



**Universidad Nacional Mayor de San Marcos**

**Universidad del Perú. Decana de América**

**Facultad de Ciencias Biológicas**

**Escuela Profesional de Genética y Biotecnología**

**Identificación y distribución de genes asociados a la  
resistencia a arsénico en bacterias de ambientes  
marinos a nivel *in silico***

**TRABAJO DE INVESTIGACIÓN**

Para optar el Grado Académico de Bachiller en Genética y  
Biotecnología

**AUTOR**

César Bryan REYES MORENO

**ASESOR**

Dr. Pablo Sergio RAMÍREZ ROCA

Lima, Perú

2021



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

## Referencia bibliográfica

---

Reyes, C. (2021). *Identificación y distribución de genes asociados a la resistencia a arsénico en bacterias de ambientes marinos a nivel in silico*. [Trabajo de investigación de bachiller, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Biológicas, Escuela Profesional de Genética y Biotecnología]. Repositorio institucional Cybertesis UNMSM.

---

## Metadatos complementarios

<b>Datos de autor</b>	
Nombres y apellidos	César Bryan Reyes Moreno
Tipo de documento de identidad	DNI
Número de documento de identidad	78202738
URL de ORCID	<a href="https://orcid.org/0000-0001-7346-2917">https://orcid.org/0000-0001-7346-2917</a>
<b>Datos de asesor</b>	
Nombres y apellidos	Pablo Sergio Ramírez Roca
Tipo de documento de identidad	DNI
Número de documento de identidad	06183797
URL de ORCID	<a href="https://orcid.org/0000-0001-9309-7021">https://orcid.org/0000-0001-9309-7021</a>
<b>Datos del jurado</b>	
<b>Presidente del jurado</b>	
Nombres y apellidos	Jaime Roberto Sánchez Venegas
Tipo de documento	DNI
Número de documento de identidad	06120091
<b>Miembro del jurado 1</b>	
Nombres y apellidos	Edith Fanincia Rodriguez Quispe
Tipo de documento	DNI
Número de documento de identidad	09202308
<b>Miembro del jurado 2</b>	
Nombres y apellidos	Miguel Angel Francisco Talledo Rivera
Tipo de documento	DNI
Número de documento de identidad	25625144
<b>Datos de investigación</b>	

Línea de investigación	Biorremediación y Bioconversión
Grupo de investigación	Genómica Funcional de Microorganismos y Biorremediación (MICROGEN)
Agencia de financiamiento	Perú. Vicerrectorado de Investigación y Posgrado de la Universidad Nacional Mayor de San Marcos. “Promoción de trabajo de Investigación para optar el grado de Bachiller”. B20100050a .
Ubicación geográfica de la investigación	Universidad Nacional Mayor de San Marcos Latitud: -12.059736 Longitud: -77.082051
Año o rango de años en que se realizó la investigación	2020 - 2021
URL de disciplinas OCDE	Biología celular, Microbiología <a href="https://purl.org/pe-repo/ocde/ford#1.06.01">https://purl.org/pe-repo/ocde/ford#1.06.01</a> Bioquímica, Biología molecular <a href="https://purl.org/pe-repo/ocde/ford#1.06.03">https://purl.org/pe-repo/ocde/ford#1.06.03</a> Genética, Herencia <a href="https://purl.org/pe-repo/ocde/ford#1.06.07">https://purl.org/pe-repo/ocde/ford#1.06.07</a>



**Universidad Nacional Mayor de San Marcos**  
(Universidad del Perú, Decana de América)

**FACULTAD DE CIENCIAS BIOLÓGICAS**

**ACTA DE SESIÓN PARA OPTAR AL GRADO ACADÉMICO DE  
BACHILLER EN GENÉTICA Y BIOTECNOLOGÍA  
(MODALIDAD: SUSTENTACIÓN VIRTUAL DE TRABAJO DE INVESTIGACIÓN)**

Siendo las 10:06 horas del 06 de octubre de 2021, en el Salón de Grados Virtual de la Facultad de Ciencias Biológicas cuya dirección electrónica fue <https://meet.google.com/bhb-xcpx-tww>, y en presencia del Jurado formado por los profesores que suscriben, se inició la sesión para optar al **Grado Académico de Bachiller en Genética y Biotecnología** de **CÉSAR BRYAN REYES MORENO**.

Luego de dar lectura y conformidad al expediente N° UNMSM-20210042796, el graduando expuso su Trabajo de Investigación: **“IDENTIFICACIÓN Y DISTRIBUCIÓN DE GENES ASOCIADOS A LA RESISTENCIA A ARSÉNICO EN BACTERIAS DE AMBIENTES MARINOS A NIVEL *IN SILICO*”**, y el Jurado efectuó las preguntas del caso calificando la exposición con la nota **19**, calificativo: **APROBADO CON MÁXIMOS HONORES**.

Finalmente, el expediente será enviado a la Escuela Profesional de Genética y Biotecnología y al Consejo de Facultad para que se apruebe otorgar el **Grado Académico de Bachiller en Genética y Biotecnología** a **CÉSAR BRYAN REYES MORENO** y se eleve lo actuado al Rectorado para conferir el respectivo grado, conforme a ley.

Siendo las 11:55 horas se levantó la sesión.

Ciudad Universitaria, 06 de octubre de 2021.

**Dr. JAIME SANCHEZ VENEGAS**  
(PRESIDENTE)



**UNMSM**

Firmado digitalmente por RAMIREZ  
ROCA Pablo Sergio FAU  
20148092282 soft  
Motivo: Soy el autor del documento  
Fecha: 06.10.2021 11:44:03 -05:00

**Dr. PABLO RAMIREZ ROCA**  
(ASESOR)

**Mg. EDITH RODRIGUEZ QUISPE**  
(MIEMBRO)

**Mg. MIGUEL TALLEDO RIVERA**  
(MIEMBRO)

## DEDICATORIA

El presente trabajo está dedicado a todas las personas que me apoyaron, motivaron y/u orientaron en realizar este trabajo de investigación. Con todo mi ser y emoción hago énfasis en dedicar este trabajo a mis padres, María Esther Moreno Espejo y César Américo Reyes Huamán, quienes me han criado y animado a seguir una carrera de ciencias básicas. Asimismo, dedico este trabajo a mis mentores en investigación quienes me han guiado con ahínco y amabilidad en esta investigación: Dr. Pablo Ramírez Roca, que también es mi asesor en esta investigación; al Dr. Ramiro Logares y al estudiante de doctorado Fran Latorre, quienes me asesoraron en mi estancia en el Instituto de Ciencias del Mar – Barcelona. Adicionalmente, dedico este trabajo a los miembros del Laboratorio de Microbiología Molecular y Biotecnología de la Facultad de Ciencias Biológicas de la UNMSM por cederme un ambiente agradable en donde realizar investigación por pasión y curiosidad. Finalmente, termino esta dedicatoria a todos los lectores de este trabajo con una cita de María Salomea Skłodowska-Curie, conocida como Marie Curie, ya que refleja mi experiencia vivida en el transcurso de esta investigación:

*“La vida no es fácil para ninguno de nosotros. Pero... ¡qué importa! Hay que perseverar y, sobre todo, tener confianza en uno mismo. Hay que sentirse dotado para realizar alguna cosa y que esa cosa hay que alcanzarla, cueste lo que cueste.”*

## **AGRADECIMIENTO**

Agradezco al programa de “Promoción de trabajo de Investigación para optar el grado de Bachiller”, organizado por el Vicerrectorado de Investigación y Posgrado de la Universidad Nacional Mayor de San Marcos en el 2020 con Resolución Rectoral N.º 015156-R-20, por el apoyo financiero brindado mediante el proyecto B20100050a para el desarrollo de este trabajo de investigación.

Agradezco a mi asesor y mentor en investigación, el Dr. Pablo Ramírez Roca, por sus consejos, orientación y apoyo mostrado durante todo el desarrollo de este trabajo de investigación.

Agradezco a mis mentores de investigación durante mi estancia en el ICM – Barcelona, al Dr. Ramiro Logares y al estudiante de doctorado Fran Latorre, quienes me asesoraron y apoyaron amablemente en la iniciativa de esta investigación.

Agradezco a mis padres, César Américo Reyes Huamán y María Esther Moreno Espejo, por la motivación y apoyo que me han brindado para realizar satisfactoriamente mis estudios de pregrado en la Escuela Profesional de Genética y Biotecnología, reflejando una muestra de lo aprendido en este trabajo de investigación. Asimismo, agradezco el apoyo emocional, económico y mental de mi familia en general: mi hermano Junior Tetsuhiko Reyes Moreno, mi tía Nancy Moreno Espejo, mi tía Maritza Moreno Espejo, mi tía Ana Moreno Espejo, mi tío Oscar Moreno Espejo, mi primo Gianfranco Moreno Vega, mi primo Joseph Moreno Vega, mi tía Elsa Vega, mi tía Marlene Reyes Huamán, a mi tía Raissa Reyes Huamán, a mis abuelos Lorenzo Reyes, Paulina Huamán, Oscar Moreno y Leocadia Espejo.

Agradezco también a mis amigos y compañeros del Laboratorio de Microbiología Molecular y Biotecnología: Carito, Gustavo, María, Hedersson, Mario, Wilma, Jaime, Danayra y Jefferson por su amistad incondicional y el agradable ambiente de trabajo en el que me permitieron desarrollar mi trabajo de investigación.



## CONTENIDO

	Pág.
Resumen .....	1
I. INTRODUCCIÓN.....	2
1.1. Planteamiento del problema de investigación .....	2
1.2. Pregunta del problema .....	3
1.3. Justificación de la investigación .....	3
1.4. Objetivo del trabajo .....	4
1.4.1. Objetivos del documento.....	4
1.5. Hipótesis.....	5
II. ESTADO DE CONOCIMIENTO.....	6
III. MATERIALES Y MÉTODOS .....	13
3.1. Materiales .....	13
3.2. Métodos.....	13
3.2.1. Diseño de base de datos de genes de interés.....	13
3.2.2. Análisis de secuencias nucleotídicas usando <i>clusters</i> bioinformáticos basados en Bash/Linux.....	13
3.2.3. Búsqueda y anotación de genes vía MMseqs2 en base de datos de metagenomas .....	13
3.2.4. Filtrado de las secuencias alineadas con la base de datos de metagenomas mediante Python.....	14
3.2.5. Relación de la distribución y abundancia de los genes de interés en las estaciones muestreadas del océano global usando Rstudio. ....	15
3.2.6. Relación de la abundancia de genes de interés y profundidad del ecosistema marino según las estaciones de Malaspina usando Rstudio.....	16
IV. RESULTADOS .....	17
4.1. Base de datos creada de genes de interés.....	17
4.2. Relación de la distribución y abundancia de los genes de interés en las estaciones muestreadas en océano global usando Rstudio. ....	19
4.3. Relación de la abundancia de genes de interés y profundidad del ecosistema marino según las estaciones de Malaspina usando Rstudio.....	25
V. DISCUSIÓN.....	27
5.1. Identificación de genes asociados a la resistencia a arsénico .....	27
5.2. Distribución y abundancia de genes asociados a la resistencia al arsénico en los datos de Tara Oceans.....	28
5.3. Distribución y abundancia de genes asociados a la resistencia al arsénico en los datos de Malaspina 2010.....	30
VI. CONCLUSIONES.....	32

VII.	REFERENCIAS BIBLIOGRÁFICAS .....	33
VIII.	ANEXOS .....	39

## Resumen

La toxicidad y ubicuidad del arsénico ha influenciado en la evolución de los microorganismos modernos, desarrollando mecanismos de respuesta al estrés a arsénico. Sin embargo, hay limitada información de dichos mecanismos a nivel genético en bacterias y su distribución en los ecosistemas marinos y oceánicos a nivel global. Por lo cual, este estudio tiene como objetivo analizar los patrones de distribución de genes asociados a la resistencia a arsénico en bacterias de ambientes marinos y oceánicos usando la base de datos de Tara Oceans y Malaspina 2010. Para ello, se construyó una base de datos de genes asociados a la resistencia a arsénico a partir de la base de datos de NCBI con el fin de compararlo con la base de datos de bacterias de ambientes marinos de las expediciones de Tara Oceans y de Malaspina 2010. Mediante la aplicación de herramientas bioinformáticas como usearch/10.0.240, MMseqs2, python y Rstudio, se identificó 4 tipos de genes asociados a la resistencia a arsénico (*arsJ*, *arsI*, *arsA* y *arsB*), destacando el gen *arsJ* con mayor abundancia génica y mayor distribución en diferentes océanos y mares. Según los datos de Tara Oceans, el gen *arsJ* de *Marinobacter* sp. se distribuye en 3 océanos (Pacífico, Atlántico e Índico) y en 9 mares de las tres capas oceanográficas de profundidad (capa de agua superficial, capa de máxima clorofila profunda y capa mesopelágica). Adicionalmente, según los datos de Malaspina 2010, el gen *arsJ* también incrementó su presencia a mayor profundidad oceánica hasta el batipelágico, distribuyéndose en el Océano Pacífico y en el Océano Índico. La presente investigación contribuye a determinar posibles lugares donde hay potencial de biotransformación bacteriana del arsénico a partir de sus genes identificados, que aún debe ser confirmado por futuros ensayos experimentales de biorremediación.

