

#### UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO MAESTRÍA Y DOCTORADO EN CIENCIAS BIOQUÍMICAS INSTITUTO DE BIOTECNOLOGÍA

#### "ANÁLISIS DE ENZIMAS INVOLUCRADAS EN LA UTILIZACIÓN DE CARBOHIDRATOS EN METAGENOMAS DE SEDIMENTOS MARINOS."

**TESIS** QUE PARA OPTAR POR EL GRADO DE: DOCTOR EN CIENCIAS

> PRESENTA: RAFAEL LÓPEZ SÁNCHEZ

> > TUTOR PRINCIPAL

**LORENZO P. SEGOVIA FORCELLA** INSTITUTO DE BIOTECNOLOGÍA UNAM

MIEMBROS DEL COMITÉ TUTOR

ROSA MARÍA GUTIÉRREZ RÍOS INSTITUTO DE BIOTECNOLOGÍA UNAM

ERIA A. REBOLLAR CAUDILLO CENTRO DE CIENCIAS GENÓMICAS UNAM

CUERNAVACA MORELOS, OCTUBRE 2024



Universidad Nacional Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

#### DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

### Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca de posgrado que me fue concedida para llevar a cabo mis estudios. Este trabajo fue financiado a través de los proyectos IN209921 e IV200322 aprobados por la Dirección General de Asuntos del Personal Académico (DGAPA) de la UNAM y por el proyecto 269435 "Fortalecimiento de la Infraestructura del grupo de Biocatálisis en el IBt/UNAM" aprobado dentro del "Programa de Apoyos al Fortalecimiento y Desarrollo de la Infraestructura Científica y Tecnológica de CONACyT".

Al Dr. Lorenzo P. Segovia Forcella por permitirme formar parte del grupo de trabajo del Laboratorio 12 del Instituto de Biotecnología de la UNAM.

A las Dras. Rosa María Gutiérrez Ríos y Eria Rebollar Caudillo por conformar mi comité tutoral el cual fue esencial para la elaboración de este trabajo.

A la M.C. Blanca Ramos Cerrillo por el apoyo en la parte técnica de este proyecto, así como al M.C. Jérôme Jean Verleyen por el apoyo bioinformático y el uso del clúster del Instituto de Biotecnología de la UNAM "Teopanzolco". A la Dra. Grisel Alejandra Escobar Zepeda y la Dra. Elizabeth Ernestina Godoy Lozano por sus aportaciones en conocimientos en bioinformática para la realización de este trabajo.

A la Unidad de Secuenciación Masiva y Bioinformática del Instituto de Biotecnología de la UNAM, conformada por el Dr. Fidel Alejandro Sánchez Flores, el Dr. Ricardo Grande, la M.C. Verónica Jiménez Jacinto, el M.C Karel Estrada, el M.C. Jérôme Jean Verleyen, la Biol.Gloria Tanahiry Vazquez Castro, la Ing.Ilse Salinas Peralta, la I.Q. Lizeth Arely Matías Valdez, la Biol. Estefania Herrera Herrera y la I.S.C. Karla Leslie Matías Valdez.

A la Unidad de Cómputo de la UNAM, conformada por el M.C. Arturo Ocádiz Ramírez M. en T.I. Juan Manuel Hurtado Ramírez, el Mtro. Servando Aguirre Cruz, el Ing. Roberto Pablo Rodríguez Bahena y el M.E.M. David Santiago Castañeda Carreón.

### Dedicatoria

A Inés y a Andrea que no sólo son mi motor, sino que me hacen ser cada día la mejor versión de mí. A mis papás por todo su apoyo y cariño siempre en todos mis proyectos académicos y de vida. Al Negro y a Marina por no quitar nunca el dedo del renglón y estar siempre ahí. A mi tío Gonzalo quién me dio valor para salirme de mi zona de confort y probar nuevos retos en nuevos campos. A Margarita por darnos la oportunidad de crecer como familia.

A mi hermana May y a su esposo Santiago. A mis tíos María Isabel, Enrique, Conchita, Loli, Antonio, Isa, Julia y Manuel, a mis primos María Isabel, Alberto, Elena, Sara, Juanto y Saúl y a mis sobrinos Gonzalo, Jimena y Emilio por sus consejos y cariño. A Chata por ser mi compañía en numerosas sesiones de escritura de este manuscrito, así como recibirme siempre regresando del laboratorio.

A Fer, R2, Brazos, Gus, Enriquito, Pau, Carmen, Nat, Memo, Manu, Toñito, Pato y Leo y a todos los Vikingos de Caravaca quiénes han sido un apoyo y mis grandes amigos. Al Almería a quién le deseo muchos títulos como este en los próximos años.

A Lorenzo, Blanca y Alejandro por todo lo que aprendí en el laboratorio 12, así como al resto de integrantes de este en estos últimos 8 años. A Delia y a Omar por no dejarme volverme "legionario". A Alejandro, Jérôme, Leslie, Vero, Leti y Tina, Diana, Rafa y Julián por siempre ayudarme con cualquier problema bioinformático y por hacer en los cursos de la Unidad. Al grupo de "Tardes de bioinfo" con los cuales he ido creciendo en la ciencia.

Esta tesis es una parte de todos ustedes.

Gracias.



#### PROTESTA UNIVERSITARIA DE INTEGRIDAD Y HONESTIDAD ACADÉMICA Y PROFESIONAL (Graduación con trabajo escrito)

De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción I, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académico.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado:

Análisis de enzimas involucradas en la utilización de carbohidratos en metagenomas de sedimentos marinos.

que presenté para obtener el grado de ---Doctorado---- es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi programa de posgrado, citando las fuentes de ideas, textos, irmágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia, acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad de los actos de carácter académico administrativo del proceso de graduación.

> Atentamente Batael Vopez Sánchez 406059534

•

(Nombre, firma y Número de cuenta de la persona alumna)

## Índice

1	Intro	oducción1
2	Ante	ecedentes4
	2.1	Metagenómica4
	2.2	El ciclo del carbono en los océanos6
	2.3	Tipos de sedimento marino9
	2.4	Metabolismo en los sedimentos12
	2.5	Carbohidratos14
	2.6	CAZymes15
	2.7	CAZy16
	2.8	Óxido reductasas19
	2.9	Estudios in silico en sedimentos marinos20
3	Just	ificación22
4	Hipo	ótesis
5	Obje	etivos
6	Met	odología
	6.1	Programas
		0
	6.2	Selección de datos genómicos; muestras metagenómicas
	6.2 6.3	Selección de datos genómicos; muestras metagenómicas
	6.2 6.3 6.4	Selección de datos genómicos; muestras metagenómicas23 Control de calidad y preprocesamiento de las muestras24 Análisis gen-céntrico
	6.2 6.3 6.4 6.4.1	Selección de datos genómicos; muestras metagenómicas
	6.2 6.3 6.4 6.4.1 6.4.2	Selección de datos genómicos; muestras metagenómicas       23         Control de calidad y preprocesamiento de las muestras       24         Análisis gen-céntrico       24         Análisis taxonómico de las lecturas metagenómicas       24         Ensamble de las lecturas metagenómicas       24         Anélisis taxonómico de las lecturas metagenómicas       24         Anélisis de las lecturas metagenómicas       24         Anélisis de las lecturas metagenómicas       24
	6.2 6.3 6.4 6.4.1 6.4.2 6.4.3 6.4.4	Selección de datos genómicos; muestras metagenómicas       23         Control de calidad y preprocesamiento de las muestras       24         Análisis gen-céntrico       24         Análisis taxonómico de las lecturas metagenómicas       24         Ensamble de las lecturas metagenómicas       24         Análisis funcional del metagenoma       24         Análisis estadísticos de los metagenomas       25
	6.2 6.3 6.4 6.4.1 6.4.2 6.4.3 6.4.4 6.5	Selección de datos genómicos; muestras metagenómicas       23         Control de calidad y preprocesamiento de las muestras       24         Análisis gen-céntrico       24         Análisis taxonómico de las lecturas metagenómicas       24         Ensamble de las lecturas metagenómicas       24         Análisis funcional del metagenoma       24         Análisis estadísticos de los metagenomas       25         Análisis genoma-céntrico       25
	6.2 6.3 6.4 6.4.1 6.4.2 6.4.3 6.4.4 6.5 6.5.1	Selección de datos genómicos; muestras metagenómicas23Control de calidad y preprocesamiento de las muestras24Análisis gen-céntrico24Análisis taxonómico de las lecturas metagenómicas24Ensamble de las lecturas metagenómicas24Análisis funcional del metagenoma24Análisis estadísticos de los metagenomas25Análisis genoma-céntrico25Reconstrucción de genomas a partir de metagenomas (MAG)25
	6.2 6.3 6.4 6.4.1 6.4.2 6.4.3 6.4.4 6.5 6.5.1 6.5.2 6.5.3	Selección de datos genómicos; muestras metagenómicas       23         Control de calidad y preprocesamiento de las muestras       24         Análisis gen-céntrico       24         Análisis taxonómico de las lecturas metagenómicas       24         Ensamble de las lecturas metagenómicas       24         Análisis funcional del metagenoma       24         Análisis estadísticos de los metagenomas       25         Análisis genoma-céntrico       25         Reconstrucción de genomas a partir de metagenomas (MAG)       25         Anotación taxonómica y análisis filogenético MAG       25
	6.2 6.3 6.4 6.4.1 6.4.2 6.4.3 6.4.4 6.5 6.5.1 6.5.2 6.5.3 6.6	Selección de datos genómicos; muestras metagenómicas       23         Control de calidad y preprocesamiento de las muestras       24         Análisis gen-céntrico       24         Análisis taxonómico de las lecturas metagenómicas       24         Ensamble de las lecturas metagenómicas       24         Análisis funcional del metagenoma       24         Análisis estadísticos de los metagenomas       25         Análisis genoma-céntrico       25         Reconstrucción de genomas a partir de metagenomas (MAG)       25         Anotación taxonómica y análisis filogenético MAG       25         Anotación funcional MAG       26
7	6.2 6.3 6.4 6.4.1 6.4.2 6.4.3 6.4.3 6.4.4 6.5 6.5.1 6.5.2 6.5.3 6.6	Selección de datos genómicos; muestras metagenómicas       23         Control de calidad y preprocesamiento de las muestras       24         Análisis gen-céntrico       24         Análisis taxonómico de las lecturas metagenómicas       24         Ensamble de las lecturas metagenómicas       24         Análisis funcional del metagenoma       24         Análisis genoma-céntrico       25         Análisis genoma-céntrico       25         Reconstrucción de genomas a partir de metagenomas (MAG)       25         Anotación funcional MAG       26         Graficación       26
7	6.2 6.3 6.4 6.4.1 6.4.2 6.4.3 6.4.4 6.5 6.5.1 6.5.2 6.5.3 6.6 <i>Res</i>	Selección de datos genómicos; muestras metagenómicas23Control de calidad y preprocesamiento de las muestras24Análisis gen-céntrico24Análisis taxonómico de las lecturas metagenómicas24Ensamble de las lecturas metagenómicas24Análisis funcional del metagenoma24Análisis genoma-céntrico25Análisis genoma-céntrico25Reconstrucción de genomas a partir de metagenomas (MAG)25Anotación taxonómica y análisis filogenético MAG25Anotación funcional MAG26Graficación26Oraficación28Outlados y Discusión20
7	6.2 6.3 6.4 6.4.1 6.4.2 6.4.3 6.4.4 6.5 6.5.1 6.5.2 6.5.3 6.6 <b>Res</b> 7.1	Selección de datos genómicos; muestras metagenómicas       23         Control de calidad y preprocesamiento de las muestras       24         Análisis gen-céntrico       24         Análisis taxonómico de las lecturas metagenómicas       24         Ensamble de las lecturas metagenómicas       24         Análisis funcional del metagenoma       24         Análisis genoma-céntrico       24         Análisis genoma-céntrico       25         Reconstrucción de genomas a partir de metagenomas (MAG)       25         Anotación taxonómica y análisis filogenético MAG       25         Anotación funcional MAG       26         Graficación       28         Sedimentos Marinos y su localización       28
7	6.2 6.3 6.4 6.4.1 6.4.2 6.4.3 6.4.4 6.5 6.5.1 6.5.2 6.5.3 6.6 Res 7.1 7.2 7.2 1	Selección de datos genómicos; muestras metagenómicas23Control de calidad y preprocesamiento de las muestras24Análisis gen-céntrico24Análisis taxonómico de las lecturas metagenómicas24Ensamble de las lecturas metagenómicas24Análisis funcional del metagenoma24Análisis genoma-céntrico25Análisis genoma-céntrico25Reconstrucción de genomas a partir de metagenomas (MAG)25Anotación taxonómica y análisis filogenético MAG25Anotación funcional MAG26Graficación26Ultados y Discusión28Sedimentos Marinos y su localización31Taxonomía de muestras metagenómicas31
7	6.2 6.3 6.4 6.4.1 6.4.2 6.4.3 6.4.4 6.5 6.5.1 6.5.2 6.5.3 6.6 <b>Res</b> 7.1 7.2 7.2.1 7.2.1 7.2.2	Selección de datos genómicos; muestras metagenómicas       23         Control de calidad y preprocesamiento de las muestras       24         Análisis gen-céntrico       24         Análisis taxonómico de las lecturas metagenómicas       24         Análisis taxonómico de las lecturas metagenómicas       24         Análisis funcional del metagenoma       24         Análisis genoma-céntrico       24         Análisis genoma-céntrico       25         Reconstrucción de genomas a partir de metagenomas (MAG)       25         Anotación taxonómica y análisis filogenético MAG       25         Anotación funcional MAG       26         Graficación       26         Matisis Gen-Céntrico       28         Sedimentos Marinos y su localización       28         Análisis Gen-Céntrico       31         Taxonomía de muestras metagenómicas       31         Carcelación de diversidad taxonómica de los sedimentos con los metadatos asociados       31
7	6.2 6.3 6.4 6.4.1 6.4.2 6.4.3 6.4.4 6.5 6.5.1 6.5.2 6.5.3 6.6 <b>Res</b> 7.1 7.2.1 7.2.1 7.2.1	Selección de datos genómicos; muestras metagenómicas       23         Control de calidad y preprocesamiento de las muestras       24         Análisis gen-céntrico       24         Análisis taxonómico de las lecturas metagenómicas       24         Ensamble de las lecturas metagenómicas       24         Análisis funcional del metagenoma       24         Análisis genoma-céntrico       25         Reconstrucción de genomas a partir de metagenomas (MAG)       25         Anotación taxonómica y análisis filogenético MAG       25         Anotación funcional MAG       26         Graficación       26         Sedimentos Marinos y su localización       28         Análisis Gen-Céntrico       31         Taxonomía de muestras metagenómicas       31         Correlación de diversidad taxonómica de los sedimentos con los metadatos asociados 33       31

