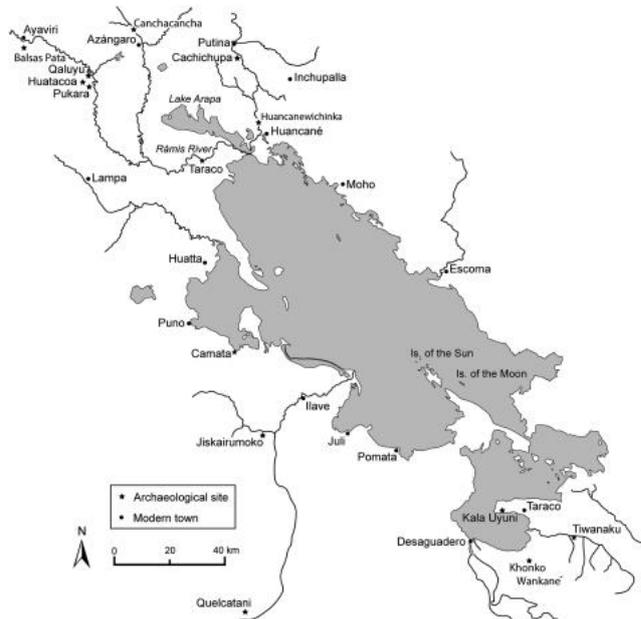
2.2 MARCO TEÓRICO

2.2.1. Lago Titicaca

El lago Titicaca es el lago de agua potable más grande de Sudamérica, con una superficie de agua de 8300 km², además de una longitud aproximada de 19km, con profundidad de 285m y una anchura de 50km (Figura 1), es el lago navegable más alto del mundo (3808 msnm). Tras la Convención de Ramsar sobre Humedales de 1971, las regiones de Perú y Bolivia abarcan más de 13 provincias y 93 distritos. Más de la mitad del territorio del Lago pertenece al Perú esto es el 55% (460 000 has) del espejo de agua y el 45% (379 000 has) pertenece a Bolivia, datos que fueron incluidos en la lista de Humedales de Importancia Internacional (Beltrán et al., 2015).

Figura 1

Cuenca del lago Titicaca



Fuente: (Stanish & Levine, 2011)

Una de las singularidades del lago es la temperatura del agua, que es (-10 a 12 °C) estas temperaturas son óptimas para la altura en la que se encuentra el lago, permitiendo la presencia de una rica flora y fauna acuática. La Bahía de Puno está situada a 15°50'34"S 69°59'43"W, cuenta con un área de 16.1 km² y una profundidad media de 2.7 m. La poca profundidad y el bajo caudal de la Bahía de Puno, combinados con el fuerte aporte del sustancioso sistema de drenaje de la ciudad de Puno (población 125663 habitantes), han resultado en una grave contaminación del cuerpo de agua, causando inconvenientes en las principales actividades en la bahía interior que son la pesca, la recolección de totora, el turismo, el transporte de embarcaciones comerciales y ganaderas (INEI, 2018).

2.2.1.1 Contaminación del lago Titicaca

Este gran cuerpo de agua recibe aguas residuales de 25 ríos, el principal de los cuales es el Río Ramis con un territorio de 14700 km², y



el Río Desaguadero con una superficie de 29843 km², además del lago Titicaca que incluye 8400 km² más de aguas efluentes. La cuenca del Titicaca tiene naturaleza endorreica, se divide en las siguientes subcuencas: La primera subcuenca se denomina “Lago Chucuito” y se ubica en la zona norte con una profundidad de 135m. La zona sur cuenta con una segunda subcuenca denominada “Lago Huíñaimarca” con una superficie de 1470 km². Con el tiempo la población del Lago Titicaca ha seguido aumentando, lo que ha tenido dos efectos, uno la demanda de agua potable y el otro el aumento en la producción de aguas residuales domésticas, que son una de las fuentes de contaminación en el lago, afectando el agua del lago. Ante esta situación, diversas autoridades nacionales e internacionales tomaron medidas para limpiar las aguas del lago (Medrano et al., 2020).

La ciudad de Puno cuenta con una laguna estabilizada para el tratamiento de aguas residuales (Tabla 1), pero la eficiencia del tratamiento es baja debido a la obstrucción de las aguas residuales, los largos tiempos de residencia en los pozos de tratamiento y las condiciones climáticas, además, las aguas residuales se descargan directamente en las acequias de drenaje de los residentes a lo largo del lago a través de tuberías secretas.



Tabla 1

Listado de plantas de tratamiento

| Empresa prestadora | Localidad | Nombre de PTAR | Año de registro | Estado de PTAR |
|---------------------------|------------------|-----------------------|------------------------|-----------------------|
| EMSAPUNO S.A. | PUNO | PTAR ALTO PUNO | 2023 | Operativo |

Nota: PTAR (Planta de tratamiento de aguas residuales)

Fuente: (DIGESA, 2023)

Tabla 2

Límites legales permitidos en aguas utilizadas para el consumo humano según normativa peruana DS N°031-2010-SA

| Físico-químico | Valores |
|-----------------------|----------------|
| Turbiedad | 5 UNTs |
| Cloro residual | 250 mgCl/L |
| Temperatura | n/a |
| Conductividad | 6.5 – 8.5 |

Nota: UNT (Unidad de turbidez nefelométrica)

Fuente: (DIGESA, 2011)

Tabla 3

Análisis microbiológico de aguas para el consumo humano

| Microbiológico | Valores |
|--|------------------------|
| Coliformes totales | 0 UFC/100ml a 35 °C |
| Coliformes fecales y <i>E. coli</i> | 0 UFC/ 100ml a 44.5 °C |

Nota: UFC (Unidades formadoras de colonias)

Fuente: (DIGESA, 2011)

Tabla 4

Análisis de metales pesados de aguas para el consumo humano

| Metales pesados | Valores (mg/l) |
|------------------------|-----------------------|
| Zn | 3.0 |
| Mn | 0.4 |
| Ni | 0.02 |
| Fe | 0.5 |
| Cu | 2.0 |

Nota: Zn (Zinc), Mn (Manganeso), Ni (Níquel), Fe (Hierro), Cu (Cobre)

Fuente: (DIGESA, 2011)

2.2.2. Microbiomas bacterianos

Los microbiomas bacterianos son comunidades complejas de microorganismos, principalmente bacterias, que coexisten en un ambiente específico, tales como agua, suelo, aire y en el humano y desempeñan roles cruciales.

Estos microbiomas bacterianos no solo influyen en la salud de sus respectivos entornos, sino que también afectan la salud humana y la producción de alimentos. Las bacterias del suelo pueden promover el crecimiento de plantas y aumentar la fertilidad en el suelo, mientras que las bacterias presentes en cuerpos de agua pueden participar en la purificación del agua, reciclaje de nutrientes y la descomposición de residuos (Beltran, 2017).

2.2.2.1 Microbiomas Ambientales

Los microbiomas ambientales se refieren a comunidades microbianas que habitan entornos no humanos, como suelos, aguas, aire y plantas, los cuales engloban la diversidad de microorganismos, como



bacterias, hongos, y virus, presentes en entornos naturales y artificiales. Contribuyen significativamente a procesos ecológicos, como la descomposición de materia orgánica, ciclos biogeoquímicos y la salud de los ecosistemas, utilizan diversas tecnologías ómicas para estudiar microbiomas ambientales, permitiendo la caracterización de comunidades microbianas y su interacción con el entorno (Belasco, 2022).

2.2.2.2. Proteobacterias

Las proteobacterias son un filo de bacterias gramnegativas, muy diverso y amplio, que se clasifica en cinco clases principales: Alfa, Beta, Gamma, Delta y Épsilon. Estas bacterias incluyen una variedad de formas y metabolismos, y muchas son importantes desde el punto de vista ecológico y médico (Weging et al., 2021).

Los nombres Alphaproteobacteria y Betaproteobacteria se derivan del sistema de clasificación que utiliza letras griegas para distinguir las diferentes clases dentro del filo Proteobacteria.

Este sistema de nomenclatura fue propuesto para facilitar la identificación y clasificación de estas bacterias basándose en sus características genéticas, fisiológicas y ecológicas. La elección de "Alpha" y "Beta" (así como "Gamma", "Delta", y "Epsilon" para otras clases) sigue un esquema alfabético sencillo que no necesariamente refleja una jerarquía o relación evolutiva específica, sino que sirve para organizar la diversidad dentro del filo de manera estructurada y comprensible (Weging et al., 2021).



Los términos alfa, beta y gamma en el contexto de aminoácidos y proteobacterias se refieren a diferentes conceptos, pero hay una relación indirecta entre ellos debido a su uso en biología y bioquímica:

Alfa-, Beta- y Gamma-Aminoácidos

Estos términos describen la posición del grupo amino en relación con el grupo carboxilo en la estructura molecular de un aminoácido.

- Alfa-aminoácidos: Tienen el grupo amino unido al carbono adyacente al grupo carboxilo. Son los componentes básicos de las proteínas.
- Beta-aminoácidos: Tienen el grupo amino en el segundo carbono a partir del carboxilo.
- Gamma-aminoácidos: Tienen el grupo amino en el tercer carbono a partir del carboxilo.

Proteobacterias

Las proteobacterias son un filo de bacterias que se clasifican en cinco clases principales: Alfa, Beta, Gamma, Delta y Epsilon. Esta clasificación no tiene relación directa con la estructura de aminoácidos, sino con las características genéticas y fisiológicas de las bacterias (Weging et al., 2021).

- Alfa-proteobacterias: Incluyen bacterias como *Rhizobium*, que fijan nitrógeno.



- Beta-proteobacterias: Incluyen bacterias como *Neisseria*, que pueden ser patógenas.
- Gamma-proteobacterias: Incluyen bacterias como *Escherichia coli*, comúnmente estudiada en microbiología.

Relación Indirecta:

La relación entre estos términos radica en su uso para clasificar y describir diferentes aspectos de la biología. Aunque se usan en contextos diferentes (estructura molecular de aminoácidos vs. clasificación de bacterias), ambos sistemas de nomenclatura ayudan a organizar y entender la diversidad biológica (Weging et al., 2021).

2.2.2.3. Microbiomas de Agua Dulce

Los microbiomas de agua dulce se refieren a las comunidades microbianas que habitan en entornos acuáticos que contienen poca o ninguna sal. Estos microbiomas son dinámicos y pueden cambiar con variaciones en las condiciones ambientales, como la temperatura y la calidad del agua, estos contribuyen a procesos ecológicos esenciales, como la descomposición de materia orgánica y la regulación de ciclos biogeoquímicos el cual albergan una gran diversidad de organismos, desde bacterias y hongos hasta plantas y animales acuáticos, contribuyendo a la biodiversidad (Belasco, 2022).

2.2.3 Metagenoma

El vocablo metagenoma hacía mención de un enfoque que pretendía analizar una recopilación de genes secuenciados de unas muestras ambientales



como si se pretenderá de un único genoma. Kevin Chen y Lior Pachter (investigadores de la Universidad de California, en Berkeley), puntualizaron la metagenómica como «la utilización de técnicas genómicas actuales para el estudio continuo de comunidades de microorganismos en su entorno natural, evitando la necesidad de aislar y cultivar cada una de las especies (Chen & Pachter, 2005).

La metagenómica es un nuevo enfoque que nos permite comprender la biología microbiana a un nivel superior (el descubrimiento, modelado y comprensión a nivel molecular de la gestión de relaciones dinámicas entre moléculas que definen las comunidades microbianas en la biosfera). Se centra en el genoma de un organismo, centrándose en los genes de la comunidad y en cómo estos genes influyen en determinadas actividades metabólicas o funciones colectivas del ecosistema. El concepto de clonación de ADN directamente del medio ambiente fue propuesto e implementado por primera vez por Pace y colaboradores en 1985 y Schmidt y colaboradores 1991, quienes construyeron bibliotecas ambientales de ADNr 16S a partir de muestras de plancton marino utilizando fagos lambda como vectores de clonación (Alraddadi et al., 2013).

Años más tarde, Healy y colaboradores en 1995 lograron secuenciar genes que codifican enzimas celulolíticas a partir de una biblioteca metagenómica de digestores anaeróbicos. En 1996 se obtuvo la primera biblioteca ambiental de agua de mar, que permitió el análisis de la organización y composición del genoma de arqueas marinas no cultivables. El gran potencial de la rodopsina, un sistema de concentración de luz que se cree que es exclusivo de las arqueas, fue reconocido en 2000 cuando se identificó su homólogo bacteriano en una biblioteca metagenómica marina. El primer gran proyecto de metagenómica, denominado GOS (Global Ocean Sampling Expedition), incluyó la exploración de genomas

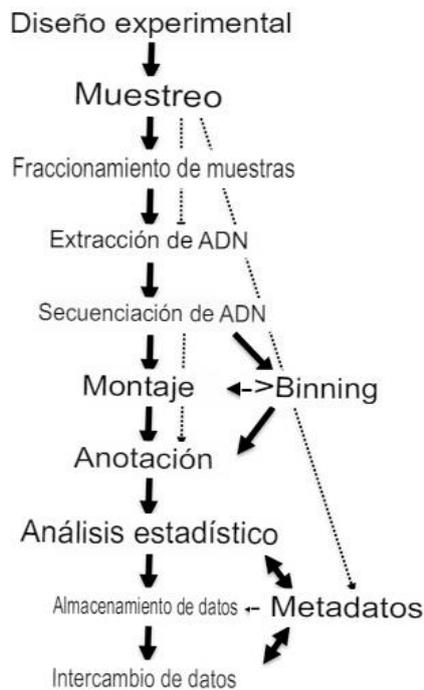


oceánicos y tenía como objetivo evaluar la diversidad genética de las comunidades microbianas marinas para comprender su papel en procesos fundamentales de la naturaleza (Chen & Pachter, 2005).

Los avances recientes en la investigación metagenómica, en particular el análisis de secuenciación de ADN basado en amplicones de genes marcadores taxonómicos, han llevado a la caracterización de comunidades microbianas en diversos ecosistemas. Aunque inicialmente limitado por la falta de herramientas de análisis adecuadas y profundidad de secuenciación, se han desarrollado nuevas técnicas computacionales y estadísticas para manejar conjuntos de datos de secuenciación de próxima generación de alta dimensión. Ambas herramientas permiten un análisis sólido de la estructura de la microbiota y los patrones de distribución, que son fundamentales para comprender la ecología y la biogeografía microbianas. Además, un consorcio de varios laboratorios que trabajan en proyectos de investigación como el Proyecto de Microbioma Humano y Terrestre ha desarrollado protocolos estandarizados para la extracción de ADN, procedimientos de secuenciación y bioinformática. Estos métodos proporcionan un gran depósito de datos disponibles públicamente que pueden servir como referencia para estudios basados en hipótesis en curso y futuros para caracterizar mejor las funciones de las comunidades microbianas en diversos ecosistemas (Staley & Sadowsky, 2018).

Figura 2

Diagrama de flujo de un proyecto metagenómico típico



Fuente: (Gao et al., 2021)

2.2.3.1 Secuenciación metagenómica de alto rendimiento

En una secuenciación de alto rendimiento se verificó que los primeros estudios usaron de manera equidistante tecnologías para el muestreo ambiental como lo son Ion Torrent Personal Genome Machine, Illumina MiSeq o HiSeq y Applied Biosystems SOLiD. Rodrigue et al., (2010), estas técnicas que son usadas para la secuenciación generan fragmentos más cortos que normalmente serían de ~400 pares de bases. Gracias a las nuevas tecnologías y el crecimiento de investigación en base a secuenciación de Metagenomas bacterianos estas técnicas evolucionaron y ahora la metagenómica de Shotgun son una de los secuenciadores más actuales y precisos entre estas son MetaPhlan 3, MetaPhlan 2, mOTUs2,

y Kraken, Bracken son los 5 HMP Metagenomas (Beghini et al., 2021).

Una ventaja adicional de la secuenciación de alto rendimiento es que esta técnica no requiere clonar el ADN antes de secuenciar, evitando así uno de los procesos más largos y complicados del muestreo ambiental (Rubio et al., 2020).

Tabla 5

Comparación de tiempo de ejecución y memoria "Consumo de MetaPhlAn 3, MetaPhlAn2, mOTUs2 y Kraken+Bracken en los 5 metagenomas de HMP "

| Herramienta | Tiempo trascorrido (h) | Tiempo trascorrido (sd/h) | Pico de memoria (media GB) | Pico de memoria (sd/GB) | Lectura por segundo (media) |
|------------------------------|---------------------------------------|--|---|--|--|
| MetaPhlAn v3.0 | 3.1120 | 2.65 | 2.614 | 0.01 | 10031.2 |
| MetaPhlAn2 v2.7 | 8.2183 | 3.58 | 2.079 | 0.27 | 2911.2 |
| mOTUs251_precisión | 10.6094 | 4.92 | 4.034 | 1.30 | 2283 |
| mOTUs251_recall | 11.5189 | 6.18 | 4.035 | 1.30 | 2186 |
| Bracken_208_25_refseq | 2.3504 | 1.40 | 32.529 | 0.27 | 11163 |

Nota: H (hora), GB (Gigabyte), sd (tarjeta digital segura).

Fuente: (Beghini et al., 2021).

2.2.3.2 Extracción de ADN metagenómico

Los protocolos de extracción de DNA genómico son cruciales para un análisis confiable posterior y, por lo tanto, la necesidad de aislamiento y extracción de DNA genómico de alta calidad se considera un paso limitante en la construcción de bibliotecas metagenómicas (Forero et al., 2021).



Analizar toda la diversidad microbiana presentes en la muestra original. En este sentido, diversos estudios han desarrollado y reportado la implementación de protocolos químicos para la extracción de DNA genómico a partir de diferentes tipos de muestras (Bag et al., 2016). Sin embargo, se sigue buscando optimizar aún más el proceso de extracción para garantizar que el DNA extraído tiene buena concentración y pureza (relación proteína-ácido nucleico), que es la base principal para los procesos de secuenciación y análisis bioinformáticos. Aunque utilizar buenos protocolos de extracción de DNA es importante para los estudios metagenómicos, se han realizado pocos estudios en muestras de origen ambiental, por tanto, el éxito de la investigación metodológica metagenómica en muestras ambientales depende principalmente de la obtención de DNA metagenómico de alta calidad, para lo cual es necesario contar con el protocolo más accesible, rentable y con resultados de gran calidad (Forero et al., 2021).

2.2.3.3 Cuantificación y análisis de la integridad de DNA

Sabemos que la obtención de DNA a partir de una muestra ya sea humana, celular, bacteriana, tejido entre otra, es útil para analizar la expresión genética, diagnosticar enfermedades, analizar mutaciones o estudiar tratamientos en el laboratorio. También sabemos que existen varios métodos de extracción de ácidos nucleicos que nos permiten descomponer las células y obtener su material genético, en primer lugar, debemos comprobar su integridad, es decir, si la calidad tras la extracción es suficiente para el análisis, de lo contrario, si no tenemos cuidado, lo destruiremos. Además, en la mayoría de las técnicas que utilizan ácidos

nucleicos como materiales de partida, es necesario conocer la concentración en la que se encuentran para poder trabajar en determinadas cantidades (Iranzo, 2022).

2.2.3.4 Espectrofotometría

La espectrometría es un mecanismo de alta precisión preparados para calcular, evaluar y analizar la luz en labor de su longitud de onda. Estas sofisticadas herramientas proporcionan calcular y determinar con exactitud los atributos de la luz, como la intensidad, el color y la transmitancia. Ya sea para el campo de la investigación (Diaz, 2023).

Tabla 6

Valores indicativos de pureza y muestras de ADN previamente filtradas

| Técnica de análisis | Análisis | Parámetros | Criterios de validez |
|---|---|------------|---|
| Espectrofotometría | A260/280 | Pureza | $\geq 1.8 - 2.1$ Pureza optima |
| | | | $\geq 1.6 - 1.7$ Pureza aceptable |
| | < 1.6 ADN contaminado con compuestos aromáticos | | |
| | > 2.1 ADN contaminado con ARN | | |
| Espectrofotometría | A260/230 | Pureza | $> 2 - 2.2$ Pureza optima |
| | | | > 1.8 Pureza aceptable |
| | | | < 1.8 ADN contaminado con sales, fenol, hidratos de carbono, entre otros. |
| < 1.5 ADN altamente contaminado con sales, fenol, hidratos de carbono | | | |

Nota: ADN (Ácido desoxirribonucleico)

Fuente: (Iranzo, 2022).



2.2.3.5 Fluorimetría

Para abordar las limitaciones de la espectrofotometría, surgieron los métodos fluorométricos. En este caso se utilizan compuestos fluorescentes que se unen específicamente a las cadenas de ácidos nucleicos. Es fácil entender que cuando mayor es la cantidad de DNA, más compuestos fluorescentes se unen y mayor es la fluorescencia emitida y detectada por el dispositivo fluorómetro (Lee et al., 2012).

2.2.3.6 Electroforesis

La electroforesis es una técnica que separa moléculas en su función de su movilidad en un campo eléctrico, la separación se puede realizar sobre la superficie hidratada de un soporte sólido, a través de una matriz porosa o en solución. Para ello, deben migrar a través de los poros del gel de agarosa, esta migración variara dependiendo del tamaño del ácido nucleico, cuando más grande es la molécula, más difícil es atravesar el poro, cuando más pequeño sea el ácido nucleico, más rápido migrara hacia el electrodo positivo (Lee et al., 2012).

Tabla 7

Criterios de validez y puntuación asignada según la integridad de ADN observada en gel de agarosa.

| Criterios validez | Puntuación |
|---|-------------------|
| Banda definida en la parte superior del gel. Integridad alta. | 3 |
| Presencia simultanea de la banda en la parte superior del gel y un ligero smear. Integridad adecuada. | 2 |
| Ausencia de banda definida u presencia de smear concentrado en la parte superior del gel. Parcialmente degradado. | 1 |
| Smear concentrado en la parte inferior del gel. Totalmente degradado. | 0 |

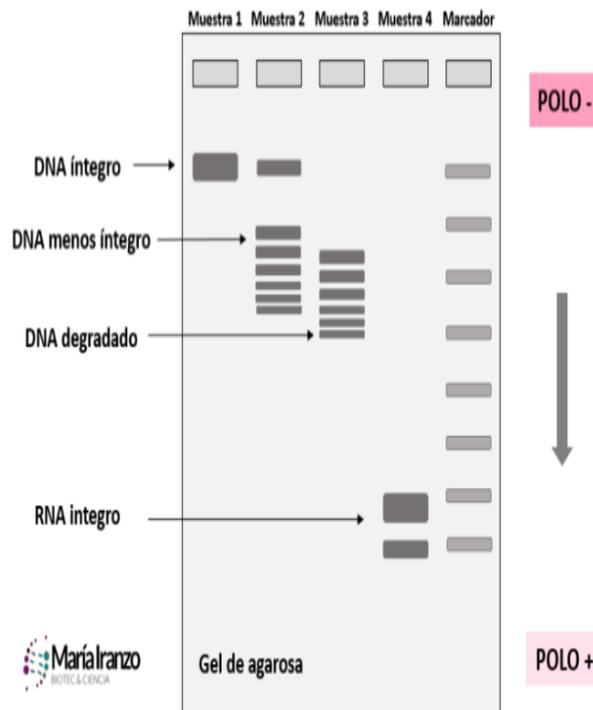
Nota: Smear (método de separación utilizado para analizar fragmentos de ADN)

Fuente: (Lee et al., 2012)

En este caso, dado que la electroforesis es un método cualitativo, no es posible asignar un valor real a la integridad de la muestra. Sin embargo, una muestra de DNA se considera intacta cuando su contorno en la electroforesis en gel de agarosa corresponde a bandas discretas. El nivel de degradación de la muestra estuvo determinado por la pérdida de definición de la banda principal y las pistas que la acompañan a lo largo del gel (DIGESA, 2020).

Figura 3

Representación del gel con lectura de barrido electroforético



Fuente: (Iranzo, 2022)

2.2.4 Secuenciación Shotgun

La secuenciación shotgun es una técnica de secuenciación de ADN que se utiliza para obtener información genética detallada.

Desde años atrás y hasta la actualidad ha sido complicado de cultivar los microbios en el laboratorio, lo que puede complicar la catalogación de los miembros y su función en el microbioma, la secuenciación de DNA ambiental ha revelado una amplia biodiversidad microbiana y ha dilucidado en la relación entre las comunidades microbianas asociadas al huésped y los fenotipos del huésped. La secuenciación de DNA metagenómico de Shotgun es un método de secuenciación ambiental relativamente nuevo y poderoso que proporciona información sobre la biodiversidad y la función de la comunidad para ello se necesitan criterios computacionales complejos para los perfiles metagenómicos



cumpliendo con ensamblaje y mapeo (Sharpton, 2014).

Sin embargo, el análisis de secuenciación metagenómica también es complejo debido a estructuras de datos complejas. Afortunadamente se han desarrollado nuevas herramientas y recursos de datos para sortear estas complejidades y permitir que los investigadores determinen que microbios están presentes en una comunidad y que podrían estar haciendo. Esta revisión describe estrategias y herramientas de análisis específicas que se pueden aplicar a los datos metagenómicos, junto con advertencias y advertencias asociadas a su uso (Quince et al., 2017).

La metagenómica de Shotgun especifica como analizar metagenomas para cuantificar la estructura y diversidad en una muestra, ensamblar nuevos genomas, identificar nuevos taxones y genes e identificar vías metabólicas codificadas en comunidades. También analiza varios métodos que se pueden usar para comparar la metagenómica para identificar taxones y funciones que distinguen las comunidades (Sharpton, 2014).

Los métodos de secuenciación basados en ampliaciones generalmente se dirigen a un solo gen, la secuenciación metagenómica de Shotgun permite la secuenciación aleatoria de todo el metagenoma de una muestra sin cebadores específicos, lo que mitiga el sesgo de selección de cebadores (Quince, et al., 2017). En comparación con el análisis comunitario basado en genes marcadores (Gen 16S), la secuenciación metagenómica de Shotgun agrega una capa de detalle a la caracterización taxonómica de la comunidad al proporcionar información sobre la composición genética y las capacidades funcionales del microbioma de la muestra, aunque es más costosa y consume más tiempo que los marcadores. El



estudio del metagenoma de Shogun consta de los siguientes pasos (Gao et al., 2021).

- La recolección, procesamiento y secuenciación de las muestras.
- Procesamiento previo de las lecturas de secuenciación.
- Análisis de secuencias para perfilar características taxonómicas, funciones u genómicas del microbioma.
- Análisis de posprocesamiento estadístico y biológico.
- Validación.

La ampliación de genes, debido a su capacidad para examinar organismos de todos los dominios de la vida, la secuenciación metagenómica de Shotgun sigue siendo el método más eficiente y completo para obtener datos estructurales y funcionales. La composición genética también puede utilizar para formular rutas funcionales putativas (Gao et al., 2021).

2.2.4.1 El ARN ribosomal 16S (ARNr 16S) y la secuenciación Shotgun

El ARNr 16S y la secuenciación shotgun son dos técnicas importantes en el campo de la genómica y la microbiología, ambas técnicas tienen aplicaciones importantes en la investigación científica y médica, como en la caracterización de microbiomas, estudio de la evolución y la ecología microbiana.

El ARN ribosomal 16S (ARNr 16S) se utiliza ampliamente como cronómetro molecular y patrón universal en la identificación y



clasificación de bacterias debido a su conservación y variabilidad en ciertas regiones (Rubio et al., 2020).

La secuenciación de ARN 16S es excelente para obtener perfiles taxonómicos a nivel de género y especie en comunidades microbianas, ya que se enfoca específicamente en una región conservada del genoma que permite identificar y clasificar bacterias. Además, es más económica y requiere menos secuenciación. (Rubio et al., 2020).

Por otro lado, la secuenciación shotgun proporciona una visión más completa de la microbiota al secuenciar todo el ADN presente en una muestra, lo que incluye no solo bacterias, sino también otros microorganismos, así como el genoma del huésped en muestras biológicas. Esto ofrece una resolución taxonómica más detallada y una mayor capacidad para identificar especies raras o desconocidas. Sin embargo, la secuenciación shotgun es más costosa y requiere más poder computacional y recursos para el análisis de datos.

- **Objetivo de la Secuenciación Shotgun**

La secuenciación Shotgun está diseñada para secuenciar todo el genoma de un organismo, no solo un fragmento específico como el ARNr 16S. Esta técnica fragmenta el ADN en pequeñas piezas que se secuencian de forma aleatoria y se ensamblan posteriormente para reconstruir el genoma completo (Rubio et al., 2020)

- **Complejidad del Genoma**

La secuenciación Shotgun aborda la secuenciación de todo el



genoma, lo cual es crucial para estudiar la genética completa del organismo, identificar todos los genes y comprender su función y organización. El ARNr 16S no proporciona esta información completa del genoma (Rubio et al., 2020).

- **Diversidad de Datos**

Mientras que el ARNr 16S es excelente para identificar y clasificar bacterias, la secuenciación Shotgun permite obtener datos de todas las partes del genoma, incluidos genes de virulencia, resistencia a antibióticos, y otros elementos móviles que no estarían presentes en la secuenciación dirigida al ARNr 16S (Rubio et al., 2020).

Por estas razones, aunque el ARNr 16S es una herramienta valiosa en la microbiología, la secuenciación Shotgun se utiliza cuando se requiere una visión más completa del genoma (Rubio et al., 2020).

2.2.5 Perfil Taxonómico

El perfil taxonómico es un análisis crucial en taxonomía, especialmente en metagenómica. Se utiliza para identificar y cuantificar la abundancia de taxones en una muestra biológica. Diversas herramientas, como kASA, QIIME, Centrifuge, MetaPhlAn, y metagenomeSeq, facilitan este proceso mediante pruebas estadísticas y visualizaciones exploratorias. El perfil taxonómico es esencial para comprender la diversidad biológica en muestras complejas, siendo fundamental en la clasificación y estudio de organismos a nivel molecular (Weging et al., 2021).