**Palabras clave:** arsénico, genes, bacterias resistentes a arsénico, ambiente marino, bioinformática.

## I. INTRODUCCIÓN

### 1.1. Planteamiento del problema de investigación

El arsénico es un metaloide ubicuo en la biosfera marina y terrestre (Fru et al., 2018). No obstante, la concentración del arsénico en un medio puede ser tóxico para los seres vivos, dependiendo de su estado de oxidación y metilación (Rahman & Hassler, 2014). Esto es una problemática mundial, debido a ciertas actividades humanas e industriales que tienden a aumentar el arsénico móvil al desechar cargas considerables del metaloide soluble en cuerpos hídricos (Mirza et al., 2010). Con respecto al estado actual del océano, no se ha considerado como un sistema cargado de arsénico; no obstante, hay presencia de estas especies de arsénico e incluso un desequilibrio termodinámico entre ellas que puede deberse a las biotransformaciones microbianas (Saunders et al., 2019). Debido a que los microorganismos han desarrollado varios tipos de mecanismos para reducir los efectos dañinos del arsénico, se han descrito diferentes genes involucrados en la resistencia a arsénico, principalmente, en ambientes altamente contaminados con metales pesados como estromatolitos, lagos y ríos (Bhattacharjee & Rosen, 2007). Sin embargo, hay limitada información sobre qué genes están presentes en bacterias marinas involucradas en este proceso y cómo se distribuye en los ambientes marinos y oceánicos a nivel global (Chen & Shao, 2009). Un factor que ha contribuido a esta limitación de información es la fracción de bacterias en el medio ambiente que permanece sin cultivar mediante técnicas de cultivo de laboratorio (Das et al., 2017; Chauhan et al., 2009). Ante esta problemática, el presente estudio busca identificar los genes asociados a la resistencia a arsénico en microorganismos marinos y visualizar la distribución de dichos genes en los océanos y mares, usando las bases de datos de Tara Oceans (Karsenti et al., 2011) y de la Expedición de la circunnavegación Malaspina 2010 (Duarte, 2015; Ortega et al., 2019) que son iniciativas a gran escala que exploran la inmensa riqueza de diversidad genética y funcional mediante secuenciación de metagenomas.

## **1.2. Pregunta del problema**

¿Cómo están distribuidos los genes asociados a la resistencia a arsénico de bacterias marinas, en los océanos y mares según los datos de Malaspina y Tara Oceans?

## **1.3. Justificación de la investigación**

El arsénico es un componente natural en los ecosistemas terrestre y acuáticos del mundo (Kurth et al., 2017). Incluso, desde el comienzo de la vida, el arsénico ha estado ampliamente distribuido en el ambiente, cuya toxicidad ha influenciado en la evolución de los organismos vivos modernos, respondiendo al estrés de arsénico (Chen et al., 2016). Reportando que el arsénico puede ser un agente responsable de graves problemas de salud pública (Andres & Bertin, 2016). No obstante, dependiendo de las condiciones fisicoquímicas y de las actividades biológicas, el arsénico se puede encontrar en ambientes acuáticos en varios estados diferentes de oxidación y metilación, mostrando varios niveles de toxicidad y biodisponibilidad (Rahman & Hassler, 2014). Frente a la toxicidad del arsénico, los microorganismos han desarrollado múltiples mecanismos para la detoxificación de arsénico orgánico e inorgánico (Chen et al., 2016). Por lo tanto, los procesos microbiológicos presentan un rol importante en el ciclo biogeoquímico del arsénico (Gadd, 2010). La tecnología de secuenciación del genoma ha contribuido en elucidar los genes que le confieren al organismo la capacidad de sobrevivir y/o crecer en un ambiente determinado (McHardy & Kloetgen, 2017). Por ejemplo, el análisis del metagenoma bacteriano de la información de Tara Oceans y de la Expedición de circunnavegación Malaspina 2010 permite obtener una gran cantidad de datos de genes de microorganismos marinos; sin embargo, el análisis en estos estudios está principalmente orientado a genes asociados al ciclo de carbono, metilación de mercurio, fijación de nitrógeno, entre otros; mas no ha sido analizado desde la perspectiva de los genes de resistencia a arsénico presentes en las bacterias marinas ni su distribución en ecosistemas marinos-oceánicos a nivel global, a pesar de la posible toxicidad del arsénico soluble en cuerpos de agua (Delmont et al., 2018; Sunagawa et al., 2015; Villar et al., 2020) . Por ello, en el presente estudio *in silico*, se desea identificar

los genes que confieren a los microorganismos, específicamente las bacterias, la capacidad de afectar la movilización-toxicidad del arsénico, ya que es importante desde la biología básica para elucidar los mecanismos de resistencia a la toxicidad del arsénico presentes en las bacterias marinas. Además, estudiar la distribución geográfica de estos genes, permitirá tener indicios de posibles lugares donde haya mayor potencial que ocurra biotransformación de arsénico por parte de las bacterias, que, a su vez, permitirá rastrear posibles lugares de muestreo de bacterias candidatas para el desarrollo de estrategias para biorremediar ambientes contaminados con arsénico en futuras investigaciones experimentales en ambientes marinos (Saunders et al., 2019; Tsai et al., 2009).

#### **1.4. Objetivo del trabajo**

##### **1.4.1. Objetivos del documento**

###### **1.4.1.1. Objetivo General:**

- Visualizar la distribución de genes asociados a la resistencia de arsénico de bacterias marinas en los océanos y mares del mundo.

###### **1.4.1.2. Objetivos específicos:**

- Construir una base de datos propia de genes asociados a la resistencia de arsénico a partir de la base de datos de NCBI.
- Identificar posibles genes asociados a la resistencia a arsénico en datos de metagenomas de microorganismos marinos según los datos de Tara Oceans y de Malaspina.
- Determinar la abundancia de los genes asociados a la resistencia a arsénico en los mares del mundo según los datos de Tara Oceans y de Malaspina.

- Evaluar la relación de la abundancia de los genes asociados a la resistencia a arsénico con la profundidad del ecosistema marino según los datos de Malaspina.

### **1.5. Hipótesis**

**H<sub>0</sub>:** Los genes de resistencia a arsénico de bacterias marinas están distribuidos en los océanos y mares a nivel global según los datos de Tara Oceans y de Malaspina 2010.

**H<sub>1</sub>:** Los genes de resistencia a arsénico de bacterias marinas no están distribuidos en los océanos y mares a nivel global según los datos de Tara Oceans y de Malaspina.

## II. ESTADO DE CONOCIMIENTO

El arsénico es el vigésimo elemento más abundante en la superficie de la Tierra. Es un metaloide que puede existir en la naturaleza en 4 estados de valencia:  $-3$  (arseniuros, arsina),  $0$  (arsénico elemental),  $+3$  (arsenitos) y  $+5$  (arseniatos). El arseniuro puede formar parte de minerales que constan de compuestos de un metal con arsénico, cuya fórmula es  $MA_s_n$ , donde M es un metal y  $n = 1, 2$  o  $3$  (Flora, 2015). La arsina ( $AsH_3$ ) es un compuesto volátil conocido como un gas incoloro. El arsénico elemental rara vez se encuentra en la naturaleza, sin embargo, puede formar sólidos en depósitos hidrotermales a ciertas temperaturas ( $50^\circ C$  a  $200^\circ C$ ) bajo condiciones muy anóxicas y de bajo sulfuro. Los arsenitos y arseniatos son los estados de oxidación más relevantes del arsénico y son los más frecuentes en el agua natural como en los suelos. Estos también pueden identificarse como As (III) y As (V), respectivamente (Henke, 2009).

Actualmente el arsénico todavía es un metaloide ubicuo, presente en la litosfera (corteza terrestre, suelo, rocas y sedimentos), hidrosfera (aguas superficiales, acuíferos, pozos y océanos profundos), atmósfera y biósfera (cadena alimentaria). Los niveles de arsénico en los suelos y rocas naturales varían desde 1 a 150 mg/kg, encontrando principalmente arsenito ( $As^{3+}$ ), arseniato ( $As^{5+}$ ) y arsénico orgánico. Con respecto a la hidrosfera, el arsénico en el agua puede estar en forma de partícula o disuelto, cuya forma disuelta son frecuentemente el arsenito y arseniato en ambientes acuáticos (Henke, 2009). En el mar, el arseniato es la forma predominante a comparación del arsenito. Sin embargo, el arsénico orgánico como dimetil-arsénico y metil-arsénico, y el arsenito inorgánico están presentes en pequeñas cantidades en el mar, pero son más prevalente en zonas fóticas. Además, el arsénico total en el medio marino está presente en una concentración muy uniforme de  $2 \mu g/L$ . Por otro lado, la presencia del arsénico en la atmósfera es mayor en el hemisferio norte que en el hemisferio sur, que puede ser atribuido a la mayor actividad industrial presente en el hemisferio norte (Jang et al., 2016).



El arsénico puede existir tanto en forma inorgánica como orgánica. Se describe su forma orgánica cuando está asociada con carbono e hidrógeno, mientras que en su forma inorgánica puede estar asociado con hierro, cobalto, níquel, sulfuro, oxígeno, hidrógeno y a sí mismo (Henke, 2009; Jang et al., 2016). En el cual, la forma del arsénico orgánico e inorgánica varía dependiendo del pH y el potencial redox (Oremland et al., 2009). Por ejemplo, el As (III) existe principalmente en aguas subterráneas y aguas hidrotermales con bajo contenido de oxígeno (reductor); adicional a ello, en aguas anóxicas y ricas en sulfuros, el azufre puede sustituir a uno o más oxígenos para formar especies tioarsénicas. En cambio, el As (V) es más común en la oxidación de aguas subterráneas y superficiales (Henke, 2009).

El arsénico puede ser introducido en el ambiente de manera natural o antropogénicamente. Por ejemplo, el arsénico en la litosfera puede ser liberado a la atmósfera mediante actividad humana asociada a la quema de combustibles fósiles y materiales de desecho que contengan arsénico o por fuentes naturales como actividad volcánica y volatilización microbiana. Asimismo, se puede liberar el arsénico a la hidrosfera mediante la meteorización de rocas oceánicas que contienen arsénico, por actividad de volcanes submarinos y por calamidades naturales como fuertes lluvias, inundaciones y deslizamientos de tierra. Además, por fuentes antropogénicas como el uso de pesticidas que contengan arsénico, la minería, vertedero de residuos electrónicos o eliminación de aguas residuales industriales, también se contaminan el suelo y el agua. Por otro lado, el arsénico en aguas contaminadas puede ingresar a la litosfera a través de la volatilización, irrigación y usos domésticos o naturalmente por aguas geotérmicas como las aguas termales. Además, el arsénico en forma de polvo puede precipitar nuevamente en la litosfera y la hidrosfera después de una lluvia (Jang et al., 2016).