	7.2.4	Beta Diversidad de las muestras metagenómicas en cuanto a su disponibilidad de	
	oxígen 7.2.5	o 35 Anotación funcional de CAZymes en muestras de sedimentos	38
7	7.3 A	Análisis Genoma-Céntrico	45
	7.3.1	Genomas ensamblados a partir de metagenomas (MAG)	45
	7.3.2	Anotación de CAZymes en los MAG	46
	7.3.3	CAZymes en MAG de sedimento vs CAZymes en MAG de suelo	50
8	Conc	lusiones	. 56
9	<ul> <li>Perspectivas</li> <li>Referencias bibliográficas</li> </ul>		. 57
10			58
11	Bib	liografía	. 68

#### Abstract

Marine sediments are the largest long-term storage for carbon on Earth. Microorganisms residing in these sediments facilitate the conversion of fixed oceanic carbon, but their exact role in the carbon cycle remains unclear. Earlier studies using culture-independent methods have highlighted the crucial role of carbohydrates in the carbon cycle. In this research, we use two metagenomic approaches; first, a gene-centric approach to examine the distribution and abundance of carbohydrate-active enzymes (CAZymes) across 37 marine sediment sites. These sediments show different levels of oxygen availability and were collected from various regions of the world. Our comparative analysis focusses on the metabolic potential for oxygen utilization, based on genes identified in both oxic and anoxic environments. We found that secreted CAZyme families involved in the degradation of plant and algal detritus, necromass, and host glycans were prevalent in all metagenomic samples. The analysis suggests that oxic/anoxic conditions not only influence the taxonomic composition of microbial communities, but also affect the presence of CAZyme families related to the breakdown of necromass, algae, and plant detritus (López-Sánchez et al., 2024). The second approach was genome-centric to better understand the microbial taxa in the sediments, so we reconstructed metagenome-assembled genomes (MAG) and analyzed the presence of primary secreted carbohydrate-active enzyme (CAZyme) families. Our findings indicate that the primary CAZyme families and CAZyme Gene Clusters (CGC) identified in our metagenomes were prevalent in the Bacteroidia, Gammaproteobacteria, and Alphaproteobacteria classes. When comparing these MAGs with organisms of the same taxonomic classes in the soil, we found similar CAZyme repertoires, although the soil MAG contained a more abundant and diverse CAZyme content. Furthermore, the data suggest that the abundant classes in our metagenomic samples, specifically Alphaproteobacteria, Bacteroidia, and Gammaproteobacteria, play a crucial role in carbohydrate transformation within the upper few meters of the sediments (López-Sánchez et al., 2024).

#### Resumen

Los sedimentos marinos son el mayor reservorio de almacenamiento de carbono a largo plazo en la Tierra. Los microorganismos que residen en estos sedimentos facilitan la conversión del carbono oceánico fijado, pero su papel exacto en el ciclo del carbono aún no está claro. Estudios anteriores que utilizaron métodos independientes del cultivo han destacado el papel crucial de los carbohidratos en el ciclo del carbono. En esta investigación, utilizamos dos enfoques metagenómicos; primero, un enfoque centrado en los genes para examinar la distribución y abundancia de enzimas activas en carbohidratos (CAZymes) en 37 sitios de sedimentos marinos. Estos sedimentos muestran diferentes niveles de disponibilidad de oxígeno y fueron recolectados en varias regiones del mundo. Nuestro análisis comparativo se centra en el potencial metabólico para la utilización de oxígeno, basado en genes identificados en ambientes tanto óxicos como anóxicos. Descubrimos que las familias CAZyme secretadas involucradas en la degradación de detritos de plantas y algas, necromasa y glicanos del huésped prevalecían en todas las muestras metagenómicas (López-Sánchez et al., 2024). El análisis sugiere que las condiciones óxicas/anóxicas no solo influyen en la composición taxonómica de las comunidades microbianas, sino que también afectan la presencia de familias CAZyme

relacionadas con la descomposición de necromasa, algas y detritos vegetales. El segundo enfoque se centró en el genoma, para comprender mejor los taxones microbianos en los sedimentos, por lo que reconstruimos genomas ensamblados en metagenomas (MAG) y analizamos la presencia de familias de enzimas primarias activas en carbohidratos secretadas (CAZyme). Nuestros hallazgos indican que las familias primarias de CAZyme y los grupos de genes CAZyme (CGC) identificados en nuestros metagenomas prevalecían en las clases Bacteroidia, Gammaproteobacteria y Alphaproteobacteria. Al comparar estos MAG con organismos de las mismas clases taxonómicas en el suelo, encontramos repertorios de CAZyme similares, aunque el MAG del suelo contenía un contenido de CAZyme más abundante y diverso. Además, los datos sugieren que las clases abundantes en nuestras muestras metagenómicas, específicamente Alphaproteobacteria, Bacteroidia y Gammaproteobacteria, desempeñan un papel crucial en la transformación de carbohidratos dentro de los pocos metros superiores de los sedimentos (López-Sánchez et al., 2024).

### 1 Introducción

El océano constituye un ambiente que impone numerosos desafíos para la investigación científica. Este ecosistema se caracteriza por una escasez de compuestos orgánicos y una tasa metabólica microbiana reducida, lo cual es consecuencia directa de las condiciones extremas prevalecientes, que incluyen altas presiones hidrostáticas y temperaturas considerablemente bajas. Estos factores intrínsecos de la biósfera de los océanos hacen que el estudio y la comprensión de sus dinámicas y poblaciones sean una tarea compleja (Moran et al., 2017).

A pesar de ser ambientes pobres en nutrimentos, es decir oligotróficos, los sedimentos son el repositorio de más del 70% del carbono secuestrado en la materia orgánica disuelta y es un reservorio global muy grande en el ciclo del carbono (Moran et al., 2017). Debido a esto, los sedimentos marinos son críticos para el almacenaje a largo plazo de carbono y el reciclado de la materia orgánica en estos nichos. Los microorganismos que viven en los sedimentos marinos controlan el almacenaje de cantidades masivas de carbono y tienen un importante rol en los ciclos biogeoquímicos (Orsi 2018).

Los grados de sedimentación dependen de la producción primaria de la superficie del océano, que es controlada por la entrega de nutrimentos como lo son el nitrato, el fosfato y el hierro que limitan el crecimiento del fitoplancton. El fitoplancton es la primera fuente de materia orgánica y de carbohidratos en los sistemas marinos. Alrededor del 5-30% de la producción del fitoplancton es directamente liberada como carbono orgánico disuelto o particulado (Biersmith & Benner 1998). Cerca de la costa, en los márgenes continentales, las corrientes transportan agua fría rica en nutrientes del océano profundo a la superficie y esto incrementa la producción de fitoplancton y así las tasas de sedimentación (Orsi 2018).