La taxonomía ha ampliado la teoría y la práctica de la clasificación



biológica y ha apoyado y nutrido muchas otras ramas de la biología. La práctica de la disciplina en su forma moderna se remonta a mediados del siglo XVIII, cuando el botánico sueco Carlos Linneo (1753, 1758) introdujo el concepto de sistema, que se refiere al ordenamiento de los seres vivos. La práctica de la taxonomía incluye la identificación, denominación y descripción de especies (microtaxonomía) y la clasificación de grupos más grandes (macrotaxonomía). Hoy, la disciplina enfrenta el desafío de integrar nuevos métodos. Los métodos y técnicas para obtener datos categóricos, compararlos e interpretarlos están cambiando drásticamente. Por lo tanto, la identificación, caracterización, clasificación y análisis de especies y sus relaciones filogenéticas se apoyan y mejoran mediante el uso de tecnologías digitales e infraestructura de red (Fernández et al., 2013).

La sistemática filogenética actual consta de un "núcleo" de conceptos básicos, así como de varios métodos basados en diferentes criterios de optimización (p. ej., parsimonia, máxima verosimilitud y métodos bayesianos) que permiten el análisis de datos moleculares. Las hipótesis filogenéticas resultantes modifican clasificaciones anteriores, sugiriendo en ocasiones relaciones que tradicionalmente no han sido consideradas y, a su vez, permiten analizar diferentes procesos y patrones evolutivos. La formación en la comprensión y el uso correcto de estos métodos no está exenta de dificultades (Mata et al., 2017).

2.2.6 Herramientas computacionales para el análisis metagenómico

2.2.6.1 Lenguaje de programación Python

Python puede leer y escribir varios formatos de secuencias de genes



comunes, incluidos FASTA, FASTQ, GenBank, Clustal, PHYLIP y NEXUS. Al leer un archivo, los miembros de las clases de Biopython, como SeqRecord, se pueden completar utilizando la información de descripción del archivo. Esto permite convertir archivos de un formato a otro.

Los archivos de flujo grandes pueden exceder los límites de los recursos de memoria de su computadora, por lo que Python ofrece varias opciones para acceder a registros de archivos grandes. Estos se pueden cargar directamente en la memoria desde las estructuras de datos de Python, como listas o diccionarios, lo que brinda un acceso rápido a ellos, pero a un costo de memoria más alto. Por otro lado, el archivo se puede leer desde el disco duro si es necesario con un rendimiento más lento, pero con menos requisitos de memoria (Chang et al., 2023).

Con el módulo de Bio.Entrez, los usuarios de Biopython pueden descargar datos biológicos de las bases de datos del NCBI. Todas las funciones proporcionadas por el motor de búsqueda Entrez se pueden realizar a través de las funciones de este módulo, incluida la búsqueda y descarga de registros.

El módulo Bio.Phylo proporciona herramientas para ver y manipular árboles filogenéticos. Se pueden leer y escribir varios topos de formatos, incluidos los formatos Newick, NEXUS y PhyloXML. Con objetos Tree y Clade, se pueden realizar operaciones transversales comunes. Por ejemplo, pueden transformar y fusionar archivos de árbol, extraer subconjuntos de árboles, cambiar la raíz de un árbol y analizar las

características de las ramas, como la longitud o la puntuación de las ramas (Chang et al., 2023).

2.2.6.2 Ensamblaje metagenómico por MetaPhlAn

MetaPhlAn integra información de ensamblajes de metagenomas y genomas de aislados microbianos para obtener perfiles taxonómicos más íntegros.

MetaPhAn utiliza un conjunto de aproximadamente 1 millón de marcadores para la asignación taxonómica, por el que no puede mapear todas las lecturas. El conjunto de marcadores se basa en secuencias específicas de clado, donde los clados pueden ser específicos como tan amplio como una especie o como un filo, y la secuencia del marcador debe estar muy conservada dentro del clado. Localmente diferente a las secuencias fuera del clado. La asignación de la clase se logra asignando todas las lecturas a un grupo usando los marcadores de Bowtie (Lindgreen et al., 2016).

MetaPhlAn es una herramienta computacional para analizar la composición de comunidades microbianas (bacterias, arqueas y eucariotas) a partir de datos metagenómicos (es decir, no 16S) de secuenciación de Shotgun a nivel de especie. Con StrainPhlAn, es posible un análisis microbiano preciso a nivel de cepa (Quince et al., 2017).

También incluye toda la base de datos de secuenciación de Shotgun metagenómica para la generación de perfiles de metagenoma (los formatos de entrada incluyen: fasta, fast.tar.bz2, ... entre otros) (Pasolli et al., 2016).



2.2.6.3 Mejoramiento en el perfil taxonómico metagenómico con especies no caracterizadas con MetaPhlAn4

MetaPhlAn4 se basa en ~5.1 millones de genes marcadores específicos de clado únicos identificados a partir de ~1 millón de genomas microbianos (~23660 de referencia y ~771500 genomas ensamblados en metagenoma), que abarcan 26970 contenedores genómicos a nivel de (SGB), de los cuales 4992 no eran taxonómicamente identificadas a nivel de especie, lo que permite (Blanco et al., 2023).

- Asignaciones taxonómicas inequívocas.
- Una estimación precisa de la abundancia relativa de organismos.
- Resolución de nivel de SGB para bacterias, arqueas y eucariotas,
- Identificación y seguimiento de cepas.

MetaPhlAn amplía y mejora las capacidades existentes para el análisis taxonómico metagenómico aprovechando un marco que integra grandes componentes metagenómicos, como genomas de referencia bacterianos y arqueales. Luego se preprocesan conjuntamente para permitir un mapeo metagenómico eficiente basado en millones de genes marcadores únicos, y en última instancia cuantifican organismos aislados y ensamblados metagenómicamente en nuevas comunidades. El algoritmo adicional utilizado por versiones antiguas de la siguiente manera (Blanco et al., 2023).

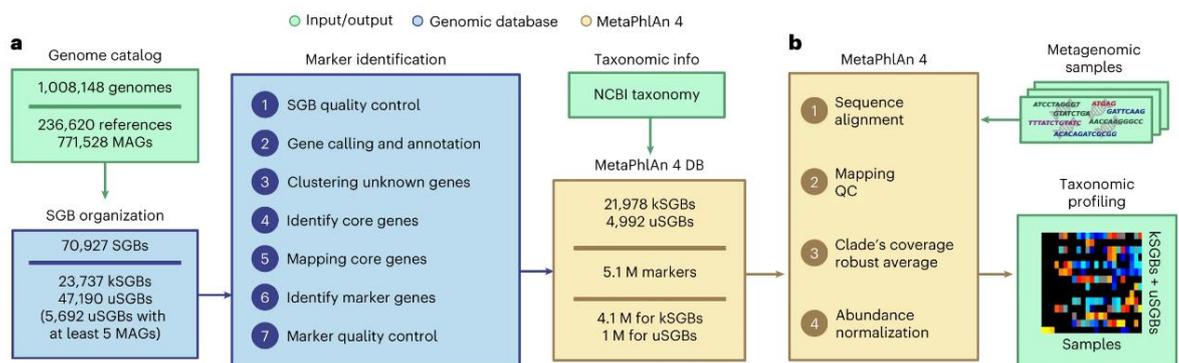
- La adopción de SGB (contenedor de genoma a nivel de especie) como unidades taxonómicas primarias, cada una de las cuales

agrupa genomas microbianos y MAG (genomas ensamblados en metagenoma) en especies existentes consistentes y grupos de genomas recién definidos de diversidad a nivel de especie aproximadamente.

- La integración de más de 1 millón de MAG y genomas en esta estructura SGB para construir una de las bases de datos más grandes de secuencias de referencia microbiomas confiables actualmente disponibles.
- La curación de unidades taxonómicas microbianas basadas en la consistencia de genomas microbianos etiquetados taxonómicamente y la asignación de nuevas etiquetas taxonómicas a SGB definidas únicamente en MAG (O’Leary et al., 2016).

Figura 4

MetaPhlAn4 integra secuencias de referencia de genomas asilados y ensamblados en Metagenomas para el perfil taxonómico del metagenoma.



Fuente: (Blanco et al., 2023)

El procedimiento mejorado para extraer genes marcadores únicos de cada SGB para el MetaPhlAn es una estrategia de mapeo en referencias NCBI, en conclusión, MetaPhlAn 4.



- Mejora la sensibilidad y especificidad del perfil taxonómico del metagenoma, amplía la diversidad microbiana observable, principalmente mediante la cuantificación de especies aun por caracterizar (uSGB).
- Permiten un perfil metagenómico preciso de microbiomas que contienen pocos taxones aislados cultivados.
- Reconstruye con precisión filogenias a nivel de cepa a gran escala de especies microbianas no caracterizadas.

2.2.6.4 Contenedores genómicos a nivel de especie (SGB)

SGB representa especies existentes (conocidas o kSGB) y especies aun no descubiertas la caracterización (desconocida, uSGB) solo se define según MAG. A partir de una colocación de 1.01 millones de MAG bacterianos y arqueas, genomas integran el ultimo catálogo de genomas y otros MAG recién ensamblados que abarcan múltiples entornos, primero ampliamos la definición a 54596 GBS y luego definimos kGBS y 4992 Ugb únicos específicos de GBS para genes marcadores (es decir, genes que caracterizan de manera única cada GBS). El conjunto de datos resultante amplía el algoritmo MetaPhlan existente para permitir el análisis taxonomico cuantitativo más profundo y preciso del microbioma ambiental, humana y relacionada con el huésped y proporciona ingormacion sobre muchos aspectos ya sean estudios que vinculan el microbioma con las condiciones del huésped (Blanco et al., 2023).

2.2.6.5 Ensamblaje metagenómico GraphlAn

GraPhlAn es una herramienta de software para producir representaciones circulares de alta calidad de árboles taxonómicos y filogenéticos. GraPhlAn se centra en representaciones concisas, integradoras, informáticas y listas para publicar de la investigación impulsada filogenética y taxonómicamente.(Blanco et al., 2023).

También puede visualizar abundancias microbianas en un árbol de la vida (también conocido como filogenético o cladograma) para capturar sus relaciones taxonómicas (o filogenias). Se usa una herramienta llamada GraPhlAn que puede representar árboles y anotarlos con datos como nombres microbianos o abundancias. Las instrucciones asumen que GraPhlAn desde la línea de comandos, como también la sección GraPhlAn en Galaxy (Gao et al., 2021).

2.2.6.6 Cladograma de GraphlAn

La herramienta GraPhlAn, desarrollada por Huttenhower et al. (2012), se erige como un valioso recurso en la representación visual de árboles taxonómicos y filogenéticos mediante representaciones circulares de alta calidad. Centrándose en la concisión y claridad, GraPhlAn se presenta como una solución efectiva para la visualización gráfica de relaciones evolutivas y taxonómicas.

GraPhlAn ha sido diseñado para producir representaciones circulares que destacan la estructura y composición de árboles filogenéticos y taxonómicos. Su enfoque se concentra en la generación de visualizaciones de calidad, facilitando la interpretación de complejas



relaciones entre entidades biológicas. La herramienta permite la inclusión de información detallada de manera intuitiva, lo que resulta crucial para la comprensión visual de datos biológicos.

Para utilizar GraPhlAn, se proporciona flexibilidad en la entrada de datos, permitiendo representaciones visuales personalizadas y adaptadas a las necesidades específicas de los usuarios. Este enfoque versátil posiciona a GraPhlAn como una herramienta valiosa en la exploración y comunicación efectiva de estructuras taxonómicas y filogenéticas en contextos científicos y de investigación biológica (Huttenhower et al., 2012)

Un cladograma es una ramificación de un linaje y esta se representa como una nueva rama donde se identifica si hay un antepasado común para los organismos, un cladograma visualiza las abundancias microbianas en un árbol de la vida (también conocido como filogenia) que captura su relación taxonómica (filogenética) (Du et al., 2022; Gao et al., 2021)

2.2.6.7 Mapas de calor por hclust2

La herramienta Hclust2 desempeña un papel crucial en la generación de mapas de calor en el ámbito de la investigación microbiómica. Este software, desarrollado por el equipo SegataLab, se presenta como una herramienta invaluable para la visualización efectiva de datos taxonómicos obtenidos a través de técnicas como MetaPhlAn (Beghini et al., 2021).

Hclust2 se erige como una solución práctica y eficiente, diseñada específicamente para la creación de mapas de calor con numerosas



opciones que garantizan la producción de figuras de alta calidad, adecuadas para su inclusión en publicaciones científicas. Entre sus características destacadas se encuentran la capacidad de realizar agrupaciones jerárquicas utilizando el algoritmo hclust, lo que permite una representación visual clara de las relaciones entre las muestras y las taxonomías identificadas.

Este software ofrece una interfaz amigable y opciones flexibles, facilitando a los investigadores la personalización de sus representaciones gráficas. Al utilizar Hclust2 en combinación con resultados de MetaPhlAn, se logra una integración sinérgica que potencia la interpretación de datos microbiómicos complejos (Blanco et al., 2023)

Dependiendo del color que nos muestra el mapa de calor producido por Hclust2 podemos determinar la cantidad de bacterias que se encuentran en una escala de colores. Y nos hace la pregunta de qué microbios son más abundantes en cada lugar del cuerpo entre estos datos (Beghini et al., 2021).



CAPITULO III

MATERIALES Y METODOS

3.1 TIPO DE ESTUDIO

El estudio es de tipo observacional y descriptivo, porque se aplicó análisis de secuenciación, herramientas moleculares y de bioinformática en la determinación de perfiles taxonómicos del microbioma bacteriano presente en el agua del lago Titicaca describiéndose las relaciones y abundancias obtenidas por Metaphlan4.

3.2 LUGAR DE ESTUDIO

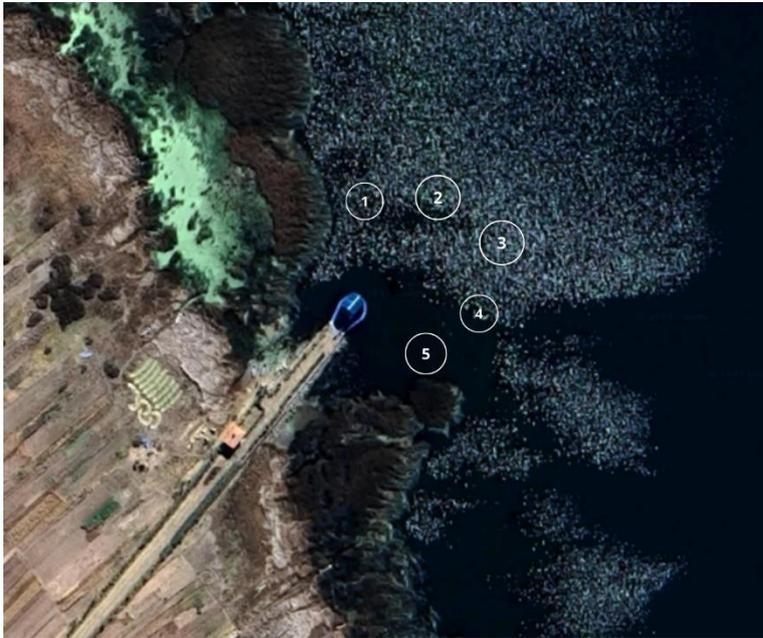
La investigación se realizó en el Laboratorio de Biología Molecular de las instalaciones del Mega laboratorio de la Universidad Nacional del Altiplano Puno, ubicada en la Avenida Floral 1153 de la ciudad de Puno (Perú) ubicado en las coordenadas -15.824263501552979 y -70.01626235614603 UTM, con la colaboración de BTS consultores para la secuenciación metagenómica Shotgun.

3.3 PUNTO DE MUESTREO

El punto de muestreo 1 para la Captación Chimú, comprendió la siguiente ubicación UTM -15.852895, -69.963400, lugar donde ocurre el procesamiento de aguas servidas, además de la potabilización del agua para la ciudad de Puno casi en su totalidad. (Figura 5 y Figura 6)

Figura 5

Puntos de muestreo en la captación Chimú.



Nota: 1,2,3,4,5=puntos de muestreo

Fuente: Imagen de Googlemaps

El punto de muestreo 2, comprendió la zona de la Bahía Interna de Puno (Perú) ubicada en las coordenadas UTM -15.834578809122428 y -70.01364981931286, la cual se caracteriza porque presenta grandes cantidades de totorales que impiden el correcto circulamiento del agua llegando a provocar el fácil acumulamiento de contaminantes y sedimentos residuales que proceden de la ciudad (ANA, 2016).

Figura 6

Puntos de muestreo de la Bahía Interna



Nota: 1,2,3,4,5=puntos de muestreo

Fuente: Imagen de Googlemaps

Tabla 8

Puntos de muestreo para la evaluación de análisis metagenómico

| Zona | Punto de muestreo | Georreferencia UTM | Tipo de muestra | Cantidad (ml) |
|-----------------|-------------------|------------------------|-----------------|---------------|
| Captación Chimú | Punto 1 | -15.848927, -70.003949 | Agua | 1000 |
| | Punto 2 | -15.848901, -70.004341 | Agua | 1000 |
| | Punto 3 | -15.848924, -70.004695 | Agua | 1000 |
| | Punto 4 | -15.849027, -70.004939 | Agua | 1000 |
| | Punto 5 | -15.849180, -70.005145 | Agua | 1000 |
| Bahía Interna | Punto 1 | -15.833426, -70.012395 | Agua | 1000 |
| | Punto 2 | -15.834273, -70.011708 | Agua | 1000 |
| | Punto 3 | -15.836058, -70.011086 | Agua | 1000 |
| | Punto 4 | -15.836987, -70.011215 | Agua | 1000 |
| | Punto 5 | -15.837524, -70.012266 | Agua | 1000 |

Nota: UTM (Mercator Transversal Universal), ml (mililitros)

Fuente: Datos obtenidos con el sistema de coordenadas geográficas (UTM)



3.4 RECOLECCIÓN Y PREPARACIÓN DE MUESTRAS

3.4.1 Recolección de muestra

A. Técnica

Muestreo directo

B. Fundamento

La muestra debe ser representativa del agua en ese sitio y momento, capturando características físicas, químicas y biológicas para una evaluación precisa donde se busca preservar la integridad de la muestra desde el momento de la recolección hasta su análisis, evitando contaminaciones o cambios en las propiedades del agua, y lo más importante que las propiedades del agua a analizar no cambien durante el proceso de extracción. (ANA, 2016).

C. Procedimiento

Se tomó muestras de agua superficial, la cantidad de 1L a una profundidad de 15-20 cm de la superficie y a una distancia de 50 m entre cada uno de los 5 puntos de muestreo alrededor de la “captación Chimú” (CH) y de la “Bahía Interna” (BI), las cuales luego se unieron, homogeneizaron para su posterior procesamiento, siguiendo las directrices del Protocolo Nacional para el Monitoreo de la Calidad de los Recursos Hídricos Superficiales (ANA, 2016).

3.4.2 Filtración de las muestras para la concentración de microorganismos

A. Técnica

Filtración de muestras ambientales (Agua)



B. Fundamento

La filtración por membrana implica el paso de una muestra líquida a través de una membrana microporosa. Los microorganismos quedan retenidos en la superficie de la membrana, permitiendo su posterior análisis. La selección cuidadosa del tamaño de poro de la membrana es crucial. Permite retener microorganismos de interés, como bacterias y hongos, mientras permite el paso de partículas más pequeñas y el agua, los microorganismos de tamaño superior al de los poros de la membrana quedan atrapados en su superficie, asegurando una concentración efectiva para su detección. La filtración es un método estéril y selectivo que permite la recuperación de microorganismos vivos, esencial para estudios microbiológicos más profundos (Forero et al., 2021).

C. Procedimiento

Una vez recolectadas las muestras de agua de la Captación Chimú como de la Bahía Interior se sometieron a un proceso de filtración al vacío para lo cual se vertieron 100 ml de muestra en un embudo en cuyo interior se colocó papel filtro estéril de 2 dobleces y se activó la bomba de vacío (Marca Pall MicroFunnel – Thermo Fisher Scientific) e inicio la filtración (Anexo 5). Cuando el filtro se observó visiblemente seco, se procedió a retirar el embudo desenroscando la base con precaución para evitar el contacto con la membrana de 0.2 μm de grosor (Marca Whatman Nucleopore – Sigma Aldrich). Posteriormente, el papel filtro del embudo se colocó en un tubo (1.5 ml) precargado con perlas de vidrio, y se almacenó en el congelador a $-80\text{ }^{\circ}\text{C}$ hasta el momento de la extracción de ADN (Forero et al., 2021).



3.5 EXTRACCIÓN DE ADN BACTERIANO

Con la finalidad de tener análisis metagenómico de calidad, se empleó el protocolo de extracción de ADN del kit genomic ADN mini kit Purelink, siguiendo las recomendaciones del fabricante.

A. Técnica

Extracción de ADN bacteriano

B. Fundamento

La extracción de ADN bacteriano es un paso crucial en la investigación microbiológica desde la ruptura de las interacciones proteína-ADN que se logra mediante agentes desnaturalizantes como el SDS (dodecilsulfato de sodio) y la proteína quinasa K. Estos agentes eliminan las proteínas asociadas al ADN, liberando el material genético bacteriano, la acción de enzimas líticas como la lisozima y la proteasa facilita la degradación de la pared celular bacteriana, permitiendo el acceso al ADN intracelular y el uso de soluciones salinas y alcoholes permite la precipitación del ADN, separándolo de otros componentes celulares. El posterior lavado asegura la eliminación de contaminantes (Forero et al., 2021).

C. Procedimiento

Para la extracción de ADN bacteriano se utilizó el Kit PureLink™® Microbiome DNA Purification (Anexo 6) para lo cual se siguieron los siguientes pasos según las recomendaciones del fabricante (ISC-III, 2011):



a. Preparación del Lisado de la Muestra

La muestra de agua, previamente filtrada, correspondiente al sedimento microbiano se resuspendió con 100µl de buffer de lisis, seguido de un pipeteo continuo para homogeneizar. Luego, la mezcla se transfirió al Bead Tube del kit al que se añadió 100µl de Lysis Enhancer, que luego se tapó firmemente, y se agitó con el vortex para obtener una solución homogénea, incubándose a 65°C por 10 minutos.

b. Adsorción del ADN

A continuación, se agregó 900 µl de Binding Buffer y se agitó brevemente con el vortex. Después, se cargaron 700 µl de la suspensión en un tubo spin column del kit y se centrifugó a 14000 rpm durante 1 minuto, luego del cual se desechó el flujo, y se repitió la centrifugación en seco.

c. Lavado de ADN

El SpinColumn se trasladó a un tubo de recolección limpio y se añadió 500µl de Wash Buffer, agitándose brevemente con el vortex, y de ahí se centrifugó durante 1 minuto a 14000rpm. Después de desechar el flujo, se realizó otra centrifugación del conjunto spin column – tubo a 14000rpm durante 30 segundos para optimizar la eliminación del Wash Buffer.

d. Elución de ADN

El SpinColumn se colocó en un tubo de microcentrífuga de 1.5 ml. Se agregaron 100µl de Elution Buffer, incubando a temperatura ambiente por 1 minuto. Posteriormente, se centrifugó a 14000 rpm durante 1 minuto, obteniendo así el ADN genómico purificado de la Bahía Interna y la captación Chimú.



e. Almacenamiento de ADN

El ADN purificado de las muestras de la Bahía Interna y la captación Chimú se almacenó a -20°C en una refrigeradora.

3.6 CALIDAD E INTEGRIDAD DEL ADN BACTERIANO

3.6.1 Calidad/cantidad del ADN

A. Técnica

Espectrofotometría

B. Fundamento

La espectroscopía UV-Visible se centra en la adsorción de radiación ultravioleta y visible por moléculas orgánicas e inorgánicas. Este fenómeno proporciona información valiosa sobre la concentración y la composición de las sustancias presentes en una muestra. La relación lineal establecida por la Ley de Lambert-Beer permite cuantificar la concentración de una sustancia absorbente en función de la absorbancia medida. La ecuación $A = \epsilon lc$, donde A es la absorbancia, ϵ es la absorptividad molar, l es la longitud del camino óptico y c es la concentración, forma la base matemática de la espectrofotometría cuantitativa. La espectrofotometría implica el uso de espectrofotómetros que generan un espectro de absorción. Estos instrumentos constan de una fuente de luz, un sistema dispersivo, una muestra y un detector, donde la variación en la absorbancia a diferentes longitudes de onda revela una firma espectral única de las moléculas presentes (ISC-III, 2011).

C. Procedimiento

Las muestras de ADN genómico fueron leídas en un espectrofotómetro de microvolúmenes UV/Vis (marca PCRmax, modelo Lambda), para esto las muestras extraídas en combinación de Buffer de Elución (BE) y ADN. El objetivo principal es evaluar la calidad del ADN obtenido de las muestras procesadas.

Para garantizar la precisión de las mediciones, se realiza un blanqueo del equipo utilizando 1 µl de buffer de elución (BE). De ahí se procedió a analizar la concentración de ADN de las muestras, expresada en unidades de microlitros por mililitro (ul/ml) para lo cual se realizaron 5 lecturas que aseguraron la calidad como la cantidad adecuada del ADN (ISC-III, 2011).

Tabla 9

Valores de pureza de ADN genómico

| 1.4 | 1.4 – 1.6 | 1.6 – 1.8 | 1.8 – 2.0 | 2.0 |
|-------------|-----------|-----------|-----------|-------------|
| Malo | regular | Bueno | Perfecto | Malo |
| Contaminado | | | | contaminado |

Fuente: Datos tomados del Departamento de control de calidad del Banco Nacional de ADN, (2020)

3.6.2 Integridad de ADN

A. Técnica

Electroforesis

B. Fundamento

La electroforesis es la aplicación de un campo eléctrico que induce la migración de las moléculas cargadas hacia el electrodo de polaridad opuesta. El uso de un gel o matriz porosa controla la velocidad de migración, permitiendo la



separación según el tamaño y la movilidad de las moléculas en la electroforesis, lo cual depende de su carga y forma, lo que posibilita su separación (Forero et al., 2021).

C. Procedimiento

En la preparación del gel de agarosa al 1%, se emplearon 0.50g de gel de agarosa (marca CSL-AG500) disuelto en 50 ml de TAE 1X (1 ml Tris-Acetato-EDTA 50X +49 ml agua destilada) según Decker et al. (2018) (Anexo 13 y Anexo 14) siendo luego colocado en una cámara con un peine para 6 pocillos, dejándose enfriar a temperatura ambiente.

Posteriormente, se efectuó la carga del gel de agarosa al 1% en la Cámara de electroforesis de la marca BIO-RAD, modelo PowerPac, cubriéndola con ~300 ml TAE 1X, luego del cual en cada uno de los pozos se colocó una mezcla de 2 μ l de ADN con 10 μ l de safe green, lo que mismo que para los marcadores de peso molecular. Posteriormente se aplicó una tensión de 100mV durante 30 minutos (Anexo 15 y Anexo 16), de ahí para evaluar la integridad del ADN en las muestras de la BI y CH se trasladó la corrida del gel de agarosa al 1%, a un transiluminador (marca Imager Camera Hood E-Gel), que permitió la observación de la migración e integridad de las bandas correspondientes al ADN genómico mediante luz UV (Anexo 17) (Decker et al., 2018).

3.7 ANÁLISIS METAGENÓMICO POR SECUENCIACIÓN SHOTGUN

A. Técnica

Secuenciación Shotgun



B. Fundamento

La secuenciación genómica Shotgun es un método de secuenciación de nueva generación que permite el análisis simultáneo de varios genomas con una mayor precisión, operando mediante la fragmentación aleatoria del ADN de interés en pequeños fragmentos, los cuales son secuenciados individualmente. Estos fragmentos, al solaparse y superponerse, permiten reconstruir la secuencia completa del genoma original. Este enfoque elimina la necesidad de secuenciar el ADN en una secuencia lineal, proporcionando una eficiencia notable en términos de tiempo y recursos. La técnica de secuenciación Shotgun ha encontrado aplicaciones cruciales en proyectos genómicos de gran envergadura, como la secuenciación de genomas completos y la identificación de variantes genéticas. Su capacidad para generar datos de alta resolución ha revolucionado la comprensión de la diversidad genética y la complejidad de los organismos estudiados (Quince et al., 2017).

C. Procedimiento

La secuenciación metagenómica Shotgun es una secuenciación de nueva generación (NGS) que requirió el envío de las muestras ambientales de ADN a BTS Consultores. Para ello después de determinar la calidad, cantidad e integridad del ADN con valores óptimos se almacenó a -80°C para su respectivo envío a BTS Consultores (Anexo 23).

Las muestras fueron enviadas en viales de microcentrífuga correspondiente a una cantidad de 20ul de ADN, cubierta en Parafilm en una caja de Tecnopor con hielo seco (-80°C) las que estuvieron debidamente codificadas haciendo referencia a la Bahía Interna (BI) y la captación Chimú (CH).



Para la secuenciación shotgun se utilizó el equipo Illumina HiSeq cuya plataforma es ampliamente utilizada en laboratorios de investigación e industria biotecnológica debido a su alta capacidad de secuenciación y su capacidad para generar datos de alta calidad, siendo uno de los sistemas de secuenciación de próxima generación (NGS) más utilizados en la actualidad, conocido también por su versatilidad, alta capacidad y calidad de datos.

Algunas especificaciones del funcionamiento del equipo Illumina HiSeq son:

- a. Capacidad de secuenciación: El HiSeq puede generar grandes volúmenes de datos, desde millones hasta miles de millones de fragmentos de ADN secuenciados en una sola corrida.
- b. Longitud de lectura: Puede variar dependiendo de la versión específica del HiSeq y de los kits de secuenciación utilizados, pero en general, puede alcanzar varios cientos de bases por lectura.
- c. Tecnología de secuenciación: El HiSeq utiliza una tecnología de secuenciación por síntesis, en la que se incorporan nucleótidos marcados fluorescentemente uno por uno durante la elongación de los fragmentos de ADN. Esto permite una alta precisión en la determinación de la secuencia de nucleótidos.
- d. Flujo de trabajo: El flujo de trabajo típico del HiSeq implica la preparación de la biblioteca de ADN, donde el ADN se fragmenta, se etiqueta y se amplifica para su secuenciación. Luego, la biblioteca preparada se carga en el instrumento, donde se secuencia y se generan los datos crudos.
- e. Análisis de datos: Una vez completada la secuenciación, los datos crudos se



procesan mediante software especializado para la eliminación de adaptadores, la alineación de lecturas y la llamada de variantes, entre otros análisis (Gupta & Verma, 2019).