Desde el punto de vista de la toxicología, los arsenitos y arseniatos son los estados de oxidación más relevantes. Las formas trivalentes de arsénico (arsenitos), tanto

inorgánico como orgánico, son generalmente más tóxicas que las formas pentavalentes (arseniato) (Henke, 2009); siendo las formas inorgánicas,  $As^{3+}$  y  $As^{5+}$ , altamente tóxicas y móviles a comparación de las formas orgánicas. En específico, el arsenito ( $As^{3+}$ ) es 10 veces más tóxico que arseniato ( $As^{5+}$ ). El arsenito es más tóxico debido a su fuerte afinidad por los grupos sulfhidrilos de biomoléculas como los grupos tiol en enzimas. Por otro lado, el arseniato es tóxico debido a su interferencia con la fosforilación oxidativa en células por sustituir al grupo fosfato (P) en la síntesis de ATP, perjudicando el almacenamiento de la energía intracelular. Además, el arsénico orgánico, formado por la metilación del arsénico por parte de los seres vivos, como el ácido monometilarsénico (MMAA) y el ácido dimetilarsénico (DMAA), son menos tóxicos, pero a medida que aumenta su concentración también genera un efecto adverso en el metabolismo celular. En síntesis, se cree que el arsénico causa cambios directamente en la formación de proteínas y enzimas, pero de manera indirecta contribuye al daño del ADN, debido a que puede interactuar con otros metales como el hierro fijado intracelularmente, liberándolo y generando estrés oxidativo (Jang et al., 2016).

Por otro lado, el arsénico no se degrada ni puede ser destruido en el ambiente. Por ello, la amplia distribución de estas formas inorgánicas y su nivel de toxicidad es un problema alarmante para la salud. Ya que hay evidencias epidemiológicas que describen su efecto carcinógeno como también asociado a otros problemas de salud, por ejemplo, enfermedades cardiovasculares, alteraciones del sistema nervioso y diabetes. Por ello, la Organización Mundial de la Salud (OMS) ha delimitado los niveles de arsénico en las aguas para consumo humano ( $10 \mu\text{g/l}$ ), sin embargo, en 42 naciones excede dicho límite. Actualmente, se conoce que la ingestión de agua, comida o tierra contaminados con arsénico y la inhalación de polvo que contenga arsénico son las dos principales rutas de exposición potencial del arsénico a los humanos (Henke, 2009; Jang et al., 2016; Oremland et al., 2009).

En el agua de mar, el arsénico puede estar en su forma disuelta y en partícula. La forma disuelta incluye al arseniato, arsenito y componentes orgánicos como los arsénicos metilados. Entre ellos, predomina el arseniato debido a su estabilidad termodinámica en el océano oxigenado. Sin embargo, en zonas eufóticas, hay mayor concentración de arsenito y arsénico orgánico debido a la excreción microbiana. En el cual, los microorganismos pueden incorporar indiscriminadamente el arseniato debido a que presenta propiedades fisicoquímicas similares al fosfato, biotransformando el arsénico en fase disuelta en material particulado, lo que hace que el comportamiento químico del arsénico sea menos móvil y tóxico (Hashihama et al., 2019). Además, la concentración de arseniato en el océano muestra un perfil de profundidad opuesto a los nutrientes como el fosfato (Wurl et al., 2015).

Estudios recientes informaron que los microorganismos marinos tienen genes de desintoxicación de arsénico ubicuos y diversos. Debido a que los microorganismos han desarrollado mecanismos de resistencia a arsénico inorgánico como, por ejemplo, la desintoxicación mediante reducción de arseniato a arsenito y su excreción fuera de la célula a través de un sistema de genes que codifica el arseniato reductasa (*arsC*), la bomba de flujo de arsenito (*arsB* o *acr3*), y el regulador represivo asociado a arsénico (*asrR*); incluso han desarrollado resistencia al arsénico por metilación, identificando genes como *arsM*, *arsI* y *arsH* que codifican proteínas para la detoxificación de órgano-arsénicos (Hashihama et al., 2019). En otras palabras, el metabolismo microbiano participa en la movilización del arsénico y la especiación del arsénico en el medio ambiente (Oremland et al., 2009; Silver & Phung, 2005). Además, ciertos microorganismos pueden usar el arseniato como un aceptor de electrones mediante la reducción a arsenito en condiciones anaeróbicas. Por el contrario, el arsenito puede ser oxidado a arseniato por bacterias aerobias que pueden ganar energía para su crecimiento a partir de la oxidación del arsenito. Estos estudios enfatizan la importancia de los microorganismos en el ciclo biogeoquímico del arsénico (Chen & Shao, 2009).

En los últimos años, las tecnologías sostenibles, rentables y eficientes han ganado más atención. Las tecnologías de remediación comúnmente utilizadas para el agua contaminada con arsénico incluyen oxidación, tratamiento con cal, precipitación química/coagulación, intercambio iónico, adsorción, biorremediación, fitorremediación, proceso de membrana híbrida (nanofiltración y ultrafiltración) y electrodiálisis. Entre ellas, la biorremediación o conversión de arsénico inorgánico en arsénico orgánico con la aplicación de microorganismos es la tecnología más simple y fácilmente disponible. Dicha biorremediación puede ser mediante la absorción de arsénico inorgánico y convertirlos en metilarsénico, dimetilarsénico e incluso, convertirlas en trimetilarsina (arsinas gaseosas como forma volátil) en condiciones anaeróbicas, con el fin de lograr la inmovilización del arsénico metálico (Jang et al., 2016).

Una visión detallada de las comunidades microbianas que controlan el ciclo biogeoquímico del arsénico en el sistema natural es desafiante debido a su extrema diversidad y, además, una gran fracción de bacterias en el medio ambiente permanece sin cultivar mediante técnicas de cultivo de laboratorio estándar. Por ello, el análisis metagenómico ha ofrecido una oportunidad sin precedentes para examinar las estrategias de respuesta y adaptación de las comunidades microbianas a la toxicidad ambiental con la ventaja de no requerir el cultivo de microorganismos o información de secuencia previa a genes conocidos, lo que representa un enfoque valioso para describir nuevos genes que codifican proteínas o enzimas con nuevas características.

La secuenciación del metagenoma ya ha sido aplicada en comunidades microbianas de varios nichos ecológicos, por ejemplo, drenaje de minas ácidas, agua marina y sedimentos, incluyendo suelos contaminados con arsénico, que han proporcionado nuevos conocimientos sobre la estructura de la comunidad microbiana, y su función junto con el patrón de evolución e identificación de nuevos genes (Chauhan et al., 2009; Das et al., 2017).

Con el fin de explorar la inmensa riqueza de diversidad genética, funcional y morfológica en los ecosistemas marinos a escala global, se ha llevado a cabo iniciativas a gran escala, como Tara Oceans (Karsenti et al., 2011) y Malaspina (Duarte, 2015; Ortega et al., 2019). La expedición Tara Oceans recolectó sistemáticamente más de 35,000 muestras de agua y plancton del océano, así como datos ambientales en 210 estaciones de muestreo y a través de métodos como secuenciación del metagenoma ha podido describir el plancton que comprende poblaciones de virus, bacterias, arqueas, eucariotas unicelulares (protistas) y animales que se mueven a la deriva con las corrientes marinas, en sus ambientes fisicoquímicos y oceanográficos a escala global y a múltiples profundidades, desde la superficie del mar hasta 1000 metros de profundidad, clasificados en epipelágico (0 a -200 m.s.n.m.) y mesopelágico (-200 m.s.n.m. a -1000 m.s.n.m.) (Sunagawa et al., 2015). Una estación de muestreo completa consistió en recolectar plancton de tres características ambientales distintas, típicamente la “capa de agua superficial” (SUR), la “capa de máxima clorofila profunda” (DCM) y la “zona mesopelágica” (MES). Por otro lado, la expedición Malaspina 2010 se propuso evaluar el estado de los océanos en 2011 y explorar 350 estaciones muestreadas a lo largo del océano, utilizando herramientas avanzadas de secuenciación de próxima generación, la diversidad de la vida en el océano, con un énfasis particular en el océano oscuro, aproximadamente hasta 4000 metros de profundidad, clasificados en 3 zonas: zona epipelágica desde la superficie hasta los 200m de profundidad, zona mesopelágica desde los 200 m hasta los 1000 m de profundidad y finalmente, la zona batipelágica que comprende la zona debajo de los 1000 m de profundidad (Sutton, 2013). Para estudiar a los procariotas en las estaciones muestreadas, se puede filtrar por fracciones de tamaño, en el cual los procariotas de vida libre varían en tamaño de 0,22 a 3  $\mu\text{m}$  (Pesant et al., 2015).

Por otro lado, debemos resaltar la importancia de la bioinformática, implicado en la aplicación de herramientas computacionales para el manejo y análisis de datos

biológicos, que ha permitido grandes avances de la investigación en genómica (Jones et al., 2006). Por lo cual, en este estudio, se aplicarán herramientas bioinformáticas orientándolas en la búsqueda de genes asociados al metabolismo del arsénico que le puedan conferir al microorganismo la capacidad de tolerar ambientes contaminados con arsénico, y obtener gráficas representativas para analizar patrones de presencia y abundancia de los microorganismos portadores de los genes de interés en ambientes marinos a nivel global. Usando como bases de datos de estudio, los catálogos de genes de microorganismos de ambientes marinos de las expediciones de Tara Oceans y Malaspina 2010.

### III. MATERIALES Y MÉTODOS

#### 3.1. Materiales

Para este estudio, se necesitó una computadora con alta capacidad de memoria RAM y memoria en el disco duro para manejar gigabases de bases de nucleótidos secuenciados. Además, programas necesarios como MMseqs2, Rstudio y Python y, el sistema operativo Bash/Linux que permiten manipular dichas secuencias para una búsqueda eficiente de genes en la base de datos de Tara Oceans y de Malaspina 2010.

#### 3.2. Métodos

##### 3.2.1. Diseño de base de datos de genes de interés

Se diseñó una base de datos de genes asociados a arsénico previamente descritos (Zhu et al., 2017) usando como referencia la base de datos de NCBI (Agarwala et al., 2018). Las secuencias se descargaron en formato FASTA. Luego, usando comandos basados en Bash/Linux (Garrels, 2004), se unió las secuencias en un solo archivo en formato FASTA para tener una propia base de datos de genes asociados a la resistencia al arsénico en microorganismos.

##### 3.2.2. Análisis de secuencias nucleotídicas usando *clusters* bioinformáticos basados en Bash/Linux

Las secuencias nucleotídicas se alinearon usando *usearch/10.0.240* (Edgar, 2010). Para ello, primero se ordenaron por longitud de secuencia mediante el comando *usearch -sortbylength*. Posteriormente, se agruparon según porcentaje de identidad mediante el comando *usearch -cluster\_fast* según su porcentaje de identidad (0.90). Finalmente, solo las secuencias representativas a noventa por ciento de identidad se guardaron en un archivo fasta (*operonArs2.clus90.fasta*) para su posterior análisis.