En estos ecosistemas de sedimento marino del océano profundo, los metabolismos microbianos son determinados por la deposición de carbono orgánico proveniente de procesos fotosintéticos de la superficie del océano (Tully et al., 2016). Las partículas y agregados ("nieve marina") están presentes en toda la columna de agua y las partículas que rápidamente se hunden jugando un papel muy importante en el transporte de la materia orgánica al sedimento. (Arnosti 2011). La deposición de nieve marina, los detritos y las partículas biodegradables proveen muchos nutrimentos para los organismos del sedimento (Flemming et al., 2019).

En general, la remineralización de materia orgánica está en su mayoría controlada por procariotes usando enzimas extracelulares que se encuentran secretadas fuera de la célula para hidrolizar la materia orgánica particulada y disuelta de alto peso molecular en moléculas de bajo peso molecular antes de la incorporación a las células. Así las actividades enzimáticas extracelulares son el paso limitante en el ciclo del carbono marino (Zhao Z. et al., 2020). Las especificidades del sustrato de las enzimas extracelulares, los factores que controlan la producción y distribución de los mismos, son centrales al reciclado de carbono en sistemas marinos (Arnosti 2011).

Los sedimentos pueden ser clasificados por su disponibilidad de oxígeno en óxicos y anóxicos. Los sedimentos marinos típicamente se convierten en anóxicos debido a que el oxígeno se consume en los primeros centímetros del sedimento y por debajo está el metabolismo anaeróbico, incluyendo la reducción del sulfato que es responsable de una cantidad sustancial de la remineralización. En una de las capas bajas, el sulfato es convertido en el altamente tóxico sulfuro de hidrógeno que se acumula y envenena el ambiente para la gran mayoría de microorganismos que consumen oxígeno (Nealson et al., 2010).

Se ha estimado que el microbioma del océano consiste en más de cien mil diferentes taxones bacterianos, arqueanos y eucarióticos con estrategias ecológicas y metabólicas diversas para la producción y el consumo de carbono fijado. (Moran et al., 2017). La limitación de energía tanto en sedimentos óxicos como anóxicos lleva a tasas bajas de actividad y reciclado de biomasa en el subsuelo marino y se ha estimado que los microorganismos del subsuelo tienen tiempos de generación que van de los meses a los millones de años. El recambio de la biomasa microbiana del subsuelo está regulado por los virus que proveen un mecanismo de regeneración de nutrimentos y son vectores potenciales para la transferencia horizontal de genes y la recombinanción (Orsi 2018).

Los restos de células microbianas (necromasa) como la lisis de productos celulares o restos de las paredes celulares de bacterias y endoesporas, se pueden acumular a través del tiempo y pueden contribuir al pool de material orgánica de sedimentos. Ya que una fracción considerable de la materia orgánica se degrada por microorganismos enterrados de hace millones de años, el total de la producción de biomasa microbiana integrada a través del tiempo puede ser una fracción significativa de la materia orgánica inicialmente enterrada (Arndt et al., 2013).

Estudios recientes de genes, transcritos y metagenomas de procariotas han destacado la importancia de los polisacáridos y sus transformaciones para el metabolismo del carbono en el océano (Teeling et al., 2012; Teeling et al., 2016). Los carbohidratos, cuyas transformaciones juegan un papel clave en los sedimentos marinos, representan una fracción significativa del carbono disuelto marino (15-30%) y son activamente liberados por el fitoplancton creciente (Biersmith & Benner 1998). Todas las clases mayores de carbohidratos ocurren en la materia orgánica disuelta y particulada, incluyendo los amino azúcares, los ácidos urónicos y las aldosas. A pesar de que los polisacáridos constituyen una gran fracción de los cuerpos de fitoplancton y macroalgas (Biersmith & Benner, 1998), así como de la materia orgánica disuelta y particulada, se sabe poco acerca de su procesamiento biogeoquímico (Lee et al., 2001).

En esas circunstancias, el entendimiento de las fuentes, transformaciones y destinos de los carbohidratos en los ambientes marinos puede darnos panorama general del reciclado fotosintético del carbono orgánico (Biersmith & Benner 1998).

Para ello existen las enzimas activas sobre carbohidratos (CAZymes) que son proteínas con actividades conocidas involucradas en la síntesis y degradación de glicoconjugados, oligoy polisacáridos. Típicamente corresponden al 1-5% de los genes de un organismo. La colección de CAZymes codificadas por el genoma de un organismo ("CAZoma") proporciona una idea de la naturaleza y el alcance del metabolismo de los carbohidratos complejos de la especie y es debido a la enorme variabilidad química, estructural y funcional de los carbohidratos, que los análisis y comparaciones de CAZomas resaltan diferencias significativas entre las mismas (Lombard et al., 2013).

Las CAZymes son enzimas clave para la degradación de carbohidratos que son la mayor parte de macromoléculas en organismos incluidos procariotes que habitan en el detrital marino orgánico que es la nieve marina (Zhao Z. et al., 2020). Una variedad de microorganismos procesan estos polisacáridos por la secreción de un conjunto de enzimas digestivas, conocido en conjunto como el secretoma. Estas enzimas degradan los compuestos complejos en el entorno extracelular y permiten la absorción del material procesado hacia el interior celular. Una vía significativa para la translocación de proteínas a través de las membranas celulares involucra secuencias específicas en los extremos amino (N) terminales de las proteínas denominados péptidos señal. Estos péptidos señal dictan la exportación de proteínas utilizando la ruta de secreción celular conocida como la vía Sec. Dichos péptidos presentan patrones de secuencia de aminoácidos conservados, que pueden ser reconocidos en proteínas teóricas mediante el uso de algoritmos de bioinformática especializados (Orsi et al., 2018). Al igual que otras enzimas que transforman la nieve marina, las CAZymes pueden ser secretadas y son una gran fracción del total de enzimas extracelulares que se encuentran en el ambiente, las cuales proveen una nueva y valiosa información de la ecología marina de procariotes y el reciclado de materia orgánica (Zhao Z. et al.,2020).

Las CAZymes han sido clasificadas y anotadas en la base de datos CAZy desde 1998. Esta es una base de datos especializada dedicada a la visualización y análisis de información genómica, estructural y bioquímica sobre enzimas activadas por carbohidratos (Lombard et al., 2014). Las CAZymes son extremadamente importantes para la investigación en la nutrición humana, la microbiota, bioenergía, las enfermedades de plantas y el reciclado global del carbono (Zheng et al., 2022).

Para esta tesis se llevó a cabo un estudio gen-céntrico de metagenomas de sedimento ubicados en varias partes del mundo, así como un acercamiento genoma-céntrico para la reconstrucción de especies abundantes en los metagenomas descritos y las anotaciones funcionales de ambos acercamientos para la búsqueda de la diversidad y abundancia de las CAZymes en los mismos.

### 2 Antecedentes

### 2.1 Metagenómica

En la ecología microbiana, las preguntas ¿qué microorganismos están ahí? y ¿qué están haciendo? (Figura 1) se pueden contestar con la metagenómica. (Escobar-Zepeda et al., 2015). La metagenómica se refiere al uso de métodos genómicos modernos para analizar comunidades de microorganismos en sus entornos naturales, lo que elimina la necesidad de aislar y cultivar especies individuales en el laboratorio (Chen et al., 2005). El análisis metagenómico, principalmente basado en la secuenciación de DNA, revela el potencial de actividades en una comunidad microbiana (De Filippis et al., 2017). Descifrar la diversidad microbiana de un metagenoma puede ser importante para entender su potencial metabólico. Esto puede obtenerse mediante un acercamiento metagenómico donde el total del DNA de una muestra se utiliza para preparar bibliotecas metagenómicas (Escobar-Zepeda et al., 2015).



# Metagenómica

Figura 1. Preguntas que contesta la metagenómica. Imagen obtenida de la página https://es.slideshare.net/slideshow/presentacinsobre-importancia-de-la-diversidad-biolgica-del-suelo/61714850#16

Un ejemplo de esto es la secuenciación del tipo shotgun que está transformando nuestra comprensión de los microbiomas asociados a humanos y otros entornos, facilitando la comparación de perfiles funcionales y taxonómicos mediante análisis computacional de comunidades microbianas. En la metagenómica shotgun, se pretende generar secuencias, o "lecturas", de todos los genomas presentes en una muestra, conjunto que denominamos metagenoma. Todos los análisis metagenómicos empiezan con el control de calidad de las lecturas en el cual se remueven secuencias que contienen bases mal anotadas y contaminantes que no son fuente de los ambientes secuenciados. (Tas et al., 2021). Típicamente la secuenciación tipo shotgun arroja resultados de una amplia cobertura de especies dominantes y menos lecturas de especies poco abundantes. (Kunath et al., 2017). Estos conjuntos de datos incluyen millones de lecturas cortas (fragmentos genómicos) que, idealmente, representan todos los microorganismos en una muestra ambiental (Segata et al., 2014). A pesar de la complejidad y el desafío de analizar estas bibliotecas metagenómicas, pueden ser utilizadas para determinar qué microorganismos están presentes y sus actividades, mediante la cuantificación de taxones o genes específicos (Nayfach et al., 2016).

Existen dos enfoques a la hora de hacer metagenómica del tipo shotgun; el análisis gencéntrico y el análisis genoma-céntrico.

El análisis gen-céntrico es la identificación de los genes a partir de datos de los metagenomas, directamente de la diversidad de la comunidad microbiana (Boyd et al., 2018). Este enfoque se centra en comprender la biología a un nivel colectivo, avanzando más allá del estudio de un solo organismo para explorar cómo los genes dentro de una comunidad interactúan y afectan las actividades de otros, contribuyendo a funciones conjuntas (Handelsman et al., 2007). La información que nos da el análisis gen-céntrico provee un entendimiento de los procesos microbianos que gobiernarn a los ciclos biogeoquímicos en los ecosistemas. Investigar los ecosistemas terrestres y marinos con análisis metagenómicos es una ardua tarea, pero estos análisis pueden dar un panorama muy completo de la función microbiana en la ecología (Tas et al., 2021).

Usualmente, el primer paso en el estudio de estas comunidades implica la identificación de los organismos presentes y su abundancia, proceso que se realiza mediante anotación taxonómica. La anotación taxonómica es un proceso computacional que permite inferir los clados taxonómicos presentes en una comunidad microbiana y determinar su abundancia relativa. Para obtener una visión completa del material genómico de una muestra ambiental, se necesitaría recuperar los genomas completos de cada individuo de la muestra. Claramente esto es idealista, pero mediante la utilización de técnicas de ensamblado de novo, las lecturas del metagenoma tipo shotgun pueden ser primeramente ensambladas y posteriormente analizadas. Después del proceso de ensamblado, las secuencias reconstruidas son regiones largas consenso de DNA (contigs) del pool de secuencias cortas y se buscan los marcos de lectura abiertos (ORFs por sus siglas en inglés), esto con la finalidad de asignar funciones a los genes. Los acercamientos basados en ensamblado son particularmente buenos para microbiomas que incluyen una amplia proporción de microorganismos no secuenciados (Segata et al., 2014).