Funcionamiento del equipo

- a. Preparación de la biblioteca: El ADN de interés se extrae y se fragmenta en fragmentos más pequeños. Luego, se añaden adaptadores a los extremos de estos fragmentos y se amplifican para generar la biblioteca de ADN preparada para la secuenciación.
- b. Carga de la biblioteca: La biblioteca de ADN preparada se carga en el instrumento HiSeq.
- c. Secuenciación: Dentro del instrumento, los fragmentos de ADN se unen a una superficie sólida y se amplifican en clusters. Luego, se inicia el proceso de secuenciación por síntesis, donde se añaden nucleótidos marcados fluorescentemente uno por uno y se detecta la señal de fluorescencia para determinar la secuencia de nucleótidos.
- d. Generación de datos: A medida que se incorporan los nucleótidos, se capturan imágenes para registrar la señal de fluorescencia y se generan los datos crudos de secuenciación.
- e. Análisis de datos: Una vez completada la secuenciación, los datos crudos se procesan mediante software especializado para generar la secuencia de nucleótidos, identificar variantes genéticas y realizar otros análisis bioinformáticos (Gupta & Verma, 2019).

En este punto BTS Consultores realizó la secuenciación hasta la fase 4



suministrando la data de secuenciamiento con las siguientes características:

Data de la Bahía Interna (BI)

-. 20008D-10-01_S204_L003_R1_001.fastq Peso de 437,019 KB

-. 20008D-10-01_S204_L003_R2_001.fastq Peso de 447,053 KB

Data de la Captación Chimú (CH):

-. 20008D-10-02_S205_L003_R1_001.fastq Peso de 453,873 KB

-. 20008D-10-02_S205_L003_R2_001.fastq Peso de 461,377 KB

Con una longitud de lectura PE (paired end) 2x150 pb y 10 millones de lecturas por muestra (5 millones en cada dirección) es que se procedió, a la obtención del árbol filogenético, que se proyectó en el sistema operativo Linux, usando los programas MEGAHit para los contigs (segmentos de ADN superpuestos), MetaPhlAn (versión 3.0.13), MetaPhlAn (versión 4.0.0), metAML, MetaLAFFA y GraphlAn para los gráficos (Blanco et al., 2023).

3.8 CALIDAD DE LA DATA SECUENCIADA

La calidad se evaluó mediante la plataforma informática Galaxy versión 0.74+galaxy0, para lo cual se analizó la data secuenciada en el formato FastQC tanto para la Bahía interna como para la Captación Chimú.

3.9 ANÁLISIS BIOINFORMÁTICO

Tabla 10

Recursos bioinformáticos empleados para el análisis metagenómico Shotgun

| Software y algoritmos | Identificador | Fuente |
|----------------------------|---|----------------------|
| MetaPhlAn (versión 3.0.13) | https://github.com/biobakery/MetaPhlAn/ | Beghini et al., 2021 |
| MetaPhlAn (versión 4.0.0) | https://github.com/biobakery/MetaPhlAn/ | Blanco et al., 2023 |
| Metaml | https://github.com/segatalab/metaml | Pasolli et al., 2016 |
| Base de datos SILVA | https://www.arb-silva.de/ | Quast et al., 2013 |
| Base de datos RefSeq | https://www.ncbi.nlm.nih.gov/refseq/ | O’Leary et al., 2016 |

Fuente: Elaboración propia

3.9.1 Herramientas bioinformáticas para metagenómica de Shotgun

Esta investigación requirió de una cantidad de recursos bioinformáticos relacionadas a la metagenómica de Shotgun para poder determinar el perfil taxonómico, relaciones filogenéticas y abundancia, los cuales fueron vitales para analizar los datos de manera eficiente y correcta, como también para visualizar de manera comprensible los resultados obtenidos.

A. Técnica

Metaphlan4



B. Fundamento

El fundamento de MetaPhlAn se basa en el análisis de secuencias metagenómicas mediante el uso de marcadores genéticos específicos. Utiliza perfiles de clado únicos para estimar las abundancias relativas de los diferentes organismos presentes en una muestra. MetaPhlAn aborda la complejidad de las comunidades microbianas al proporcionar una clasificación taxonómica de alta resolución y discriminación (Chang et al., 2023).

C. Procedimiento

Se llevó a cabo la aplicación de los software y algoritmos de MetaPhlAn en sus versiones 3.0.13 y 4.0.0, que es una herramienta especializada en perfilar la composición de comunidades microbianas a partir de datos de secuenciación por Shotgun de metagenomas. Para la utilización efectiva de MetaPhlAn, se estableció que es necesario cumplir con ciertos requisitos previos, según lo indicado por O’Leary et al. (2016).

Los requisitos incluyen la presencia de Python (versión ≥ 2.7), pajarita2, numpy, Pandas (opcional, requerido solo por scripts de utilidad), BioPython (opcional, solo requerido por scripts de utilidad), SciPy (opcional, solo requerido por utilidad de guiones), Matplotlib (opcional, solo requerido por scripts de utilidad), y biom (opcional, solo requerido para el formato de entrada/salida biom) (Home “BioBakery,” s.f.).

El lenguaje de programación usado para poder trabajar con MetaPhlAn fue Python que es una herramienta muy poderosa en el ámbito de la ciencia de datos y el análisis de datos. Bibliotecas como NumPy, Pandas, Matplotlib y SciPy proporcionaron las herramientas para manipular datos, realizar análisis



estadísticos, visualizar datos y realizar cálculos científicos. Para ello se descargó la versión más actualizada para el sistema operativo Linux.

a. Descarga de Python

Se descargó desde el sitio web oficial de Python en <https://www.python.org/downloads/> la última versión estable para el sistema operativo Linux.

b. Ejecución del instalador

Una vez que se completó la descarga, se ejecutó el archivo de instalación.

c. Verificación de la instalación

Se aseguro que Python se haya instalado correctamente, para lo cual se abrió en la terminal y escribió el siguiente comando: `python --version`

Se siguió con la implementación de bioBakery VM, MetaPhlAn que se fundamenta en BowTie2 (versión 2.3 o superior) para realizar el mapeo de lecturas contra genes marcadores. Se especifico que la existencia de Bowtie2 en la ruta del sistema con permisos adecuados es esencial para este proceso. En esta situación se utilizó Conda para la instalación de MetaPhlAn, la cual no requiere requisitos previos adicionales, escribiendo el siguiente comando.

```
$ conda install -c bioconda metaphlan=2.7
```

Esto también instaló todas las dependencias de MetaPhlAn, alternativamente, se pudo instalar MetaPhlAn manualmente desde la fuente y luego instalar manualmente las dependencias con el siguiente código.

```
$ tar xzvf biobakery-metaphlan-<versionid>.tar.gz
```



```
$ cd biobakery-metaphlan-<versionid>/
```

```
$ ls
```

Se agrego este comando `export PATH=$MDIR:$MDIR/utils/:$PATH` a `$HOME/.bashrc` archivo donde se reemplazo `$MDIR` con la ruta completa a la carpeta `MetaPhlAn`. Para obtener el valor de `$MDIR` se ejecutó `$ pwd` desde dentro de la carpeta `MetaPhlAn`.

Se obtuvo el `$HOME/.bashrc` archivo, luego el `$ source $HOME/.bashrc`. Esto solo se hizo una vez para actualizar la RUTA en su sesión actual, de ahí todas las sesiones futuras se obtuvieron automáticamente cuando se inició sesión.

Alternativamente, en lugar de agregar los scripts a la RUTA, se puede proporcionar la ruta completa al script cada vez que lo ejecute. El siguiente comando imprime la pantalla de ayuda de `MetaPhlAn`.

```
$ $MDIR/metaphlan.py -help
```

Instalación de las dependencias de MetaPhlAn

Se instaló `bowtie2` en una carpeta en la RUTA y se especificó la ruta al ejecutable de `bowtie2` con cada ejecución de `MetaPhlAn` usando la opción `--bowtie2_exe <bowtie2>`.

Cabe destacar que `MetaPhlAn` se integra de manera sinérgica con `hclust2` para la generación avanzada de mapas de calor y `GrPhlAn` para la visualización de cladogramas. La metodología que se usó para la instalación de `MetaPhlAn` fue



a través de conda, utilizando el canal Bioconda., en caso de no haber configurado la instalación de Anaconda para restaurar paquetes de Bioconda, se deben seguir los pasos indicados por Chang et al. (2023).

3.9.2 Uso de Metaphlan4 para el perfil taxonómico por medio de árboles filogenéticos y el mapa de calor

A. Técnica

MetaphAn

B. Fundamento

El fundamento de MetaPhlAn se basa en el análisis de secuencias metagenómicas mediante el uso de marcadores genéticos específicos. Utiliza perfiles de clado únicos para estimar las abundancias relativas de los diferentes organismos presentes en una muestra. MetaPhlAn aborda la complejidad de las comunidades microbianas al proporcionar una clasificación taxonómica de alta resolución y discriminación (Chang et al., 2023).

C. Procedimiento

MetaPhlAn acepta pequeñas lecturas de una sola escopeta (Shotgun) como entrada y genera una lista de microorganismos detectados junto con su relativa abundancia. La herramienta también es capaz de procesar secuencias genómicas extraídas

Archivos de entrada

Están en formatos como `.fasta`, `.fastq` y `.tar.bz2`. Para ver todos los formatos de entrada, se escribió `metaphlan.py -h | less`.



Utilizando la tecla de flecha nos desplazamos hacia arriba y hacia abajo.

Escribiendo `q` sirve para salir y volver al mensaje.

Teniendo ya nuestra base de datos crudos se empieza ejecutando en la terminal los siguientes códigos:

Para la data de la Bahía interior (BI)

```
$ curl -LO https://github.com/biobakery/biobakery/raw/master/test_suite/biobakery_tests/data/metaphlan/input/20008D-10-01_S204_L003_R1_001.fastq
```

```
$ curl -LO https://github.com/biobakery/biobakery/raw/master/test_suite/biobakery_tests/data/metaphlan/input/20008D-10-01_S204_L003_R2_001.fastq
```

Para la data de la Captación Chimú (CH):

```
$ curl -LO https://github.com/biobakery/biobakery/raw/master/test_suite/biobakery_tests/data/metaphlan/input/20008D-10-02_S205_L003_R1_001.fastq
```

```
$ curl -LO https://github.com/biobakery/biobakery/raw/master/test_suite/biobakery_tests/data/metaphlan/input/20008D-10-02_S205_L003_R2_001.fastq
```

Teniendo los 4 archivos de entrada, creamos una nueva carpeta donde se



trasladó todos los archivos subidos usando los siguientes comandos de la terminal:

```
$ mkdir metaphlan_analysis
```

Permite mover los archivos de la carpeta de descargas con el siguiente código.

```
$ mv ~/Downloads/SRS*.fasta.gz  
metaphlan4_analysis/
```

Teniendo la ubicación y los nombres de los archivos ejecutamos muestra por muestra a partir de lecturas sin procesar para perfilar un metagenoma.

Para Bahía interior (BI)

```
$ metaphlan4.py 20008D-10-  
01_S204_L003_R1_001.fasta.gz --input_type fasta >  
20008D-10-01_S204_L003_R1_001.txt
```

```
$ metaphlan4.py 20008D-10-  
01_S204_L003_R2_001.fasta.gz --input_type fasta >  
20008D-10-01_S204_L003_R2_001.txt
```

Para Captación Chimú (CH)

```
$ metaphlan4.py 20008D-10-  
01_S205_L003_R1_001.fasta.gz --input_type fasta >  
20008D-10-01_S205_L003_R1_001.txt
```

```
$ metaphlan4.py 20008D-10-  
01_S205_L003_R2_001.fasta.gz --input_type fasta >  
20008D-10-01_S205_L003_R2_001.txt
```



Archivos de salida

Ejecutando MetaPhlAn, siguiendo el ejemplo de la sección anterior, se crearon dos archivos de salida. Las alineaciones se enumeran una por línea en columnas de marcador de lectura y referencia separadas por tabulaciones:

Archivo 1: 20008D-10-01_S205_L003_R1_001.fasta.gz.bowtie2out.txt

Este archivo contiene los resultados del mapeo intermedio a marcadores de secuencia únicos:

```
$ less -S 20008D-10-01_S205_L003_R2_001.fasta.gz.bowtie2out.txt
```

Archivo 2: 20008D-10-01_S205_L003_R1_001.txt

Este archivo contiene las abundancias finales de los organismos, donde las abundancias de los organismos se enumeran en un clado por línea, separadas por tabulaciones del porcentaje de abundancia del clado:

```
$ less -S 20008D-10-01_S205_L003_R1_001.txt
```

El archivo de salida resultante contiene los resultados de asignaciones intermedias a "etiquetas" de secuencias individuales. Cada línea del archivo presenta alineaciones enumeradas en columnas separadas por listas de etiquetas de lectura y referencia, proporcionando la abundancia final calculada del organismo correspondiente.

Las abundancias de los organismos se enumeran por clado y linaje, separadas mediante tablas de abundancia porcentual de clado. La primera columna del archivo enumera los clados, que abarcan desde reinos taxonómicos como



Bacteria y Archaea hasta especies individuales. Cada nivel taxonómico de clado está precedido por un prefijo indicativo de su nivel, por ejemplo, Kingdom: k__, Phylum: p__, Class: c__, Order: o__, Family: f__, Genus: g__, Species: s__s__.

Es relevante señalar que MetaPhlAn tiene la capacidad de ejecutar múltiples núcleos, adaptándose a la capacidad de procesamiento de cada ordenador, según Blanco et al. (2023).

Ejecución en múltiples núcleos

Esta opción tuvo lugar gracias al computador usado que correspondió a un servidor hp z440 de 32gb de RAM con un procesador Intel Xeon, 2 discos solidos de 550gb y 500gb ´de 8 núcleos y 4 procesadores lógicos. Todo esto para aprovechar sus múltiples núcleos usando la cantidad de nproc con el fin de agilizar y trabajar varias muestras a la vez.

```
$ metaphlan.py 20008D-10-01_S205_L003_R2_001--  
input_type fasta --nproc 4 > 20008D-10-  
01_S205_L003_R2_001_profile.txt
```

A continuación, se trabajó con bash para la sintaxis del shell, con lo que se pudo recorrer todo el conjunto de muestra con el siguiente código:

```
$ for f in SRS*.fasta.gz  
> do  
> metaphlan2.py $f --input_type fasta --nproc  
4 > ${f%.fasta.gz}_profile.txt
```



> done

Finalmente, la distribución MetaPhlAn incluye un script de utilidad que creo una única tabla delimitada por tabulaciones a partir de estos archivos:

```
$ merge_metaphlan_tables.py *_profile.txt >  
merged_abundance_table.txt
```

La tabla resultante se puede abrir en Excel, cualquier programa de análisis de expresión genética con `less` (Anexo 2 y Anexo 3)

```
$ less -S merged_abundance_table.txt
```

3.9.3 Obtención de árbol filogenético

A. Técnica

GraPhlAn

B. Fundamento

GraPhlAn ha sido diseñado para producir representaciones circulares que destacan la estructura y composición de árboles filogenéticos y taxonómicos. Su enfoque se concentra en la generación de visualizaciones de calidad, facilitando la interpretación de complejas relaciones entre entidades biológicas. La herramienta permite la inclusión de información detallada de manera intuitiva, lo que resulta crucial para la comprensión visual de datos biológicos (Blanco et al., 2023).



C. Procedimiento

Se utilizo la guía de bioBakery GraPhlan y que al insertar el siguiente código:

```
$ conda install -c biobakery graphlan
```

Esto instalo GraPhlAn y todas sus dependencias. Alternativamente, se instalaron GraPhlAn manualmente desde el código fuente descargando GraPhlAn y luego instalando las dependencias de GraPhlAn (numpy, pandas, biopython, scipy y matplotlib).

Creación de los archivos de entrada de GraPhlAn

Se aplico el algoritmo de GraPhlAn que al usarla, demando dos inputs: (1) una estructura de árbol que fungirá como representación y (2) opciones de anotación gráfica específicas para dicho árbol.

```
$ export2graphlan.py --skip_rows 1,2 -i  
merged_abundance_table.txt --tree  
merged_abundance.tree.txt --annotation  
merged_abundance.annot.txt --most_abundant 100 --  
abundance_threshold 1 --least_biomarkers 10 --  
annotations 5,6 --external_annotations 7 --  
min_clade_size 1
```

El comando anterior tiene opciones para omitir las líneas 1 y 2, seleccionar los 100 clados más abundantes, establecer un umbral de abundancia mínimo para que los clados anoten, extraer al menos 10 biomarcadores, seleccionar un nivel taxonómico 5 y 6 para la anotación en el árbol, seleccionar el nivel taxonómico 7

para usar en la leyenda externa y establecer el tamaño mínimo de los clados anotados como biomarcadores en 1. Los archivos de salida creados son `merged_abundance.tree.txt` y `merged_abundance.annot.txt`. (Blanco et al., 2023).

Creación del árbol filogenético

Una vez descargado los archivos de entrada, se creó una nueva carpeta y se trasladó los archivos usando los siguientes comandos en el símbolo del terminal.

Se ejecuto los siguientes comandos para generar el cladograma que proporciona los archivos de árbol (`merged_abundance.tree.txt`) y anotaciones (`merged_abundance.annot.txt`) del paso anterior, se introdujo los códigos a la terminal:

```
$ graphlan_annotate.py --annot
merged_abundance.annot.txt merged_abundance.tree.txt
merged_abundance.xml

$ graphlan.py --dpi 300 merged_abundance.xml
merged_abundance.png --external_legends
```

El primer comando crea un archivo xml a partir del árbol y las entradas de anotaciones. El segundo comando crea la imagen, configura la resolución de la imagen en 300 y solicita leyendas externas.

Las primeras líneas del archivo xml son:



```
<phyloxml xmlns="http://www.phyloxml.org"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.phyloxml.org
http://www.phyloxml.org/1.10/phyloxml.xsd">

  <phylogeny rooted="true">

    <clade>

      <clade>

        <name>k__Viruses</name>

        <branch_length>1.0</branch_length>
```

3.9.4 Obtención del mapa de calor

A. Técnica

Hclust2

B. Fundamento

Hclust2 se erige como una solución práctica y eficiente, diseñada específicamente para la creación de mapas de calor con numerosas opciones que garantizan la producción de figuras de alta calidad, adecuadas para su inclusión en publicaciones científicas. Entre sus características destacadas se encuentran la capacidad de realizar agrupaciones jerárquicas utilizando el algoritmo hclust, lo que permite una representación visual clara de las relaciones entre las muestras y las taxonomías identificadas (Blanco et al., 2023).

C. Procedimiento

En este proyecto se llevó a cabo la creación de un mapa de calor para todas las muestras. Utilizando el tutorial de bioBakery se instaló hclust2 y otras herramientas que de no ser así, se insertó el siguiente código:

```
$ conda install -c biobakery hclust2
```

Esto instalo hclust2 y todas sus dependencias. Alternativamente, se puede instalar hclust2 manualmente desde el código fuente descargando hclust2 y luego instalando las dependencias de hclust2 (numpy, pandas, biopython, scipy y matplotlib).

Generación de la tabla de abundancia exclusiva de especies.

Se ejecutó un comando específico con el propósito de generar una tabla de abundancia exclusivamente para especies, utilizando como entrada la tabla de abundancia (merged_abundance_table.txt).

```
$ grep -E "(s__)|(^ID)" merged_abundance_table.txt  
| grep -v "t__" | sed 's/^.*/s__//g' >  
merged_abundance_table_species.txt
```

Dicho comando se compone de tres partes. En primer lugar, el comando 'grep' número 1 busca en el archivo la expresión regular "(s__)|(^ID)", coincidiendo con las líneas que contienen información sobre especies. El 'grep' número 2 identifica las líneas con información acerca de la cepa, etiquetada como t__. Posteriormente, el comando 'sed' (editor de secuencias) elimina la taxonomía completa de cada línea, dejando solo el nombre de la especie en la primera columna.

La tabla resultante de abundancia (merged_abundance_table_species.txt) contiene únicamente las abundancias de especies, representadas por los nombres de las especies sin incluir la taxonomía completa, siguiendo la metodología propuesta por (Beghini et al. 2021).

Generación del mapa de calor

Luego se generó el mapa de calor solo de especies ejecutando el siguiente comando:

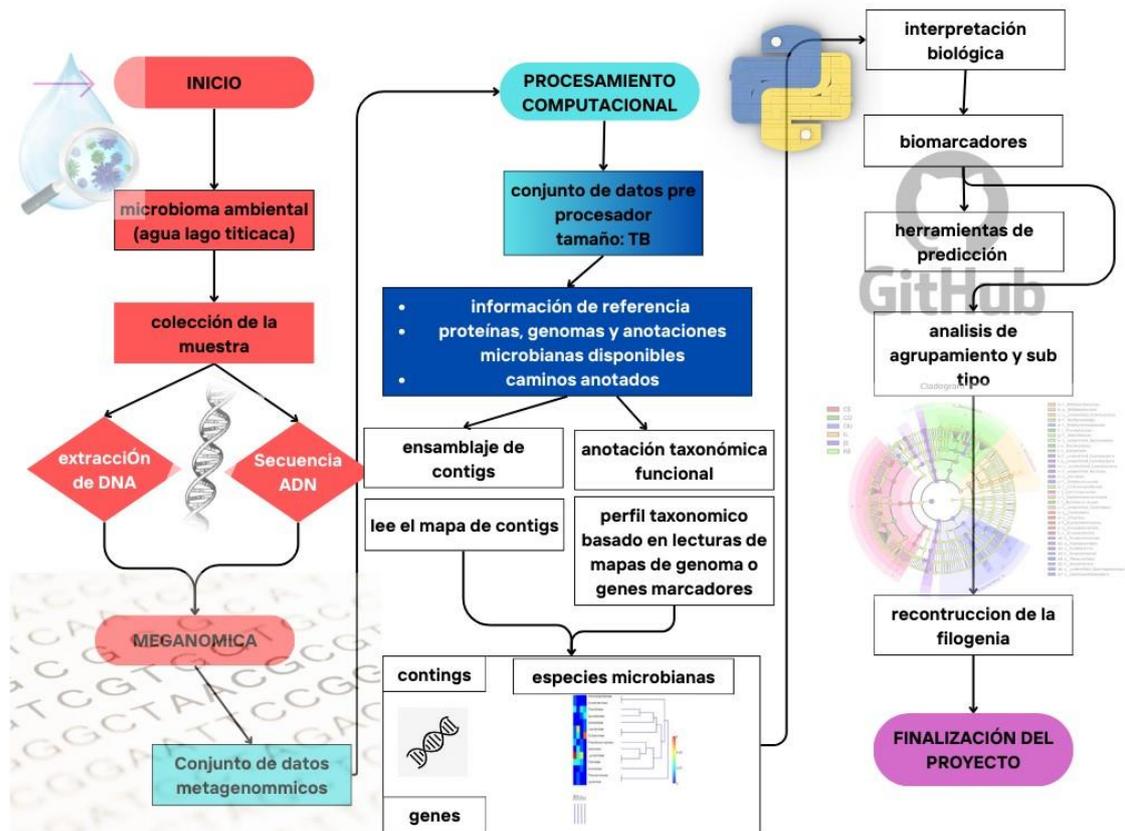
```
$ hclust2.py -i  
merged_abundance_table_species.txt -o  
abundance_heatmap_species.png --ftop 25 --f_dist_f  
braycurtis --s_dist_f braycurtis --cell_aspect_ratio  
0.5 -l --flabel_size 6 --slabel_size 6 --max_flabel_len  
100 --max_slabel_len 100 --minv 0.1 --dpi 300
```

El comando anterior incluye opciones para seleccionar 25 características principales, para lo cual se usó Bray-Curtis como medida de distancia entre muestras y entre características (microorganismos), además se estableció la relación ancho/alto de la celda en 0,5 y se empleó usar una escala logarítmica para escalar el mapa de calor. También se estableció el tamaño de las etiquetas de muestra y características en 6, además de la longitud máxima de las muestras y etiquetas de características en 100; asimismo se seleccionó el valor mínimo mostrado como 0.1 y se seleccionó la resolución de la imagen como 300.

Se abrió el mapa de calor resultante (affectity_heatmap_species.png) para verlo.

Figura 7

Diagrama de flujo de trabajo metagenómico



Fuente: Elaboración propia



CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1 EXTRACCIÓN DE ADN CALIDAD, CONCENTRACIÓN E INTEGRIDAD

A. Calidad de ADN de las muestras previamente filtradas

Entre los resultados de las lecturas de calidad de ADN obtenidas a través del espectrofotómetro de microvolúmenes con la unidad de concentración de $\mu\text{g/ml}$.

Tabla 11

Lectura de valores que resaltan la calidad de DNA de la muestra de Bahía interna

| Calidad de ADN Bahía Interna | |
|------------------------------|-------|
| Repeticiones | Valor |
| Lectura 1 | 1.692 |
| Lectura 2 | 1.751 |
| Lectura 3 | 1.702 |
| Lectura 4 | 1.732 |
| Lectura 5 | 1.724 |

Fuente: elaboración propia

En la Tabla 11 muestra datos del espectrofotómetro sobre la calidad de la muestra de ADN extraído de la Bahía Interna estando dentro de los valores aptos no siendo menor a 1.4 y no siendo mayor a 2.0, donde se realizaron 5 repeticiones para demostrar la precisión de la calidad por ende se escogió la última lectura siendo el valor final para la aplicación al gel de agarosa.



Tabla 12

Lectura de valores que resaltan la calidad de DNA de la muestra de Captación chimu

| Calidad de ADN Captación Chimú | |
|---------------------------------------|--------------|
| Repeticiones | Valor |
| Lectura 1 | 1.601 |
| Lectura 2 | 1.543 |
| Lectura 3 | 1.532 |
| Lectura 4 | 1.531 |
| Lectura 5 | 1.571 |

Fuente: elaboración propia

En la Tabla 12 muestra datos del espectrofotómetro sobre la calidad de la muestra de ADN extraído de la Captación Chimú estando dentro de los valores aptos no siendo menor a 1.4 y no siendo mayor a 2.0, donde se realizaron 5 repeticiones para demostrar la precisión de la calidad por ende se escogió la última lectura siendo el valor final para la aplicación al gel de agarosa.

B. Concentración de ADN de las muestras previamente filtrada

Resultados de las lecturas de concentración de ADN a través del espectrofotómetro de microvolúmenes con la unidad de concentración de $\mu\text{g/ml}$.

Tabla 13

Concentración de ADN bacteriano de las muestras de agua de la Bahía Interior

| Concentración de ADN (ng/μl) Bahía Interior | |
|--|--------------|
| Repeticiones | Valor |
| Lectura 1 | 39.415 |
| Lectura 2 | 44.781 |
| Lectura 3 | 36.487 |
| Lectura 4 | 40.935 |
| Lectura 5 | 41.512 |

Fuente: Elaboración propia

En la Tabla 13 se presenta datos del espectrofotómetro sobre la concentración de ADN extraído de la Bahía Interna donde se realizaron 5 repeticiones para demostrar la precisión de la concentración de ADN, de las cuales se escogió la última lectura siendo el valor final para la aplicación al gel de agarosa

Tabla 14

Concentración de ADN bacteriano de las muestras de agua de la Captación Chimú

| Concentración de ADN (ng/μl) Captación Chimú | |
|---|--------------|
| Repeticiones | Valor |
| Lectura 1 | 21.741 |
| Lectura 2 | 22.973 |
| Lectura 3 | 22.012 |
| Lectura 4 | 21.412 |
| Lectura 5 | 22.321 |

Fuente: Elaboración propia

En la Tabla 14 se presenta datos del espectrofotómetro sobre la concentración de la muestra de ADN extraído de la Captación Chimú donde se realizaron 5 repeticiones

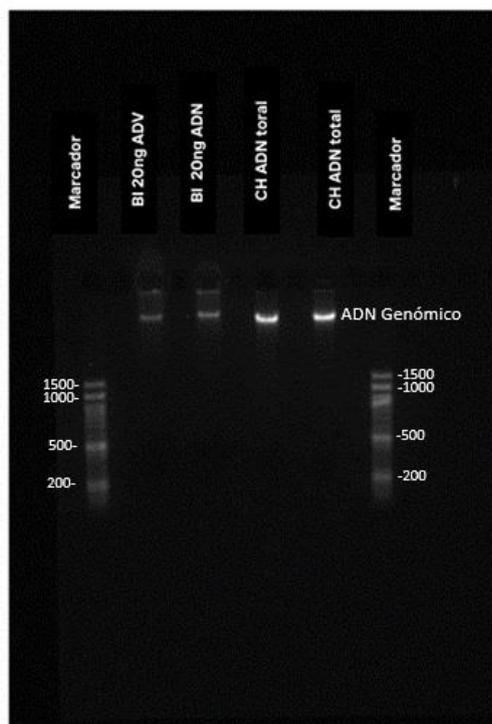
para demostrar la precisión de la concentración de ADN, de las cuales, se escogió la última lectura siendo el valor final para la aplicación al gel de agarosa.

C. Integridad del ADN

La integridad obtenida de las muestras de la BI y CH por medio del gel de agarosa a 1% nos indica la presencia de una banda entera de 20ng de ADN, la que, al no estar fragmentado, señala que el ADN estuvo de una forma íntegra y sin la presencia de contaminantes en la corrida.

Figura 8

Integridad de ADN genómico bacteriano de las muestras filtradas de la Bahía Interna y Captación Chimú, en gel de agarosa con TAE 1%



Fuente: Elaboración propia

En la Figura 8 nos detalla una integridad perfecta de ADN donde nos muestra que conserva su longitud y no hay presencia de fragmentación que evidencie que el ADN ha sufrido roturas significativas.



4.1.2 Análisis de secuenciamiento Shotgun

El análisis de calidad se llevó a cabo mediante el recurso en línea Galaxy (www.usegalaxy.org) con la versión Galaxy Version 0.74+galaxy0 para los reportes de calidad de cada lectura secuenciada donde se obtuvo un análisis estadístico que representa el resumen de la secuenciación genómica de la Bahía interna y la Captación Chimú, el cual indica el tipo de codificación del secuenciamiento, el total de secuencias, cantidad de bases nitrogenadas, tamaño de la secuencia y porcentaje de Guanina (G), Citosina (C).