##### 3.2.3. Búsqueda y anotación de genes vía MMseqs2 en base de datos de metagenomas

Las secuencias nucleotídicas del *operonArs2.clus90.fasta*, se buscaron en el catálogo de genes no redundantes de la base de datos de Tara Oceans denominado *OM-RGC.v2*. Por otro lado, también se buscaron las secuencias del *operonArs2.clus90.fasta*

en el catálogo de genes de la base de datos de la Malaspina 2010. Ambas búsquedas se realizaron mediante el uso del programa *MMseqs2* (Steinegger & Söding, 2017). Para ello, se diseñó un proceso en Bash/Linux para que *MMseqs2* permita comparar ambas bases de datos tomando como consulta (*query*) a cada secuencia del *operonArs2.clus90.fasta* y como blanco (*target*) a las secuencias de *OM-RGC.v2* mediante el comando *mmseqs createdb*. Para luego buscar secuencias del *query* en la base de datos del *target* mediante similitud y alineamiento de Smith-Waterman mediante el comando *mmseqs search*. Finalmente, el resultado de la búsqueda puede ser transformado en un formato legible mediante el comando *mmseqs convertalis*, el cual dará un resultado tabular con 12 columnas: (1,2) identificadores para *query* y *target*, (3) identidad de secuencia, (4) longitud de alineación, (5) número de diferencias (*mismatches*), (6) número de vacíos (*gaps*), (7-8, 9-10) posición de inicio y posición final en *query* y *target*, (11) Valor E y (12) puntaje de bit (Steinegger & Söding, 2017). El archivo resultante se denominó *AvsT.m8*.

Por otro lado, también se buscó las secuencias nucleotídicas del *operonArs2.clus90.fasta* en el catálogo de genes de la base de datos de metagenomas de Malaspina 2010 mediante el programa *MMseqs2* con el mismo proceso descrito anteriormente en Bash/Linux. Denominando al archivo resultante *AvsM.m8*.

### **3.2.4. Filtrado de las secuencias alineadas con la base de datos de metagenomas mediante Python**

Las secuencias resultantes del alineamiento entre las secuencias *query* y *target* (*AvsT.m8* o *AvsM.m8*) fueron filtradas y seleccionadas por el porcentaje de identidad, cobertura y longitud de secuencia usando Python (Perkel, 2015). Estos tres parámetros fueron modificados y evaluados para finalmente seleccionar el más adecuado. Seleccionando un porcentaje de identidad de 90%; una cobertura de 60% y una longitud mínima de 900pb. El archivo resultante se nombró *genesArs.filter.txt*, guardados en directorios distintos según el catálogo usado (Tara Oceans o Malaspina 2010),



obteniendo los genes posibles asociados a la resistencia a arsénico en bacterias marinas.

### **3.2.5. Relación de la distribución y abundancia de los genes de interés en las estaciones muestreadas del océano global usando Rstudio.**

Se usó la versión 3.5.3 de R y se instaló los paquetes `tydiverse`, `maps`, `mapplots` y `ggplot2` en Rstudio (Murrell & Murrell, 2016) para manipular, explorar y visualizar los datos. Usando Rstudio, se cargó en una tabla los datos de los genes de interés proveniente de los catálogos de genes y sus respectivas abundancias. La información de la abundancia de cada gen fue obtenida a partir de los datos de “*Tara Oceans Microbial Reference Gene Catalog version 2*” (*OM-RGC.v2*) (Sunagawa et al., 2015). Esta abundancia de cada gen en *OM-RGC.v2* fue obtenida usando *MOCAT* (Kultima et al., 2012). Obteniendo finalmente abundancias relativas y corregidas para diferentes profundidades de muestreo. Por otro lado, la abundancia de los genes identificados en Malaspina fue obtenida de la base de datos de genes de Malaspina 2010 (Acinas et al., 2021).

Con respecto a los genes de interés en base a los metadatos de Tara Oceans, se cargó la información de las estaciones y localización de muestreo correspondientes: muestra, capa, latitud y longitud; para convertirlo en una tabla de datos que permita seleccionar cada una de estas variables. Con el fin de correlacionar los genes y su abundancia, filtrándolos según los datos de *layer* o capa oceánica de los metadatos de Tara Oceans: superficie (SUR), clorofila profunda máxima (DCM) o mesopelágica (MES) (Kirchman, 2008). Se ordenó la información en una tabla Excel con el fin de visualizar lo más relevante de la distribución y abundancia de los genes. Posterior a ello, se visualizó los genes más relevantes con amplia distribución y gran abundancia génica en un mapa global de los océanos según las coordenadas de longitud y latitud de los metadatos de Tara Oceans y de los datos de un mapa de referencia mediante la función `geom_map`, usando el paquete `maps` y el paquete `ggplot2`.

Un similar proceso se realizó para los genes de interés en base a los metadatos de Malaspina 2010. Con el fin de correlacionar los genes y solamente su distribución en las estaciones de Malaspina, filtrándolos según los datos de “layer” o capa oceánica de los metadatos de Malaspina 2010: epipelágica, mesopelágica y batipelágica (Sutton, 2013). Visualizándolo en un mapa global las estaciones de muestreo según las coordenadas de longitud y latitud de los datos de Malaspina y de los datos de un mapa de referencia mediante la función `geom_map`, usando el paquete `maps` y el paquete `ggplot2`, y resaltando de un color diferentes (rojo) las estaciones donde se distribuyen los genes de interés identificados.

### **3.2.6. Relación de la abundancia de genes de interés y profundidad del ecosistema marino según las estaciones de Malaspina usando Rstudio.**

En Rstudio, se cargaron los metadatos de las estaciones y muestras de Malaspina 2010 que contenían los genes de interés, la información de las estaciones localización de muestreo correspondientes (capa, latitud y longitud) y la información de los niveles de profundidad clasificadas por zonas: zona epipelágica, zona mesopelágica y zona batipelágica (Sutton, 2013).

## IV. RESULTADOS

### 4.1. Base de datos creada de genes de interés

Se creó una base de datos de genes asociados a arsénico previamente descritos (Zhu et al., 2017). Usando la base de datos de NCBI (Agarwala et al., 2018), se descargó 200 secuencias nucleotídicas diferentes provenientes de 28 tipos de genes asociados a la resistencia y metabolismo de arsénico (Zhu et al., 2017). Posteriormente, estas secuencias fueron filtradas y alineadas con dos diferentes catálogos de genes (Tara Oceans o Malaspina 2010), obteniendo una lista de posibles genes asociados a la resistencia a arsénico en bacterias marinas (Tabla N°1). Los genes identificados fueron nombrados según las etiquetas en la Tabla N°1 para resaltar el catálogo usado (Malaspina 2010 o Tara Oceans); los cuadros negros indican ausencia de dicho gen y el %Id hace referencia al porcentaje de identidad.

Se evidencia en la tabla, que hay genes asociados a la resistencia a arsénico presentes tanto en los datos de Malaspina 2010 (*G1*, *G2*, *G3*) como en los datos de Tara Oceans (*J6*, *J7*, *J1*). Todos estos genes son identificados como genes *arsJ*. Los datos de Tara Oceans, también permite identificar el género, familia o inclusive la especie del cual proviene el gen identificado. Entre los genes encontrados en Tara Oceans, hay una concordancia entre las bacterias identificadas con respecto a la base de datos diseñado a partir de NCBI, a excepción del gen *I* y *J7*. Por otro lado, también se identificaron genes únicos en los datos de Malaspina 2010 (*G4*, *G5*) y en los datos Tara Oceans (*J3*, *J4*, *J5*, *I*).

**Tabla N°1. Presencia de genes posibles asociados a la resistencia a arsénico en microorganismos marinos en ambas bases de datos.**

En Malaspina 2010		Genes asociados a arsénico		En Tara Oceans		
Etiqueta	%Id	Gen	Microorganismo	%Id	Etiqueta	Microorganismo
G1	95	<i>arsJ</i>	<i>Thalassospira profundimaris</i>	94	J6	<i>Thalassospira</i> sp.
G2	91	<i>arsJ</i>	<i>Sulfitobacter indolifex</i>	92	J7	Rhodobacteraceae
G3	93	<i>arsJ</i>	<i>Marinobacter hydrocarbonoclasticus</i>	95	J1	<i>Marinobacter</i> sp.
		<i>arsJ</i>	<i>Vibrio campbellii</i>	98	J2	Vibrionaceae
		<i>arsJ</i>	<i>Shewanella frigidimarina</i>	99	J3	<i>Shewanella frigidimarina</i>
		<i>arsJ</i>	<i>Pseudoalteromonas ruthenica</i>	96	J4	<i>Pseudoalteromonas</i> sp.
		<i>arsJ</i>	<i>Thalassospira xiamenensis</i>	97	J5	<i>Thalassospira</i> sp.
		<i>arsI</i>	<i>Cupriavidus metallidurans</i>	98	I	<i>Idiomarina</i> sp.
G4	96	<i>arsB</i>	<i>Agrobacterium fabrum</i>			
G5	98	<i>arsA</i>	<i>Staphylococcus hominis</i>			

#### **4.2. Relación de la distribución y abundancia de los genes de interés en las estaciones muestreadas en océano global usando Rstudio.**

Los genes asociados a la resistencia a arsénico, se relacionaron con su localización y distribución en el océano global según su latitud y longitud, identificando su distribución en diferentes cuerpos de agua (mar u océano) en cada capa oceánica (SUR, DCM y MES) (Tabla N°2), expresando la frecuencia en la que se presencia cada gen de interés en los distintos cuerpos de agua (océano o mar) y especificando la mayor abundancia génica encontrado para cada gen y el respectivo cuerpo de agua que lo contiene. Adicionalmente, se detalla la codificación de la estación de muestreo de Tara Oceans que contiene la mayor abundancia génica para cada gen en cada capa oceánica. Los datos presentados en la Tabla N°2 es un resumen de lo datos encontrados en la Tabla S1, Tabla S2 y Tabla S3 (ver Anexos).

En la capa SUR, la mayoría de los genes asociados a la resistencia a arsénico estaban en más de un cuerpo de agua, en el cual el gen *J1* presentó mayor versatilidad ambiental ya que es el que presenta mayor distribución en el océano global, identificándose en 15 cuerpos de agua diferentes y en 55 estaciones de Tara Oceans. Además de ello, el gen *J2* presentó la mayor abundancia máxima en la capa SUR con un valor de 49,3693 en la estación 94 de Tara Oceans que se ubica en el Océano Pacífico Sur. Este patrón de amplia distribución y mayor abundancia del gen *J1*, se repitió en la capa DCM y MES (tabla N°2), encontrándose en 12 y 8 cuerpos de agua diferentes, respectivamente, con mayor abundancia en el Mar Rojo (DCM) y en el Océano Pacífico Norte (MES). Ningún otro gen asociado a la resistencia a arsénico presento esta amplia distribución (en más de 10 estaciones) con una gran abundancia mayor a 5.

Con el fin de correlacionar la ubicación de las estaciones de muestreo de Tara Oceans que contienen los genes asociados a la resistencia de arsénico de interés con abundancia génica máxima, se plasmó una imagen de las 210 estaciones de muestreo en un mapa global, en donde las estaciones de muestreo con abundancia génica

máxima de la Tabla N°2 fueron encerrados en azul claro (Figura N°1), modificado de Sunagawa et al. (2020)

En la Figura N°2, considerando solo las estaciones con mayor abundancia génica para cada gen de interés, se puede observar que hay más presencia de genes asociados a la resistencia a arsénico en el Océano Pacífico (6 estaciones de muestreo), posteriormente le sigue el Mar Rojo (3 estaciones de muestreo), luego el Océano Atlántico (3 estaciones de muestreo, incluyendo al Estrecho de Davis), finalmente el Golfo de México y Mar Siberiano del Este, cada uno con una estación de muestreo. No obstante, las tablas suplementarias del anexo (Tabla S1, Tabla S2 y Tabla S3) dan evidencia que estos genes están en la mayoría de los mares y océanos de la Tierra.

El gen J1 fue el más destacado por su amplia distribución y abundancia, por lo cual se graficó en un mapa global tanto su distribución en cada capa oceánica (SUR, DCM y MES) con sus respectivas abundancias aproximadas representadas en círculo morado (Figura N°2).