Con los recientes avances en las tecnologías de secuenciación y las herramientas computacionales, ha llegado el otro método de análisis de la metagenómica tipo shotgun; el genoma-céntrico. Este análisis se ha hecho mediante el ensamble de genomas a partir de metagenomas (MAGs por sus siglas en inglés) (Yang et al., 2021). Los contigs que derivan de los ensambles están altamente fragmentados y pueden ser redundantes, lo cual hace que no puedan ser agrupados en genomas. Sin embargo por algoritmos de binning que usan una variedad de propiedades genómicas como la composición del DNA, el contenido de GC y la frecuencia de tetranucleótidos en combinación con la cobertura secuenciada y la abundancia, se pueden agrupar los contigs en MAGs (Tas et al., 2021). Un MAG se refiere a un grupo de contigs que representan un genoma microbiano. Los MAGs que tienen un alto grado de completitud y bajos niveles de contaminación en base a la cantidad de genes marcadores de copia única que tienen, son usados para anotación funcional y taxonómica (Yang et al., 2021). La incorporación de los MAGs al árbol de la vida ha incrementado el número de fila microbianos conocidos y también ha revelado nuevas perspectivas en la diversidad metabólica microbiana y la diferenciación por nichos. (Yang et al., 2021).

Es mediante estos enfoques que la metagenómica comparativa permite a los investigadores identificar la abundancia diferencial de características (vías metabólicas, subsistemas o roles funcionales) entre dos o más condiciones siguiendo un procedimiento estadístico con alguna etapa de normalización. Esto es posible para proyectos completos de metagenomas shotgun, donde la información sobre la codificación de genes proteicos está disponible.

### 2.2 El ciclo del carbono en los océanos

El carbono es un bloque esencial de construcción para moléculas orgánicas y es uno de los elementos más abundantes en el ambiente. La abundancia de carbono y la diversidad única en la que se presenta en compuestos orgánicos, permiten que este elemento sirva como un elemento común de toda la vida conocida (Frieden 1972).

"Considerado de forma global, el carbono recircula por todos los reservorios principales de la Tierra como son la atmósfera, los suelos, el mar, los ambientes de agua dulce, las rocas, la biomasa y los sedimentos y lo hace en forma de CO<sub>2</sub>. Adicionalmente, todos los ciclos de nutrientes se enlazan al ciclo del carbono. Las principales áreas de interés del ciclo del carbono son la magnitud de los depósitos de carbono en la Tierra, la velocidad del reciclado del carbono dentro de los reservorios y entre reservorios distintos, y cómo se acopla el ciclo del carbono a los ciclos de otros elementos nutritivos" (Madigan et al., 2010).

"La manera más rápida de transferencia de carbono a la atmósfera es mediante el CO<sub>2</sub>: este se retira de la atmósfera principalmente por la fotosíntesis de las plantas terrestres y de organismos marinos y vuelve a la atmósfera por la respiración de los animales y de microorganismos quimioorganótrofos. El aporte más importante de CO<sub>2</sub> a la atmósfera procede de la descomposición microbiana de la materia orgánica muerta" (Madigan et al., 2010).

En la Tierra, los nuevos compuestos orgánicos sintetizados biológicamente proceden solo de la fijación de CO<sub>2</sub> por los organismos fotótrofos y quimilitótrofos. La fotosíntesis es uno de los procesos biológicos más importantes de la tierra para obtener energía de la luz, requiere de ATP y consiste en convertir el CO<sub>2</sub> en agua y materia orgánica (Kappelman et.,2019). "La mayor parte de los compuestos orgánicos derivan de la fotosíntesis; por tanto, los organismos fotótrofos se encuentran en la base del ciclo del carbono. Sin embargo estos organismos abundan en la naturaleza solo en los hábitats donde hay luz disponible. De ahí que en las profundidades marinas, en el subsuelo profundo y en otros hábitats permanentemente oscuros no haya fotótrofos autóctonos. Hay dos grupos de organismos fotótrofos que predominan en los ambientes terrestres, mientras que los microorganismos fotótrofos predominan en ambientes acuáticos" (Madigan et al., 2010). La mitad de la producción neta de esta materia orgánica generada a partir de la fotosíntesis es llevada a cabo principalmente por el fitoplancton marino (Kappelman et., 2019).

La materia orgánica disuelta en los océanos es uno de los más grandes reservorios de carbono en la Tierra, comparable en tamaño al reservorio de  $CO_2$  atmosférico. Un gran número de componentes que están presentes en la materia orgánica disuelta (DOM por sus siglas en inglés) juegan un papel importante en los mayores elementos de los ciclos, contribuyendo al almacenaje de  $CO_2$  atmosférico en el océano, dando soporte a los ecosistemas marinos y facilitando las interacciones entre organismos. El flujo de carbono a través del cúmulo de DOM marina es mediado principalmente por la actividad microbiana. (Moran et. al 2016). El aporte más importante de  $CO_2$  a la atmósfera procede de la descomposición microbiana de la materia orgánica muerta. Por consiguiente, casi todo el carbono de los compuestos orgánicos termina convirtiéndose de nuevo en  $CO_2$  con lo que se completa el ciclo (Madigan et al., 2010).

El carbono orgánico puede entrar al océano en un ciclo rápido o en un ciclo lento. El ciclo rápido consiste principalmente en la fotosíntesis y la subsecuente DOM y la materia orgánica particulada para sedimentación a través de la bomba biológica, mientras que el ciclo lento es mediado por la bomba microbiana de carbono, prolongando la residencia del carbono. El loop microbiano es una ruta en la que la red acuática de alimento en donde bacterias y arqueas toman la DOM y la materia orgánica particulada (POM por sus siglas en inglés), que es consumido por protistas y que estos a su vez son consumidos por metazoarios. La deriva viral provoca la lisis de los microorganismos lo que devuelve el carbono particulado y disuelto al océano (Jiao et al. 2010; Figura 2).



Figura 2. Interacciones entre los ciclos rápidos y lentos del carbono orgánico en el océano y sedimentos (Jiao et. al. 2010).

La biomasa marina representa una fuente rica de nutrimentos para los microorganismos. Esta es la principal razón por la que comunidades microbianas juegan un papel clave en el ciclo del carbono mediante la transformación de esta amplia variedad de moléculas orgánicas (Andrade et al., 2017).

Un conductor crítico del ciclo del carbono en el océano es el flujo hacia el fondo de las partículas orgánicas que se hunden, cuya función es bajar la concentración de dióxido de carbono atmosférico. Este flujo hacia abajo se reduce en más del 70% en la zona mesopelágica (100 a 1000 metros de profundidad), pero esta pérdida no puede contarse de manera completa con las medidas actuales. Por décadas se ha supuesto que la pérdida desconocida se puede explicar mediante la fragmentación de grandes agregados en pequeñas partículas, a pesar de que faltan datos sobre esta hipótesis (Briggs et al., 2020).

La eficiencia de transferencia de materia orgánica a través de la zona mesopelágica está controlada por la temperatura y el oxígeno, con una mayor eficiencia en regiones con latitudes altas y zonas mínimas de oxígeno. En contraste, la eficiencia de transferencia en el océano (debajo de los 1000 m) está controlada por la velocidad del hundimiento de partícula con la mayor eficiencia en el océano profundo debajo de los giros subtropicales (De Vries and Weber 2017).

### 2.3 Tipos de sedimento marino

En las últimas dos décadas, los científicos han explorado la naturaleza y la extensión de la vida del subsuelo a través del taladrado de varios sitios oceanográficos. El número total de células microbianas en el sedimento es aproximadamente del 0.18 al 3.6% del total de la biomasa celular. La abundancia de las células normalmente decrece con la profundidad y el incremento de la edad del sedimento. El sedimento del subsuelo incluye ecosistemas aeróbicos y anaeróbicos, que persisten con bajos flujos de energía biodisponible a través de los tiempos geológicos. La riqueza microbiana del sedimento es comparable a la riqueza del agua y del suelo (Hoshino et al., 2020).

Los sedimentos marinos pueden ser divididos en tres zonas de arriba hacia abajo; óxica, anóxica y reducida. La respiración aerobia y la nitrificación toman lugar en la zona óxica. La reducción de sulfato, toma lugar en la zona anóxica. La reducción disimilitoria de sulfato toma lugar en la parte de arriba de la zona reducida y después de eso se lleva a cabo la reducción del dióxido de carbono por medio de la metanogénesis (Raggi et al., 2020; Figura 3).



Figura 3. Tipos de sedimento marino en base a la profundidad y cercanía a la costa (Orsi 2018).

Los sedimentos pueden ser clasificados por su disponibilidad de oxígeno en óxicos y anóxicos. El microbioma del océano ha sido estimado en consistir en más de cien mil diferentes taxas bacterianos, arqueanos y eucarióticos con estrategias ecológicas y metabólicas diversas para la producción y el consumo de carbono fijado (Orsi 2018) . Las concentraciones celulares usualmente son órdenes de magnitud mayores en los sedimentos anóxicos ricos en materia orgánica de los márgenes continentales que en los sedimentos óxicos pobres en materia orgánica del océano abierto. Los factores que pueden limitar la biósfera sedimentaria no están restringidos a la falta de nutrientes o la escasez de sustratos. Los factores limitantes pueden incluir la temperatura, la presión, el pH, la salinidad, la disponibilidad de agua, la porosidad del sedimento y su permeabilidad (Hoshino et al., 2020). El oxígeno de la capa superior de agua se difunde en la primera capa del sedimento (la zona óxica), donde es reducido por bacterias para formar agua. Los sedimentos marinos típicamente se convierten en anóxicos porque las bacterias entre estos consumen oxígeno mientras se difunde hacia abajo en la capa suprayacente del agua. Esto produce una serie de capas sedimentarias en las que diferentes tipos de moléculas como hierro aceptan electrones del oxígeno que es consumido para obtener productos reducidos (Orsi 2018). En la siguiente capa el sulfato se convierte en sulfuro de hidrógeno que se acumula y envenena el ambiente para los microorganismos consumidores de oxígeno. Este sulfuro de hidrógeno se puede convertir en sulfato y otras formas oxidadas del azufre por bacterias que utilizan ese material como fuente de energía. (Nealson 2010).