Para esta investigación, la cantidad de secuencias totales en la BI fueron de 5643902, y en la CH es de 5863208. Las secuencias etiquetadas como baja calidad fueron de 0 lo cual nos da a entender que el secuenciamiento fue de alto rendimiento.

Tabla 15

Resumen de datos de la secuenciación genómica de la Bahía Interna

| Medida | Valor |
|--|-----------------------|
| Nombre | Bahía Interior |
| Codificación | Sanger / Illumina 1.9 |
| Secuencias Totales | 5643902 |
| Bases Totales | 852.2 Mbp |
| Secuencias etiquetadas como de baja calidad | 0 |
| Tamaño de las secuencias | 151 |
| %GC | 46 |

Fuente: Elaboración propia

En la Tabla 15 se presenta un resumen sobre la secuenciación genómica de la muestra de la Bahía Interna, indicando de que no se tienen secuencias etiquetadas de baja calidad.

Tabla 16

Resumen de datos de la secuenciación genómica de la Captación Chimú

| Medida | Valor |
|--|-----------------------|
| Nombre | Captación Chimú |
| Codificación | Sanger / Illumina 1.9 |
| Secuencias Totales | 5863208 |
| Bases Totales | 885.3 Mbp |
| Secuencias etiquetadas como de baja calidad | 0 |
| Tamaño de las secuencias | 151 |
| %GC | 48 |

Fuente: Elaboración propia

En la Tabla 16 muestra un resumen sobre la secuenciación genómica de la muestra de la Bahía Interna, dándonos a conocer que no se tienen secuencias etiquetadas de baja calidad.

4.2 ÁRBOL FILOGENÉTICO DE BACTERIAS PRESENTES EN LA BAHÍA INTERNA (BI)

Clados identificados por Metaphlan

En el análisis de las 5643902 lecturas obtenidas mediante secuenciamiento metagenómico Shotgun, se empleó la herramienta computacional Metaphlan para perfilar la composición bacteriana. Las muestras de la Bahía Interna revelaron clados microbiológicos significativos, especialmente los contenedores de genoma a nivel de especie (SGB). Entre las especies conocidas (kSGB) presentes, se destacaron

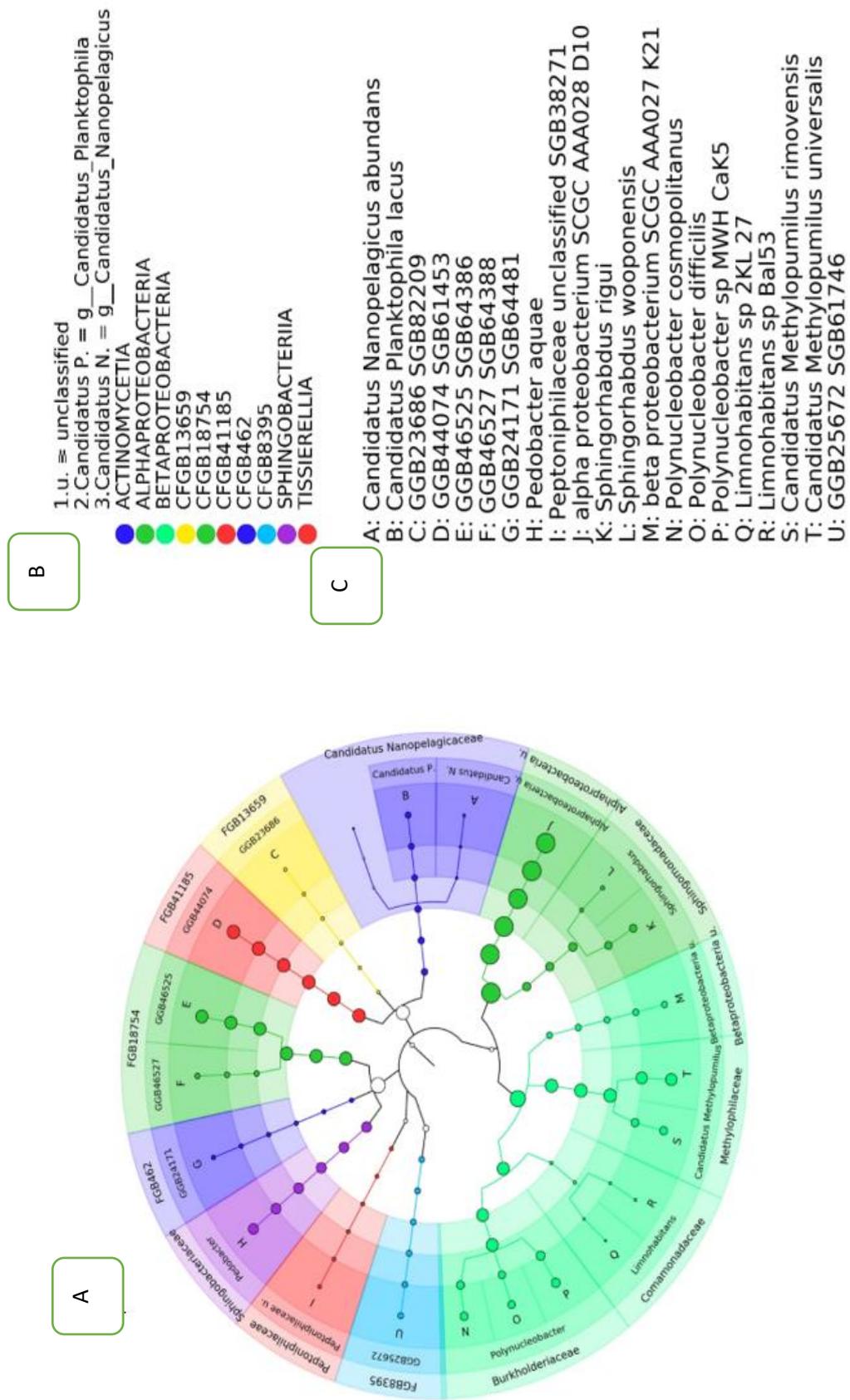


actinomicetia, alphaproteobacteria, betaproteobacteria, sphingobacteria y tissierellia y las especies *Polynucleobacter sp MWH CaK5*, *Candidatus methylopumilus universalis*, *Pedobacter aquae* y *Alpha proteobacterium SCGC AAA028 D10*. Además, se identificaron especies aún por caracterizar o desconocidas (uSGB) como CFGB13659, CFGB18754, CFGB41185, CFGB462 y CFGB8395 y géneros, especies; *GGB46525_SGB64386*, *GGB44074_SGB61453*. En el estudio de Ahmad et al. (2021) sobre la composición y distribución del microbioma en el lago Dal, Srinagar, India, se observaron comunidades microbiológicas distintivas en los puntos de muestreo. El filo bacteriano más abundante, confirmado también por Du et al. (2022), fue Proteobacteria, con un total de 3097 especies.

De otro lado, en la investigación del sedimento del lago Daihai, China, se encontraron los filos Proteobacteria, Actinobacteria, Firmicutes y Bacteroidetes (Li et al., 2015), el cual subraya la globalidad de la contaminación lacustre, evidenciada por la rápida diseminación de bacterias portadoras de carbapenemasas en *Klebsiella pneumoniae* y metalo- β -lactamasas en Nueva Delhi, abarcando regiones de Asia, Europa y América del Norte.

Figura 9

Árbol filogenético del metagenoma bacteriano aislado de la Bahía Interna del lago Titicaca



Nota: (A) Árbol filogenético de microorganismos que conforman en consorcio microbiano de la Bahía Interna de lago Titicaca del departamento de Puno, (B) Cladograma generada por merged_abundance, (C) anotaciones de especies uSGB y kSGB.



4.3 ÁRBOL FILOGENÉTICO DE BACTERIAS PRESENTES DE LAS MUESTRAS DE CAPTACIÓN CHIMÚ (CH)

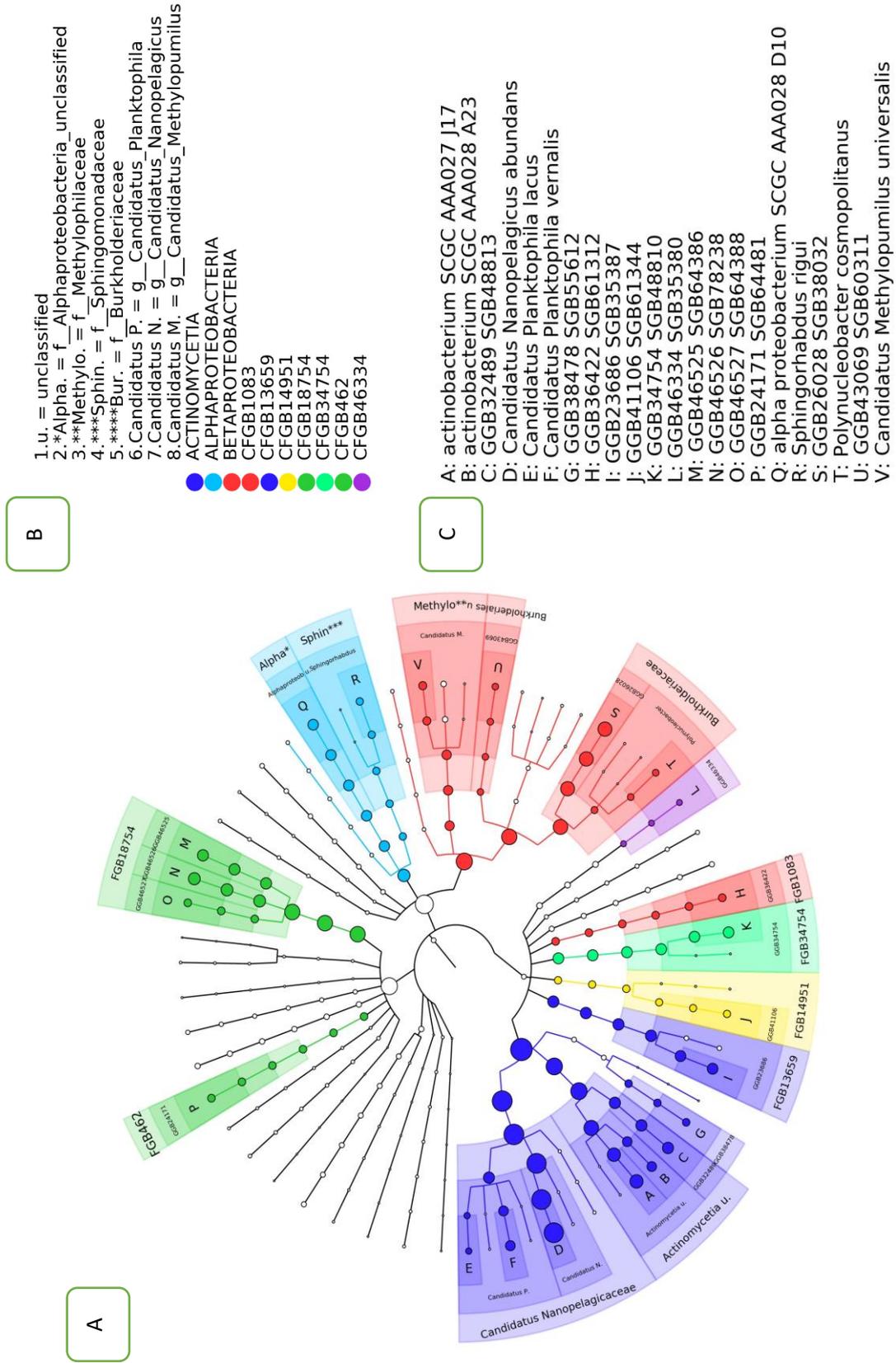
Clados descubiertas por Metaphlan

De las 5863208 lecturas procesados mediante secuenciamiento metagenómico Shotgun, donde se utilizó Metaphlan para perfilar la composición bacteriana. En las muestras de la captación Chimú, se identificaron los contenedores de genomas completos (SGB), con especies conocidas (kSGB) pertenecientes a los clados de actinomycetia, alphaproteobacteria y betaproteobacteria; se determinaron las especies *Candidatus nanopelagicus abundans*, *Actinobacterium SCGC_AAA027_J17*, y *Candidatus planktophila vernalis*, Asimismo, se encontraron especies aún por caracterizar o desconocidas (uSGB) como CFGB1083, CFGB13659, CFGB14951, CFGB18754, CFGB34754, CFGB462 y CFGB46334; género y especies *GGB46525_SGB64386* y *GGB26028_SGB38032*.

Al respecto, Quillaguamán et al. (2021), en su análisis en el lago Alalay, Cochabamba, Bolivia, revelaron que las proteobacterias dominaban la diversidad bacteriana. Mientras que Wei et al. (2023), en el lago Bamucho Tibet, identificaron 74 genomas sin caracterizar (uSGB) y 75 genomas completos (kSGB), destacando los filos de proteobacteria y actinobacteria. Estos resultados se complementan con los hallazgos de Wang & Bao (2022) en el lago Qarham, Tibet, quienes demostraron que las betaproteobacterias y halobacterias fueron predominantes en esa área.

Figura 10

Árbol filogenético del metagenoma bacteriano aislado de la Captación Chimú del lago Titicaca



Nota: (A) Árbol filogenético de microorganismos que conforman en consorcio microbiano de la captación Chimú de lago Titicaca del departamento de Puno, (B) Leyenda del cladograma generada por (merged_abundance), (C) anotaciones de especies uSGB y kSGB.



4.4 COMPARACIÓN DE ABUNDANCIA DE BACTERIAS PRESENTES EN LA BAHÍA INTERNA (BI) Y LA CAPTACIÓN CHIMÚ (CH) POR MEDIO DE MAPA DE CALOR

La mayoría de los métodos de agrupamiento se basan en emparejar los grupos más similares con comportamiento homogéneo. Se basa en la matriz de similitud entre cada par de puntos. Le permite representar visualmente las similitudes (o diferencias) entre diferentes muestras. Cuando encuentra diferencias, genera un nuevo grupo. La agrupación jerárquica utiliza métodos de distancia mínima, máxima o promedio para agrupar e iterar en etapas.

Del análisis de conglomerados jerárquico se puede visualizar que existen 4 conglomerados, una de color verde con 13 especies (con GGB23686_SGB35387 y GGB34754_SGB48810 las más relacionadas), otra de color naranja con 10 especies (*Candidatus_Methilopumilus_rimovensis* y *Polynucleobacter_cosmopolitanus* como las más relacionadas) y dos de color azul con 1 especie cada una.

Respecto a las 2 zonas de estudio, se puede mencionar lo siguiente:

Captación Chimú, comprende 4 conglomerados de los cuales el cluster con mayor abundancia es el de *Candidatus Nanopelagicus abundans*, asimismo las especies más abundantes a continuación de la mencionada serían GGB26028_SGB38032 y *actinocatreium_SCGC_AAA027_J17* del cluster color verde. En general destaca por su abundancia el cluster de color verde.

Bahía Interna, comprende también 4 conglomerados de el cluster con mayor abundancia es el de *Alpha_proteobacterium_SCGC_AAA028_D10*, de igual modo las especies más abundantes son GGB46525_SGB64386 y GGB46525_SGB64386 pertenecientes al cluster de color naranja. En general destaca por su abundancia el cluster



naranja.

Ahora, comparando los cuatro conglomerados según la Bahía Interna con la Captación Chimú, se observa que la especie presente con más abundancia para los 2 lugares es *Alpha_proteobacterium_SCGC_AAA028_D10* seguido de *Candidatus Nanopelagicus abundans*

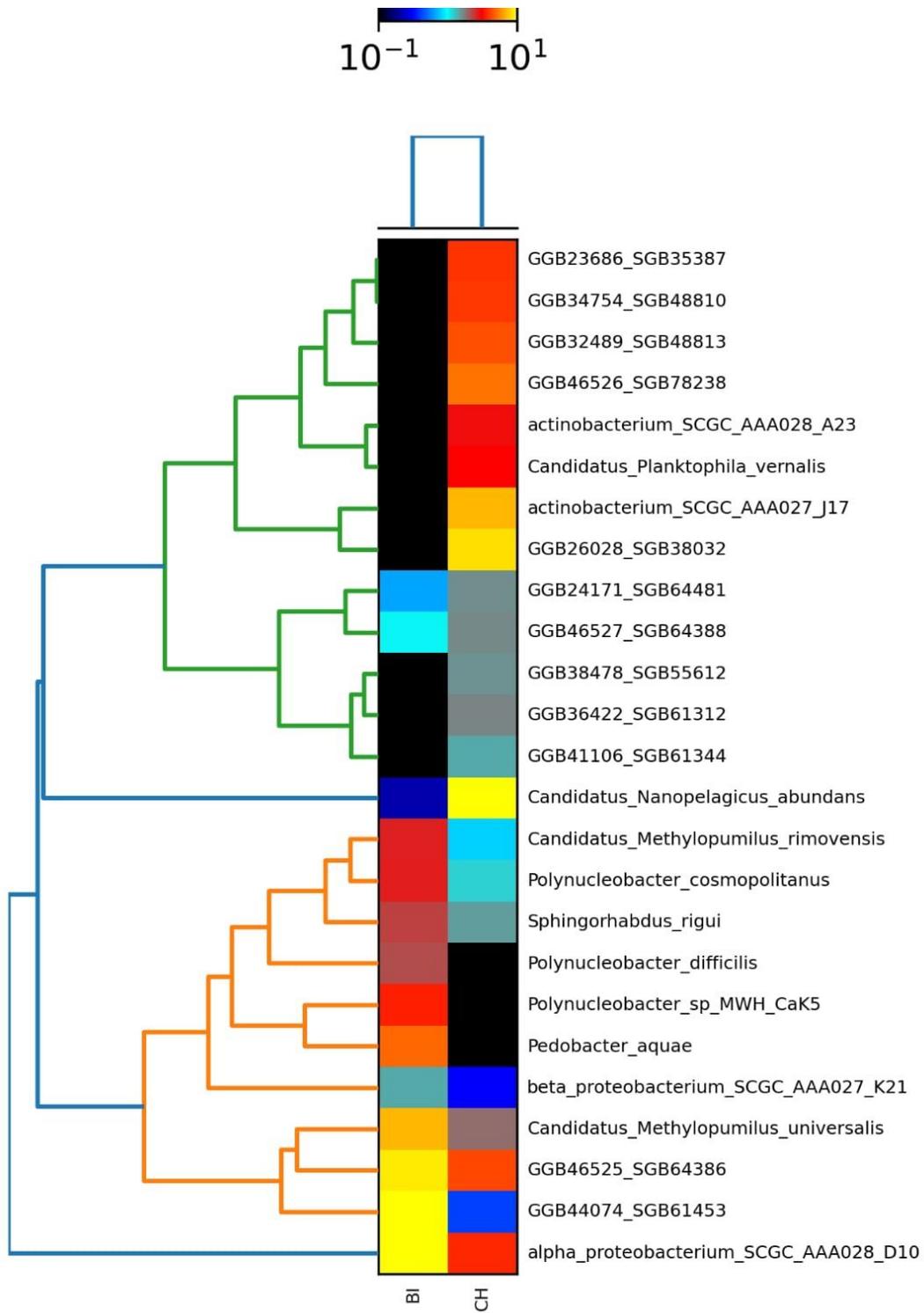
Al respecto, Xu et al., (2014) en la caracterización de suelos como microbiomas ambientales descubrieron que una serie de biomarcadores metagenómicos eran significativamente diferentes y podrían usarse como indicadores para diferenciar comunidades de suelos. Esto indicaría que tendrían comportamientos similares más el potencial metabólico entre bacterias y comunidades.

Asimismo, Santiago, (2018) coincide con los resultados obtenidos en este estudio (proteobacterias y actinobacterias) con respecto a la de microbiomas ambientales en el golfo de México, que mostraron que los filos más abundantes fueron las Proteobacterias, Thaumarchaeota, Planctomycetes, Actinobacteria y Chloroflexi. Correlativamente Du et al., (2022) nos mencionan que en el sedimento de lago Daihai en China, las bacterias preponderantes son Proteobacteria, Actinobacteria, Firmicutes y Bacteroidetes.

Figura 11

Relaciones genómicas y abundancia en el metagenoma bacteriano aislado de la Bahía

Interior y Captación Chimú del lago Titicaca





V. CONCLUSIONES

- Para el árbol filogenético de bacterias presentes en la Bahía Interior, se encontró a los grupos actinomicetia, alphaproteobacteria, betaproteobacteria, sphingobacteria y tissierellia; además, se identificaron especies aún por caracterizar o desconocidas tales como *GGB46525_SGB64386* y *GGB44074_SGB61453*.
- Del mismo modo en el árbol filogenético de la Captación Chimú, los grupos más relevantes fueron actinomycetia, alphaproteobacteria y betaproteobacteria. Asimismo, se identificó bacterias aún por caracterizar, entre los más relevantes, *GGB46525_SGB64386* y *GGB26028_SGB38032*.
- El mapa de calor permitió determinar que en la Bahía Interna y la Captación Chimú hay 4 conglomerados de microorganismos que permitió determinar la abundancia de las bacterias estando presente principalmente en ambas zonas *Alpha proteobacterium SCGC AAA028 D10*. Respecto a la Bahía Interna, *Alpha proteobacterium SCGC AAA028 D10* y *GGB46525_SGB64386* y *GGB46525_SGB64386* serían los más abundantes. Y respecto a la captación Chimú, las especies más representativas son *Candidatus nanopelagicus abundans* *GGB26028SGB38032* y *actinocatreium_SCGC_AAA027_J17*.



VI. RECOMENDACIONES

- Continuar con estudios de poblaciones y evaluar la determinación de diversidad alfa y beta de metagenomas bacterianos presentes tanto de la Bahía Interna como la captación Chimú.
- Determinar los genes de resistencia a antibióticos y las rutas metabólicas de metagenomas bacterianos de la Bahía Interna y la captación Chimú.
- Ampliar con estudios metagenómicos bacterianos en otros cuerpos de lagos de altura usando los tutoriales de GitHub.



VII. REFERENCIAS BIBLIOGRAFICAS

- Ahmad, T., Gupta, G., Sharma, A., Kaur, B., El-Sheikh, M. A., & Alyemeni, M. N. (2021). Metagenomic analysis exploring taxonomic and functional diversity of bacterial communities of a Himalayan urban fresh water lake. *PLOS ONE*, *16*(3), e0248116. <https://doi.org/10.1371/journal.pone.0248116>
- Alraddadi, B., Al-Azri, S., & Forward, K. (2013). Influence of 16S Ribosomal RNA Gene Polymerase Chain Reaction and Sequencing on Antibiotic Management of Bone and Joint Infections. *Canadian Journal of Infectious Diseases and Medical Microbiology*, *24*(2), 85–88. <https://doi.org/10.1155/2013/747145>
- Autoridad Nacional del Agua. (2016). *Protocolo nacional para el monitoreo de la calidad de los recursos hídricos superficiales*. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ana.gob.pe/sites/default/files/normatividad/files/r.j._010-2016-ana_0.pdf
- Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., & Segata, N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *ELife*, *10*. <https://doi.org/10.7554/eLife.65088>
- Belasco Moreno Julian. (2022). *El microbioma ambiental: Una herramienta para evaluar el impacto de los contaminantes clasicos y emergentes en areas costeras*. <https://www.csic.es/es/investigacion/proyectos-de-investigacion/el-microbioma-ambiental-una-herramienta-para-evaluar-el>
- Beltrán Farfán, D. F., Palomino Calli, R. P., Moreno Terrazas, E. G., Peralta, C. G., & Montesinos-Tubée, D. B. (2015). Calidad de agua de la bahía interior de Puno, lago Titicaca durante el verano del 2011. *Revista Peruana de Biología*, *22*(3), 335–340. <https://doi.org/10.15381/rpb.v22i3.11440>
- Beltran, M. (2017). *Microbiota autóctona* (2nd ed., Vol. 31).



- Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., Manghi, P., Dubois, L., Huang, K. D., Thomas, A. M., Nickols, W. A., Piccinno, G., Piperni, E., Punčochář, M., Valles-Colomer, M., Tett, A., Giordano, F., Davies, R., Wolf, J., ... Segata, N. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology*, 41(11), 1633–1644. <https://doi.org/10.1038/s41587-023-01688-w>
- Chang, J., Chapman, B., Friedberg, I., Hamelryck, T., De Hoon, M., Cock, P., Antao, T., Talevich, E., & Wilczyński, B. W. (2023). *Biopython Tutorial and Cookbook*.
- Chen, K., & Pachter, L. (2005). Bioinformatics for Whole-Genome Shotgun Sequencing of Microbial Communities. *PLoS Computational Biology*, 1(2), e24. <https://doi.org/10.1371/journal.pcbi.0010024>
- Costa Aponte, F., Morán Flores, G., & Organización Y Funciones Del, L. DE. (2018). *Instituto Nacional de Estadística e Informática*.
- Departamento de control de calidad del Banco Nacional de ADN. (2020). *programa de control de calidad de muestras de ADN y ARN*.
- Díaz, P. (2023, June 29). *Espectrofotómetro: ¿Qué es y para qué sirve?* Tecnología Del Plástico.
- Dirección General de Salud Ambiental del Ministerio de Salud. (2011). *Reglamento de la Calidad del Agua para Consumo Humano* . 2011.
- Du, C., Yang, F., Li, X., Liao, H., Li, Z., Gao, J., & Zhang, L. (2022). Metagenomic analysis of microbial community structure and distribution of resistance genes in Daihai Lake, China. *Environmental Pollution*, 302, 119065. <https://doi.org/10.1016/j.envpol.2022.119065>
- Fernández, M. S., María, F. B., Damborenea, C., Dellapé, P. M., & Gallardo, F. E. (2013). *Introducción a la Taxonomía Manual de ejercitaciones facultad de ciencias naturales y museo Libros de Cátedra*. www.editorial.unlp.edu.ar



- Forero Pineda, N., Marín - Suárez, J., Forero- Ulloa, F. E., & Gómez-Palacio, A. (2021). Extracción de ADN bacteriano a partir de cuerpos de agua de uso agrícola. *Ciencia y Agricultura*, 18(1), 36–45. <https://doi.org/10.19053/01228420.v18.n1.2021.11703>
- Gao, B., Chi, L., Zhu, Y., Shi, X., Tu, P., Li, B., Yin, J., Gao, N., Shen, W., & Schnabl, B. (2021). An Introduction to Next Generation Sequencing Bioinformatic Analysis in Gut Microbiome Studies. *Biomolecules*, 11(4), 530. <https://doi.org/10.3390/biom11040530>
- Gupta, N., & Verma, V. K. (2019). *Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality* (pp. 313–341). https://doi.org/10.1007/978-981-13-8844-6_15
- Home “BioBakery.” (n.d.). *BioBakery*.
- Iranzo, M. (2022, May 11). *Cuantificación y análisis de la integridad de los ácidos nucleicos ADN y ARN*. BIOTEC Y CIENCIA.
- ISC-III. (2011). *Red Nacional de Biobancos*. www.redbiobancos.es
- Lee, P. Y., Costumbrado, J., Hsu, C.-Y., & Kim, Y. H. (2012). Agarose Gel Electrophoresis for the Separation of DNA Fragments. *Journal of Visualized Experiments*, 62. <https://doi.org/10.3791/3923>
- Lindgreen, S., Adair, K. L., & Gardner, P. P. (2016). An evaluation of the accuracy and speed of metagenome analysis tools. *Scientific Reports*, 6(1), 19233. <https://doi.org/10.1038/srep19233>
- Mata López, R., Nitxin Castañeda Sortibrán, A., Alejandro García Trejo, E., Blanca Isabel Honey Escandón, M., Book, M., Mata, R., & Alejandro García-Trejo, E. (2017). *"Sistemática Molecular y Bioinformática Guía práctica*. <https://www.researchgate.net/publication/325707985>
- Medrano, M. ;, Mamani, A. ;, Muñoz, E., Díaz, R. ;, & Medrano, E. (2020). *Operatividad de las Plantas de Tratamiento de Aguas Residuales Domésticas circunlacustres al lago Titicaca-Sector Perú y el marco legal en defensa de los ecosistemas Operation of the Domestic Wastewater Treatment Plants*



circumlacteustres to Lake Titicaca-Sector Peru and the framework in defense of ecosystems. <http://revistas.uap.edu.pe/ojs/index.php/CYD/index>

- Meslier, V., Quinquis, B., Da Silva, K., Plaza Oñate, F., Pons, N., Roume, H., Podar, M., & Almeida, M. (2022). Benchmarking second and third-generation sequencing platforms for microbial metagenomics. *Scientific Data*, 9(1), 694. <https://doi.org/10.1038/s41597-022-01762-z>
- O’Leary, N. A., Wright, M. W., Brister, J. R., Ciuffo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., ... Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1), D733-45. <https://doi.org/10.1093/nar/gkv1189>
- Pasolli, E., Truong, D. T., Malik, F., Waldron, L., & Segata, N. (2016). Machine Learning Meta-analysis of Large Metagenomic Datasets: Tools and Biological Insights. *PLOS Computational Biology*, 12(7), e1004977. <https://doi.org/10.1371/journal.pcbi.1004977>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(Database issue), D590-6. <https://doi.org/10.1093/nar/gks1219>
- Quillaguamán, J., Guzmán, D., Campero, M., Hoepfner, C., Relos, L., Mendieta, D., Higdon, S. M., Eid, D., & Fernández, C. E. (2021). The microbiome of a polluted urban lake harbors pathogens with diverse antimicrobial resistance and virulence genes. *Environmental Pollution*, 273, 116488. <https://doi.org/10.1016/j.envpol.2021.116488>
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9), 833–844. <https://doi.org/10.1038/nbt.3935>



- Rodrigue, S., Materna, A. C., Timberlake, S. C., Blackburn, M. C., Malmstrom, R. R., Alm, E. J., & Chisholm, S. W. (2010). Unlocking Short Read Sequencing for Metagenomics. *PLoS ONE*, 5(7), e11840. <https://doi.org/10.1371/journal.pone.0011840>
- Rubio, S., Pacheco-Orozco, R. A., Gómez, A. M., Perdomo, S., & García-Robles, R. (2020). Secuenciación de nueva generación (NGS) de ADN: presente y futuro en la práctica clínica. *Universitas Médica*, 61(2). <https://doi.org/10.11144/Javeriana.umed61-2.sngs>
- Santiago, C. B. (2018). *Caracterización de las comunidades microbianas presentes en los sedimentos de Perdido y Coatzacoalcos del Golfo de México mediante análisis metagenómicos.*
- Sharpton, T. J. (2014a). An introduction to the analysis of shotgun metagenomic data. *Frontiers in Plant Science*, 5. <https://doi.org/10.3389/fpls.2014.00209>
- Sharpton, T. J. (2014b). An introduction to the analysis of shotgun metagenomic data. *Frontiers in Plant Science*, 5. <https://doi.org/10.3389/fpls.2014.00209>
- Staley, C., & Sadowsky, M. J. (2018). Practical considerations for sampling and data analysis in contemporary metagenomics-based environmental studies. *Journal of Microbiological Methods*, 154, 14–18. <https://doi.org/10.1016/j.mimet.2018.09.020>
- Stanish, C., & Levine, A. (2011). War and early state formation in the northern Titicaca Basin, Peru. *Proceedings of the National Academy of Sciences*, 108(34), 13901–13906. <https://doi.org/10.1073/pnas.1110176108>
- Usyk, M., Peters, B. A., Karthikeyan, S., McDonald, D., Sollecito, C. C., Vazquez-Baeza, Y., Shaffer, J. P., Gellman, M. D., Talavera, G. A., Daviglius, M. L., Thyagarajan, B., Knight, R., Qi, Q., Kaplan, R., & Burk, R. D. (2023). Comprehensive evaluation of shotgun metagenomics, amplicon sequencing, and harmonization of these platforms for epidemiological studies. *Cell Reports Methods*, 3(1). <https://doi.org/10.1016/j.crmeth.2022.100391>



- Wang, Y., & Bao, G. (2022). Diversity of prokaryotic microorganisms in alkaline saline soil of the Qarhan Salt Lake area in the Qinghai-Tibet Plateau. *Scientific Reports*, *12*(1), 3365. <https://doi.org/10.1038/s41598-022-07311-3>
- Weging, S., Gogol-Döring, A., & Grosse, I. (2021). Taxonomic analysis of metagenomic data with kASA. *Nucleic Acids Research*, *49*(12), e68–e68. <https://doi.org/10.1093/nar/gkab200>
- Wei, C., Sun, D., Yuan, W., Li, L., Dai, C., Chen, Z., Zeng, X., Wang, S., Zhang, Y., Jiang, S., Wu, Z., Liu, D., Jiang, L., & Peng, S. (2023). Metagenomics revealing molecular profiles of microbial community structure and metabolic capacity in Bamucuo lake, Tibet. *Environmental Research*, *217*, 114847. <https://doi.org/10.1016/j.envres.2022.114847>
- Xu, Z., Hansen, M. A., Hansen, L. H., Jacquiod, S., & Sørensen, S. J. (2014). Bioinformatic Approaches Reveal Metagenomic Characterization of Soil Microbial Community. *PLoS ONE*, *9*(4), e93445. <https://doi.org/10.1371/journal.pone.0093445>
- Zubairi, O., & Weber, F. (2018). The Linux/Unix operating system. In *Introduction to Computational Physics for Undergraduates* (pp. 1-1-1–38). Morgan & Claypool Publishers. <https://doi.org/10.1088/978-1-6817-4896-2ch1>



ANEXOS



ANEXO 1. Solicitud de acceso y uso del laboratorio de Biología Molecular Megalaboratorio de la Universidad Nacional del Altiplano.