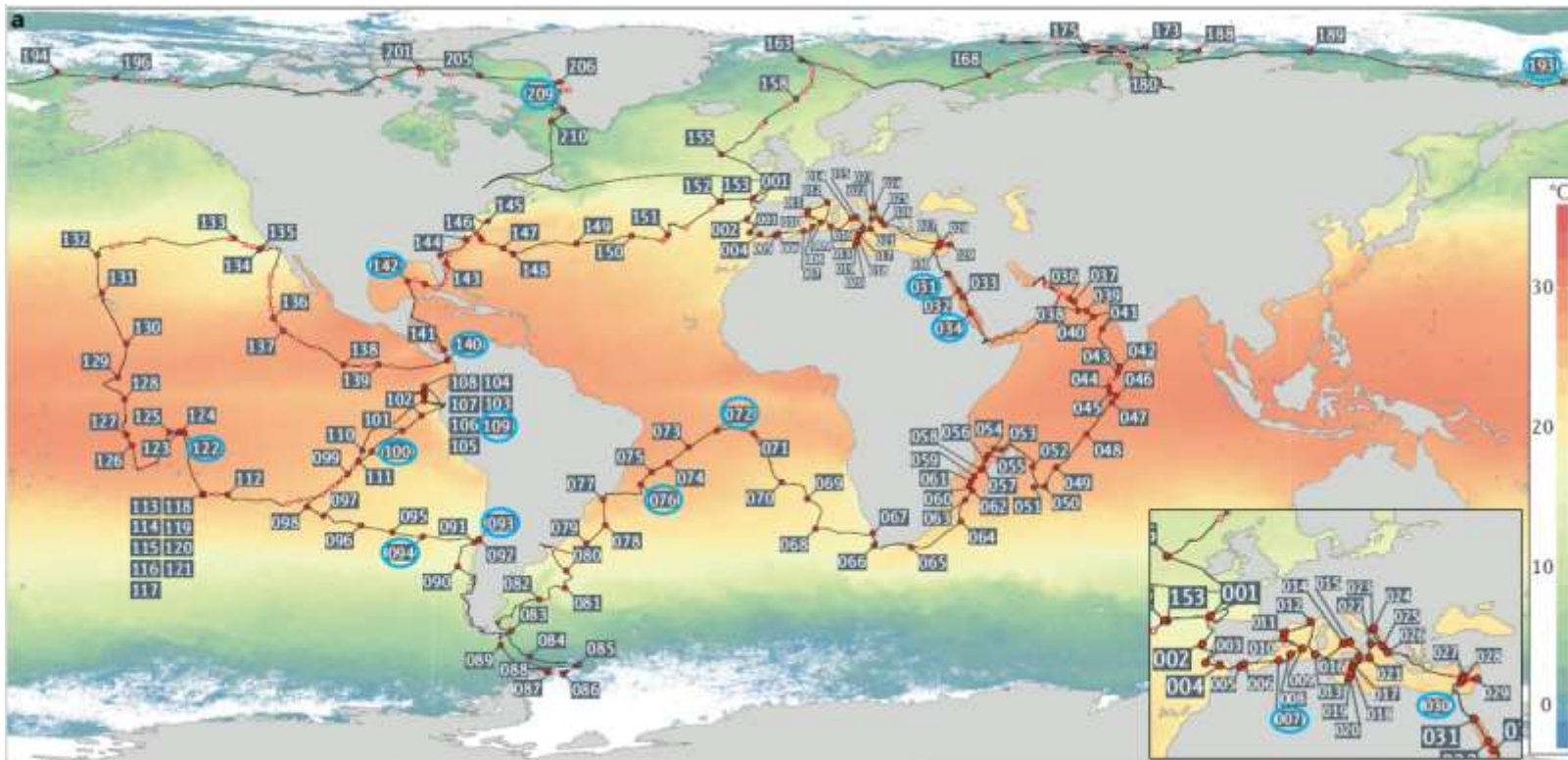
Los datos de Malaspina (Figura N°3), evidencian que los genes de interés están ubicados en los diferentes océanos muestreados (Océano Pacífico, Océano Atlántico, Océano Índico). Observando una mayor abundancia génica para el gen G3.

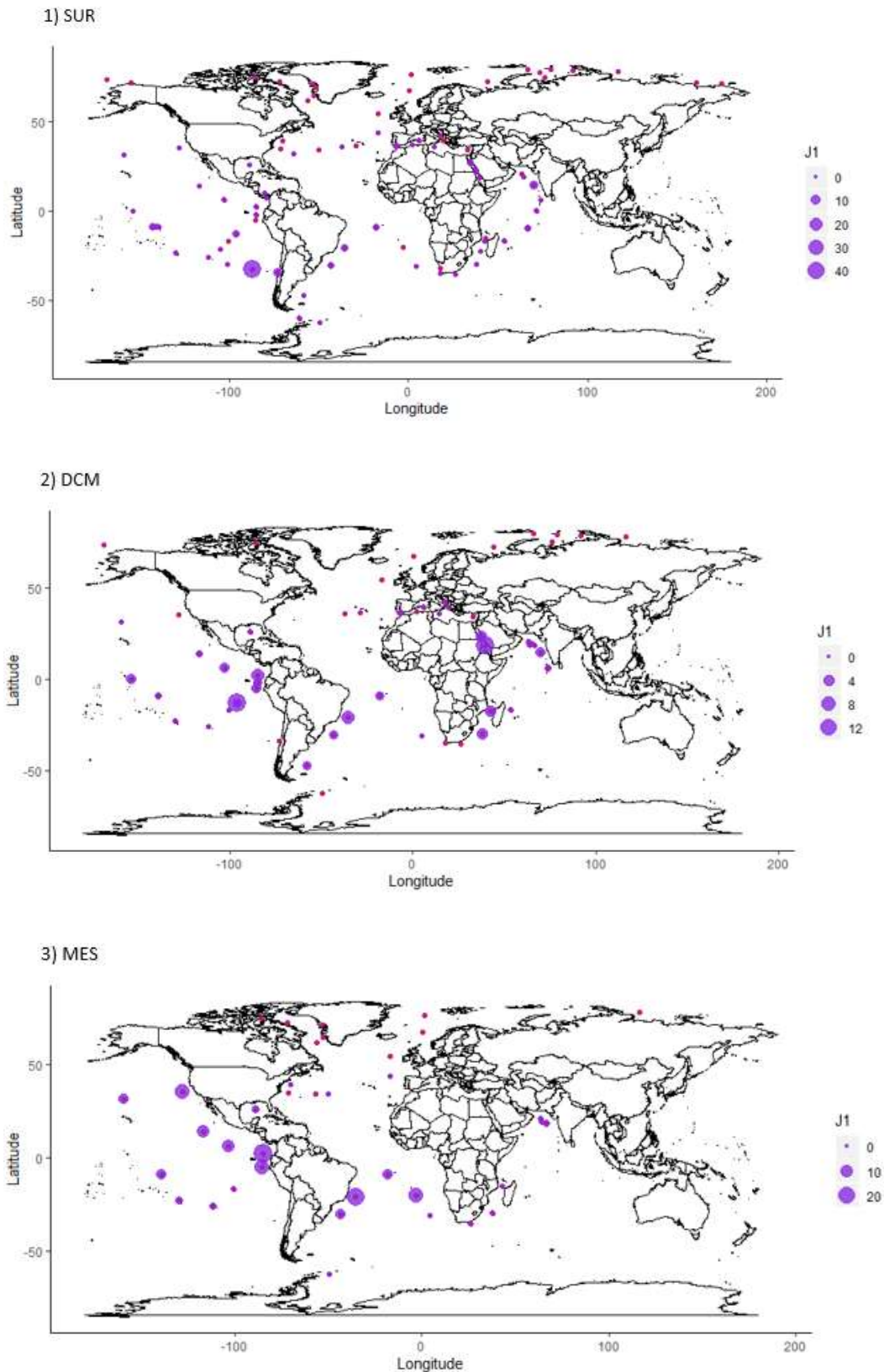
**Tabla N°2. Resumen de la distribución y abundancia de los genes asociados a la resistencia arsénico identificados en los datos de Tara Oceans.**

	Gen	Frecuencia en diferentes cuerpos de agua	Frecuencia total en estaciones de muestreo	Abundancia génica máx. (A.G.M.)	Codificación de la Estación de muestreo con A.G.M	Nombre del cuerpo de agua con A.G.M
Capa de Agua Superficial (SUR)	I	5	8	0.3589	7	Mar Mediterráneo
	J1	15	55	49.3693	94	Océano Pacífico Sur
	J2	4	10	0.3161	30	Mar Mediterráneo
	J3	1	1	0.1200	193	Mar Siberiano del Este
	J4	1	1	0.9749	140	Océano Pacífico Norte
	J5	7	16	1.0162	93	Océano Pacífico Sur
	J6	2	4	0.2629	72	Océano Atlántico Sur
	J7	9	18	0.5148	31	Mar Rojo
Capa de máxima de clorofila profunda (DCM)	Gen	Frecuencia en diferentes cuerpos de agua	Frecuencia total en estaciones de muestreo	Abundancia génica máx. (A.G.M.)	Codificación de la Estación de muestreo con A.G.M	Nombre del cuerpo de agua con A.G.M
	I	4	5	7.2431	93	Océano Pacífico Sur
	J1	12	32	15.7282	34	Mar Rojo
	J2	3	6	0.2438	93	Océano Pacífico Sur
	J3	1	1	0.1346	93	Océano Pacífico Sur
	J4	2	2	18.2002	122	Océano Pacífico Sur
J5	6	10	1.0957	100	Océano Pacífico Sur	

	J6	2	3	0.7765	76	Océano Atlántico Sur
	J7	8	17	0.6271	142	Golfo de México
Zona Mesopelágica (MES)	Gen	Cantidad de su distribución en diferentes cuerpos de agua	Frecuencia total en estaciones de muestreo	Abundancia génica máx. (A.G.M.)	Codificación de la Estación de muestreo con A.G.M	Nombre del cuerpo de agua con A.G.M
	I	6	11	2.2105	109	Océano Pacífico Norte
	J1	8	27	27.6558	109	Océano Pacífico Norte
	J2	4	6	1.0517	122	Océano Pacífico Sur
	J3	2	2	0.4417	209	Estrecho de Davis
	J4	4	5	0.5553	122	Océano Pacífico Sur
	J5	6	12	0.9351	76	Océano Atlántico Sur
	J6	4	0	2.3161	76	Océano Atlántico Sur
	J7	6	23	4.8062	142	Golfo de México

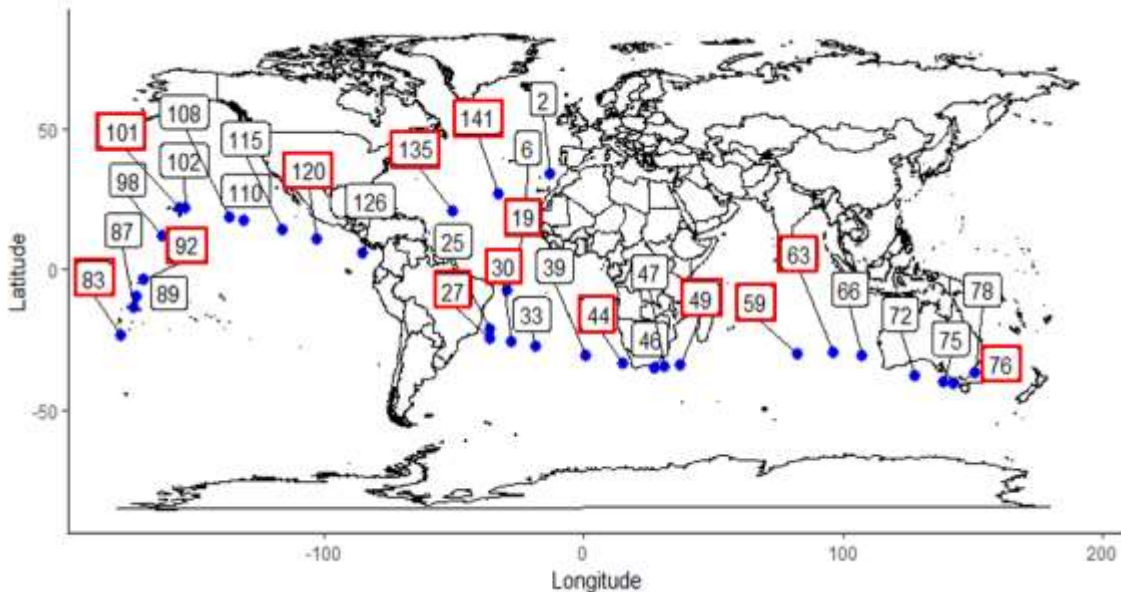






**Figura N°2. Distribución y abundancia del gen J1 en los datos de Tara Oceans.** Se observa la distribución del gen J1 en los diferentes ambientes marinos en gráficas que representan un mapa global con las coordenadas de latitud y longitud de los continentes, océanos y mares. Cada