Como ya se mencionó, los sedimentos de subsuelo anóxicos tienen alto grado de entrega de materia orgánica, usualmente cerca de la costa, donde la estimulación del fitoplancton por la surgencia de agua rica en nutrientes es común. Por contraste el sedimento del subsuelo óxico tiene comunidades aerobias de aguas subyacentes con bajas tasas de sedimentación de materia orgánica y que tienden a estar lejos de la costa, donde la disponibilidad de nutrientes a fitoplancton es limitada. En algunos casos el O<sub>2</sub> se predice que penetra más de 100 metros en el suelo marino extendiéndose en la corteza oceánica subvacente. Así de manera similar a la distribución de sulfato en sedimentos anóxicos, la profundidad de penetración del oxígeno en el lodo sedimentado óxico abisal es controlada por la tasa de sedimentación, que correlaciona con la distancia de la costa. En las regiones con altas tasas de sedimentación, el metabolismo de los microorganismos bénticos agota la disponibilidad del oxígeno rápidamente y estos sedimentos se convierten en anóxicos. Sin embargo, en el plano abisal donde el océano suprayacente es extremamente oligotrófico, el oxígeno penetra muy profundamente, a veces hasta la corteza subyacente del océano. La abundancia del O<sub>2</sub> en este lodo abisal está directamente conectada con un grado de sedimentación muy bajo y la cantidad de materia orgánica que se deriva de las regiones con pocos nutrimentos. (Orsi 2018).

Estudios de mediciones de oxígeno disuelto en el sedimento en el Giro del Sur del Pacífico (SPG por sus siglas en inglés) encontraron una baja tasa de respiración y un sedimento muy delgado, lo cual hace que este sea óxico en toda la columna del sedimento en casi toda la región. Consecuentemente, la comunidad sedimentaria del SPG es predominantemente aeróbica, a diferencia de comunidades previamente exploradas. Las de respiración y las concentraciones de células en el sedimento del subsuelo a través de la mitad del mundo pueden ser iguales a los del SPG (D'Hondt et al., 2009). La implicación de este nuevo descubrimiento es que entre el 9 y el 37% de los sedimentos a nivel global están oxigenados y por lo tanto soportan comunidades microbianas aeróbicas. (Orsi 2018).

Los sedimentos anóxicos por su parte, típicamente tienen una zona de transición de sulfatometano, por debajo donde el sulfato es generalmente indetectable y se acumulan microorganismos metano, metanogénicos y fermentativos. Los sedimentos anóxicos ocupan un volumen del suelo marino más grande que los óxicos. Como contraste los sedimentos anóxicos de aguas subyacentes con alta productividad, tienen concentraciones mucho más altas de carbón (Orsi 2018). Las reacciones redox están ampliamente separadas por capas de sedimentos marinos que están acopladas entre ellas. Esto sugiere que son los microorganismos los que medían el flujo de electrones entre capas. (Nealson 2010).

La diversidad microbiana de sedimentos óxicos y anóxicos difiere de manera marcada. Los sedimentos anóxicos tienen células con órdenes de magnitud mucho mayores que los sedimentos óxicos y están enriquecidas con grupos estrictamente anaeróbicos, como lo son las bacterias sulfato-reductoras y también las arqueas metanogénicas y metanotróficas. La producción primaria en el suelo del océano óxico está soportada por arqueas amoniooxidantes, mientras que en los sedimentos anóxicos, grupos nuevos y no cultivables tienen el potencial de producir hidrógeno y metano que potencían la fijación del carbono (Orsi 2018). El análisis estadístico revela correlaciones significativas entre la taxonomía composicional, el carbono sedimentario orgánico y la presencia o ausencia de oxígeno disuelto. Estudios de secuencias del 16S ribosomal han demostrado que diversas taxa de arqueas y bacteria son ubicuas en sedimentos anóxicos orgánicos, las comunidades microbianas están estratificadas por profundidad de sedimento y las propiedades geoquímicas y sedimentológicas influyen en la composición de la comundidad. En el caso de las arqueas, comunidades anaeróbicas Crenarchaeota, que incluye Bathyarchaeia prevalecen. En contraste a las comunidades anaeróbicas, las comunidades aqueanas aeróbicas son dominadas por Thaumarchaeota (Orsi 2018, Hoshino et al., 2020). Miembros de Hadesachaeaota forman parte de más del 70% de las comunidades arqueanas y su abundancia relativa incrementa según el incremento en la profundidad del sedimento.

La composición de la comunidad bacteriana pone a miembros de Proteobacteria, incuyendo Alphaproteobacteria y Betaproteobacteriales (Gammaproteobacteria), como predominantes junto con miembros de Firmicutes en el sedimento óxico del subsuelo. En contraste, miembros de Atribacteria, Chloroflexi y Planctomycetes son prevalecientes en el sedimento anóxico (Orsi 2018, Hoshino et al., 2020).

Los sedimentos de la placa continental han sido objeto de muchos estudios de la vida del subsuelo y son generalmente anóxicos inmediatamente abajo del suelo del océano. Nuestra comprensión de la biogeoquímica de los sedimentos anóxicos se ha expandido dramáticamente en las décadas recientes, en parte debido a que son de acceso mucho más fácil que los sedimentos que residen en las aguas relativamente poco profundas (Orsi 2018).

A pesar de que la disponibilidad de aceptores de electrones y donadores de electrones es importante para la supervivencia de los microorganismos en el suelo marino, hay ciertos sedimentos donde las comunidades depositadas originalmente, también conocidas como paleomas, permanecen sin cambios. Estudios recientes de las relaciones composicionales entre los sedimentos poco profundos y el agua de mar han demostrado que los sedimentos del subsuelo profundo están poblados por descendientes directos de las comunidades del subsuelo que se vuelven predominantes a través de supervivencia preferencial, ya que las comunidades son enterradas por miles a cientos de años (Hoshino et al., 2020).Los paleomas surgen de la supervivencia de comunidades completas después de estar enterradas en una particular subsuperficie estratificada, por ejemplo, en las lutitas negras. Mientras los microorganismos se entierran más profundo en el suelo marino, el tamaño de la población decrece exponencialmente (Orsi 2018).

### 2.4 Metabolismo en los sedimentos

Las tasas de sedimentación se relacionan de manera positiva con las tasas de metabolismo microbiano en la superficie. El flujo resultante de materia orgánica que se hunde y su entrega al suelo del océano se refiere a la bomba biológica y la eficiencia de esa bomba es clave para la producción béntica.

La eficiencia de la bomba biológica describe cuánta de esa materia orgánica últimamente alcanza el suelo marino y es controlada por la disponibilidad de nutrimentos al fitoplancton y la subsecuente remineralización de lo que resulte de la producción primaria fotosintética (Orsi 2018).

Los microorganismos que utilizan los nutrimentos orgánicos pueden tener una ventaja significativa sobre aquellos que no (Moran et al., 2016). Se cree que el recambio de carbono está impulsado mediante la secreción de enzimas de organismos microbianos que pueden descomponer macromoléculas en monómeros constitutivos que pueden ser transportados a las células (Orsi et. al., 2018). En general la remineralización del carbono orgánico disuelto y particulado (DOC y POC por sus siglas en inglés) está en su mayoría controlada por procariotes usando enzimas extracelulares para hidrolizar POC y DOC de alto peso molecular en moléculas de bajo peso molecular antes de la incorporación a las células. Así las actividades enzimáticas extracelulares (EEAs por sus siglas en inglés) son el paso limitante en el ciclo del carbono marino (Zhao Z. et al., 2020). Estas comunidades producen enzimas extracelulares con la correcta especificidad estructural para hidrolizar sustratos para que tengan tamaños suficientemente pequeños para ingresar en la célula. La hidrólisis extracelular enzimática así inicia el ciclo del carbono heterotrófico (Arnosti 2011; Figura 4).



Figura 4. Metabolismo extracelular en los sedimentos marinos (Arnosti 2014).

Las enzimas extracelulares se encuentran principalmente disueltas en el ambiente más que asociadas a la célula. En estudios cultivo dependientes se ha visto que las enzimas libres pueden ser originadas por la actividad de liberación de las células (Antranikian et al., 1987, Albertson et al., 1900, Alderkamp et al., 2007). Vetter et al., (1998) exploró las circunstancias bajo las que las enzimas libres pueden ser una estrategia beneficiosa para los microorganismos en ambientes ricos en partículas como son los agregados de los sedimentos y la nieve marina, demostrando que las enzimas libres pueden proveer suficiente sustrato para crecimiento en células microbianas. Un estudio de las zonas meso y batipelágicas del Océano Atlántico encontraron que del 50 al 100% de la actividad enzimática total se debe a enzimas libres (Baltar et al., 2010). Se supone que las enzimas de origen microbiano se pegan a las partículas o coloides del océano profundo (Arnosti 2011). La abundancia, porcentaje y diversidad de genes que codifican para procesos secretorios como lo son las enzimas disueltas, consistentemente aumenta en aguas del epipelágico y el batipelágico, indicando que el corte de materia orgánica, y por ello el metabolismo es mediado principalmente por procariotes asociados a partícula que liberan sus enzimas extracelulares en el reino del batipelágico de difusión límitada de partículas (Zhao Z. et al., 2020).

En sedimentos, sin embargo, la heterogenicidad de organismos y sustratos es mucho más grande (Burdige 2006), así que la variabilidad natural en actividades microbianas correspondientes es mayor. La gran mayoría de las medidas reportadas en sedimentos son, de hecho, realizadas en sedimentos mezclados con agua de mar (Arnosti 2011). A pesar de estas limitaciones, las actividades enzimáticas han sido medidas utilizando sustitutos en sedimentos de la costa, así como sedimentos profundos y polares (Hoppe et al., 2022) Típicamente los grados de hidróliss son más altos en la superficie o cerca de la superficie del sedimento y estos decrecen con la profundidad (Meyer-Reil 1986, Poremba & Hoppe 1995), debido posiblemente a la afluencia de materia orgánica fresca que estimula las actividades de las superficies del océano (Boeticus & Lochte 1994). Las tasas de hidrólisis en los sedimentos son entre dos y tres órdenes de magnitud superiores a las del agua (Hoppe et al., 2002), lo que se explica por un factor de escalamiento que refleja las diferencias en los conteos de células que comúnmente se observan entre el agua de mar y los sedimentos. La observación indica que el espectro de actividades enzimáticas en sedimentos (incluyendo los anóxicos) es en muchos casos mucho más amplia que en el agua (ejemplo, sustratos no hidrolizados en agua de mar, se hidrolizan en el sedimento suprayacente (Arnosti 2000, 2008)).

Para que una enzima pueda ser secretada desde su sitio de síntesis que son los ribomosomas al exterior de la célula debe pasar por un sistema de secreción en el cual estas enzimas son sintetizadas con una secuencia de 15 a 20 aminoácidos al principio (en el extremo N-terminal) llamada péptido señal (Madigan et al., 2010). Los péptidos señal se encuentran en muchas cadenas de polipéptidos nacientes en prácticamente todos los organismos y dirigen proteínas a las máquinas de exportación incrustadas en membranas en Bacteria, Archaea y Eukarya (Armenteros et al., 2019). Los péptidos señal dirigen las correspondientes enzimas a la ruta de secreción para la translocación transmembrananal. Así podemos distintinguir entre las enzimas secretadas o no secretadas (ejemplo entre las que son libres de célula y las que son asociadas a la célula) (Zhao Z. et al., 2020).