UNIVERSIDAD NACIONAL DEL ALTIPLANO PUNO

Dra. MARÍA TRINIDAD ROMERO TORRES
JEFE DEL LABORATORIO DE BIOLOGIA MOLECULAR
MEGALABORATORIO – UNA -PUNO

AUNTO. Solicita acceso y uso de laboratorio de Biología Molecular Megalaboratorio de la
Universidad Nacional del Altiplano

Tengo el agrado de dirigirme a Ud. Por intermedio del presente y poner en su conocimiento y al mismo tiempo solicitar que se autorice la elaboración y ejecución del proyecto de tesis "PERFIL TAXONÓMICO DE MICROBIOMAS BACTERIANOS CON LA SECUENCIACIÓN GENÓMICA SHOTGUN PRESENTES EN LA CAPTACIÓN CHIMU Y LA BAHÍA INTERNA DEL LAGO TITICACA". Con el fin de obtener el Grado de Licenciado en Biología El trabajo se realizará de lunes a viernes de 9AM a las 1PM desde el 3 de abril del 2023 hasta 31 de Julio del 2023, comprometiéndome a cumplir y hacer cumplir las políticas y normas establecidas en el reglamento del laboratorio.

POR LO EXPUESTO

Se le ruega a usted que nuestra solicitud sea aprobada

PUNO, 03 de abril del 202

Abel Vidal Mayta Mendizábal
DNI: 74484205

Dra. María Trinidad Romero Torres
Directora de Departamento Académico
F. CC. BB - UNAP



ANEXO 2. Bacterias descubiertas con el análisis bioinformático Metaphlan4 de la Bahía Interna.

```
#mpa_v0ct22_CHOCOfhlanSGB_202212
#/home/meglab/centro/anaconda3/envs/mpa/bin/metaphlan 20008D-10-01_S204_L003_R1_001.fastq,20008D-10-01_S204_L003_R2_001.fastq --bowtie2out metagenome.bowtie2.bz2 --nproc 4 -
-input_type fastq -o profiled_metagenome.txt
#1287804 reads processed
#SampleID Metaphlan_Analysis
#clade_name NCBI_tax_id relative_abundance additional_species
k_Bacteria 2 100.0
k_Bacteria|p_Proteobacteria 2|1224 70.10081
k_Bacteria|p_Bacteroidetes 2|976 15.76992
k_Bacteria|p_Actinobacteria 2|201174 12.3499
k_Bacteria|p_Verrucomicrobia 2|74201 1.27027
k_Bacteria|p_Firmicutes 2|1239 0.5091
k_Bacteria|p_Proteobacteria|c_Alphaproteobacteria 2|1224|28211 49.74166
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria 2|1224|28216 20.35915
k_Bacteria|p_Actinobacteria|c_CFBG41185 2|201174| 10.32336
k_Bacteria|p_Bacteroidetes|c_CFBG18754 2|976| 10.07537
k_Bacteria|p_Bacteroidetes|c_Sphingobacteriia 2|976|117747 5.02433
k_Bacteria|p_Actinobacteria|c_Actinomycetia 2|201174|1760 1.65817
k_Bacteria|p_Verrucomicrobia|c_CFBG8395 2|74201| 1.27027
k_Bacteria|p_Bacteroidetes|c_CFBG462 2|976| 0.67023
k_Bacteria|p_Firmicutes|c_Tissierellia 2|1239|1737404 0.5091
k_Bacteria|p_Actinobacteria|c_CFBG13659 2|201174| 0.36837
k_Bacteria|p_Proteobacteria|c_Alphaproteobacteria|c_Alphaproteobacteria_unclassified 2|1224|28211| 46.82273
k_Bacteria|p_Actinobacteria|c_CFBG41185|c_CFBG41185 2|201174| 10.32336
k_Bacteria|p_Bacteroidetes|c_CFBG18754|c_CFBG18754 2|976| 10.07537
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Nitrosomonadales 2|1224|28216|32003 9.92375
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales 2|1224|28216|80840 8.62749
k_Bacteria|p_Bacteroidetes|c_Sphingobacteriia|c_Sphingobacteriales 2|976|117747|200666 5.02433
k_Bacteria|p_Proteobacteria|c_Alphaproteobacteria|c_Sphingomonadales 2|1224|28211|204457 2.91893
k_Bacteria|p_Actinobacteria|c_Actinomycetia|c_Candidatus_Nanopelagicales 2|201174|1760|2039638 1.65817
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Betaproteobacteria_unclassified 2|1224|28216| 1.4622
k_Bacteria|p_Verrucomicrobia|c_CFBG8395|c_CFBG8395 2|74201| 1.27027
k_Bacteria|p_Bacteroidetes|c_CFBG462|c_CFBG462 2|976| 0.67023
k_Bacteria|p_Firmicutes|c_Tissierellia|c_Tissierellales 2|1239|1737404|1737405 0.5091
k_Bacteria|p_Actinobacteria|c_CFBG13659|c_CFBG13659 2|201174| 0.36837
k_Bacteria|p_Proteobacteria|c_Alphaproteobacteria|c_Alphaproteobacteria_unclassified|f_Alphaproteobacteria_unclassified 2|1224|28211| 46.82273
k_Bacteria|p_Actinobacteria|c_CFBG41185|f_CFBG41185 2|201174| 10.32336
k_Bacteria|p_Bacteroidetes|c_CFBG18754|c_CFBG18754|f_CFBG18754 2|976| 10.07537
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Nitrosomonadales|f_Methylophilaceae 2|1224|28216|32003|32011 9.92375
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Burkholderiaceae 2|1224|28216|80840|119060 8.62749
k_Bacteria|p_Bacteroidetes|c_Sphingobacteriia|c_Sphingobacteriales|f_Sphingobacteriaceae 2|976|117747|200666|84566 5.02433
k_Bacteria|p_Bacteroidetes|c_Sphingobacteriia|c_Sphingobacteriales|f_Sphingobacteriaceae 2|976|117747|200666|84566 5.02433
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Sphingomonadales|f_Sphingomonadales 2|1224|28211|204457|41297 2.91893
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Betaproteobacteria_unclassified|f_Betaproteobacteria_unclassified 2|1224|28216| 1.4622
k_Bacteria|p_Verrucomicrobia|c_CFBG8395|c_CFBG8395|f_CFBG8395 2|74201| 1.27027
k_Bacteria|p_Bacteroidetes|c_CFBG462|f_CFBG462 2|976| 0.67023
k_Bacteria|p_Firmicutes|c_Tissierellia|c_Tissierelliales|f_Peptoniphilaceae 2|1239|1737404|1737405|1570339 0.5091
k_Bacteria|p_Actinobacteria|c_CFBG13659|c_CFBG13659|f_CFBG13659 2|201174| 0.36837
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Comonadaceae 2|1224|28216|80840|80864 0.34572
k_Bacteria|p_Proteobacteria|c_Alphaproteobacteria|c_Alphaproteobacteria_unclassified|f_Alphaproteobacteria_unclassified|g_Alphaproteobacteria_unclassified 2|1224|28211| 46.82273
k_Bacteria|p_Actinobacteria|c_CFBG41185|c_CFBG41185|f_CFBG41185|g_GGB44074 2|201174| 10.32336
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Nitrosomonadales|f_Methylophilaceae|g_Candidatus_Methylophilus 2|1224|28216|32003|32011|1679002 9.92375
k_Bacteria|p_Bacteroidetes|c_CFBG18754|c_CFBG18754|f_CFBG18754|g_GGB46525 2|976| 9.03199
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter 2|1224|28216|80840|119060|44013 8.62749
k_Bacteria|p_Bacteroidetes|c_Sphingobacteriia|c_Sphingobacteriales|f_Sphingobacteriaceae|g_Pedobacter 2|976|117747|200666|84566|84567 5.02433
k_Bacteria|p_Proteobacteria|c_Alphaproteobacteria|c_Sphingomonadales|f_Sphingomonadales|g_Sphingorhabdus 2|1224|28211|204457|41297|1434046 2.91893
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Betaproteobacteria_unclassified|f_Betaproteobacteria_unclassified|g_Betaproteobacteria_unclassified 2|1224|28216| 1.4622
k_Bacteria|p_Actinobacteria|c_Actinomycetia|c_Candidatus_Nanopelagicales|f_Candidatus_Nanopelagicaceae|g_Candidatus_Flanktophila 2|201174|1760|2039638|2162846|622681 1.33847
k_Bacteria|p_Verrucomicrobia|c_CFBG8395|c_CFBG8395|f_CFBG8395|g_GGB25672 2|74201| 1.27027
k_Bacteria|p_Bacteroidetes|c_CFBG18754|c_CFBG18754|f_CFBG18754|g_GGB46527 2|976| 1.04338
k_Bacteria|p_Bacteroidetes|c_CFBG462|c_CFBG462|f_CFBG462|g_GGB24171 2|976| 0.67023
k_Bacteria|p_Firmicutes|c_Tissierellia|c_Tissierelliales|f_Peptoniphilaceae|g_Peptoniphilaceae_unclassified 2|1239|1737404|1737405|1570339| 0.5091
k_Bacteria|p_Actinobacteria|c_CFBG13659|c_CFBG13659|f_CFBG13659|g_GGB23686 2|201174| 0.36837
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Comonadaceae|g_Limnochitans 2|1224|28216|80840|80864|665874 0.34572
k_Bacteria|p_Actinobacteria|c_Actinomycetia|c_Candidatus_Nanopelagicales|f_Candidatus_Nanopelagicaceae|g_Candidatus_Nanopelagicus 2|201174|1760|2039638|2162846|2039639 0.21749
k_Bacteria|p_Actinobacteria|c_Actinomycetia|c_Candidatus_Nanopelagicales|f_Candidatus_Nanopelagicaceae|g_GGB46331 2|201174|1760|2039638|2162846| 0.10221
k_Bacteria|p_Proteobacteria|c_Alphaproteobacteria|c_Alphaproteobacteria_unclassified|f_Alphaproteobacteria_unclassified|g_Alphaproteobacteria_unclassified|s_alpha_protobacterium_SCG_AAA028_D10 2|1224|28211| 938641 46.82273
k_Bacteria|p_Actinobacteria|c_CFBG41185|c_CFBG41185|f_CFBG41185|g_GGB44074|s_GGB44074_SGB61453 2|201174| 10.32336
k_Bacteria|p_Bacteroidetes|c_CFBG18754|c_CFBG18754|f_CFBG18754|g_GGB46525|s_GGB46525_SGB64386 2|976| 9.03199
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Nitrosomonadales|f_Methylophilaceae|g_Candidatus_Methylophilus|s_Candidatus_Methylophilus_universalis 2|1224|28216|32003|32011|1679002|2589536 7.20997
k_Bacteria|p_Bacteroidetes|c_Sphingobacteriia|c_Sphingobacteriales|f_Sphingobacteriaceae|g_Pedobacter_aquae 2|976|117747|200666|84566|84567|2605747 5.02433
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter|s_Polynucleobacter_sp_MNH_CaK5 2|1224|28216|80840|119060|44013|2689107 3.61266
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter|s_Polynucleobacter_cosmopolitanus 2|1224|28216|80840|119060|44013|351345 2.76979
```



k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter|s_Polynucleobacter_difficilis
2|1224|28216|80840|119060|44013|56054| 2.24504

k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Betaproteobacteria|c_Betaproteobacteria_unclassified|f_Betaproteobacteria_unclassified|g_Betaproteobacteria_unclassified|s_beta_protectobacterium_SGCC_AAA027_K21| 2|1224|28216|11938785| 1.4622

k_Bacteria|p_Actinobacteria|c_Actinomycetia|c_Candidatus_Nanopelagiales|f_Candidatus_Nanopelagiales|g_Candidatus_Planktophila|s_Candidatus_Planktophila_lacus
2|201174|1760|2039638|2162846|622681|1884913| 1.33847

k_Bacteria|p_Verrucomicrobia|c_CFBG8395|f_CFBG8395|g_GGB25672|s_GGB25672_SGB61746| 2|174201|11111| 1.27027

k_Bacteria|p_Bacteroidetes|c_CFBG18754|f_CFBG18754|g_GGB46527|s_GGB46527_SGB64388| 2|1976|11111| 1.04338

k_Bacteria|p_Bacteroidetes|c_CFBG462|f_CFBG462|g_GGB24171|s_GGB24171_SGB64481| 2|1976|11111| 0.67023

k_Bacteria|p_Proteobacteria|c_Alphaproteobacteria|c_Sphingomonadales|f_Sphingomonadales|g_Sphingorhabdus|s_Sphingorhabdus_wooponensis
2|1224|28211|204457|41297|1434046|940136| 0.58146

k_Bacteria|p_Firmicutes|c_Tissierelliales|f_Tissierelliales|g_Peptoniphilaceae|g_Peptoniphilaceae_unclassified|s_Peptoniphilaceae_unclassified_SGB38271
2|1239|1737404|1737405|1570339| 0.5091

k_Bacteria|p_Actinobacteria|c_CFBG13659|f_CFBG13659|g_GGB23686|s_GGB23686_SGB82209| 2|201174|11111| 0.36837

k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Comamonadaceae|g_Limnochabittans|s_Limnochabittans_sp_Ba153| 2|1224|28216|80840|80864|665874|1977910| 0.21901

k_Bacteria|p_Actinobacteria|c_Actinomycetia|c_Candidatus_Nanopelagiales|f_Candidatus_Nanopelagiales|g_Candidatus_Nanopelagicus|s_Candidatus_Nanopelagicus_abundans
2|201174|1760|2039638|2162846|2039639|1884916| 0.21749

k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Comamonadaceae|g_Limnochabittans_sp_2KL_27| 2|1224|28216|80840|80864|665874|1100705| 0.12671

k_Bacteria|p_Actinobacteria|c_Actinomycetia|c_Candidatus_Nanopelagiales|f_Candidatus_Nanopelagiales|g_GGB46331|s_GGB46331_SGB64109| 2|201174|1760|2039638|2162846| 0.10221

k_Bacteria|p_Proteobacteria|c_Alphaproteobacteria|c_Alphaproteobacteria_unclassified|f_Alphaproteobacteria_unclassified|g_Alphaproteobacteria_unclassified|s_alpha_protectobacterium_SGCC_AAA028_D10|t_SGB50895_group| 2|1224|28211|11938641| 46.82273

k_Bacteria|p_Proteobacteria|c_Alphaproteobacteria|c_Alphaproteobacteria_unclassified|f_Alphaproteobacteria_unclassified|g_Alphaproteobacteria_unclassified|s_alpha_protectobacterium_SGCC_AAA027_C06|k_Bacteria|p_Proteobacteria|c_Alphaproteobacteria|c_Alphaproteobacteria_unclassified|f_Alphaproteobacteria_unclassified|g_Alphaproteobacteria_unclassified|s_alpha_protectobacterium_SGCC_AAA027_D15
2|201174|1760|2039638|2162846|622681|1884913| 1.33847

k_Bacteria|p_Actinobacteria|c_CFBG41185|f_CFBG41185|g_GGB44074|s_GGB44074_SGB61453|t_SGB61453| 2|201174|11111| 10.32336

k_Bacteria|p_Bacteroidetes|c_CFBG18754|f_CFBG18754|g_GGB46525|s_GGB46525_SGB64386|t_SGB64386| 2|1976|11111| 9.03199

k_Bacteria|p_Bacteroidetes|c_Betaproteobacteria|c_Nitrospomonadales|f_Candidatus_Methylophilum|s_Candidatus_Methylophilum_universale|t_SGB38034
2|1224|28216|32003|32011|1679002|2588536| 7.20997

k_Bacteria|p_Bacteroidetes|c_Sphingobacteriales|c_Sphingobacteriales|f_Sphingobacteriales|g_Pedobacter|s_Pedobacter_aquae|t_SGB3123
2|1976|117747|200666|84566|84567|2605747| 5.02433

k_Bacteria|p_Bacteroidetes|c_Sphingobacteriales|c_Sphingobacteriales|f_Sphingobacteriales|g_Pedobacter|s_Pedobacter_glucoacidilyticus
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter|s_Polynucleobacter_sp_MWH_CaK5|t_SGB94461
2|1224|28216|80840|119060|44013|2689107| 3.61266

k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter|s_Polynucleobacter_cosmopolitanus|t_SGB13556
2|1224|28216|80840|119060|44013|351345| 2.76979

k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Nitrospomonadales|f_Methylophilaceae|g_Candidatus_Methylophilum|s_Candidatus_Methylophilum_rimovensis|t_SGB80878
2|1224|28216|32003|32011|1679002|2588535| 2.71377

k_Bacteria|p_Bacteroidetes|c_Sphingobacteriales|c_Sphingobacteriales|f_Sphingobacteriales|g_Pedobacter|s_Pedobacter_aquae|t_SGB3123
2|1976|117747|200666|84566|84567|2605747| 5.02433

k_Bacteria|p_Bacteroidetes|c_Sphingobacteriales|c_Sphingobacteriales|f_Sphingobacteriales|g_Pedobacter|s_Pedobacter_glucoacidilyticus
k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter|s_Polynucleobacter_sp_MWH_CaK5|t_SGB94461
2|1224|28216|80840|119060|44013|2689107| 3.61266

k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter|s_Polynucleobacter_cosmopolitanus|t_SGB13556
2|1224|28216|80840|119060|44013|351345| 2.76979

k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Nitrospomonadales|f_Methylophilaceae|g_Candidatus_Methylophilum|s_Candidatus_Methylophilum_rimovensis|t_SGB80878
2|1224|28216|32003|32011|1679002|2588535| 2.71377

k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter|s_Polynucleobacter_difficilis|t_SGB30629
2|1224|28211|204457|41297|1434046|1282858| 2.33748

k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter|s_Polynucleobacter_difficilis|t_SGB30629
2|1224|28216|80840|119060|44013|56054| 2.24504

k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Betaproteobacteria_unclassified|f_Betaproteobacteria_unclassified|g_Betaproteobacteria_unclassified|s_beta_protectobacterium_SGCC_AAA027_K21|t_SGB57451| 2|1224|28216|11938785| 1.4622

k_Bacteria|p_Actinobacteria|c_Actinomycetia|c_Candidatus_Nanopelagiales|f_Candidatus_Nanopelagiales|g_Candidatus_Planktophila|s_Candidatus_Planktophila_lacus|t_SGB5706| 2|201174|1760|2039638|2162846|622681|1884913| 1.33847

k_Bacteria|p_Verrucomicrobia|c_CFBG8395|f_CFBG8395|g_GGB25672|s_GGB25672_SGB61746|t_SGB61746| 2|174201|11111| 1.27027

k_Bacteria|p_Bacteroidetes|c_CFBG18754|f_CFBG18754|g_GGB46527|s_GGB46527_SGB64388|t_SGB64388| 2|1976|11111| 1.04338

k_Bacteria|p_Bacteroidetes|c_CFBG462|f_CFBG462|g_GGB24171|s_GGB24171_SGB64481|t_SGB64481| 2|1976|11111| 0.67023

k_Bacteria|p_Proteobacteria|c_Alphaproteobacteria|c_Sphingomonadales|f_Sphingomonadales|g_Sphingorhabdus|s_Sphingorhabdus_wooponensis|t_SGB31754
2|1224|28211|204457|41297|1434046|940136| 0.58146

k_Bacteria|p_Firmicutes|c_Tissierelliales|f_Tissierelliales|g_Peptoniphilaceae|g_Peptoniphilaceae_unclassified|s_Peptoniphilaceae_unclassified_SGB38271|t_SGB38271
2|1239|1737404|1737405|1570339| 0.5091

k_Bacteria|p_Actinobacteria|c_CFBG13659|f_CFBG13659|g_GGB23686|s_GGB23686_SGB82209|t_SGB82209| 2|201174|11111| 0.36837

k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Comamonadaceae|g_Limnochabittans|s_Limnochabittans_sp_Ba153|t_SGB22499
2|1224|28216|80840|80864|665874|1977910| 0.21901

k_Bacteria|p_Actinobacteria|c_Actinomycetia|c_Candidatus_Nanopelagiales|f_Candidatus_Nanopelagiales|g_Candidatus_Nanopelagicus|s_Candidatus_Nanopelagicus_abundans|t_SGB5714| 2|201174|1760|2039638|2162846|2039639|1884916| 0.21749

k_Bacteria|p_Proteobacteria|c_Betaproteobacteria|c_Burkholderiales|f_Comamonadaceae|g_Limnochabittans|s_Limnochabittans_sp_2KL_27|t_SGB32529
2|1224|28216|80840|80864|665874|1100705| 0.12671

k_Bacteria|p_Actinobacteria|c_Actinomycetia|c_Candidatus_Nanopelagiales|f_Candidatus_Nanopelagiales|g_GGB46331|s_GGB46331_SGB64109|t_SGB64109| 2|201174|1760|2039638|2162846| 0.10221



ANEXO 3. Bacterias descubiertas con el análisis bioinformático Metaphlan4 de la Captación Chimu.

```
#mpa_vOct22_CHOCOPhlanSGB_202212
#/#home/megalaboratorio/anaconda3/envs/mpa/bin/metaphlan 20008D-10-02_S205_L003_R1_001.fastq,20008D-10-02_S205_L003_R2_001.fastq --bowtie2out metagenome.bowtie2.bs2 --nproc 4 -
-input_type fastq --o profiled_metagenome.txt
#11726416 reads processed
#SampleID Metaphlan_Analysis
#clade_name NCBI_tax_id relative_abundance additional_species
k__Bacteria 2 100.0
k__Bacteria|p__Actinobacteria 2|201174 62.74572
k__Bacteria|p__Proteobacteria 2|1224 21.30775
k__Bacteria|p__Bacteroidetes 2|976 14.94072
k__Bacteria|p__Verrucomicrobia 2|74201 0.76333
k__Bacteria|p__Cyanobacteria 2|1117 0.18922
k__Bacteria|p__Armatimonadetes 2|67819 0.03228
k__Bacteria|p__Planctomycetes 2|203682 0.02097
k__Bacteria|p__Actinobacteria|c__Actinomycetia 2|201174|1760 48.80394
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria 2|1224|28216 14.61128
k__Bacteria|p__Bacteroidetes|c__CFGB18754 2|976| 11.33284
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria 2|1224|28211 5.66099
k__Bacteria|p__Actinobacteria|c__CFGB13659 2|201174| 4.481
k__Bacteria|p__Actinobacteria|c__CFGB34754 2|201174| 4.04941
k__Bacteria|p__Actinobacteria|c__CFGB1083 2|201174| 1.73407
k__Bacteria|p__Bacteroidetes|c__CFGB462 2|976| 1.6796
k__Bacteria|p__Actinobacteria|c__CFGB14951 2|201174| 1.54503
k__Bacteria|p__Actinobacteria|c__CFGB46334 2|201174| 0.91335
k__Bacteria|p__Bacteroidetes|c__CFGB40560 2|976| 0.76844
k__Bacteria|p__Actinobacteria|c__CFGB13153 2|201174| 0.69629
k__Bacteria|p__Verrucomicrobia|c__CFGB8395 2|74201| 0.66264
k__Bacteria|p__Proteobacteria|c__CFGB14846 2|1224| 0.55321
k__Bacteria|p__Bacteroidetes|c__CFGB35715 2|976| 0.45045
k__Bacteria|p__Actinobacteria|c__CFGB41185 2|201174| 0.42263
k__Bacteria|p__Proteobacteria|c__CFGB8596 2|1224| 0.30905
k__Bacteria|p__Bacteroidetes|c__CFGB7657 2|976| 0.29904
k__Bacteria|p__Cyanobacteria|c__Cyanobacteria_unclassified 2|1117| 0.18922
k__Bacteria|p__Bacteroidetes|c__CFGB7628 2|976| 0.1775
k__Bacteria|p__Proteobacteria|c__CFGB42731 2|1224| 0.17321
k__Bacteria|p__Bacteroidetes|c__Flavobacteriia 2|976|117743 0.14651
k__Bacteria|p__Verrucomicrobia|c__CFGB8396 2|74201| 0.10069
k__Bacteria|p__Bacteroidetes|c__CFGB32421 2|976| 0.06272
k__Bacteria|p__Armatimonadetes|c__CFGB55302 2|67819| 0.03228
k__Bacteria|p__Bacteroidetes|c__Cytophagia 2|976|1768503 0.02363

k__Bacteria|p__Planctomycetes|c__CFGB24506 2|203682| 0.02097
k__Bacteria|p__Actinobacteria|c__Actinomycetia|c__Candidatus_Nanopelagicales 2|201174|1760|2039638 32.59065
k__Bacteria|p__Actinobacteria|c__Actinomycetia|c__Actinomycetia_unclassified 2|201174|1760| 14.66375
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria|c__Burkholderiales 2|1224|28216|80840 11.49456
k__Bacteria|p__Actinobacteria|c__CFGB13659|c__CFGB13659 2|201174| 4.481
k__Bacteria|p__Actinobacteria|c__CFGB34754|c__CFGB34754 2|201174| 4.04941
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|c__Alphaproteobacteria_unclassified 2|1224|28211| 3.76603
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria|c__Nitrosomonadales 2|1224|28216|32003 2.8018
k__Bacteria|p__Actinobacteria|c__CFGB1083|c__CFGB1083 2|201174| 1.73407
k__Bacteria|p__Bacteroidetes|c__CFGB462|c__CFGB462 2|976| 1.6796
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|c__Sphingomonadales 2|1224|28211|204457 1.57361
k__Bacteria|p__Actinobacteria|c__CFGB14951|c__CFGB14951 2|201174| 1.54503
k__Bacteria|p__Actinobacteria|c__CFGB46334|c__CFGB46334 2|201174| 0.91335
k__Bacteria|p__Bacteroidetes|c__CFGB40560|c__CFGB40560 2|976| 0.76844
k__Bacteria|p__Actinobacteria|c__CFGB13153|c__CFGB13153 2|201174| 0.69629
k__Bacteria|p__Verrucomicrobia|c__CFGB8395|c__CFGB8395 2|74201| 0.66264
k__Bacteria|p__Proteobacteria|c__CFGB14846|c__CFGB14846 2|1224| 0.55321
k__Bacteria|p__Bacteroidetes|c__CFGB35715|c__CFGB35715 2|976| 0.45045
k__Bacteria|p__Actinobacteria|c__CFGB41185|c__CFGB41185 2|201174| 0.42263
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|c__Pelagibacteriales 2|1224|28211|54526 0.32135
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria|c__Betaproteobacteria_unclassified 2|1224|28216| 0.31492
k__Bacteria|p__Proteobacteria|c__CFGB8596|c__CFGB8596 2|1224| 0.30905
k__Bacteria|p__Bacteroidetes|c__CFGB7657|c__CFGB7657 2|976| 0.29904
k__Bacteria|p__Cyanobacteria|c__Cyanobacteria_unclassified|c__Synechococcales 2|1117|1890424 0.18922
k__Bacteria|p__Bacteroidetes|c__CFGB7628|c__CFGB7628 2|976| 0.1775
k__Bacteria|p__Proteobacteria|c__CFGB42731|c__CFGB42731 2|1224| 0.17321
k__Bacteria|p__Bacteroidetes|c__Flavobacteriia|c__Flavobacteriales 2|976|117743|200644 0.14651
k__Bacteria|p__Verrucomicrobia|c__CFGB8396|c__CFGB8396 2|74201| 0.10069
k__Bacteria|p__Bacteroidetes|c__CFGB32421|c__CFGB32421 2|976| 0.06272
k__Bacteria|p__Armatimonadetes|c__CFGB55302|c__CFGB55302 2|67819| 0.03228
k__Bacteria|p__Bacteroidetes|c__Cytophagia|c__Cytophagales 2|976|1768503|768507 0.02363
k__Bacteria|p__Planctomycetes|c__CFGB24506|c__CFGB24506 2|203682| 0.02097
k__Bacteria|p__Actinobacteria|c__Actinomycetia|c__Candidatus_Nanopelagicales|f__Candidatus_Nanopelagicales 2|201174|1760|2039638|2162846 32.59065
k__Bacteria|p__Actinobacteria|c__Actinomycetia|c__Actinomycetia_unclassified|f__Actinomycetia_unclassified 2|201174|1760| 14.66375
k__Bacteria|p__Bacteroidetes|c__CFGB18754|c__CFGB18754|f__FGB18754 2|976| 11.33284
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria|c__Burkholderiales|f__Burkholderiales 2|1224|28216|80840|119060 9.87441
k__Bacteria|p__Actinobacteria|c__CFGB13659|c__CFGB13659|f__FGB13659 2|201174| 4.481
k__Bacteria|p__Actinobacteria|c__CFGB34754|c__CFGB34754|f__FGB34754 2|201174| 4.04941
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|c__Alphaproteobacteria_unclassified|f__Alphaproteobacteria_unclassified 2|1224|28211| 3.76603
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria|c__Nitrosomonadales|f__Methylophilaceae 2|1224|28216|32003|32011 2.8018
```