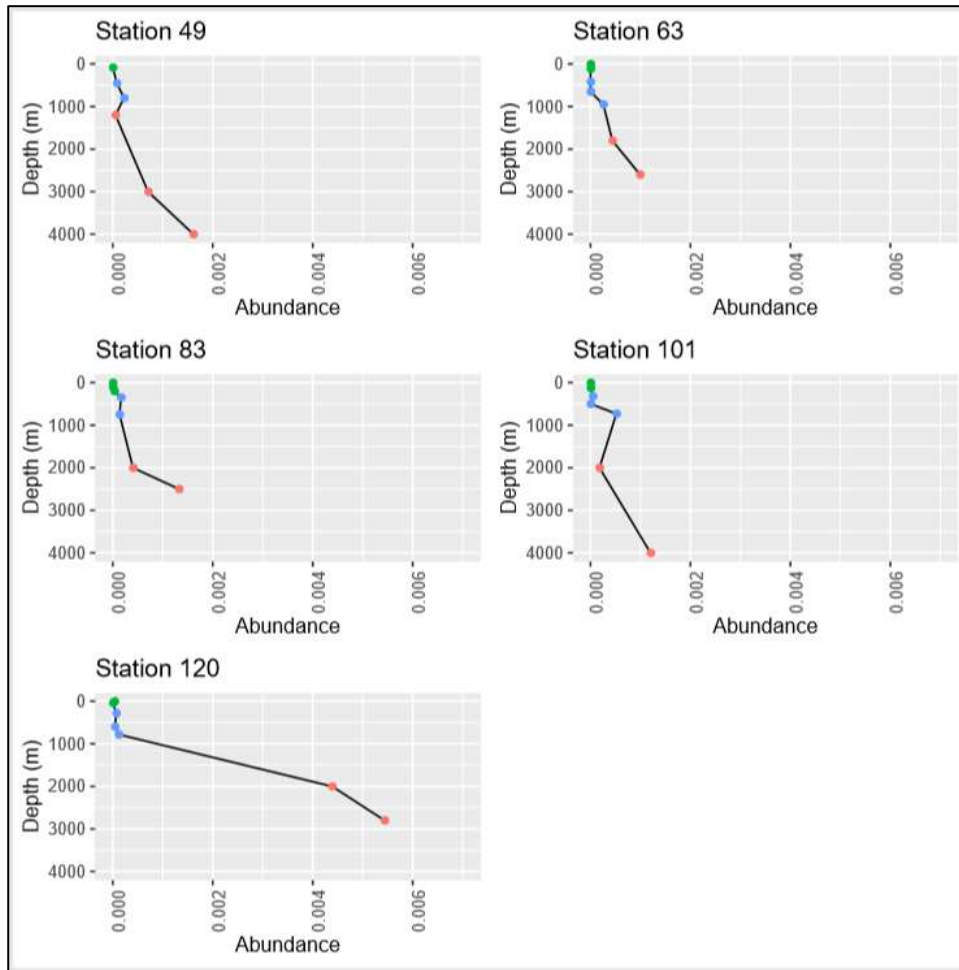
gráfica representa una capa oceanográfica según los datos de Tara Oceans. SUR: superficie, DCM: clorofila profunda máxima y MES: mesopelágica. Así mismo, se observa la abundancia de dicho gen como indica el círculo violeta según su tamaño, a mayor tamaño mayor abundancia, pero cada capa oceanográfica tiene su propio rango de abundancia y no hay una estandarizada para las tres capas. Dicho gen *J1* es el gen *arsJ* de *Marinobacter hydrocarbonoclasticus* con 95% de identidad.



**Figura N°3. Distribución de las estaciones de los datos de Malaspina 2010.** La enumeración de las estaciones es según fueron nombradas en los datos de Malaspina 2010 y los cuadros en rojo son las estaciones de muestreo que contienen la presencia de genes asociados a la resistencia a arsénico.

#### **4.3. Relación de la abundancia de genes de interés y profundidad del ecosistema marino según las estaciones de Malaspina usando Rstudio.**

De todos los genes de interés analizados en los datos de Malaspina, el gen G3 fue el que presenta una abundancia mayor y un patrón repetitivo en las gráficas de profundidad, por lo cual, en el presente trabajo se plasma las gráficas del gen 3 en relación de su abundancia y profundidad en el ecosistema marino según los datos de Malaspina (Figura N°4).



**Figura N°4: Abundancia del gen G3 según la profundidad del mar en los datos de Malaspina.** Gráfica de abundancia del gen vs profundidad del mar en 5 estaciones seleccionadas. Cada punto de la recta indica una capa oceanográfica; en el cual, el verde es para epipelágico, azul para mesopelágico y rojo para batipelágica.

## V. DISCUSIÓN

### 5.1. Identificación de genes asociados a la resistencia a arsénico

La Tabla N°1 indica la presencia de los genes asociados a la resistencia a arsénico según los datos extraída de NCBI (*operonArs2.clus90.fasta*) en los catálogos de genes de la Expedición de Malaspina 2010 y Tara Oceans, respectivamente. Según los datos de Tara Oceans, los genes identificados provienen de bacterias gammaproteobacterias (Rhodobacteraceae, *Marinobacter* sp., *Vibrionaceae*, *Shewanella frigidimarina*, *Pseudoalteromonas* sp., *Idiomarina* sp.) o alfa proteobacterias (*Thalassospira* sp) (Schoch et al., 2020). Se evidenció la presencia de posibles genes *arsJ* descritos en 3 microorganismos diferentes tanto en los catálogos de Tara Oceans y de Malaspina 2010, todos superando el 90% de identidad. El *arsJ* es una permeasa de eflujo que extruye organoarsénico inestable, que es el 1-arseno-3-fosfoglicerato (1As3PGA), disociándose en el ambiente extracelular a As (V) y 3 fosfoglicerato (3PGA); por lo cual, el *arsJ* forma parte de una vía de resistencia al arseniato que es el arsénico con valencia 5 (Chen et al., 2016). Por lo tanto, los microorganismos que portan el gen *arsJ* presentan una posible resistencia al arseniato, ya que disminuye su acumulación intracelular. Pero se ha descrito que el *arsJ* trabaja en conjunto con el *gapdh*, gen que codifica la enzima glucolítica gliceraldehído-3-fosfato deshidrogenasa, que permite formar el organoarsénico 1As3PGA para luego extruirlo (Zhao, 2016). Por lo cual, sería ventajoso analizar a futuro la correlación de ambos genes en los catálogos de Malaspina y Tara Oceans.

Por otro lado, hay otros genes únicos para cada catálogo. En el caso de Tara Oceans, hay más genes *arsJ* que superan el 95% de identidad, observando que el microorganismo *Shewanella frigidimarina* presenta un 99% de identidad y *Vibrio campbellii* con un 98% de identidad. Estos valores altos de porcentaje de identidad y con valor E muy pequeños (datos no mostrados) indican que los genes de referencia (*operonArs2.clus90.fasta*) están presentes con una alta confiabilidad en los datos de

Tara Oceans, incluso hay correspondencia entre los microorganismos portadores de dichos genes tanto de los datos de Tara Oceans como de los datos de referencia extraída de NCBI. Además, se encontró el gen *arsI* en *Cupriavidus metallidurans* con un 98% de identidad. Confiriendo a esta bacteria una posible resistencia a arsénico debido a que el gen *arsI* codifica una dioxigenasa dependiente de hierro no hemo con actividad de C-As liasa, es decir, rompe enlaces carbono-arsénico incorporando átomos de oxígeno (Pawitwar et al., 2017), convirtiendo el arsenito monometilado, MAS (III), a arsenito, As (III), que es un componente menos tóxico. Además, la enzima *ArsI* rompe enlaces C-As en arsénicos aromáticos (Yang & Rosen, 2016).

Así mismo, hay otros genes únicos para el catálogo de Malaspina, como el gen *arsA* con un 98% de identidad y el gen *arsB* con un 96% de identidad, para *Staphylococcus hominis* y *Agrobacterium fabrum*, respectivamente. Las proteínas que codifican, *ArsA* y *ArsB*, pueden formar un complejo de translocación de oxianión, en el que *ArsA* funciona como una ATPasa, mientras que *ArsB* es un transportador de eflujo de membrana. Pero encontrar cada gen en microorganismos diferentes no nos brinda indicios de la posibilidad de que dichos microorganismos sean resistentes a arsénico mediante la formación de un complejo. No obstante, el *ArsB* puede funcionar solo (sin *ArsA*) en la extrusión de arsenito (Wang et al., 2009). Dando la posibilidad de que *Agrobacterium fabrum* sea resistente a arsenito. Incluso, *ArsB* es una bomba transmembrana que puede dar lugar a la extrusión de As (III) y Sb (III) (Castillo & Saier, 2010).

## **5.2. Distribución y abundancia de genes asociados a la resistencia al arsénico en los datos de Tara Oceans**

Al relacionar los genes identificados y su abundancia máxima con su distribución geográfica, se pudo identificar que la mayoría de los genes se encuentran ubicados en el Océano Pacífico (Anexos: Tabla S1, S2 y S3 y Figura N°1). Resaltando que el único gen con mayor abundancia en las tres capas oceánicas de Tara Oceans (SUR, DCM y

MES) lo presenta el gen *J1* (Tabla N°2), por lo que posteriormente se precisará en analizar su distribución y abundancia de dicho gen.

En la Figura N°2 se evidencia que, en las tres capas oceanográficas, el gen *J1* presenta una amplia distribución a nivel global en los diferentes océanos y mares (Sunagawa et al., 2015). Con respecto a la capa SUR, se observa la mayor abundancia del gen *J1* con respecto a las otras capas oceanográficas, llegando incluso a una abundancia representando con el círculo 40 (abundancia real de 49,3693) en un punto de muestreo ubicado en el océano Pacífico Sur; además, se observa la presencia del gen *J1* con una abundancia presentada con el círculo de 10 (abundancia real menor a 10, oscilando entre 0,07 a 5,6) en todos los océanos y mares. Por otro lado, con respecto a la capa DCM, la abundancia máxima del gen *J1* es menor a las otras dos capas, pero que presenta puntos de muestreo con una abundancia representando con el círculo 12 tanto en el océano Pacífico Sur como en el Mar Rojo (abundancia real de 15,7282 y 15,1690, respectivamente). Finalmente, con respecto a la capa MES, su abundancia máxima representado con el círculo de 20 (abundancia real igual a 27,6558) está presente en el Océano Pacífico Norte, pero también presenta una abundancia destacable en el Océano Pacífico Sur y Océano Atlántico Sur (abundancia real de 13 y 23,4, respectivamente). El gen *J1* presenta una amplia distribución oceanográfica, resaltando su mayor abundancia en el océano Pacífico Sur en las tres capas oceanográficas de Tara Oceans. La distribución y abundancia de este gen *J1* está respaldada por el microorganismo que lo contiene, *Marinobacter hydrocarbonoclasticus*, que pertenece a un género que se encuentran en localidades tan diversas como el océano profundo, el agua de mar costera, hidrotermales, entre otros ambientes acuáticos y terrestres. Además, presenta rasgos fisiológicos notables que abarcan la oxidación de Fe (II) dependiente de nitrato, el ciclo redox del arsénico y fumarato, y la oxidación de Mn (II). Considerando su ubicuidad y capacidades metabólicas del género, las especies del género *Marinobacter* pueden desempeñar un papel importante en el ciclo biogeoquímico de compuestos

orgánicos y metales, como el arsénico, en variados hábitats marinos, y abarcando gradientes redox aeróbicos a anóxicos (Handley & Lloyd, 2013). Por lo cual, se podría postular como un buen candidato a futuro para estrategias de biorremediación.