### 2.5 Carbohidratos

Los carbohidratos son moléculas que contienen átomos de carbono, hidrógeno y oxígeno. En el siglo XIX, las sustancias basadas en azúcar se referían como carbohidratos, o "hidratos de carbono", basados en la fórmula general Cx(H2O)n, que también poseen un grupo carbonilo, ya sea un aldehído o una cetona (Seeberger et al., 2017). Los carbohidratos son parte de las cuatro clases mayores de biopolímeros encontrados en las células y estos tienen estructuras muy versátiles (Zhang et al., 2018). Están ampliamente distribuidos en la naturaleza y median una multitud de funciones biológicas, como son las reservas de carbono, moléculas estructurales y como mediadores de reconocimiento intra e intercelular entre organismos y en un organismo (Lombard et al., 2013).

Como ejemplo de esto, los carbohidratos forman parte de los componentes de las paredes celulares del fitoplancton o pueden ser moléculas secretadas como sustancias extracelulares exopoliméricas (Kappelman et., 2019).

Los carbohidratos más sencillos son los denominados monosacáridos, los cuales sirven como moléculas oxidables o componentes de los seres vivos (Styer et al., 2012). Los monosacáridos son los más simples de estos compuestos carbonílicos poli-hidroxilados (sacárido se deriva de la palabra griega para azúcar o dulzura) (Seeberger et al., 2017). Los monosacáridos más pequeños están compuestos de tres moléculas de carbono. La unión de varios monosacáridos mediante los enlaces glicosídicos forma los polisacáridos. También los azúcares pueden unirse de forma covalente a las proteínas y formar polisacáridos (Stryer et al., 2012). Los monosacáridos se unen entre sí para dar lugar a oligosacáridos o polisacáridos. Típicamente, el término "oligosacárido" se refiere a cualquier glicano que contenga menos de 20 residuos de monosacáridos conectados por enlaces glicosídicos. El término "polisacárido" se utiliza típicamente para denotar cualquier polímero lineal o ramificado que consista en residuos de monosacáridos, un ejemplo de esto es la celulosa (Seeberger et al., 2017).

A pesar de tener una composición química similar, los carbohidratos pueden formar un enorme número de combinaciones a través de una variedad estereoquímica de grupos hidroxilo. Lo anterior puede llevar, a través de muchas posibilidades, al ensamblaje de monosacáridos uno tras otro y, mediante la riqueza de sustituyentes no carbohidratos, a decorar los resultantes oligo y polisacáridos (Lombard et al., 2013). El término "glicoconjugado" se utiliza a menudo para describir una macromolécula que contiene monosacáridos covalentemente ligados a proteínas o lípidos. El prefijo "glico" y los sufijos "sacárido" y "glicano" indican la presencia de componentes de carbohidratos (por ejemplo, glicoproteínas, glicolípidos y proteoglicanos). Al igual que se observa con las proteínas en la naturaleza, la diversidad estructural adicional puede ser impartida a los glicanos modificando sus grupos hidroxilo con fosfato, sulfato o ésteres de acetilo y/o sus grupos amino con grupos acetilo o sulfato. Un carbohidrato puede denominarse "complejo" si contiene más de un tipo de unidad de construcción de monosacárido. El polímero basado en glucosa, la celulosa, es un ejemplo de un carbohidrato "simple", mientras que un polisacárido galactomanano, compuesto de galactosa y manosa, es un ejemplo de un carbohidrato complejo (Seeberger et al., 2017).

### 2.6 CAZymes

En virtud de la extrema variedad de estructuras monosacáridas, la variedad de uniones inter azúcares y el hecho de que virturalmente todo tipo de moléculas pueden ser glicosiladas, la gran cantidad de enzimas actuando en estos glicoconjugados, oligo- y polisacáridos probablemente constituye uno de los más diversos sustratos estructurales de la Tierra. Colectivamente designadas como Enzimas Activas a Carbohidratos (CAZymes), estas enzimas construyen y rompen carbohidratos completos y glicoconjugados para distintos roles biológicos (Cantarel et al., 2008).

Las CAZymes se refieren a un vasto panel de actividades catalíticas que se involucran en la transformación de azúcares o derivados de azúcares para asegurar el ensamblado de monosacáridos en la forma de oligo o polisacáridos así como la conjugación de ácidos nucléicos, proteínas, lípidos, polifenoles y otros compuestos naturales.

Las CAZymes son las responsables de la síntesis (a través de las glicosil transferasas [GTs]), la degradación (glicosil hidrolasas [GHs], polisacárido liasas [PLs], carbohidrato esterasas [CEs] y enzimas de Actividades Auxiliares [AAs]) y el reconocimiento (módulos de unión a carbohidratos [CBMs]) de todos los carbohidratos de la Tierra (Huang et al., 2018; Tabla 1).

Familia	Función
Glicosil hidrolasas (GHs)	Incluyen glicosidasas y transglicosidasas. Estas enzimas constituyen familias que son responsables de la hidrólisis y translglicosilación de enlaces glicosídicos.
Glicosiltransferasas (GTs)	Estas son enzimas responsables de la biosíntesis de enlaces glicosídicos de donadores de azúcares fosfo- activados.
Polisacárido liasas (PLs)	Rompen uniones glicosídicas de polisacáridos que contienen ácido urónico por un mecanismo de beta eliminación.
Carbohidrato esterasas (CEs)	Estas remueven modificaciones basadas en esteres presentes en mono-, oligo y polisacáridos y gracias a esto facilitan la acción de GHs de polisacáridos complejos.
Módulos de Unión a Carbohidratos (CBMs).	Estos son fragmentos autónomos de plegado y funcionalidad de proteínas que no tienen actividad enzimática per se, pero que se sabe que potencian la actividad de muchas actividades enzimáticas descritas mediante la dirección y promoción de la interacción prolongada con el sustrato.
Actividades Auxiliares (AAs)	Enzimas redox que actúan en conjunto con las CAZymes

Tabla 1. Clasificaciór	n CAZymes (Cantarel et al.,2008)	
------------------------	----------------------------------	--

Las CAZymes no son solo las más importantes enzimas en las industrias de bioenergía y agricultura, sino que también son muy importantes en la salud humana, ya que la microbiota humana codifica centenares de genes CAZyme en sus genomas para la degradación de varios carbohidratos de la dieta y de los huéspedes. La razón es que un gran número de CAZymes son necesarias para construir (en plantas) y degradar (en microbios) los carbohidratos complejos de la pared celular de las plantas. En particular, los microorganismos que habitan en los intestinos de los animales codifican el mayor porcentaje de CAZymes, enzimas responsables de la degradación de diversos carbohidratos provenientes tanto de la dieta como del huésped. Modificar los carbohidratos de la dieta puede tener un impacto significativo en la estructura de la microbiota intestinal humana, lo que a su vez puede influir en la salud (Huang et al., 2018).

Las CAZymes pueden funcionar solas o como clusters de genes CAZyme (CGCs por sus siglas en inglés). Definimos como un término más general a los genes de clusters CAZyme (CGCs) por ser clusters que están físicamente. Además de las CAZymes, los CGC también deben contener al menos uno de los otros tres tipos de genes característicos: transportadores (TC), proteínas de transducción de señales (STP) y factores de transcripción (TF). Todas estas enzimas trabajan juntas para ensamblarse como una maquinaria en una cadena de procesamiento y transporte de sustratos, lo que permite una degradación eficiente de los carbohidratos complejos (Zheng et al., 2023).

### 2.7 CAZy

La base de datos CAZy es un recurso de conocimiento especializado en enzimas que construyen y rompen carbohidratos complejos y glicoconjugados. Las familias que existen son creadas basadas en proteínas experimentalmente caracterizadas y son pobladas con secuencias de las bases de datos públicas con similitud significativa. La información bioquímica de proteínas es continuamente curada basada en la literatura e información estructural. La clasificación refleja las características estructurales de estas enzimas mejor que su sola especificidad en sustrato y ayuda a revelar relaciones evolutivas entre enzimas y así proporcionar un mejor entendimiento de sus propiedades mecanísticas (Cantarel et al., 2009). La base de datos de CAZy comenzó a colectar de manera experimental proteínas caracterizadas CAZyme de la literatura y a clasificarlas en familias de proteínas basadas en similitud de secuencia desde los años 1990s. Posteriormente pobló cada familia incluida con homólogos de las proteínas semilla caracterizadas de las bases de datos del GenBank, UniProt y PDB (Cantarel et al., 2009).

La primera característica definitoria en la clasificación de las CAZyme es que las familias están definidas basadas en similitud de secuencia de aminoácidos y por lo menos un miembro fundador bioquímicamente caracterizado (Lombard et al.2013). De manera adicional es clave para la clasificación de CAZy la disección de los marcos de lectura abiertos para revelar una discreta y a veces compleja, organización modular de las CAZymes, lo que incrementa de manera significativa la precisión del análisis bioinformático (Grondin et al., 2017).

Por esta razón, una segunda carácterística definitoria es que nuestra clasificación se hace módulo por módulo. Las CAZymes son frecuentemente proteínas modulares con un módulo catalítico que albergan un número variable de otros módulos discretos, que pueden ser catalíticos o no. Así una CAZyme modular puede ser asignada a varias familias si su módulo constitutivo pertenece a familias separadas. Su tercera característica importante es que sólo se analizan secuencias de proteínas liberadas diariamente del GenBank (Lombard et al., 2013). Las tres tareas primarias de los curadores de Cazy son: mantener y actualizar la clasificación de familias de esta clase de enzimas, clasificar secuencias recién salidas de GenBank y de Protein Data Bank y capturar y presentar la información funcional de cada familia (Drula et al., 2022).

El uso de las herramientas bioinformáticas, combinado con los alineamientos múltiples de secuencia y los análisis filogenéticos se han vuelto un reflejo de la ingeniería de proteínas para investigar la variabilidad de secuencia, las relaciones ancestrales y para identificar firmas o secuencias consenso relacionadas putativamente de las propiedades deseadas (Drula et al., 2022). Diferentes herramientas para comparar proteínas han sido utilizadas para definir familias de enzimas, particularmente BLAST y HMMER utilizando modelos ocultos de Markov (HMMs) hechos para cada familia lo que contínuamente mejora la calidad de la anotación genómica, metagenómica y metatranscriptómica. Todas las secuencias correspondientes a las enzimas catalíticas o de unión activas de carbohidratos son extirpadas de la secuencia completa de proteínas y agrupadas en una librería de BLAST.