k_Bacteria|p_Bacteroidetes|c_CFBG18754|o_CFBG18754|f_FBG18754|g_GGB46526|s_GGB46526_SGB78238 2|976|1|1|1| 5.29106
k_Bacteria|p_Actinobacteria|c_Actinomycoetia|o_Actinomycoetia_unclassified|f_Actinomycoetia_unclassified|g_GGB32489|s_GGB32489_SGB48813 2|201174|1760|1|1| 4.47252
k_Bacteria|p_Bacteroidetes|c_CFBG18754|o_CFBG18754|f_FBG18754|g_GGB46525|s_GGB46525_SGB44386 2|976|1|1|1| 4.3519
k_Bacteria|p_Actinobacteria|c_CFBG34754|o_CFBG34754|f_FBG34754|g_GGB34754|s_GGB34754_SGB48810 2|201174|1|1|1| 4.00833
k_Bacteria|p_Actinobacteria|c_CFBG13659|o_CFBG13659|f_FBG13659|g_GGB13659|s_GGB23666_SGB35387 2|201174|1|1|1| 3.89242
k_Bacteria|p_Froteobacteria|c_Alphaproteobacteria|o_Alphaproteobacteria_unclassified|f_Alphaproteobacteria_unclassified|s_Alpha_proteobacterium_SGCG_AAA028_D10 2|1224|28211|1|1|938641 3.74603
k_Bacteria|p_Actinobacteria|c_Actinomycoetia|o_Candidatus_Nanopelagiales|f_Candidatus_Nanopelagiales|g_Candidatus_Planktophila|s_Candidatus_Planktophila_vernalis 2|201174|1760|2039638|2162846|622681|1884907 3.15272
k_Bacteria|p_Actinobacteria|c_Actinomycoetia|o_Actinomycoetia_unclassified|f_Actinomycoetia_unclassified|g_Actinomycoetia_unclassified|s_Actinobacterium_SGCG_AAA028_A23 2|201174|1760|1|1|932036 2.98358
k_Bacteria|p_Froteobacteria|c_Betaproteobacteria|o_Nitrosomonadales|f_Methylophilaceae|g_Candidatus_Methylophilum|s_Candidatus_Methylophilum_universalis 2|1224|28216|32003|32011|1679002|2588536 1.93257
k_Bacteria|p_Actinobacteria|c_CFBG1083|o_CFBG1083|f_FBG1083|g_GGB36422|s_GGB36422_SGB61312 2|201174|1|1|1| 1.73407
k_Bacteria|p_Bacteroidetes|c_CFBG18754|o_CFBG18754|f_FBG18754|g_GGB46527|s_GGB46527_SGB4388 2|976|1|1|1| 1.68988
k_Bacteria|p_Bacteroidetes|c_CFBG462|o_CFBG462|f_FBG462|g_GGB24171|s_GGB24171_SGB64481 2|976|1|1|1| 1.6796
k_Bacteria|p_Actinobacteria|c_Actinomycoetia|o_Corynebacteriales|f_Mycobacteriaceae|g_GGB38478|s_GGB38478_SGB55612 2|201174|1760|85007|1762|1| 1.62594
k_Bacteria|p_Froteobacteria|c_Alphaproteobacteria|o_Sphingomonadales|f_Sphingomonadaceae|g_Sphingorhabdus|s_Sphingorhabdus_rigui 2|1224|28211|204457|41297|1434046|1282858 1.54448
k_Bacteria|p_Actinobacteria|c_CFBG14951|o_CFBG14951|f_FBG14951|g_GGB41106|s_GGB41106_SGB61344 2|201174|1|1|1| 1.48287
k_Bacteria|p_Froteobacteria|c_Betaproteobacteria|o_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter|s_Polynucleobacter_cosmopolitanus 2|1224|28216|80840|119060|44013|951345 1.24093
k_Bacteria|p_Froteobacteria|c_Betaproteobacteria|o_Burkholderiales|f_Burkholderiales_unclassified|g_GGB43069|s_GGB43069_SGB60311 2|1224|28216|80840|1| 1.19033
k_Bacteria|p_Actinobacteria|c_Actinomycoetia|o_Candidatus_Nanopelagiales|f_Candidatus_Nanopelagiales|g_Candidatus_Planktophila|s_Candidatus_Planktophila_lacus 2|201174|1760|2039638|2162846|622681|1884913 1.07277
k_Bacteria|p_Actinobacteria|c_CFBG46334|o_CFBG46334|f_FBG46334|g_GGB46334|s_GGB46334_SGB35380 2|201174|1|1|1| 0.91335
k_Bacteria|p_Froteobacteria|c_Betaproteobacteria|o_Nitrosomonadales|f_Methylophilaceae|g_Candidatus_Methylophilum|s_Candidatus_Methylophilum_rimovensis 2|1224|28216|32003|32011|1679002|2588535 0.81553
k_Bacteria|p_Bacteroidetes|c_CFBG40560|o_CFBG40560|f_FBG40560|g_GGB43022|s_GGB43022_SGB60257 2|976|1|1|1| 0.76844
k_Bacteria|p_Actinobacteria|c_CFBG13153|o_CFBG13153|f_FBG13153|g_GGB38373|s_GGB38373_SGB52357 2|201174|1|1|1| 0.69629
k_Bacteria|p_Verrucomicrobia|c_CFBG8395|o_CFBG8395|f_FBG8395|g_GGB25672|s_GGB25672_SGB61746 2|74201|1|1|1| 0.66264
k_Bacteria|p_Froteobacteria|c_CFBG14846|o_CFBG14846|f_FBG14846|g_GGB26003|s_GGB26003_SGB61793 2|1224|1|1|1| 0.55321
k_Bacteria|p_Actinobacteria|c_CFBG13659|o_CFBG13659|f_FBG13659|g_GGB23686|s_GGB23686_SGB82209 2|201174|1|1|1| 0.49859
k_Bacteria|p_Bacteroidetes|c_CFBG35715|o_CFBG35715|f_FBG35715|g_GGB46715|s_GGB46715_SGB4592 2|976|1|1|1| 0.45045
k_Bacteria|p_Froteobacteria|c_Alphaproteobacteria|o_Pelagibacteriales|f_Pelagibacteriales_unclassified|g_Candidatus_Fonsibacter|s_Candidatus_Fonsibacter_ubiquis 2|201174|1|1|1| 0.42263
k_Bacteria|p_Froteobacteria|c_Betaproteobacteria|o_Betaproteobacteria_unclassified|f_Betaproteobacteria_unclassified|g_Betaproteobacteria_unclassified|s_Beta_proteobacterium_SGCG_AAA027_R21 2|1224|28211|1|938785 0.31492
k_Bacteria|p_Froteobacteria|c_CFBG8596|o_CFBG8596|f_FBG8596|g_GGB25977|s_GGB25977_SGB37971 2|1224|1|1|1| 0.30905
k_Bacteria|p_Bacteroidetes|c_CFBG7657|o_CFBG7657|f_FBG7657|g_GGB24307|s_GGB24307_SGB64774 2|976|1|1|1| 0.29904

k_Bacteria|p_Actinobacteria|c_Actinomycoetia|o_Candidatus_Nanopelagiales|f_Candidatus_Nanopelagiales|g_GGB46331|s_GGB46331_SGB64109 2|201174|1760|2039638|2162846|1| 0.21351
k_Bacteria|p_Cyanobacteria|c_Cyanobacteria_unclassified|o_Synechococcales|f_Prochlorococcales|g_Cyanobium|s_Cyanobium_usitatum 2|1117|11890424|2881426|167375|2304190 0.18922
k_Bacteria|p_Bacteroidetes|c_CFBG7628|o_CFBG7628|f_FBG7628|g_GGB73349|s_GGB73349_SGB36049 2|976|1|1|1| 0.1775
k_Bacteria|p_Froteobacteria|c_CFBG42731|o_CFBG42731|f_FBG42731|g_GGB47400|s_GGB47400_SGB65360 2|1224|1|1|1| 0.17321
k_Bacteria|p_Froteobacteria|c_Betaproteobacteria|o_Burkholderiales|f_Cammonadaceae|g_Limnochabitsa|s_Limnochabitsa_ep_63ED37_2 2|1224|28216|80840|80864|665874|1678128 0.1662
k_Bacteria|p_Bacteroidetes|c_Flavobacteriales|o_Flavobacteriales|f_Flavobacteriaceae|g_GGB46797|s_GGB46797_SGB4687 2|976|117743|200644|49546|1| 0.13632
k_Bacteria|p_Froteobacteria|c_Betaproteobacteria|o_Burkholderiales|f_Cammonadaceae|g_Limnochabitsa|s_Limnochabitsa_ep_MMS_10A_178 2|1224|28216|80840|80864|665874|1835767 0.12562
k_Bacteria|p_Actinobacteria|c_Actinomycoetia|o_Candidatus_Nanopelagiales|f_Candidatus_Nanopelagiales|g_Candidatus_Nanopelagius|s_Candidatus_Nanopelagius_limnes 2|201174|1760|2039638|2162846|622681|1884634 0.11465
k_Bacteria|p_Froteobacteria|c_Betaproteobacteria|o_Burkholderiales|f_Cammonadaceae|g_Limnochabitsa|s_Limnochabitsa_ep_2KL_51 2|1224|28216|80840|80864|665874|1977911 0.10353
k_Bacteria|p_Verrucomicrobia|c_CFBG8396|o_CFBG8396|f_FBG8396|g_GGB25677|s_GGB25677_SGB61755 2|74201|1|1|1| 0.10069
k_Bacteria|p_Froteobacteria|c_Betaproteobacteria|o_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter|s_Polynucleobacter_ep_MWH_CaX5 2|1224|28216|80840|119060|44013|2689107 0.07053
k_Bacteria|p_Bacteroidetes|c_CFBG32421|o_CFBG32421|f_FBG32421|g_GGB56701|s_GGB56701_SGB78134 2|976|1|1|1| 0.06272
k_Bacteria|p_Actinobacteria|c_CFBG14951|o_CFBG14951|f_FBG14951|g_GGB64139|s_GGB64139_SGB86549 2|201174|1|1|1| 0.06216
k_Bacteria|p_Actinomycoetia|o_Actinomycoetia_unclassified|f_Actinomycoetia_unclassified|g_Actinomycoetia_unclassified|s_Actinomycoetia_unclassified_SGB64102 2|201174|1760|1|1| 0.0585
k_Bacteria|p_Actinomycoetia|o_Candidatus_Nanopelagiales|f_Candidatus_Nanopelagiales|g_Candidatus_Planktophila|s_Candidatus_Planktophila_sulfonica 2|201174|1760|2039638|2162846|622681|1884904 0.05078
k_Bacteria|p_Froteobacteria|c_Betaproteobacteria|o_Nitrosomonadales|f_Methylophilaceae|g_Candidatus_Methylophilum|s_Candidatus_Methylophilum_planktonicus 2|1224|28216|32003|32011|1679002|1581557 0.0507
k_Bacteria|p_Actinobacteria|c_CFBG34754|o_CFBG34754|f_FBG34754|g_GGB34754|s_GGB34754_SGB35383 2|201174|1|1|1| 0.04108
k_Bacteria|p_Actinomycoetia|o_Candidatus_Nanopelagiales|f_Candidatus_Nanopelagiales|g_Candidatus_Planktophila|s_Candidatus_Planktophila_verzatis 2|201174|1760|2039638|2162846|622681|1884905 0.03477
k_Bacteria|p_Froteobacteria|c_Betaproteobacteria|o_Burkholderiales|f_Cammonadaceae|g_Limnochabitsa|s_Limnochabitsa_ep_Bal53 2|1224|28216|80840|80864|665874|1977910 0.03247
k_Bacteria|p_Armatimonadetes|c_CFBG55302|o_CFBG55302|f_FBG55302|g_GGB58711|s_GGB58711_SGB0383 2|67819|1|1|1| 0.03228
k_Bacteria|p_Froteobacteria|c_Alphaproteobacteria|o_Sphingomonadales|f_Sphingomonadaceae|g_Sphingorhabdus|s_Sphingorhabdus_wooponensis 2|1224|28211|204457|41297|1434046|940136 0.02913
k_Bacteria|p_Froteobacteria|c_Betaproteobacteria|o_Burkholderiales|f_Burkholderiaceae|g_Polynucleobacter|s_Polynucleobacter_difficilis 2|1224|28216|80840|119060|44013|556054 0.02767
k_Bacteria|p_Bacteroidetes|o_Cytophagiales|f_Cytophagiales|g_Aquirufa|s_Aquirufa_antherigenensis 2|976|768503|768507|89973|6274247|2516559 0.02363
k_Bacteria|p_Actinobacteria|c_Actinomycoetia|o_Corynebacteriales|f_Mycobacteriaceae|g_GGB43997|s_GGB43997_SGB61369 2|201174|1760|85007|1762|1| 0.02359
k_Bacteria|p_Planctomycetes|c_CFBG24506|o_CFBG24506|f_FBG24506|g_GGB2611|s_GGB2611_SGB4818 2|203662|1|1|1| 0.02097
k_Bacteria|p_Bacteroidetes|o_Flavobacteriales|f_Flavobacteriales|g_Flavobacterium|s_Flavobacterium_ep_LM09 2|976|117743|200644|49546|237|2654245 0.01019



k_Bacterialp_Actinobacteria/c_Actinomyetia/c_Candidatus_Nanopelagiales/f_Candidatus_Nanopelagicaceae/g_Candidatus_Nanopelagicus/s_Candidatus_Nanopelagicus_abundans/t_SGB5714 2|201174|1760|2039638|2162846|2039639|1884916| 27.94258

k_Bacterialp_Proteobacteria/c_Betaproteobacteria/c_Burkholderiales/f_Burkholderiaceae/g_GGB26028/s_GGB26028_SGB38032/t_GGB38032 2|1224|28216|80840|119060||| 8.53529

k_Bacterialp_Actinobacteria/c_Actinomyetia/c_Actinomyetia_unclassified/f_Actinomyetia_unclassified/g_Actinobacterium_SGOC_AAA027_017/t_SGB46820 2|201174|1760|||1932040| 7.14916

k_Bacterialp_Bacteroidetes/c_CFBP18754/c_CFBP18754/f_FGB18754/g_GGB46526/s_GGB78238/t_SGB78238 2|9761||| 5.29106

k_Bacterialp_Actinobacteria/c_Actinomyetia/c_Actinomyetia_unclassified/f_Actinomyetia_unclassified/g_GGB32469/s_GGB32469_SGB48813/t_SGB48813 2|201174|1760||| 4.47252

k_Bacterialp_Bacteroidetes/c_CFBP18754/c_CFBP18754/f_FGB18754/g_GGB46525/s_GGB46525_SGB64386/t_SGB64386 2|9761||| 4.3519

k_Bacterialp_Actinobacteria/c_CFBP34754/c_CFBP34754/f_FGB34754/g_GGB34754/s_GGB34754_SGB48810/t_SGB48810 2|201174||| 4.00833

k_Bacterialp_Actinobacteria/c_CFBP13659/c_CFBP13659/f_FGB13659/g_GGB23686/s_GGB23686_SGB35387/t_SGB35387 2|201174||| 3.98242

k_Bacterialp_Proteobacteria/c_Alphaproteobacteria/c_Alphaproteobacteria_unclassified/f_Alphaproteobacteria_unclassified/g_Alphaproteobacteria_unclassified/s_alpha_protobacterium_SGOC_AAA028_D10/t_SGB50895_group 2|1224|28211|||1938641| 3.76603

k_Bacterialp_Proteobacteria/c_Alphaproteobacteria/c_Alphaproteobacteria_unclassified/f_Alphaproteobacteria_unclassified/g_Alphaproteobacteria_unclassified/s_alpha_protobacterium_SGOC_AAA027_006/k_Bacteria/p_Proteobacteria/c_Alphaproteobacteria/c_Alphaproteobacteria_unclassified/f_Alphaproteobacteria_unclassified/g_Alphaproteobacteria_unclassified/s_alpha_protobacterium_SGOC_AAA027_L15

k_Bacterialp_Actinobacteria/c_Actinomyetia/c_Candidatus_Nanopelagiales/f_Candidatus_Nanopelagicaceae/g_Candidatus_Planktophila/s_Candidatus_Planktophila_vernalis/t_SGB5705 2|201174|1760|2039638|2162846|622681|1884907| 3.15272

k_Bacterialp_Actinobacteria/c_Actinomyetia/c_Actinomyetia_unclassified/f_Actinomyetia_unclassified/g_Actinobacterium_SGOC_AAA028_A23/t_SGB5713 2|201174|1760|||1932036| 2.98558

k_Bacterialp_Proteobacteria/c_Betaproteobacteria/c_Nitrosomonadales/f_Methylophilaceae/g_Candidatus_Methylophilum/s_Candidatus_Methylophilum_universale/t_SGB38034 2|1224|28216|32003|32011|1679002|2588556| 1.93257

k_Bacterialp_Actinobacteria/c_CFBP1083/c_CFBP1083/f_FGB1083/g_GGB36422/s_GGB36422_SGB61312/t_SGB61312 2|201174||| 1.73407

k_Bacterialp_Bacteroidetes/c_CFBP18754/c_CFBP18754/f_FGB18754/g_GGB46527/s_GGB46527_SGB4388/t_SGB4388 2|9761||| 1.68988

k_Bacterialp_Bacteroidetes/c_CFBP462/c_CFBP462/f_FGB462/g_GGB24171/s_GGB24171_SGB4481/t_SGB4481 2|9761||| 1.6796

k_Bacterialp_Actinobacteria/c_Actinomyetia/c_Corynebacteriales/f_Mycobacteriaceae/g_GGB38478/s_GGB38478_SGB55612/t_SGB55612 2|201174|1760|85007|1762||| 1.62594

k_Bacterialp_Proteobacteria/c_Alphaproteobacteria/c_Sphingomonadales/f_Sphingomonadales/g_Sphingorhabdus/s_Sphingorhabdus_riguiti/t_SGB89968 2|1224|28211|204457|41297|1434046|1282858| 1.54448

k_Bacterialp_Actinobacteria/c_CFBP14951/c_CFBP14951/f_FGB14951/g_GGB41106/s_GGB41106_SGB61344/t_SGB61344 2|201174||| 1.48287

k_Bacterialp_Proteobacteria/c_Betaproteobacteria/c_Burkholderiales/f_Burkholderiaceae/g_Polynuclibacter/s_Polynuclibacter_cosmopolitanus/t_SGB13556 2|1224|28216|80840|119060|44013|351345| 1.24093

k_Bacterialp_Proteobacteria/c_Betaproteobacteria/c_Burkholderiales/f_Burkholderiales_unclassified/g_GGB43069/s_GGB43069_SGB60311/t_SGB60311 2|1224|28216|80840||| 1.19033

k_Bacterialp_Actinobacteria/c_Actinomyetia/c_Candidatus_Nanopelagiales/f_Candidatus_Nanopelagicaceae/g_Candidatus_Planktophila/s_Candidatus_Planktophila_lacus/t_SGB5706 2|201174|1760|2039638|2162846|622681|1884916| 1.07277

k_Bacterialp_Actinobacteria/c_CFBP46334/c_CFBP46334/f_FGB46334/g_GGB46334/s_GGB46334_SGB35380/t_SGB35380 2|201174||| 0.91335

k_Bacterialp_Proteobacteria/c_Betaproteobacteria/c_Burkholderiales/f_Cammonadaceae/g_Limnochabitans/s_Limnochabitans_sp_63D03_2/t_SGB12655 2|1224|28216|80840|80864|665874|1679128| 0.1862

k_Bacterialp_Bacteroidetes/c_Flavobacteriales/c_Flavobacteriales/f_Flavobacteriaceae/g_GGB46797/s_GGB46797_SGB64687/t_SGB64687 2|9761|117743|200644|49546||| 0.13632

k_Bacterialp_Proteobacteria/c_Betaproteobacteria/c_Burkholderiales/f_Cammonadaceae/g_Limnochabitans/s_Limnochabitans_sp_MMS_10A_1761/t_SGB22489 2|1224|28216|80840|80864|665874|1835767| 0.12562

k_Bacterialp_Actinobacteria/c_Actinomyetia/c_Candidatus_Nanopelagiales/f_Candidatus_Nanopelagicaceae/g_Candidatus_Nanopelagicus/s_Candidatus_Nanopelagicus_limneis/t_SGB5711 2|201174|1760|2039638|2162846|2039639|1884934| 0.1165

k_Bacterialp_Actinobacteria/c_Betaproteobacteria/c_Burkholderiales/f_Cammonadaceae/g_Limnochabitans/s_Limnochabitans_sp_2KL_51/t_SGB22500 2|1224|28216|80840|80864|665874|1977911| 0.10353

k_Bacterialp_Verrucomicrobia/c_CFBP8396/c_CFBP8396/f_FGB8396/g_GGB25677/s_GGB25677_SGB61755/t_SGB61755 2|74201||| 0.10069

k_Bacterialp_Proteobacteria/c_Betaproteobacteria/c_Burkholderiales/f_Burkholderiaceae/g_Polynuclibacter/s_Polynuclibacter_sp_MNH_Ca5/t_SGB94461 2|1224|28216|80840|119060|44013|2689107| 0.07053

k_Bacterialp_Bacteroidetes/c_CFBP32421/c_CFBP32421/f_FGB32421/g_GGB56701/s_GGB56701_SGB78134/t_SGB78134 2|9761||| 0.06272

k_Bacterialp_Actinobacteria/c_CFBP14951/c_CFBP14951/f_FGB14951/g_GGB4139/s_GGB4139_SGB6549/t_SGB6549 2|201174||| 0.06216

k_Bacterialp_Actinobacteria/c_Actinomyetia/c_Actinomyetia_unclassified/f_Actinomyetia_unclassified/g_Actinomyetia_unclassified/s_Actinomyetia_unclassified_SGB64102/t_SGB64102 2|201174|1760||| 0.0585

k_Bacterialp_Actinobacteria/c_Actinomyetia/c_Candidatus_Nanopelagiales/f_Candidatus_Nanopelagicaceae/g_Candidatus_Planktophila/s_Candidatus_Planktophila_sulfonicea/t_SGB5709 2|201174|1760|2039638|2162846|622681|1884904| 0.0576

k_Bacterialp_Proteobacteria/c_Betaproteobacteria/c_Nitrosomonadales/f_Methylophilaceae/g_Candidatus_Methylophilum/s_Candidatus_Methylophilum_planktonicus/t_SGB3709 2|1224|28216|32003|32011|1679002|1581557| 0.0537

k_Bacterialp_Actinobacteria/c_CFBP34754/c_CFBP34754/f_FGB34754/g_GGB34754/s_GGB34754_SGB35383/t_SGB35383 2|201174||| 0.04108

k_Bacterialp_Actinobacteria/c_Actinomyetia/c_Candidatus_Nanopelagiales/f_Candidatus_Nanopelagicaceae/g_Candidatus_Planktophila/s_Candidatus_Planktophila_versatilis/t_SGB5707 2|201174|1760|2039638|2162846|622681|1884905| 0.03477

k_Bacterialp_Proteobacteria/c_Betaproteobacteria/c_Burkholderiales/f_Cammonadaceae/g_Limnochabitans/s_Limnochabitans_sp_Ba153/t_SGB22499 2|1224|28216|80840|80864|665874|1977910| 0.03247

k_Bacterialp_Amatimonadetes/c_CFBP55302/c_CFBP55302/f_FGB55302/g_GGB56711/s_GGB56711_SGB80383/t_SGB80383 2|67819||| 0.03228

k_Bacterialp_Proteobacteria/c_Alphaproteobacteria/c_Sphingomonadales/f_Sphingomonadales/g_Sphingorhabdus/s_Sphingorhabdus_wooponensis/t_SGB31754 2|1224|28211|204457|41297|1434046|940136| 0.02913

k_Bacterialp_Proteobacteria/c_Betaproteobacteria/c_Burkholderiales/f_Burkholderiaceae/g_Polynuclibacter/s_Polynuclibacter_difficilis/t_SGB30629 2|1224|28216|80840|119060|44013|356054| 0.02767

k_Bacterialp_Bacteroidetes/c_Cytophagiales/c_Cytophagiales/f_Cytophagaceae/g_Aquirufa/s_Aquirufa_antheringensis/t_SGB28348 2|9761|768503|768507|89373|2676247|2516559| 0.02363

k_Bacterialp_Actinobacteria/c_Actinomyetia/c_Corynebacteriales/f_Mycobacteriaceae/g_GGB43997/s_GGB43997_SGB61369/t_SGB61369 2|201174|1760|85007|1762||| 0.02359

k_Bacterialp_Planctomyces/c_Actinomyetia/c_Corynebacteriales/f_Mycobacteriaceae/g_GGB24506/f_FGB24506/g_GGB24506/s_GGB24506_SGB4618/t_SGB4618 2|203682||| 0.02097

k_Bacterialp_Bacteroidetes/c_Flavobacteriales/c_Flavobacteriales/f_Flavobacteriaceae/g_Flavobacterium/s_Flavobacterium_sp_LM09/t_SGB3036 2|9761|117743|200644|49546|237|2654245| 0.01019

k_Bacterialp_Bacteroidetes/c_Flavobacteriales/c_Flavobacteriales_unclassified/f_Flavobacteriales_unclassified/s_Flavobacteria_bacterium_BAL38

ANEXO 4. Toma de muestras superficiales entre 15 a 20cm de profundidad en ambas zonas.



Nota: Toma de muestra de la Bahía Interior (A) y Captación Chimú (B)

ANEXO 5. Filtración de muestras de agua de la Bahía Interior y la Captación Chimú.



ANEXO 6. Protocolo de kit de extracción de ADN microbiano by Thermo Fisher Science.

USER GUIDE

invitrogen

PureLink™ Microbiome DNA Purification Kit

Purification of high-quality microbial DNA from microbial culture and transport media samples

Catalog Number A29790

Pub. No. MAND014332 Rev. A.0

WARNING! Read the Safety Data Sheets (SDSs) and follow the handling instructions. Wear appropriate protective eyewear, clothing, and gloves. Safety Data Sheets (SDSs) are available from thermofisher.com/support.

Product description

The Invitrogen™ PureLink™ Microbiome DNA Purification Kit enables fast purification of high-quality microbial and host DNA from a wide variety of sample types. The kit uses proven PureLink™ spin-column technology for robust yields of purified DNA that is ready for downstream PCR, sequencing, or other applications.

Typical DNA recovery is 5–20 µg from 1 mL of microbial culture. For transport media, DNA is recovered in sufficient quantity for downstream PCR detection, but may be undetectable by spectrophotometric methods.

Procedure overview

This guide describes purification of DNA from microbial culture (for example, from single colonies or from broth cultures) or liquid transport media (for example, Amies or Stuart Transport Medium). In this procedure, the microorganisms are efficiently lysed by a combination of heat, chemical, and mechanical disruption with specialized beads. The sample is then applied to a PureLink™ spin column, and the DNA that is bound to the column undergoes a single wash step before elution.

Kit contents

Table 1 PureLink™ Microbiome DNA Purification Kit [Cat. no. A29790, 50 reactions]

| Component | Quantity | Storage |
|---|----------|--------------|
| S1—Lysis Buffer | 40 mL | 15°C to 30°C |
| S2—Lysis Enhancer | 5 mL | |
| S3—Cleanup Buffer ^[1] | 125 mL | |
| S4—Binding Buffer | 45 mL | |
| S5—Wash Buffer Concentrate ^[2] | 13 mL | |
| S6—Elution Buffer | 5 mL | |

| Component | Quantity | Storage |
|--|----------|--------------|
| PureLink™ Spin Columns with Collection Tubes | 50 | 15°C to 30°C |
| PureLink™ Collection Tubes | 100 | |
| Bead Tubes ^[3] | 50 | |

^[1] Not used in this procedure.

^[2] Add 13 mL of 96–100% ethanol before use. See “Before you begin” on page 2.

^[3] Ships separately.

Required materials

Unless otherwise indicated, all materials are available through thermofisher.com. MLS: Fisher Scientific (www.fisherscientific.com) or other major laboratory supplier.