### **5.3. Distribución y abundancia de genes asociados a la resistencia al arsénico en los datos de Malaspina 2010**

Las estaciones fueron seleccionados según el patrón de incremento de abundancia en cualquier zona del océano pelágico. Según los datos de Malaspina, se observa que su abundancia incrementa tanto en la zona mesopelágica y batipelágica, incluso en la mayoría de las estaciones en donde el gen G3 está presente, se evidencia que en la zona batipelática hay mayor abundancia que en la mesopelágica. Por lo cual, se plantea que la distribución del gen G3 incrementa a mayor profundidad oceánica. Además, tres de las cinco estaciones mostradas en la Figura N°4 (“station 120, 101 y 83”) están presentes en el océano Pacífico y las otras dos estaciones, en el océano Índico.

Los resultados obtenidos de los datos de Malaspina son posiblemente a que el gen G3 corresponde al gen *arsJ* de *Marinobacter hydrocarbonoclasticus* con 93% de identidad. Dicha bacteria, al pertenecer al género *Marinobacter*, puede colonizar localidades diversas e incluso el océano profundo (Handley & Lloyd, 2013) como lo respalda estos resultados de Malaspina. Por otro lado, estos datos de las cinco estaciones seleccionadas, analizándolos desde un enfoque de abundancia vs temperatura, se observa que a menor temperatura del ambiente marino mayor es la abundancia de dicho gen. Esto puede ser debido a que dicha bacteria es capaz de adaptarse a ambientes con baja temperatura y usar fuentes carbónicas (Branchu et al., 2017) pero que aún requiere más estudios en profundidades marinas.

En este estudio se han seleccionado resultados relevantes, tanto de los genes encontrados en Malaspina como en Tara Oceans. Por otro lado, no se puede comparar directamente las abundancias y distribuciones de ambos catálogos con los genes encontrados de la propia base de datos (*AvsT.m8* o *AvsM.m8*), debido a que estos datos



no presentan un mismo nivel geográfica de muestreo ni mucho menos valores estandarizados de abundancia para ambas metadatas.

Este estudio *in silico* ha permitido satisfactoriamente emplear las herramientas bioinformáticas descritos en la metodología con el fin de seguir buscando genes de interés en biorremediación en base a distintos parámetros fisicoquímicos de ambientes marinos y, así mismo, en otros catálogos de genes.

## VI. CONCLUSIONES

1. Según la base de datos de genes de Tara Oceans y de Malaspina 2010, se identificó al gen *arsJ*, que codifica para una permeasa de eflujo que extruye organoarsénico, al gen *arsI*, que codifica una dioxigenasa con actividad de C-As liasa, al gen *arsA* y al gen *arsB*, que forman parte de un complejo de translocación de oxianión, distribuidos en los océanos y/o mares del globo terráqueo.
2. Según los datos de Tara Oceans, el gen con mayor abundancia génica fue el gen *J1* de *Marinobacter* sp. con un 95% de identidad con el gen *arsJ* de *Marinobacter hydrocarbonoclasticus*, presentando una distribución en 3 océanos (Pacífico, Atlántico e Índico) y en 9 mares, inclusive en las tres capas oceanográficas de profundidad (SUR, DCM, MES) de los 3 océanos, en el Mar Árábigo y en el Golfo de México.
3. Según los datos de Malaspina 2010, el gen con mayor abundancia fue el gen *G3* con un 93% de identidad con el gen *arsJ* de *Marinobacter hydrocarbonoclasticus*, presentando una distribución en el Océano Pacífico y en el Océano Índico, incrementando su distribución a mayor profundidad oceánica (desde el epipelágico hasta el batipelágico).

## VII. REFERENCIAS BIBLIOGRÁFICAS

- Acinas, S. G., Sánchez, P., Salazar, G., Cornejo-Castillo, F. M., Sebastián, M., Logares, R., ... Gasol, J. M. (2021). Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Communications Biology*, 4(1). <https://doi.org/10.1038/s42003-021-02112-2>
- Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., ... Zbicz, K. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkx1095>
- Andres, J., & Bertin, P. N. (2016). The microbial genomics of arsenic. *FEMS Microbiology Reviews*. <https://doi.org/10.1093/femsre/fuv050>
- Bhattacharjee, H., & Rosen, B. P. (2007). Arsenic Metabolism in Prokaryotic and Eukaryotic Microbes. In *Molecular Microbiology of Heavy Metals*. [https://doi.org/10.1007/7171\\_2006\\_086](https://doi.org/10.1007/7171_2006_086)
- Branchu, P., Canette, A., Fernandez, S. M., Mounier, J., Meylheuc, T., Briandet, R., ... Naïtali, M. (2017). Impact of temperature on *Marinobacter hydrocarbonoclasticus* SP17 morphology and biofilm structure during growth on alkanes. *Microbiology (United Kingdom)*. <https://doi.org/10.1099/mic.0.000466>
- Castillo, R., & Saier, M. H. (2010). Functional promiscuity of homologues of the bacterial ArsA ATPases. *International Journal of Microbiology*. <https://doi.org/10.1155/2010/187373>
- Chauhan, N. S., Ranjan, R., Purohit, H. J., Kalia, V. C., & Sharma, R. (2009). Identification of genes conferring arsenic resistance to *Escherichia coli* from an effluent treatment plant sludge metagenomic library. *FEMS Microbiology Ecology*. <https://doi.org/10.1111/j.1574-6941.2008.00613.x>
- Chen, J., Yoshinaga, M., Garbinski, L. D., & Rosen, B. P. (2016). Synergistic interaction

- of glyceraldehydes-3-phosphate dehydrogenase and ArsJ, a novel organoarsenical efflux permease, confers arsenate resistance. *Molecular Microbiology*. <https://doi.org/10.1111/mmi.13371>
- Chen, S., & Shao, Z. (2009). Isolation and diversity analysis of arsenite-resistant bacteria in communities enriched from deep-sea sediments of the Southwest Indian Ocean Ridge. *Extremophiles*. <https://doi.org/10.1007/s00792-008-0195-1>
- Das, S., Bora, S. S., Yadav, R. N. S., & Barooah, M. (2017). A metagenomic approach to decipher the indigenous microbial communities of arsenic contaminated groundwater of Assam. *Genomics Data*, 12, 89–96. <https://doi.org/10.1016/j.gdata.2017.03.013>
- Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T., Rappé, M. S., ... Eren, A. M. (2018). Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology*. <https://doi.org/10.1038/s41564-018-0176-9>
- Duarte, C. M. (2015). Seafaring in the 21st century: The Malaspina 2010 circumnavigation expedition. *Limnology and Oceanography Bulletin*. <https://doi.org/10.1002/lob.10008>
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btq461>
- Flora, S. J. S. (2015). Arsenic: Chemistry, Occurrence, and Exposure. In *Handbook of Arsenic Toxicology*. <https://doi.org/10.1016/B978-0-12-418688-0.00001-0>
- Fru, E. C., Callac, N., Posth, N. R., Argyraki, A., Ling, Y. C., Ivarsson, M., ... Kiliyas, S. P. (2018). Arsenic and high affinity phosphate uptake gene distribution in shallow submarine hydrothermal sediments. *Biogeochemistry*, 141(1), 41–62. <https://doi.org/10.1007/s10533-018-0500-8>

- Gadd, G. M. (2010). Metals, minerals and microbes: Geomicrobiology and bioremediation. *Microbiology*. <https://doi.org/10.1099/mic.0.037143-0>
- Garrels, M. (2004). Bash Guide for Beginners. *Linux*.
- Handley, K. M., & Lloyd, J. R. (2013). Biogeochemical implications of the ubiquitous colonization of marine habitats and redox gradients by *Marinobacter* species. *Frontiers in Microbiology*. <https://doi.org/10.3389/fmicb.2013.00136>
- Hashihama, F., Suwa, S., Kanda, J., Ehama, M., Sakuraba, R., Kinouchi, S., ... Hamasaki, K. (2019). Arsenate and microbial dynamics in different phosphorus regimes of the subtropical Pacific Ocean. *Progress in Oceanography*, 176(May). <https://doi.org/10.1016/j.pocean.2019.05.007>
- Henke, K. R. (2009). Arsenic: Environmental Chemistry, Health Threats and Waste Treatment. In *Arsenic: Environmental Chemistry, Health Threats and Waste Treatment*. <https://doi.org/10.1002/9780470741122>
- Jang, Y.-C., Somanna, Y., & Kim, H. (2016). Source, Distribution, Toxicity and Remediation of Arsenic in the Environment – A review. *International Journal of Applied Environmental Sciences ISSN*, 11(2), 973–6077. Retrieved from <http://www.ripublication.com>
- Jones, M. B., Schildhauer, M. P., Reichman, O. J., & Bowers, S. (2006). The new bioinformatics: Integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution, and Systematics*. <https://doi.org/10.1146/annurev.ecolsys.37.091305.110031>
- Karsenti, E., Acinas, S. G., Bork, P., Bowler, C., de Vargas, C., Raes, J., ... Zingone, A. (2011). A holistic approach to marine Eco-systems biology. *PLoS Biology*. <https://doi.org/10.1371/journal.pbio.1001177>
- Kirchman, D. L. (2008). Microbial Ecology of the Oceans: Second Edition. In *Microbial*

*Ecology of the Oceans: Second Edition*. <https://doi.org/10.1002/9780470281840>

- Kultima, J. R., Sunagawa, S., Li, J., Chen, W., Chen, H., Mende, D. R., ... Bork, P. (2012). MOCAT: A Metagenomics Assembly and Gene Prediction Toolkit. *PLoS ONE*, 7(10), 1–6. <https://doi.org/10.1371/journal.pone.0047656>
- Kurth, D., Amadio, A., Ordoñez, O. F., Albarracín, V. H., Gärtner, W., & Farías, M. E. (2017). Arsenic metabolism in high altitude modern stromatolites revealed by metagenomic analysis. *Scientific Reports*. <https://doi.org/10.1038/s41598-017-00896-0>
- McHardy, A. C., & Klotgen, A. (2017). Finding genes in genome sequence. *Methods in Molecular Biology*, 1525, 271–291. [https://doi.org/10.1007/978-1-4939-6622-6\\_11](https://doi.org/10.1007/978-1-4939-6622-6_11)
- Mirza, N., Mahmood, Q., Pervez, A., Ahmad, R., Farooq, R., Shah, M. M., & Azim, M. R. (2010). Phytoremediation potential of *Arundo donax* in arsenic-contaminated synthetic wastewater. *Bioresource Technology*, 101(15), 5815–5819. <https://doi.org/10.1016/j.biortech.2010.03.012>
- Oremland, R. S., Saltikov, C. W., Wolfe-Simon, F., & Stolz, J. F. (2009). Arsenic in the evolution of earth and extraterrestrial ecosystems. *Geomicrobiology Journal*, 26(7), 522–536. <https://doi.org/10.1080/01490450903102525>
- Ortega, A., Geraldí, N. R., Alam, I., Kamau, A. A., Acinas, S. G., Logares, R., ... Duarte, C. M. (2019). Important contribution of macroalgae to oceanic carbon sequestration. *Nature Geoscience*. <https://doi.org/10.1038/s41561-019-0421-8>
- Pawitwar, S. S., Nadar, V. S., Kandegedara, A., Stemmler, T. L., Rosen, B. P., & Yoshinaga, M. (2017). Biochemical Characterization of Arsl: A Novel C-As Lyase for Degradation of Environmental Organoarsenicals. *Environmental Science and Technology*. <https://doi.org/10.1021/acs.est.7b03180>
- Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., ...