Aproximádamente del 1 al 5% de las proteínas codificadas en un genoma típico corresponden a las CAZymes (Lombard et., al 2013). Esto ha permitido avances para empezar a decodificar las relaciones entre la especificidad del sustrato y la secuencia CAZyme y crear un número creciente de subfamilias en las bases de datos de CAZy. El uso de herramientas computacionales combinado con la información estructural 3D de las secuencias de datos es otra alternativa que fortalece la predicción de modelos de guía y diseño de enzimas (Andre et al., 2014). Menos del 1% de las secuencias referenciadas en las bases de datos de CAZy, CAT y dbCan provienen de organismos no cultivables deseados (Drula et al., 2022).

En cualquier familia, los sitios activos claves de residuos, el mecanismo catalítico y el plegamiento tridimensional general están estrictamente conservados (excepto unas pocas excepciones (Grondin et al.,2017). Así, el poder predictivo de la clasificación de CAZy ha probado ser bastante robusto ya que la información funcional y estructural es añadida y curada regularmente basándose en la literatura disponible (Lombard et al., 2013).

El sistema de clasificación de CAZy permite en algunos casos algún tipo de predicción en la amplia categoría de sustratos de carbohidratos, basados en la asignación de una familia. Esto acarrea el potencial de inferir un perfil glicobiológico de un organismo (o comunidad) basado en su secuencia de DNA (Lombard et al., 2013). La especificidad de sustrato, sin embargo, es menos clara debido a que muchas familias grandes son poliespecíficas y contienen miembros con distintos perfiles de actividad, normalmente debido a la estructura de los carbohidratos complejos (Grondin et al., 2017). A pesar de que la especificidad del

sustrato es variable en las familias, se ha encontrado que el mecanismo catalítico es extremadamente bien conservado y así predecible una vez establecido para un miembro de la familia. En el caso de familias multifuncionales, la tarea de buscar en la literatura para miembros funcionalmente caracterizados puede hacer que estas se dividan en subfamilias para acotar la especificidad e incrementar el poder de predicción funcional (Drula et al., 2022).

La base de datos de CAZy describe dos tipos de componentes comunes en las CAZymes:

- Módulos Catalíticos (Enzimas): Subdivididos en varias clases de enzimas y sus familias que catalizan la degradación, biosíntesis y/o modificación de glicoconjugados, oligosacáridos y polisacáridos (www.cazy.org).
- Módulos Asociados: Familias de módulos que se encuentran adheridas a los módulos catalíticos (www.cazy.org).

La clasificación de CAZy está dividida en tres niveles: clases, familias y subfamilias (Figura 5).

- Nivel 1: Las clases se clasifican en las seis categorías de CAZymes (GHs, GTs, PLs, CEs, CBMs y AAs).
- **Nivel 2:** Las familias típicamente se designan utilizando una forma simple que incluye la clase o categoría del módulo y un número que refleja el orden de creación.
- Nivel 3: Las subfamilias son subgrupos dentro de una familia que comparten un ancestro más reciente y que, por lo general, son más uniformes en su función molecular.



Figura 5: Distintos niveles de la base de datos de CAZy. www.cazy.org

### 2.8 Óxido reductasas

Las oxidasas terminales de cadenas de transferencia de electrones unidas a la membrana de los organismos aeróbicos catalizan la reducción de dioxígeno a agua, acoplando la energía redox a la translocación de protones a través de la mitocondría o la membrana citoplasmática. Muchas de las oxidasas terminales pertenecen a la superfamilia de oxidoreductasas de hemo-cobre (Pereira et al., 2001).

La superfamilia oxidoreductasas de hemo-cobre (HCO) es una gran superfamilia de enzimas respiratorias terminales que están ampliamente distribuidas en los tres dominios de la vida. Las HCOs catalizan la reducción de 4 electrones de dioxígeno a agua en un proceso acoplado a la translocación a través de la membrana. Debido a que los electrones y los protones son necesarios en los lados opuestos de la membrana, estas enzimas promueven la formación de una diferencia de potencial por separación de carga. Así los protones son sustratos químicos de todas estas enzimas y deben estar presentes en la subunidad I (Sousa et al., 2012). Esta familia incluye diversas familias de oxígeno reductasas y óxido nítrico reductasas que son enzimas pivotales en las rutas de respiración aeróbica y denitrificación. En la respiración aerobia y la denitrificación, la reducción de O<sub>2</sub> o NO están usualmente canalizadas del pool de quinol a través del complejo III que reduce reductasas terminales de la superfamilia de HCO. Sin embargo la adaptación de HCOs a usar quinol como donador de electrones en lugar de citocromo c tiene una implicación en la flexibilidad respiratoria y eficiencia de las cadenas respiratorias que la incluyen. La superfamilia HCO es bioquímicamente diversa e incluye muchas oxígeno reductasas muy bien caracterizadas (Murali et al., 2022).

Análisis filogenómicos y datos metagenómicos han identificado por lo menos seis nuevas familias que pertenecen a la superfamilia HCO. Estos datos nos muestran que son de organismos no caracterizados y se encuentran en una gran variedad de ambientes (Murali et al., 2021). Todos los miembros de la superfamilia HCO tienen en común homólogos en cada reductasa que es donde ocurre la reacción catalítica en el sitio activo (Murali et al., 2022).

Se ha establecido una clasificación mediante herramientas bioinformáticas de HCOs y óxido nítrico reductasas (NORs) que permite clasificarlas con un alto grado de precisión (99.8%) y ésta se encuentra disponible en la base de datos de proteínas de <u>www.evocell.org/hco</u> (Sousa et al., 2012).

### 2.9 Estudios in silico en sedimentos marinos

El desarrollo de métodos cultivo-independientes para realizar estudios in silico como lo son las ciencias genómicas permite a los investigadores estudiar las comunidades microbianas directamente de sus muestras ambientales y presenta una plataforma en la que genes que codifican enzimas de interés pueden ser estudiados (Kunath et al.2017).

Especialmente en ecosistemas como los sedimentos marinos donde los grupos microbianos no son todavía cultivables en su mayoría y sólo son conocidos por sus secuencias de DNA, y además cuentan con una gran diversidad, los estudios in silico tienen el poder de mejorar rápidamente el entendimiento de la microbiología ambiental ya que es una técnica para caracterizar la diversidad genética y microbiana así como la predicción de las funciones biológicas potenciales asociadas a las comunidades microbianas (Kerkhof & Goodman 2009).

Debido a los avances en la secuenciación y el abaratamiento de los costos, se han empezado a estudiar a mayor profundidad los sedimentos marinos y sus comunidades.

Como ejemplos de estos estudios se encuentran los perfiles taxonómicos de los sedimentos marinos como el del Consorcio de Investigación del Golfo de México (CIGOM) cuyos objetivos fueron determinar un primer perfil taxonómico de las muestras de sedimento del Suroeste del Golfo de México (Godoy-Lozano et al., 2018; Raggi et al., 2020). El Golfo de México está expuesto a una presencia continua de hidrocarburos lo que influye en la composición de la comunidad microbiana.

De manera análoga están también los estudios de la Universidad de Delaware que examinó la respuesta de las comunidades microbianas a las filtraciones naturales de hidrocarburos en los sedimentos del Golfo de México, a través de un análisis centrado tanto en genes como en genomas de los datos de secuenciación metagenómica (Zhao R. et al, 2020).

Adicionalmente, estudios han caracterizado la diversidad de microorganismos que habitan en los sedimentos de la Cuenca de Guaymas (Golfo de California), y muchos taxones eran distintos de aquellos cuyas fisiologías han sido descritas. Se realizaron estudios metagenómicos de muestras de sedimentos de la Cuenca de Guaymas y se reconstruyeron genomas microbianos a partir de los datos de secuenciación generados (Dombrowski et al., 2017).

De manera global se investigaron por medio de 47 millones de secuencias del gen 16S ribosomal RNA 299 muestras de núcleos de sedimentos distribuidos globalmente de 40 sitios diferentes a profundidades de 0.1 a 678 m bajo el fondo marino en donde se obtuvieron correlaciones significativas entre la composición taxonómica de los sedimentos, la concentración de carbono orgánico del mismo y si son óxicos o anóxicos (Hoshino et al., 2020).

Se ha visto también la diversidad taxonómica de los virus del sedimento. Algunos de los genomas identificados similares a pithovirus y marseillevirus pertenecen a ramas profundas

en el árbol filogenético de los genes. Estudios de los grandes virus de DNA nucleocitoplasmático centrales, han ampliado sustancialmente la diversidad y la profundidad filogenética de los respectivos grupos (Bäckström et al., 2019). También se hicieron análisis de genes 16S ribosomales de RNA y el análisis de lípidos biomarcadores que proporcionaron información sobre las comunidades microbianas que prosperaban dentro de las estructuras porosas de sulfuro de las fuentes hidrotermales profundas, tanto activas como inactivas (Jaeschke et al., 2012).

Se han hecho perfiles metabólicos de los sedimentos como el realizado en el Giro del Pacífico Sur, un ambiente oligotrófico donde se investigó el impacto de la geoquímica local del carbono por medio del potencial de los microorganismos presentes en él (Tully et al., 2016) y por último, el realizado con muestras del sedimento anóxico del Margen del Perú en el cual se hizo una predicción del secretoma mediante estudios metatranscriptómicos (enzimas secretadas por microorganismos) y se encontraron muchos genes que expresan para CAZymes (la mayoría de ellas que apuntan para bacterias), siendo éste el primer análisis que involucra la predicción de péptidos señal en metatranscriptomas de sedimento marino. (Orsi et al., 2018).

Otros trabajos en los que se involucraba el estudio del secretoma son realizados para la búsqueda de peptidasas y CAZymes en el océano y sedimento, dos grupos clave de enzimas que inician la asimilación de materia orgánica, utilizando un enfoque integrado de metagenómica, metatranscriptómica y metaproteómica, es mediado principalmente por procariotas asociados a partículas que liberan sus enzimas extracelulares al ambiente (Zhao Z. et al., 2018).

Y por último está un ejemplo en el cual el estudio in silico se hizo para ver el impacto ecológico de la estación de Davis mediante la secuenciación de metagenomas de sedimentos marinos con diferentes proximidades a la fuente de efluentes de aguas residuales en la estación Davis (Antártida), además de las heces de animales marinos cercanos y así determinar el impacto que tienen las aguas residuales de la estación de investigación sobre las comunidades microbianas ambientales cercanas y el tracto gastrointestinal de las especies de fauna nativa (Leeming et al., 2015).

### 3 Justificación

El reciclado de la materia orgánica en sedimentos es un componente importante de los ciclos biogeoquímicos, debido a que los sedimentos son el principal reservorio para el almacenaje a largo plazo de carbono. Por otra parte, el papel de las CAZymes en la comunidad microbiana de los sedimentos no está bien definido. Suponemos que las CAZymes de los microorganismos presentes podría influir en la función de la comunidad que vive en los sedimentos.