Table 2 Required materials not included with the kit

| Item | Source |
|---|--|
| Heat block, dry bath, or water bath, 65°C | MLS |
| <i>[Optional]</i> For dry bath, LabArmor™ Beads | Cat. no. A12543 |
| Microcentrifuge capable of 14,000 × g | MLS |
| Vortex mixers, 2 ^[1] | MLS |
| For vortex bead homogenization: hands-free adapter for vortex mixer, with horizontal tube orientation | Fisher Scientific NC0070788 ^[2] |
| <i>[Optional; alternative to vortex bead homogenization]</i> Bead mill homogenizer | Omni 19-040, or equivalent |
| Adjustable pipettors, 100–1000 µL | MLS |
| Microcentrifuge tubes, DNase-free, 1.5 mL or 2.0 mL | MLS |
| Ethanol, 96–100% | MLS |

^[1] For vortex bead homogenization, we recommend using two mixers, one dedicated to the hands-free adapter.

^[2] Cat. no. AM10024 (not available for sale) can also be used.

For Research Use Only. Not for use in diagnostic procedures.

ThermoFisher
SCIENTIFIC

ANEXO 7. Extracción de DNA empleando el inserto del kit DNA by Thermo Fisher Science.



ANEXO 8. Centrifugado de muestras con Buffer de lysis, Buffer de lavado, Buffer de proteinasa K, Buffer de RNAs.



ANEXO 9. Incubación a 65 °C para ayudar a una correcta lysis



ANEXO 10. Muestra de ADN completamente purificado de las muestras de la Bahía
Interna BI y la Captación Chimú CH.



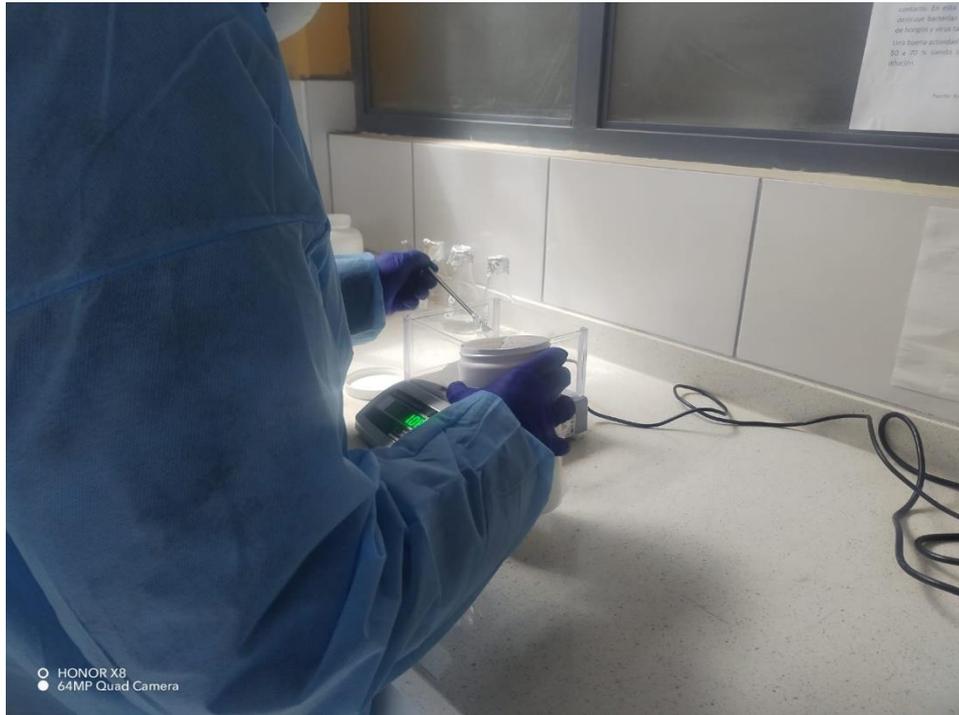
ANEXO 11. Cuantificación y calidad de DNA con el Buffer de elución (BE).



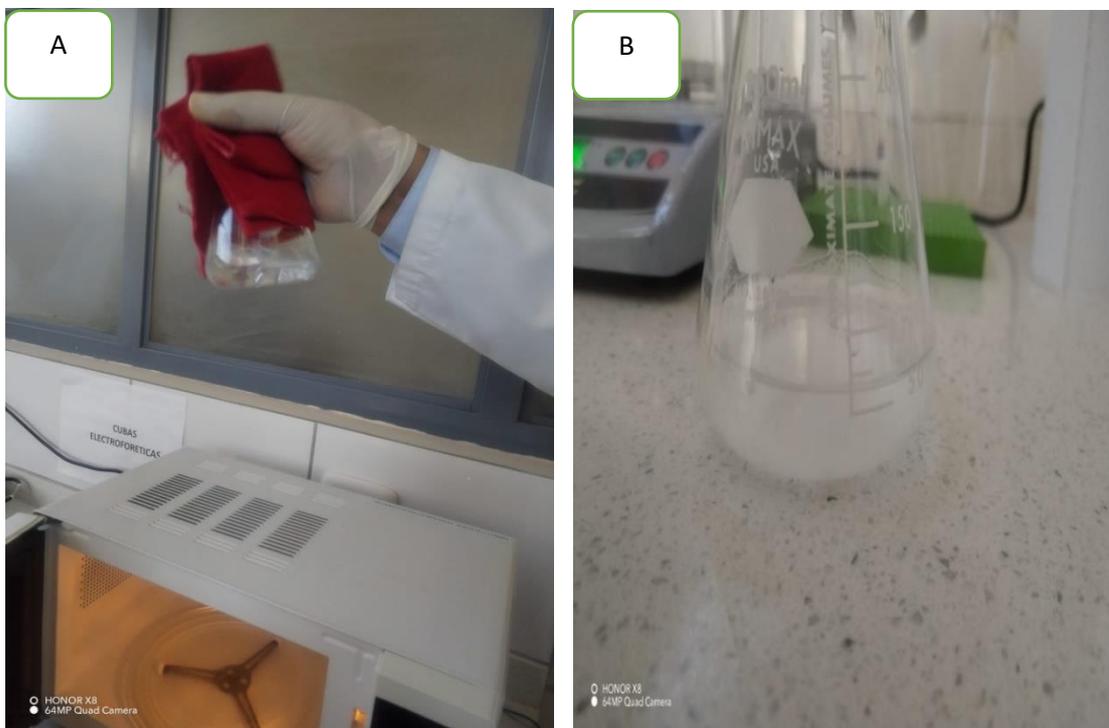
ANEXO 12. Datos de la cuantificación y calidad de las muestras.



ANEXO 13. Preparado del el de agarosa al 1% para la observación de bandas de ADN

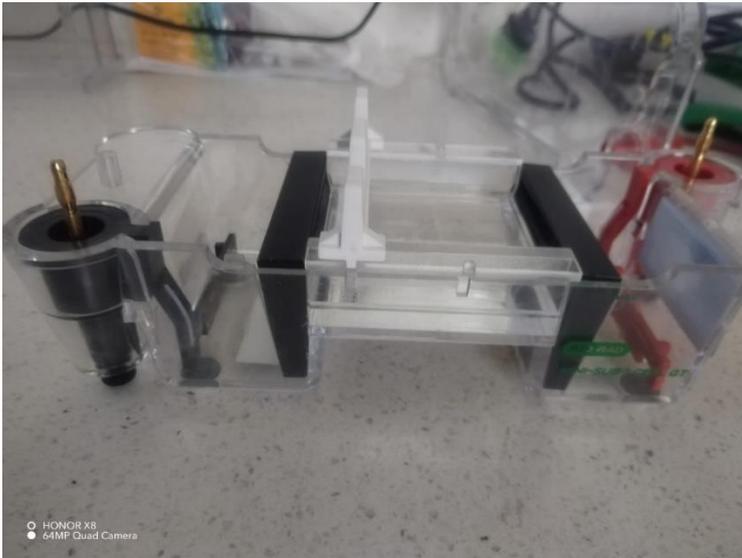


ANEXO 14. Calentamiento de la solución TAE 1X con gel agarosa al 1%.



Nota: (A) Homogenización de la solución de agarosa al 1%. (B) solución final del gel de agarosa.

ANEXO 15. Gel de agarosa al 1% en la cámara electroforética.

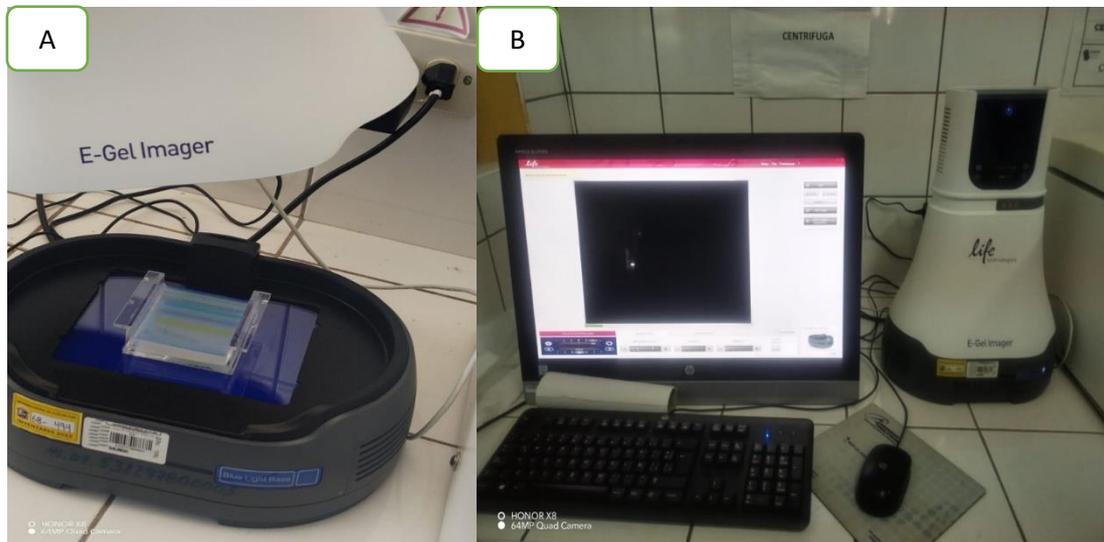


ANEXO 16. Inoculación de ADN de las muestras para el corrido electroforético (A),
voltaje de corrido para la cámara electroforética (B).



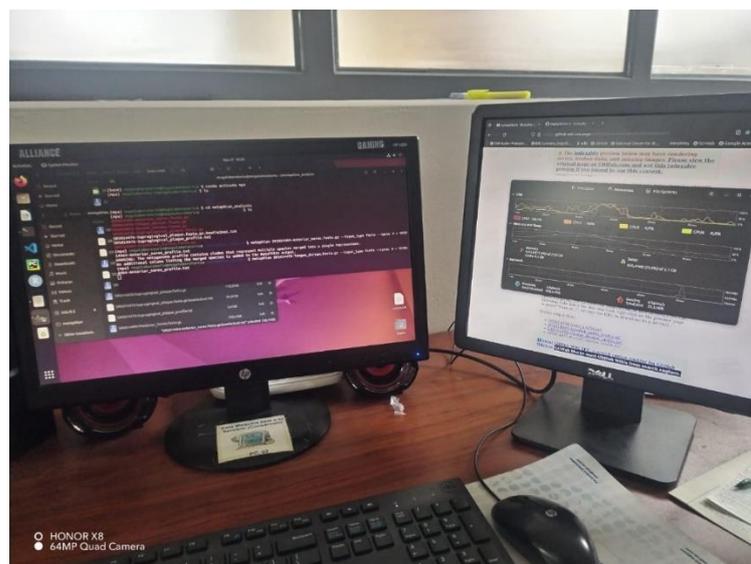
Nota: (A) Inoculación de ADN de las muestras para el corrido electroforético, (B) voltaje de corrido para la cámara electroforética.

ANEXO 17. Lectura de la integridad de DNA de las muestras de la bahía interna y la captación chimú del gel de agarosa a 1%

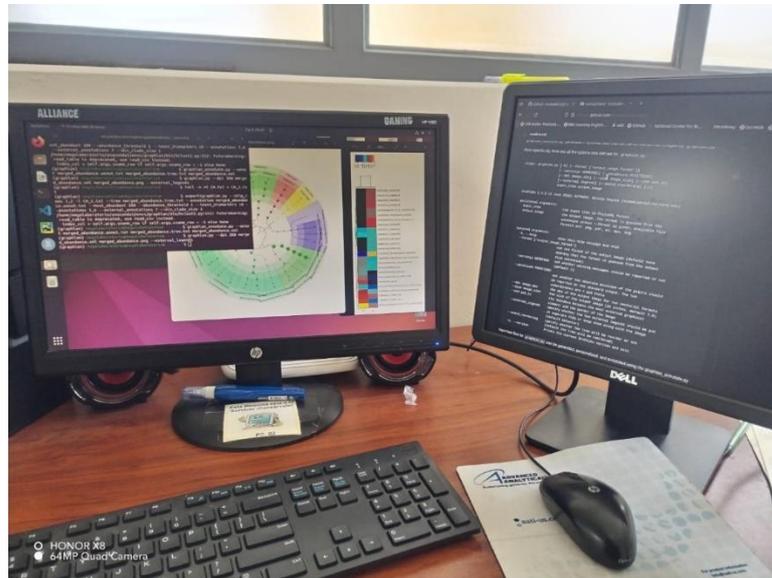


Nota: (A) gel de electroforesis en el equipo transluminador Imager (B), camera Hood E-Gel para observar la migración de las bandas con luz UV.

ANEXO 18. Instalación de herramientas bioinformáticas para el secuenciamiento metagenómico.



ANEXO 19. Lectura de los datos crudos del secuenciado por tecnología Shotgun.





ANEXO 21. Data secuenciada correspondiente a la parte inicial de los contigs del ADN bacteriano extraído de la muestra de agua en la Bahía Interior de Puno en formato FASTQ

```

>A01940:43:HT3V3DSX3:3:1101:11713:1047 1:N:0:GTAGAGGAAT+TATGCAGTGT
ANTCAGAGTCGTATCGCTAAACGGAACAAGAAAATCCCCAGACGTCGGTTTGTATTTTGTTCGTCGGTCCCATGAATATGTTGCCCGC
ATCGTGATATTCGTGGCCGTGATCTGGTTGTAATCGTTGGTCAGTGTGAATACGTTGTTGAG
>A01940:43:HT3V3DSX3:3:1101:22001:1047 1:N:0:GTAGAGGAAT+TATGCAGTGT
GNTCTGGACATAGCCAAGCGCTCGTTCAGCTTCTTCGCGCTTTTCTCAACCGCCGCGCCCACTCGGCTTGTTCGCCGCAAGGGG
AGCGCTCCGAGGATGCTTTCGGCGCGACCAAGAGTTTTGGTTGCTCGTTCGATTTGGCCCTTG
>A01940:43:HT3V3DSX3:3:1101:30101:1047 1:N:0:GTAGAGGAAT+TATGCAGTGT
GNCCCTAATTACGCTCATTTATTTCTCCTTTTGTCTTTTATATTAGTTTACCATAGACTGTATTGATTTGTCAAACCTCAAACCTCAA
GTTCTTTAATTGACTTATCGTCCATATCGCCTATATCTTTATATCTGTCTCTTATACACATC
>A01940:43:HT3V3DSX3:3:1101:22263:1094 1:N:0:GTAGAGGAAT+TATGCAGTGT
CTTATATTATTCCTTTGTGGCAGAGTCGGCTCGGTTTCCGGCGACTGTTTGGACCTTGAGAATCTGGATTGCTCCAAAAAACTTA
CGTGGTTAATCGCCTGACCGTTAAGCGCACCTCGAAATGCAGATGGCTTCCGGTTGACCGA
>A01940:43:HT3V3DSX3:3:1101:26621:1094 1:N:0:GTAGAGGAAT+TATGCAGTGT
GCTTTAGGTTGTATGCTGTAGAATCCAATGCTTTTTTGTCTACTACGACAACACCACGACTATCCCATTCAGCCTTAGCCTGTGCAC
CAATTTTACCATCAACAGGGCAAGGTGTACCTGCAGCATCATAGCAGCAAACACACGTTTCGT
>A01940:43:HT3V3DSX3:3:1101:31340:1094 1:N:0:GTAGAGGAAT+TATGCAGTGT
GCTTAGACACGACCTATACGATGGCACAACCTATGACTGAATCGGCGACGCCCCCGCACATGCCCATCGAACATCCCTCCCTTTCCG
TATACAGCGCGCTTGGCTGCCGATATTGAGTCCAAATGGCAAGATTTCGTGGGAGACCGAAGGC
>A01940:43:HT3V3DSX3:3:1101:2483:1125 1:N:0:GTAGAGGAAT+TATGCAGTGT
TAGATATGCGGATTACCGTTGAACAGTGGGACGGACAGATATTGCTTCACCAGAATCGGAGAGTCCAGTGTGTGATTGCAGAGCG
AGTGCTCGCTGCAAGAGTCAATGCTCGTCAACGTTTTTCCGGCAGAGGTTTTGAACTTAATAG
>A01940:43:HT3V3DSX3:3:1101:3061:1125 1:N:0:GTAGAGGAAT+TATGCAGTGT
GCTCAATCGCGTCACGCGGGAACCTCGAAATTGTAACACGTCGCGGCTTTCGCTGTTGCCCGCCCGCCGCATAACCACTCTGGG
CTTTTTCATCATCAGCGTCGCAATCTGGCGCGCAGGTAGTTGATGTCTCCGGGACGAGGTA
>A01940:43:HT3V3DSX3:3:1101:3495:1125 1:N:0:GTAGAGGAAT+TATGCAGTGT
GTTCCGAGAACTTTCCTCGTCAAATGATACCTTGTGTCTACTACTAGGCTTTTATCCTTTCTTAACTCCTTTCCGCGCTTGGC
TTCGAGTGTTTTAACTGCTCGTCCAATGGCTCGTAGTTTTTCTTTCGGCATTCACCTTCTC
>A01940:43:HT3V3DSX3:3:1101:26549:1188 1:N:0:GTAGAGGAAT+TATGCAGTGT
TCATTAAGTTGGTCTGTGGCTTCATCAACTACCAGTGTAGGATGTGCAGAAATAGGTAATGGCATTGTAGATCTCTGCCATGTACCGT
AGATACTGCGTTGGATCTGTGCAACGCTTGAATAACTGTGGTGCCAATATTGCTGTAAATCT
>A01940:43:HT3V3DSX3:3:1101:27344:1188 1:N:0:GTAGAGGAAT+TATGCAGTGT
TCATTAACAGGTTGGGACTCCACCTTATACGGTGAATGGTTAAACCTGTATTATCTCCTGATATTGTAACCACAGCACCTGCAGTTA
AAACAGGTTTAAAGTGCCACAACATACCGGTTAGGGTTAATGATAGTTTACCTACCGAACAACG
>A01940:43:HT3V3DSX3:3:1101:27995:1188 1:N:0:GTAGAGGAAT+TATGCAGTGT
TCATCACTTAAATCCGATCCGATAAATTCGACAAGGTAATTCGTCACTCGCTTCTTTGCCGCGCCATATATTTCCCTCCACATCTCTT
CATCAAACCATTGAAATCTTAAACGCATAATGTTGCTTTTCGATATTGGAAGGAATACCCG
>A01940:43:HT3V3DSX3:3:1101:11623:1204 1:N:0:GTAGAGGAAT+TATGAAGTGT
AGGAAGGTGCGCAAGTGCCAGCAGGCACTGACCTCATGCGATTTAAAGGCCATGCACGAGCCAAGATTACCGCCAGTCACTTATACAC
ATCTCCGAGCACAGCAGCTAGAGGAATCTCGGCGCGCCGAGTTGGGAGTTATAGGGGGTG
>A01940:43:HT3V3DSX3:3:1101:6171:1219 1:N:0:GTAGAGGAAT+TATGCAGTGT
GTTTTACGATCATGATTTATATTATGTTTGTTTTTTAAATAGTGTACCTGTCTTCCCGTCTTGTGATGTCTAACGTATAAATATG
GATTCATCGTTAGGCATCTCTTTAGCTCGTCTATGAATAAGTTAGCCCTCCTGTACCCTA
>A01940:43:HT3V3DSX3:3:1101:7211:1235 1:N:0:GTAGAGGAAT+TATGCAGTGT
GGTCTGGAGGCGAGCGTGGCGTAACCTGGTCGCGCAGGAGCGCTCGTCGCGGATCCTGTATTGGTCCCTCCTGATGCGCCTGCAC
GCTCAAGCCGCTTCAGGGTTCGCGCAGTGGGATTCTCGTGGTCTGGACAGCGGGCAGGAGC
>A01940:43:HT3V3DSX3:3:1101:20338:1235 1:N:0:GTAGAGGAAT+TATGCAGTGT
TGTATACCCTGTTGGTAGGCTGATGCAGATAGTGAAGGCTTACTTATTTTAAACAAACGATACCTTTTAAACCATAAATATTGCAT
CCAAGTTCTGTCTCTTATACACATCTCCGAGCCACAGACGTAGAGGAATCTCGGGGGGCG
>A01940:43:HT3V3DSX3:3:1101:31620:1235 1:N:0:GTAGAGGAAT+GATGCAGTGT
CCAAAAGCCCGTGGCGTTATAATGAACGAGGGGGTGTAGTCCGAACCTGACTAAGCGAGATATTGGTGGAGCTATCAGACACCCCG
GAGGTGCAGGTAGAGCTTAAACGATACCTCGATAGAGTCCGAAGTAAGCACCAACCAACATAC
>A01940:43:HT3V3DSX3:3:1101:14633:1251 1:N:0:GTAGAGGAAT+TATGCAGTGT
ATCTACGAGCCTTTCTTCTTAGTAATTGAGACCTAATAAAGTAGATTCTATCCCTTAAATCAAAGTTTATGGCAAAGAACTCTACTCT
ATTATCAGCCAAATCCAGATGGACTACCGTCTCTTTCTGACTCAATCCACATAACTTCATCTTT
>A01940:43:HT3V3DSX3:3:1101:15772:1251 1:N:0:GTAGAGGAAT+TATGCAGTGT
ATTAGATGCTGCCATGCAGATAAGAATGCGCCAGCAGTTAGTACGATTGGATTCTTCATATTCAATTTTGGCACCTAACATGGGGG
TATTAAGAGCGGAGCATTGTCTTTGCTCTGTTTTTTATTGAATGAACATGATGTGATGCG
>A01940:43:HT3V3DSX3:3:1101:27163:1251 1:N:0:GTAGAGGAAT+TATGCAGTGT
GATTAAGATACACTTTTTCTTAGAGATTTTTACATCGATTCCAAAAAGTTTTCTGTTCAAGTTGCAACACTGTCATTCTTTTTTGATG
TGCTAACTCACTATAAAGTCCACATATCTCTGTCTCTTATACACATCTCCGAGCCACGAG
>A01940:43:HT3V3DSX3:3:1101:1081:1297 1:N:0:GTAGAGGAAT+TATGCAGTGT
GACTAATTTGGTCTGATAGGGTTAGGCTGTATCTGTAAAGTCCAACCTGCTGTGCGTATCCCAATTTATCTCCTTCTACGGTTTTCTAC
TTCACAAATCAAACTGAAATCATTAGTTTACCATCTACCACCGACACATACCAAACTCATT
>A01940:43:HT3V3DSX3:3:1101:14678:1297 1:N:0:GTAGAGGAAT+TATGCAGTGT
ATGTAATCACCTGCTGCAACAGTAACAACGTAAGGCCATGAAGGAACCTGCTTGGCATCATTCAAGGTCAAGGTCAGCGGTTGTTAG
AGTAAGGAACGCTACACCGTTTTTACGCAACCAGACATTGACATCATCCTGTCTCTTATACA

```



>A01940:43:HT3V3DSX3:3:1101:22327:1297 1:N:0:GTAGAGGAAT+TATGCAGTGT
 ATTTATATACCTGAGGGATTTGCTCATGGTTTTGTTGCTTAGATAAAGAGAATATAGTTCACTATTGCAACTCAAATATAGAAACA
 AAAATTATGAGACAGGAATTTCTTTGGAATGATCCTGATCTTAACATCGACTGCGCTATTA
 >A01940:43:HT3V3DSX3:3:1101:24569:1297 1:N:0:GTAGAGGAAT+TATGCAGTGT
 GTTTTGAAGGTTGCCCGTGCCTGGTGAAGCTGGTGTGTTGCCGTAATCCAACCAAACATCGGTGTGCAGTCGCAGCACTGCCCTG
 TGGCGATGTTGAAAGACGGCGAGCGGTAGCCGCTTCGCTCAGCAGCAAGGTGCCCGGTTGA
 >A01940:43:HT3V3DSX3:3:1101:27407:1297 1:N:0:GTAGAGGAAT+TATGCAGTGT
 CTTTCATGTTTACCACCAATTTATTTCAATCGTAGACTTTGCGTGTGTGAGCGGTACCTCTTTTTCTTTTTCTTTATAGGTACTGTCT
 AAGTTATTTTCAAGTACAGGTTGAATGAGTTCTGTGCTTTACCTATGCTCATAGCTTCTCT
 >A01940:43:HT3V3DSX3:3:1101:5358:1313 1:N:0:GTAGAGGAAT+TATGCAGTGT
 CTCACCTCTAGAGAAAATCGTGGAGCCAGTGGCATTTCGAGCATAATTGAATCTACGTTCTCGTGATATTCAGTAGCCATCCGCGGA
 ATCGGCGGAATCTTAATTGAAGGCGATTTCAATTTAATTGGTGATCACAATCTTGATAATCTC
 >A01940:43:HT3V3DSX3:3:1101:11071:1313 1:N:0:GTAGAGGAAT+TATGCAGTGT
 ATTTTATTTATCTTTAGTGTTTTTTAACTATCCAAGCATATCATTGCGAGGCAATTTGGCAAATGGGTTTTAATGAGTCAGTAACT
 TCTGTATGAGCAGCATTTGTTTATCTACATGACCTGATTTTTACTTCCAATTTATAGTCGCAATA
 >A01940:43:HT3V3DSX3:3:1101:12933:1313 1:N:0:GTAGAGGAAT+TATGCAGTGT
 GCTCGATCATGGCTTCGAACGTGGCATTACCCGAAAAGCCACGAGGAATCGCCACTTTGAGCGTTGTTGGCTGACGGCGCCCATCG
 GGCAGCTCAAACGCGGGCGGAGGTGGCCATGGTGTGAGTACCGCGTGGCGTGAACGAAAC
 >A01940:43:HT3V3DSX3:3:1101:19605:1313 1:N:0:GTAGAGGAAT+TATGCAGTGT
 ACGATAACCAGAGCCGATATCGAAAGTGGCGGTAGTTAATGTGAGTGGAGATCAATCGCTCTCGAATTGATCTAACATGAGAAACGG
 CTGAGTATGTGGCTGCAACTGTGCTAAAGAGATGAAGGCTCCGGAATATAGATAAATTCGGAG
 >A01940:43:HT3V3DSX3:3:1101:20726:1313 1:N:0:GTAGAGGAAT+TATGCAGTGT
 GACGTATCAAGCGCCGCTTCAAAGCTGGCCGCTAGTTCGCGCAATGGAGCGCAAGCAGGGCGCAGCGTCAAACAACTTGCCC
 AACAGCTTAAAAAACAGCCCGGACACAGGCTTATCAAACTACAGCGAAGAAGCTGGCCGTC
 >A01940:43:HT3V3DSX3:3:1101:2474:1329 1:N:0:GTAGAGGAAT+TATGCAGTGT
 TACTGAAAGCATGTGTCAATAATGCCATCAACAATTTTATCTGCACTTAGACAACCACTAGCCACAGCACTCTCAACAGTTGCAGGCA
 ACGTTTACAGTTTCTGTCGCCGAGTCAGTAATTCACAGCAGTCTGTTATTTGTTGTTCTCTTA
 >A01940:43:HT3V3DSX3:3:1101:15085:1344 1:N:0:GTAGAGGAAT+TATGCAGTGT
 GCCCGAGGACAGCGGGAAGCCTTCAAGGGAGAGATGGACGCTTACCAGCAACAGGAATACATGAAGTGGTTTTATGACTGTCTCTTAT
 ACACATCTACGAGCCACGAGACGTAGAGGAATCTCGGGTTCGGGGTGGGGTGGTTATGTGG
 >A01940:43:HT3V3DSX3:3:1101:22625:1344 1:N:0:GTAGAGGAAT+TATGCAGTGT
 CTCGTAGATCGAGCCAAATCGGCGCAGACCCGTGCGTGCACCTGATCACCAGCAGGTAATCTGTGCTGTACCCAAAGCGGGGGCC
 CACTTCCGAATCCCTGAATACTGCACCTGGAGGGCAATCCGCGACTCGTTTACAGTCAAGCGCG
 >A01940:43:HT3V3DSX3:3:1101:27145:1344 1:N:0:GTAGAGGAAT+TATGCAGTGT
 CTATTATCCAACCTTCAACAAAGACTGGGTAGAAGGCACTAACCTAAAGTTTTAAATCCAACTTTCAATCCGAAGTATGACGATTAT
 CTGTACTATTTTGTATGCGGACACCAATGTTTCAAACCTTTTGGTGTATTGGTTTATCGGGAT
 >A01940:43:HT3V3DSX3:3:1101:13937:1360 1:N:0:GTAGAGGAAT+TATGCAGTGT
 GATAAGGGTTATTAACAATCTCTCGCATCGTGGTGTATGTAATGGGTAATACCACCCACACCGGTTGACACCCGATTTACGACTCTG
 TAGTTGTTTAGCTAGCATTGGTTCGCTTACCTACTTATTCGATGGAAATGGAAACTTCG
 >A01940:43:HT3V3DSX3:3:1101:32344:1360 1:N:0:GTAGAGGAAT+TATGCAGTGT
 GGGTAGCGCGACCGAATAAGTGTAAATAGTTAATTCGCCAGAGGAGAACAGATATGTAAGAACACCTGATGTTTCTTACATTACGG
 GCTTTTCGGAAGTTGTTTACCAGCAATTCACAAATGTCTTCAAATTTATTCGGATTTGATG
 >A01940:43:HT3V3DSX3:3:1101:29894:1376 1:N:0:GTAGAGGAAT+TATGCAGTGT
 GCCTCAAGGTATGTAACAGCCTTACGATATGCGCTTGCATCTGGGAGTTACGCTCATCTTGATTCATCGTATTCAACGCACGATTGA
 CGTGTGCTGGCATTAGACGGGACACTAGGCTTGTATCTACTGCTACTTGGCAAGTGTGTATC
 >A01940:43:HT3V3DSX3:3:1101:31684:1376 1:N:0:GTAGAGGAAT+TATGCAGTGT
 CTATACGGTTCGATAGAAGCCGGCACCAGTCCCAATTTTTGATCGACTGGGGCAGTGGCAGTGGCCCGGATACCTTACACCAGTACC
 GGCGAAAAGACGACGATTTACGACACTGCGCGGTCTTACACCGTCAAGCTTTTGGGTTTCATT
 >A01940:43:HT3V3DSX3:3:1101:8241:1391 1:N:0:GTAGAGGAAT+TATGCAGTGT
 GTTATAAATTAAGTAGCGGTTCTGAACATATAAAGGCTTCTCTCAGCGCAAGCGGAGGAAAGCCTTTTAAATTAGATTAAATTA
 CTTACTATAAAAACCCAGTTGGTTGAAATCTCTCGTTACTGGCTGATTCATTAGTAAAGC
 >A01940:43:HT3V3DSX3:3:1101:5484:1407 1:N:0:GTAGAGGAAT+TATGCAGTGT
 AGATACTCTGCATTGCTGATTTTTTCAACATTTATTTGGATTAGTCTTCAAGTCAATTGCATTTCTGACTGCGAGTATTTTCTGTTTTCA
 ATATAAGCCCTTCTTATCAATTTGATACTGGTTAATGATTGCCATAGTACCCCGTTAATC
 >A01940:43:HT3V3DSX3:3:1101:9896:1407 1:N:0:GTAGAGGAAT+TATGCAGTGT
 GGTCAAGTGTGATGGCGCATCTGGTAGCGCGGGCTGAAGCTGAACCTCACGCAACAGCAGGGAGTGAAGCTCGCCCGCACACCCGG
 CGCAGCTGGAAGCCAGCAATGAGCGCGTCTGCTGAACCTGTCTTTATACACATCTCCGAG
 >A01940:43:HT3V3DSX3:3:1101:14018:1438 1:N:0:GTAGAGGAAT+TATGCAGTGT
 ATACTTTACGTTTCAAGTAAACATTTCAAGTTCTCCCTGTGCATATACCATTCTTAAACTTGCCAAATGATTTTTTAAATTTGAGT
 ATTAACATCTCAGGTTATTTCTTTTATGAATATACTCTCTCTCTTTATACACATCTCC
 >A01940:43:HT3V3DSX3:3:1101:16107:1454 1:N:0:GTAGAGGACT+TATGCAGTGT
 CAACGTAGGTTGATGTCACCCAGGAATACGTGACCGCAATCGTGGGGTCACTGTGAGGCCACAATCACGGCAGAACGTATCATCTAC
 CCAACACACATGCTCACTCATCTACTCTCCACATGCTACCAATTCGTGAGAGTACG
 >A01940:43:HT3V3DSX3:3:1101:24542:1470 1:N:0:GTAGAGGAAT+TATGCAGTGT
 TCGCTGGACCAATCCCAGCGCGGACCCACGCTCCGGAGGAGATCATCCATCTGCGGTATCTGGCGGATCCGGCCAACCCCATCATGG
 GCATCGGCCCGCTTACGGCGGGCAAAATCCGTGGATGTGGACAACGCCAGAGGGAGCACA
 >A01940:43:HT3V3DSX3:3:1101:5385:1485 1:N:0:GTAGAGGAAT+TATGCAGTGT
 TAAGAAATCGGGCAGCAATTTGCGCATCCTCGTATCTCAACCCAGTAAGCCAAAGCCAAATTTCTTATTGGGGTCTGAGACCAAGTT
 TATATAGATGGCTTGAATAGTGACCTTACTTTTGGCGATGCTCCAGCAAACCAATGGAGC
 >A01940:43:HT3V3DSX3:3:1101:22101:1501 1:N:0:GTAGAGGAAT+TATGCAGTGT
 GTCTAGACCAAGTACCACAGAGTATGCCCCATTTTTTCAAACCTATATCAATTTATACCAGTGGTAAGGATTATGAAATTTTAGTTAAA
 CAATACCATGACCGTTTAGTCGATTTCATGGAATTTCAATACCTCCAGAAAAAATTAAGTATGCC
 .
 .
 .