- Searson, S. (2015). Open science resources for the discovery and analysis of Tara Oceans data. *Scientific Data*, 2(Lmd), 1–16. <https://doi.org/10.1038/sdata.2015.23>
- Rahman, M. A., & Hassler, C. (2014). Is arsenic biotransformation a detoxification mechanism for microorganisms? *Aquatic Toxicology*, 146, 212–219. <https://doi.org/10.1016/j.aquatox.2013.11.009>
- Saunders, J. K., Fuchsman, C. A., McKay, C., & Rocap, G. (2019). Complete arsenic-based respiratory cycle in the marine microbial communities of pelagic oxygen-deficient zones. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1818349116>
- Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., ... Karsch-Mizrachi, I. (2020). NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database*, Vol. 2020. <https://doi.org/10.1093/database/baaa062>
- Silver, S., & Phung, L. T. (2005). Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. *Applied and Environmental Microbiology*. <https://doi.org/10.1128/AEM.71.2.599-608.2005>
- Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*. <https://doi.org/10.1038/nbt.3988>
- Sunagawa, S., Acinas, S. G., Bork, P., Bowler, C., Acinas, S. G., Babin, M., ... de Vargas, C. (2020). Tara Oceans: towards global ocean ecosystems biology. *Nature Reviews Microbiology*, 18(8), 428–445. <https://doi.org/10.1038/s41579-020-0364-5>
- Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., ... Bork, P. (2015). Structure and function of the global ocean microbiome. *Science*. <https://doi.org/10.1126/science.1261359>

- Sutton, T. T. (2013). Vertical ecology of the pelagic ocean: Classical patterns and new perspectives. *Journal of Fish Biology*. <https://doi.org/10.1111/jfb.12263>
- Tsai, S. L., Singh, S., & Chen, W. (2009). Arsenic metabolism by microbes in nature and the impact on arsenic remediation. *Current Opinion in Biotechnology*. <https://doi.org/10.1016/j.copbio.2009.09.013>
- Villar, E., Cabrol, L., & Heimbürger-Boavida, L. E. (2020). Widespread microbial mercury methylation genes in the global ocean. *Environmental Microbiology Reports*. <https://doi.org/10.1111/1758-2229.12829>
- Wang, L., Jeon, B., Sahin, O., & Zhang, Q. (2009). Identification of an arsenic resistance and arsenic-sensing system in *Campylobacter jejuni*. *Applied and Environmental Microbiology*. <https://doi.org/10.1128/AEM.00149-09>
- Wurl, O., Shelley, R. U., Landing, W. M., & Cutter, G. A. (2015). Biogeochemistry of dissolved arsenic in the temperate to tropical North Atlantic Ocean. *Deep-Sea Research Part II: Topical Studies in Oceanography*. <https://doi.org/10.1016/j.dsr2.2014.11.008>
- Yang, H. C., & Rosen, B. P. (2016). New mechanisms of bacterial arsenic resistance. *Biomedical Journal*. <https://doi.org/10.1016/j.bj.2015.08.003>
- Zhao, F. J. (2016). A novel pathway of arsenate detoxification. *Molecular Microbiology*. <https://doi.org/10.1111/mmi.13395>
- Zhu, Y. G., Xue, X. M., Kappler, A., Rosen, B. P., & Meharg, A. A. (2017). Linking Genes to Microbial Biogeochemical Cycling: Lessons from Arsenic. *Environmental Science and Technology*. <https://doi.org/10.1021/acs.est.7b00689>



## VIII. ANEXOS

Para más información de los scripts usados en cada sección de la metodología, puede visitar el siguiente enlace: <https://drive.google.com/drive/folders/174JtD6uswVb-kFz7CnSVsDv-nbdHxTQM?usp=sharing>

**Tabla S1. Distribución y abundancia de los genes asociados a la resistencia a arsénico en la capa SUR de Tara Oceans**

Gen	Enumeración	Cuerpo de agua (C.A.)	Frecuencia del C.A.	Abundancia génica máx. (A.G.M.)	Estación de muestreo con A.G.M
Gen I	1	Mar Mediterráneo	2	0.3589	7
	2	Mar Rojo	1	0.1551	31
	3	Océano Pacífico Norte	3	0.1493	109
	4	Océano Pacífico Sur	1	0.0638	102
	5	Océano Atlántico Norte	1	0.0523	146
Gen J1	1	Océano Pacífico Sur	13	49.3693	94
	2	Mar Árabe	2	5.6422	41
	3	Océano Atlántico Sur	7	4.7350	76
	4	Mar Rojo	4	2.7896	31
	5	Océano Atlántico Norte	4	2.0969	4
	6	Océano Índico	5	1.9512	48
	7	Mar de Hoces	1	1.6658	84
	8	Océano Pacífico Norte	7	1.2822	109
	9	Mar Mediterráneo	4	0.7640	9
	10	Mar de Laquedivas	1	0.5487	42
	11	Canal de Mozambique	3	0.5332	57
	12	Golfo de México	1	0.4117	142
	13	Mar Caribe	1	0.3985	141
	14	Mar de Kara	1	0.1139	188
	15	Mar Jónico	1	0.0710	25
Gen J2	1	Mar Mediterráneo	3	0.316	30
	2	Océano Pacífico Norte	2	0.246	133
	3	Océano Pacífico Sur	4	0.178	100
	4	Mar de Laquedivas	1	0.122	42
Gen J3	1	Mar Siberiano del Este	1	0.1200	193
Gen J4	1	Océano Pacífico Norte	1	0.9749	140
Gen J5	1	Océano Pacífico Sur	8	1.016	93
	2	Océano Pacífico Norte	1	0.418	109
	3	Mar Árabe	1	0.281	41
	4	Océano Atlántico Sur	2	0.247	72
	5	Mar Rojo	1	0.137	32

	6	Mar Mediterráneo	2	0.081	9
	7	Océano Índico	1	0.006	45
Gen J6	1	Océano Atlántico Sur	3	0.2629	72
	2	Océano Pacífico Norte	1	0.1297	138
Gen J7	1	Mar Rojo	1	0.5148	31
	2	Golfo de México	1	0.2692	142
	3	Océano Atlántico Sur	1	0.1552	82
	4	Océano Atlántico Norte	6	0.1520	145
	5	Mar Mediterráneo	3	0.1314	18
	6	Mar Jónico	1	0.1251	25
	7	Océano Pacífico Norte	3	0.1124	109
	8	Mar Caribe	1	0.0578	141
	9	Océano Pacífico Sur	1	0.0459	93

**Tabla S2. Distribución y abundancia de los genes asociados a la resistencia a arsénico en la capa DCM de Tara Oceans**

Gen	Enumeración	Cuerpo de agua (C.A.)	Frecuencia del C.A.	Abundancia génica máx. (A.G.M.)	Estación de muestreo con A.G.M
Gen I	1	Océano Pacífico Sur	1	7.2431	93
	2	Mar Jónico	1	1.1252	25
	3	Océano Pacífico Norte	1	0.3069	109
	4	Mar Mediterráneo	2	0.2442	18
Gen J1	1	Mar Rojo	2	15.7282	34
	2	Océano Pacífico Sur	8	15.1690	100
	3	Océano Atlántico Sur	5	7.1617	76
	4	Océano Pacífico Norte	5	6.8789	109
	5	Océano Índico	2	3.7547	64
	6	Canal de Mozambique	1	3.6671	58
	7	Mar Árabe	3	2.8980	41
	8	Mar de Laquedivas	1	1.0343	42
	9	Océano Atlántico Norte		0.7459	4
	10	Mar Jónico	1	0.5074	25
	11	Golfo de México	1	0.2888	142
	12	Mar Mediterráneo	3	0.2525	9
Gen J2	1	Océano Pacífico Sur	3	0.2438	93
	2	Océano Atlántico Sur	1	0.1560	76
	3	Mar Mediterráneo	2	0.1100	7
Gen J3	1	Océano Pacífico Sur	1	0.1346	93
Gen J4	1	Océano Pacífico Sur	1	18.2002	122
	2	Mar Árabe	1	0.1600	36
Gen J5	1	Océano Pacífico Sur	4	1.0957	100
	2	Mar Rojo	2	0.3706	32

	3	Mar Árabe	1	0.2620	38
	4	Mar Mediterráneo	1	0.1168	30
	5	Océano Atlántico Sur	1	0.0870	76
	6	Golfo de México	1	0.0251	142
Gen J6	1	Océano Atlántico Sur	2	0.7765	76
	2	Golfo de México	1	0.0206	142
Gen J7	1	Golfo de México	1	0.6271	142
	2	Océano Pacífico Sur	5	0.6081	122
	3	Mar Rojo	2	0.5505	34
	4	Océano Atlántico Sur	4	0.3420	72
	5	Mar Jónico	1	0.2621	25
	6	Mar Mediterráneo	1	0.2309	23
	7	Océano Índico	1	0.1876	52
	8	Océano Pacífico Norte	2	0.1433	132

**Tabla S3. Distribución y abundancia de los genes asociados a la resistencia a arsénico en la capa MES de Tara Oceans**

Gen	Enumeración	Cuerpo de agua (C.A.)	Frecuencia del C.A.	Abundancia génica máx. (A.G.M.)	Estación de muestreo con A.G.M
Gen I	1	Océano Pacífico Norte	3	2.2105	109
	2	Océano Atlántico Norte	2	0.1444	146
	3	Océano Pacífico Sur	3	0.1378	98
	4	Océano Atlántico Sur	1	0.0517	85
	5	Mar de Noruega	1	0.0299	158
	6	Golfo de México	1	0.0015	142
Gen J1	1	Océano Pacífico Norte	5	27.6558	109
	2	Océano Atlántico Sur	6	23.3872	76
	3	Océano Pacífico Sur	6	13.0455	102
	4	Golfo de México	1	2.1517	142
	5	Mar Árabe	3	1.5475	39
	6	Océano Índico	2	0.7855	64
	7	Océano Atlántico Norte	3	0.2302	152
	8	Canal de Mozambique	1	0.1403	56
Gen J2	1	Océano Pacífico Sur	2	1.0517	122
	2	Océano Pacífico Norte	2	0.7668	138
	3	Océano Atlántico Sur	1	0.2159	72
	4	Océano Atlántico Norte	1	0.0805	152
Gen J3	1	Estrecho de Davis	1	0.4417	209
	2	Fiordo de Uummannaq	1	0.1593	206
Gen J4	1	Océano Pacífico Sur	1	0.5553	122
	2	Océano Pacífico Norte	2	0.1650	138
	3	Mar Árabe	1	0.1549	38

	4	Golfo de México	1	0.1491	142
Gen J5	1	Océano Atlántico Sur	4	0.9351	76
	2	Océano Pacífico Sur	2	0.1638	112
	3	Océano Pacífico Norte	3	0.1598	138
	4	Golfo de México	1	0.1500	142
	5	Océano Atlántico Norte	1	0.1038	148
	6	Mar Árabe	1	0.0795	39
Gen J6	1	Océano Atlántico Sur	3	2.3161	76
	2	Golfo de México	1	0.2297	142
	3	Océano Pacífico Sur	1	0.1298	112
	4	Océano Pacífico Norte	1	0.0394	133
Gen J7	1	Golfo de México	1	4.8062	142
	2	Océano Atlántico Norte	6	3.2036	149
	3	Océano Atlántico Sur	6	1.5201	70
	4	Océano Pacífico Norte	3	0.8682	133
	5	Océano Pacífico Sur	6	0.6207	112
	6	Canal de Mozambique	1	0.1433	56