Paginas restantes 4,515,415



ANEXO 22. Data secuenciada correspondiente a la parte inicial de los contigs del ADN

bacteriano extraído de la muestra de agua en la Captación Chimú en

formato FASTQ

```
>A01940:43:HT3V3DSX3:3:1101:5403:1517 1:N:0:GTAGAGGACT+TATGCAGTGT
GTATGGGCGGTGAATGCTGAGAAGGTTTCGAGTAGCTGCCCTCGACGTAGCCTTTACCGCCGGTGGGACCGGATGCGTTCTCGCGTTCCG
GCACTGTTGGGAAGACGATCAACGGCTCTCGCCACCTAACTTTGACGAGGTCCTGACCATCT
>A01940:43:HT3V3DSX3:3:1101:25536:3098 1:N:0:GTAGAGGAAT+TATGCAGTGT
CTCTTAATTAATAAGCGTTAACTTAAAAAGTACGAAAGTACGAAAGTTTTAGCTATTTCTTTCACGTGAGTAATAACTTTTTATATT
TATTTACTCCTGTCTCTTATACACATCTCCGAGCCACGAGACGTAGAGGAATCTCGGTTGG
>A01940:43:HT3V3DSX3:3:1101:27887:3098 1:N:0:GTAGAGGAAT+TATGCAGTGT
GACTAACAGCGATGTTTGATGTCATCCAGCACCCACCTGGTAAAAAGACATCAATTTAACGTTTACACTAAAAGTATATCTGAAATTCC
GTGAACAGAATAACCAAAATTCAAATGTGTCCCACTTGATATCCAGGATTACCAATAATTGG
>A01940:43:HT3V3DSX3:3:1101:3757:3114 1:N:0:GTAGAGGAAT+TATGCAGTGT
GAAGAAAAGGCATTCTTGACAAACATCTTTCGTTTCTTTACGCAAGGTGACATTGACGTAGCTGGTGGTTATGTCAAGAAGTATCTAC
CACACTTCCACAACCAGAGATTCGTATGATGTTGTCTGGTTTGTCTGCACGTGAAGCATTAC
>A01940:43:HT3V3DSX3:3:1101:1217:3129 1:N:0:GTAGAGGAAT+TATGCAGTGT
TATTAGAAGCATATAAAGATATAAAAAAGCCACAAGTTGTGGTGCATTTTGTAGAAGATAAAATAAAGATAACAATGAAACTAAC
AATAAAAAACAGATTCGGAGTTGCTCCAAATGATCTATTAACAATACCAATATATCATTAAA
>A01940:43:HT3V3DSX3:3:1101:20491:3129 1:N:0:GTAGAGGAAT+TATGCAGTGT
GGACTGATGAGTGGACTCAGTTTTATATGCCGTCCGAATTTGAGTGTGACGAATGTCACATAGAATGTTGGAAGAAGACAACAGATG
TATTGACTGTGGCAAGAAATCTGTCATAAATTTGTGTACAGACAAGGAGCACACATGACCACA
>A01940:43:HT3V3DSX3:3:1101:25283:3129 1:N:0:GTAGAGGAAT+TATGCAGTGT
ACTAACTAACTTAAATAAAATGAAAAATAGTTGCAATTAACCTTCATAGATAACCATAAATGAATCAGTGTAAACCATTATCCTCATT
CTTAGTCTGTCTCTTATACACATCTCCGAGCCACGAGACGTAGAGGAATCTCGGTTGGGGG
>A01940:43:HT3V3DSX3:3:1101:6542:3145 1:N:0:GTAGAGGAAT+TATGCAGTGT
GTTGAGGTTTGAATGCGGTGATTGGGAAATCGGACCACTCCTGAGCGACAGATGGTTATTAACCTTCCGGCAACAGTTGAGATGAT
CACACCTAACGTTTACGACAGCTCAATCCTGTCTCTTATACACATCTCCGAGCCACGAGACG
>A01940:43:HT3V3DSX3:3:1101:9200:3145 1:N:0:GTAGAGGAAT+TATGCAGTGT
ACCTACTACAAGCTGAGCATCAACGGGATCGAATTTGGTGGAGATCGACCCTGTGAACATGATCCGCAAGATCAATGGCGTCGATCAGG
TAGCATCCATCCGAGCAGCGATCGGGGTTTGATCTATGGGCGCATCCGCTACGGCTCCGGAGC
>A01940:43:HT3V3DSX3:3:1101:12418:3145 1:N:0:GTAGAGGAAT+TATGCAGTGT
CCATGTGGATAAATGTTAGGTAATTTGCATTACTTTCATCATCTCGTAAAGTATGTCGGTATTCGGTTATTAGTGAATTTGAAATCCC
GTTATCGTATAGTACTGCTTTTTTCCACCTGTCTTGTATACCATAGCTGTCTCTTATAAACATA
>A01940:43:HT3V3DSX3:3:1101:30047:3145 1:N:0:GTAGAGGAAT+TATGCAGTGT
CTGCGGATTACTCTCAGATTGAAATGCGAATCATGGCTCATTGTCTGAGGACCCGTGGCCTGATTGAGGCATTCAACGCAGGAGAGGA
TCTCCATAGATCTGTCTCTTATACACATCTCCGAGCCACGAGACGTAGAGGAATCGCGGGTG
>A01940:43:HT3V3DSX3:3:1101:26277:3161 1:N:0:GTAGAGGAAT+TATGCAGTGT
ATCTCAAACCCAGGCACTAATGAGTGGGAAGTAGGCATTGGTACAGTAAAGTGTGGTCAATTAGCCCGTACAACGATTCTTTCGCTTT
CTAATAGCGGTTCAATTTGTTAACTTTAGTGTGGTACTAAGGATGTGTTTGTACTTACCAG
>A01940:43:HT3V3DSX3:3:1101:27073:3161 1:N:0:GTAGAGGAAT+TATGCAGTGT
GGCCCAATCCGCTCGATGGCCATCATGACCAAGAGCACCCGGACGACCGGTGCACGCGTGGCGTTGAGCAGCGACAGCCAAAGAACCG
GCTCGGGCACCAAGCCGAAGCCAAAGGCCTGAGCAACACAACTGCAGCAACAAAAACCCG
>A01940:43:HT3V3DSX3:3:1101:6036:3176 1:N:0:GTAGAGGAAT+TATGCAGTGT
ACCGTTTCTGATGATTAACCTGTCCCTGAATGACTAAATAAAGACTGACGTAAGTTTGTGGTATTGAAGATAACGTAACGAACTG
CCTGAAAAGACTGTAGTACTAATAAAGTCAATTCACCACGAAATGTCACCCCATGCAGGA
>A01940:43:HT3V3DSX3:3:1101:7934:3176 1:N:0:GTAGAGGAAT+TATGCAGTGT
ATGTGGATAAGATTAATAAATATGTTACCGTACCTTATACCGCCATGCAAGTGGTTCGTAACCGCGCGCATGGAAGAAGAAACAGATAC
AAATTAACCTAGGATAAAAAATGGCGATTGCAAAACACAACACAAGTATTTGGTTGATACAAAATA
>A01940:43:HT3V3DSX3:3:1101:10140:3176 1:N:0:GTAGAGGAAT+TATGCAGTGT
CGCTTGCAGCCACCCAGAGGCCAATGCGGACACCTCTTTTCCGGCGCCGACCCAGCCAGCCGCAAGCGCATCATCGAGGCAATCGCC
GCGATTGAAAATGCACCATTGGCTAAGCCCAATAAAAAATACGTTAAACGCAGTGGCCACTCA
>A01940:43:HT3V3DSX3:3:1101:14443:3176 1:N:0:GTAGAGGAAT+TATGCAGTGT
CAATATGCAGGTGCAATAATGTTCTGAACCAAGTTTAAATCCTGGATTCTGGAAGTGGAAAAAGTACTGGAATACGAACACTTCCAC
CAGAACAACATTTTTATTTAATGTTTGGGGGAGGACTTGCCTTTTTAAAGGAGCACGTAATA
>A01940:43:HT3V3DSX3:3:1101:21694:3176 1:N:0:GTAGAGGAAT+TATGCAGTGT
GATCTGGACATAGCCAAAGCGCTCGTTTCAGCTTCTTCGCGCTTTCTTCAACCGCCGCCCCACTCGGCTTGTGTTCCGCGCAAGGGG
AGCGCTCCGAGGATGCTTTCCGCGCGACCAAGAGTTTTGGTTGCTCGTTCGAGTTGGCCCTTG
>A01940:43:HT3V3DSX3:3:1101:29776:3176 1:N:0:GTAGAGGAAT+TATGCAGTGT
AAACAGCACCATTGCTGCCATGTTGTTAATTATACAGTTATTTAGTTCTCTTGTCAACCTCTTTTTAATCTTTTGTACTAAAAAGTTTT
TAACAACTGAACCTTCTTAACGTATGCCTTAATTTGTACAGTTATTTAGTCTCTTTGTCAACCAC
>A01940:43:HT3V3DSX3:3:1101:5593:3192 1:N:0:GTAGAGGAAT+TATGCAGTGT
ATGTTTACCATCCGCGTCGCCACTCCTGCTGCCAACCGCAAGCTTCGTCCGATCTTTGTGAGAAGAAGTCCATGGCAGATCAATA
CGATTGGATTCAAAAACCTTGAAGTTGAAGGCTTGTGACACCATCGGTGAGCATTTGCTGCA
>A01940:43:HT3V3DSX3:3:1101:9263:3192 1:N:0:GTAGAGGAAT+TATGCAGTGT
TAGCGTCTGTGTTGAGTTCGTTTGGTCCATCGTGTATAATTTGTTGTATGGGCTCTTTTGTACTCTTAAACATATGCTTGAACATA
CAAAAATTAATCAATGCAATTCCTGTCTTATACACATCTCCGAGCCACGAGACGTAGAGGA
>A01940:43:HT3V3DSX3:3:1101:10818:3192 1:N:0:GTAGAGGAAT+TATGCAGTGT
TTATTGAAGTAGCGCAACAAAGTTTTACTCGAAGAAAATGAACGCATCAAAAACAGAAAACGGAGAATCACTGTTGACGTATTTACGACG
TAATGTTGGTGGTTCAAAATTTAGGACAAACGGGGAATCACGTTGGAAATCTTTCGGTTTT
>A01940:43:HT3V3DSX3:3:1101:12970:3192 1:N:0:GTAGAGGAAT+TATGCAGTGT
```



GTTTGAGTTACAGGGACGTTGCTACTAATTCATTTAAAAATTAATTCGGCTATGTTGATACGTTATCGTAAAATGAGTTTTCCAGGAG
AAAAATTCCTGCGTTTCTAGAAAATTTGAATGGTTAAGGATCAAAATACGAAATTTCCCTG
>A01940:43:HT3V3DSX3:3:1101:27181:3192 1:N:0:GTAGAGGAAT+TATGCAGTGT
AGCTTGACACGCATTTCCCGTGTCTCAGGACTGAGCTGTGGAAAACGGATGTAATGGTCCACATCCCCATGAAGACATGCCATAGC
CCCCTTACGAATTTCCAGACAGCGATCAAAGAACTCATCTGAGCAGCACCCCTGTTTCAGAGTG
>A01940:43:HT3V3DSX3:3:1101:1244:3208 1:N:0:GTAGAGGAAT+TATGCAGTGT
AAATACTGTTTTAGAGATTGCATCCGTTTACGATTCATTTTGGCTTCAACAAAATTAACCTAACGATAAAGGCTGTCTCTTATAC
ACATCTCCGAGCCCACGAGACGTAGAGGAATCTCGGGGGGGGGGGGGGGGGGGAAGGGGG
>A01940:43:HT3V3DSX3:3:1101:10212:3208 1:N:0:GTAGAGGAAT+TATGCAGTGT
CGTTGCGCGCCACCCAGAGGCCAATGCGGACACCCCTTTTCCGGCGCCGCCAGCCACCGCAAGCGCATCATCGAGGCAATCGCG
GCGATTGAAAATGACCCATTGGCTAAGCCCAATAAAAATACGTTAAAACGCGAGTGGCCACTCA
>A01940:43:HT3V3DSX3:3:1101:24496:3208 1:N:0:GTAGAGGAAT+TATGCAGTGT
GTTCTGTAGTGCCTAATGATTTTCAGAACGTAGACCTGCGCAGCTGAGAGTAGTGCAGAGTTGACTTCAAAGGAAGGGTTTTTCAGTGGT
CGGCCAATAAAGGTAAACAAGCCAGACTCAACATGTGGCAACAAAGCGTCTTGTGGCTTTT
>A01940:43:HT3V3DSX3:3:1101:6442:3223 1:N:0:GTAGAGGAAT+TATGCAGTGT
CATATTTGCTCTATGAAAGCCTGATTATGGCGTCAGATGCTGTGCGCGTTGGGAAAGTACGGTAAATGTGCTGTTGCAGTTTTGT
CCGCACAAAATCTAATACAGCTACAGCAGCACCGGTTGCATATTGTAGATTAAACGCGCCC
>A01940:43:HT3V3DSX3:3:1101:1208:3239 1:N:0:GTAGAGGAAT+TATGCAGTGT
AAATACTGTTTTAGAGATTGCATCCGTTTACGATTCATTTTGGCTTCAACAAAATTAACCTAACGATAAAGGCTGTCTCTTATAC
ACATCTCCGAGCCCACGAGACGTAGAGGAATCTCGTGTGGGGTGTGGGGGGGTTAATAGGG
>A01940:43:HT3V3DSX3:3:1101:11406:3239 1:N:0:GTAGAGGAAT+TATGCAGTGT
TTCACATCTTCTCCAAAACAGGTACTCAAACACGCTCTCGATCTGCTCAATGCGGAAGTCCGGCATCCCATCTTTTCGGGGTATA
CAAAAAGAGTATCTTTACCATTGTCACAAAGCCCTTCAAGTATTCTCTCGCTCAAGTCAA
>A01940:43:HT3V3DSX3:3:1101:1145:3255 1:N:0:GTAGAGGAAT+TATGCAGTGT
TATTAGAAGCATATAAAGAGTATAAAAAGAAGCCACAAGTTGTGGTGCATTTTGTAGAAGATAAATAAAGATAACAATGAACTAAC
AATAAAAACAGATTCGGAGTTGCTCCAAATGATCTATTAACAATACCAATATATCATTA
>A01940:43:HT3V3DSX3:3:1101:9625:3255 1:N:0:GTAGAGGAAT+TATGCAGTGT
CTCAACAACGAAGAGGCAGCCGAAGTCTCGTAAGAACGGTGTGAAGAGTTGGTACGCCCCGTCGAATGATGAAGTCAATTTCCG
TAAAGACCGGCACCTGTCTCTTATACACATCTCCGAGCCACGAGACGTAGAGGAATCTCGGG
>A01940:43:HT3V3DSX3:3:1101:25030:3255 1:N:0:GTAGAGGACT+TATGCAGTGT
GTGTAGCCGTTTCTGGGTGGACAACCCAGATGGCGGAAATCAAGCCGTTGGTTACAGTCCCGTGGCAAAAAGTAAATGAATCAAACCT
GTCAGCTGTTTCCATTGCTATCCAACAGCAAGCTGTTGGTGTACAACCCGGGTGGAAC
>A01940:43:HT3V3DSX3:3:1101:20211:3270 1:N:0:GTAGAGGAAT+TATGCAGTGT
CATCAATATAGTATGGATCCAGCATCAAGATAATGGCGCAGCCTTACAGAACAGCGAGTGTAGGTAATCAAGGGTAAATCGACCTA
GGTAAACCTCAATGAAATCAAGGGTTTTAAGCGGTATCGACCTAGGGCAAAACCTTAGTAAATA
>A01940:43:HT3V3DSX3:3:1101:23592:3270 1:N:0:GTAGAGGAAT+TATGCAGTGT
GTTCTGTAGTGCCTAATGATTTTCAGAACGTAGACCTGCGCAGCTGAGAGTAGTGCAGAGTTGACTTCAAAGGAAGGGTTTTTCAGTGGT
CGGCCAATAAAGGTAAACAAGCCAGACTCAACATGTGGCAACAAAGCGTCTTGTGGCTTTT
>A01940:43:HT3V3DSX3:3:1101:23683:3270 1:N:0:GTAGAGGAAT+TATGCAGTGT
AAATAACATATGATAAAGCATTTGTTTAATACTATTTTGTGTTTGTGTTGGGATACATGCTGCCATGCACAATCAGAATCAACGGTT
GTATATATTGAGCAGTATAAAAATCTGTCTTTATACACATCTCCGAGCCACGAGACGTA
>A01940:43:HT3V3DSX3:3:1101:27064:3270 1:N:0:GTAGAGGAAT+TATGCAGTGT
ATACATTTGCTGGATCGGATTTGTGAGGGGTTGTCAAAGTTATGGAATGACATCAGAATTTGGCAATGCCTCAGAATTGCAACCG
AGCGTTTTGGCTCTGTCTCTTATACACATCTACCAGCCACAGACGTAGAGGAATCGCGGG
>A01940:43:HT3V3DSX3:3:1101:1741:3286 1:N:0:GTAGAGGAAT+TATGCAGTGT
GTACCTATAGCACCGCCAATAGCGCCCCAGCAACTGTCCAATACCTGTAGATAAAGCACTACCTGTACCAATACCTACGTTTTGAG
CTGCTAACATTGCTTTGTTGTGAGCCAATTGAGGTTCCAAACTGCTTAACGTCCCAACAT
>A01940:43:HT3V3DSX3:3:1101:4634:3286 1:N:0:GTAGAGGAAT+TATGCAGTGT
GCACGGGAATCCTAACCTTGCCTTTTGAACCTTACGGACAGGTGAACCTTCCGAGAGCACTCGACTTAAGTTTATACCGCTTCAGA
CGGGCACTGGGGCAACCCAAATACAGATTTTGTGATCGCATCGGCAACTGATTGAAGCGAT
>A01940:43:HT3V3DSX3:3:1101:21178:3286 1:N:0:GTAGAGGAAT+TATGCAGTGT
TCATTGGCTGGATCTCTGATGTGACTCATATAGAACCCTTTTGGTGCATCTGCCGTGATCTTTTCTCTGTTCTATCTCTGGTCTA
TCAAATCCCAGAGACCAGCAATGTCGAAGGGCCCCAGGTTAAAACCTGAAGCCCCGCGGGGA
>A01940:43:HT3V3DSX3:3:1101:3911:3317 1:N:0:GTAGAGGAAT+TATGCAGTGT
ATTTAAACCTATGTCATAGTGTGTGCTAGTATAAACAATTTAAGAGGTTTAAAGCGATGGCAGTAAATAGAATCAATGAATTTCTGG
AAGCGCTGAAAGTTTTCTACAGACTTAGAAACCCCTTGTCAAACCTGAGTTAAATATCTTGACG
>A01940:43:HT3V3DSX3:3:1101:11849:3317 1:N:0:GTAGAGGAAT+TATGCAGTGT
GTACTAGGTTCAATCAATACACAGAAGTCAACAGTAACTAAGTACACAGATGCCGACGGCTTCGTTTCAATTTATGCTGGTTCAAATG
CTGCTGGTGAACACCATCAATTTACATCAGGAAGTCTACAACATCAGGTGCATTTCTGGA
>A01940:43:HT3V3DSX3:3:1101:13277:3317 1:N:0:GTAGAGGAAT+TATGCAGTGT
GCAGAACCCTTCTTGGTCCGCAATGGCTTTATGGCCAGGACTCCTCGGGACCGGTGGCTACAGCCCTTGGCTACTCCCACCTGGGGC
GGCCAGTACCAGCGGGTTTAGCAACGCTTTTTGATCCCATCTCTGAGTAATCTCACATCTGAC
>A01940:43:HT3V3DSX3:3:1101:14434:3317 1:N:0:GTAGAGGAAT+TATGCAGTGT
CAATATGCAGGTGCAATAATGTCTGAACCAAGTTTAACTCCTTGGATTTTCTGGAAGTGGAAAAAGTACTGGAATACGAACACTTCCAC
CAGAACAACATTTTTATTTAATGTTTCCGGGAAGGAATGCTTTTTAAAGGAGCACGTAAAA
>A01940:43:HT3V3DSX3:3:1101:23502:3333 1:N:0:GTAGAGGAAT+TATGCAGTGT
CAGCAGGTCATCGCTGATGCCGCGGACCGTGGTGGCTTATTGATTGGCGTTGCTGGCACCACGGAAAGACCCTCGACGGGAT
GGGTGCTCGATGCCCTGATGCCGTTCCGGGATTGATCTCTCTGCAATTTGTTGGCGCCCACTCG
>A01940:43:HT3V3DSX3:3:1101:7726:3380 1:N:0:GTAGAGGAAT+TATGCAGTGT
TCTCCATACCTTCTACAATCTGTGGATTGAGACAACTTCTCAATAGCCACACCCATCGCATCATCCACAGACACTGCGTTAGGGTC
AATTAAGAAGTCTTAGGATTAAGTGGTGTACTTAAACAGAAAGTAGTCTTCTCTTGTAC
>A01940:43:HT3V3DSX3:3:1101:10719:3396 1:N:0:GTAGAGGAAT+TATGCAGTGT
.
.
.
Paginas restantes 4,680,873



ANEXO 23. Constancia de secuenciamiento por Shotgun de muestras de ADN presentes en muestras de agua la Captación Chimú y Bahía Interna del Lago Titicaca por BTS consultores.



Carta N° 004-2024

Lima, 25 de enero de 2024

Señores

Megalaboratorio de Biología Molecular

Universidad Nacional del Altiplano – UNA – Puno

Ref. Secuenciamiento por shotgun de muestras de ADN

Apreciados investigadores,

Luego de su solicitud verbal para realizar una constancia de que sus muestras de ADN purificado de la Bahía interior, codificado como "BI" y la Captación Chimú codificado como "CH" del Lago Titicaca de la UNA Puno, fueron secuenciados por medio de la empresa BTS Consultores.

Como gerente de dicha empresa certifico que se realizó dicho secuenciamiento de las muestras de ADN que los investigadores del Megalaboratorio de Biología Molecular de Universidad Nacional del Altiplano – UNA – Puno llevaron a nuestras instalaciones y se obtuvieron las datas correspondientes con las siguientes características:

Data de la Bahía Interior (BI):

| | | |
|----|-------------------------------------|--------------------|
| 1. | 20008D-10-01_S204_L003_R1_001.fastq | Peso de 437,019 KB |
| 2. | 20008D-10-01_S204_L003_R2_001.fastq | Peso de 447,053 KB |

Data de la Captación Chimú (CH):

| | | |
|----|-------------------------------------|--------------------|
| 1. | 20008D-10-02_S205_L003_R1_001.fastq | Peso de 453,873 KB |
| 2. | 20008D-10-02_S205_L003_R2_001.fastq | Peso de 461,377 KB |

Agradecemos su comprensión,

Cordialmente,

LUIS ENRIQUE BARAVIA CAHUANA
GERENTE DE I+D+i - CTO R&D+i
BTS CONSULTORES S.A.C.
RUC 20562842145



ANEXO 24. Constancia de ejecución del proyecto de investigación.

CONSTANCIA DE EJECUCION DE PROYECTO DE INVESTIGACION

El que suscribe, Dra. MARIA TRINIDAD ROMERO TORRES. Jefe del laboratorio de Biología Bolecular, Megalaboratorio – UNA – PUNO Otorga la presente constancia de ejecución de proyecto de investigación a:

Abel Vidal Mayta Mendizabal, bachiller en Biología, de la Facultad de Ciencias Biológicas de la Universidad Nacional Del Altiplano Puno, código de matricula 161454 identificado con N° DNI: 74484205.

Quien ha realizado la ejecución de su proyecto de investigación denominado **PERFIL TAXONOMICO DE MICROBIOMAS BACTERIANOS CON LA SECUENCIACION GENOMICA SHOTGUN PRESENTES EN LA CAPTACION CHIMU Y LA BAHIA INTERNA DEL LAGO TITICACA, PUNO, PERU, 2022.** tendientes a la obtención del título profesional, bajo mi supervisión, desde el 03 de Abril al 30 de Julio del 2023.

El tesista Abel Vidal Mayta Mendizabal realizo su ejecución de investigación a completa satisfacción y mostro en todo momento eficiencia, puntualidad y responsabilidad y buena formación académica.

Se otorga la presente constancia para fines que el interesado considere conveniente.

Puno, 25 de Octubre del 2023.

Dra. Maria Trinidad Romero Torres



AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo Abel Vidal Mayta Mendizabal,
identificado con DNI 74484205 en mi condición de egresado de:

Escuela Profesional, Programa de Segunda Especialidad, Programa de Maestría o Doctorado
De Biología.

informo que he elaborado el/la Tesis o Trabajo de Investigación denominada:

“Perfil Taxonomico de microbiomas bacterianos con la
seuenciación shotgun presentes en la captación Chimú y
bahía Interna del lago Titicaca, Puno, Perú, 2023”

para la obtención de Grado, Título Profesional o Segunda Especialidad.

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los “Contenidos”) que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

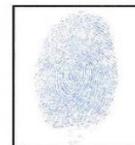
Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

En señal de conformidad, suscribo el presente documento.

Puno 27 de Mayo del 2024

FIRMA (obligatoria)



Huella



DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo Abel Vidal Mayta Mendizabal,
identificado con DNI 74484205 en mi condición de egresado de:

Escuela Profesional, Programa de Segunda Especialidad, Programa de Maestría o Doctorado
De Biología

informo que he elaborado el/la Tesis o Trabajo de Investigación denominada:
" Perfil taxonomico de microbiomas bacterianos con la
secuenciación shotgun presentes en la captación Chimú y
bahía Interna del lago Titicaca, Puno, Perú, 2023. "

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 27 de Mayo del 20 24

FIRMA (obligatoria)



Huella