### 4 Hipótesis

• Los sedimentos marinos, dada su profundidad y tipo (óxico o anóxico), determinarán la presencia y abundancia de las CAZymes.

### 5 Objetivos

### General

• Analizar los genes que codifican para CAZymes de los microorganismos que se encuentran en sedimentos marinos.

### Particulares

- Seleccionar metagenomas de buena calidad en el Centro Nacional para la Información Biotecnológica (NCBI por sus siglas en inglés) de sedimento marino.
- Proponer el tipo de metabolismo que predomina (anóxico/óxico) en cada metagenoma analizado.
- Identificar la presencia y abundancia de los genes que codifican para las familias CAZymes que se encuentran en los sedimentos marinos.
- Comparar los "MetaCAZomas" (composición de CAZymes en los metagenomas) entre sí.
- Identificar las principales CAZymes que forman el secretoma del sedimento.
- Reconstruir MAGs y hacer un análisis de su CAZoma secretado.
- Comparar CAZoma de los MAGs de suelo vs CAZoma de los MAG sedimento.

### 6 Metodología

### 6.1 Programas

- Kraken v.2 (Wood et al.,2019)
- LEfSe v.1 (Segata et al. 2011)
- FastQC v0.11.7 (Andrews, 2010)
- Prodigal v.2.6.3 (Hyatt et al., 2010).
- MEGAHIT v1.1.1-2 (Li et al., 2015)
- dbCAN v.2 (Zhang et al.,2018)
- HMMER v-3.2.1(Eddy et al., 2011)
- Trimmomatic v-0.32 (Bolger et al., 2014)
- SignalP v-5.0 (Almagro et al. 2019)
- GTDB-Tk v-2.1.0 (Chaumeil et al., 2020)
- Diamond v0.9.24.125 (Buchfink, Reuter y Drost, 2021)
- FetchMG v-1.2 (Kultima et al., 2012)
- SqueezeMeta v-1.4 (Tamames y Puente-Sanchez et al., 2019)
- Phylophlan v-3 (Asnicar et al., 2020)
- MUSCLE v3.8.31 (Edgar, 2004),
- trimAl v1.4. rev22 (Capella-Gutiérrez, Silla-Martínez y Gabaldón, 2009)
- IQ-TREE v2.0.6 (Nguyen, Schmidt, Von Haeseler y Minh, 2015)
- RaxML v.8 (Stamatakis, 2006)
- iTOL Versión 6.8.1. (2023). (Letunic et al., 2021)

R-v. 4.2.3 (R Core Team 2023).

- Vegan (Oksanen et al.,2020)
- Geosphere (Hijmans et al., 2020)

### 6.2 Selección de datos genómicos; muestras metagenómicas

Una vez identificados los BioProyectos adecuados, se descargaron las secuencias de metagenoma shotgun de la base de datos del NCBI. Se obtuvieron al final datos crudos de 37 muestras metagenómicas de 12 BioProyectos que representan muestras de sedimento marino que incluían metadatos. (Tabla 2). Los datos asociados a estas muestras fueron sus coordenadas de latitud/longitud, metros por debajo del subsuelo marino y metros por debajo del nivel del mar. Esta información adicional proporcionó un contexto importante para comprender la distribución espacial y las características ambientales de los ecosistemas marinos muestreados.

Se tomaron muestras de sedimentos de todo el mundo con un rango de profundidad de 0 a 7942 metros por debajo del nivel del mar (mbsl) y de 0 a 2.23 metros por debajo del lecho sedimentario (mbsf) (Figura 6). Para reducir el sesgo del secuenciamiento, se descartaron las muestras que no se secuenciaron con Illumina (Tabla 2).

### 6.3 Control de calidad y preprocesamiento de las muestras

Se llevaron a cabo procedimientos de control de calidad utilizando herramientas y software ampliamente utilizados, como Trimmomatic (Bolger et al., 2014) y FastQC (Andrews, 2010). Todas las lecturas de las muestras que no superaron los filtros de control de calidad (calidad de lectura  $\geq$  Q20) fueron descartadas.

### 6.4 Análisis gen-céntrico

### 6.4.1 Análisis taxonómico de las lecturas metagenómicas

Para el análisis taxonómico de las lecturas, utilizamos la base de datos específica de Kraken2 basada en espectros k-mer de genomas completos de RefSeq y la base de datos nt del NCBI (descargada el 7/09/2021) (Wood et al., 2019).

#### 6.4.2 Ensamble de las lecturas metagenómicas

Como en cualquier estudio de metagenómica, se realizó un ensamble *de novo* para cada muestra utilizando MEGAHIT v1.1.1-2 con el parámetro 'metasensitive', recomendado para muestras diversas (Li et al., 2015). Se predijeron los marcos de lectura abiertos (ORFs por sus siglas en inglés) de cada muestra utilizando la herramienta Prodigal-v2.6.3 (Hyatt et al., 2010). Se descartaron de este estudio las muestras con menos de un millón de genes.

### 6.4.3 Análisis funcional del metagenoma

Para hacer la anotación de las CAZymes, utilizamos HMMER-3.2.1 (Eddy et al., 2011) (hmmscan: valor E < 1e-15, cobertura > 0.35) contra la base de datos HMM V9 de dbCAN2 (Zhang et al.,2018).

Teniendo en cuenta la suposición de que el reciclaje de carbono en los sedimentos marinos es llevado a cabo por los microorganismos que lo habitan, mismos que usan enzimas secretadas (conocidas como "secretoma") para almacenar carbono a largo plazo (Orsi et al., 2018), decidimos buscar CAZymes secretadas. Anotamos las secuencias en las seis clases de la base de datos CAZy, que están implicadas en la creación, descomposición e identificación de carbohidratos. Estas clases son Glicosil Transferasas (GTs), Glicosil Hidrolasas (GHs), Esterasas de Carbohidratos (CEs), Módulos de Unión a Carbohidratos (CBMs), Liasas de Polisacáridos (PLs) y Actividades Auxiliares (AAs). Las especificidades de los sustratos de las CAZymes se infirieron por inspección manual de CAZy (Lombard et al., 2013; Lombard et al., 2014). Las CAZymes secretadas se anotaron utilizando SignalP V-5.0 (Almagro et al., 2019).

Se analizaron las oxidorreductasas de hemo-cobre (HCO) y las oxidorreductasas de óxido nítrico (NOR) utilizando Diamond (parámetros "ultra-sensitive") frente a la base de datos HCO (Sousa et al. 2011).

La normalización de los recuentos de genes entre las muestras de CAZymes y los genes de oxidorreductasa de hemo-cobre se llevó a cabo utilizando la ecuación: (Número de genes anotados / Número total de genes en la muestra) X  $10^6$ .

### 6.4.4 Análisis estadísticos de los metagenomas

Los análisis estadísticos se realizaron utilizando R-v. 4.2.3 (R Core Team 2023). Utilizando el índice de disimilitud de Bray-Curtis para calcular matrices de distancia relativas a los grupos de abundancia taxonómica a nivel de clase y a la composición de CAZymes, se realizó un Análisis de Coordenadas Principales (PCoA). Para buscar correlaciones entre los metadatos de las muestras y la diversidad taxonómica, se calcularon pruebas de Mantel de la matriz de abundancia frente a metros por debajo del nivel del mar (mbsl), metadatos de latitud y longitud de las muestras, y metros por debajo del lecho marino (mbsf). El análisis del tamaño del efecto de la prueba de discriminación lineal (LEfSe) de las matrices taxonómicas de arqueas y bacterias se realizó en la herramienta Galaxy de Hutlab (Segata et al., 2011) (corte LEfSe: valor alfa de Kruskal-Wallis = 0.05, valor alfa de prueba de Wilcoxon = 0.05, puntuación LDA > 3.0). Se realizó un análisis similar en la matriz de CAZymes extracelulares con una puntuación LDA > 3.5.

### 6.5 Análisis genoma-céntrico

### 6.5.1 Reconstrucción de genomas a partir de metagenomas (MAG)

Los MAG fueron anotados, reconstruidos y refinados utilizando el pipeline Squeeze-Meta v.1.4.0 (Tamames y Puente-Sanchez et al., 2019) (parámetros: modo=secuencial, ensamble=extensión, double pass, low memory). Se eliminaron las agrupaciones genómicas con una completitud baja (<75%) y una alta contaminación (>10%). Los MAG se refinaron con el programa remove\_duplicate\_markers.pl del pipeline SqueezeMeta.

### 6.5.2 Anotación taxonómica y análisis filogenético MAG

La clasificación taxonómica de estas agrupaciones se realizó mediante GTDB-Tk v2.1.0 (parámetros: classify\_wf) frente a la base de datos GTDB v-207 (Chaumeil et al., 2020). Para el análisis filogenético de los MAGs reconstruidos, así como los MAGs del suelo, utilizamos las secuencias de aminoácidos y el pipeline Phylophlan 3.0 para calcular árboles (Asnicar et al., 2020). Usamos la base de datos Phylophlan (Segata, Börnigen, Morgan y Huttenhower, 2013) que incluye 400 genes marcadores universales y Diamond v0.9.24.125 (Buchfink, Reuter y Drost, 2021) para mapear la base de datos contra nuestros proteomas. Los alineamientos de múltiples secuencias (MSA) se realizaron con MUSCLE v3.8.31 (Edgar, 2004), y el software trimAl v1.4. rev22 (Capella-Gutiérrez, Silla-Martínez y Gabaldón, 2009) para el recorte de regiones con brechas. Finalmente, para el cálculo y

refinamiento de los árboles, utilizamos la estimación de máxima verosimilitud con el software IQ-TREE v2.0.6 (Nguyen, Schmidt, Von Haeseler y Minh, 2015) y RaxML v.8 (Stamatakis, 2006), respectivamente con 100 bootstraps.

### 6.5.3 Anotación funcional MAG

Las familias de CAZymes y los clústeres de genes de CAZymes (CGC) se anotaron utilizando dbCAN2 (Zhang et al., 2018); (cortes hmmscan: valor E < 1e-15, cobertura > 0.35, cortes DIAMOND: valor E < 1e-102, Hotpep (Frecuencia > 2.6, Hits > 6), CGCFinder (Distancia <= 2, genes característicos = CAZyme+TC). Los genes marcadores (MG) se anotaron con FetchMG v-1.2 (Kultima et al., 2012). La normalización de los recuentos de CAZymes entre los MAG se llevó a cabo utilizando la ecuación: [(Número del módulo de CAZyme en el MAG / Número del módulo de CAZyme en la muestra de metagenoma) / Mediana (MGs en la muestra de metagenoma)] X 10<sup>6</sup>. La lista de MG se puede descargar desde el sitio web de mOTU <u>https://motu-tool.org/fetchMG.html</u> (López-Sanchez et al., 2024).

Se seleccionaron aleatoriamente MAG de suelo asignados taxonómicamente a las clases Alphaproteobacteria, Gammaproteobacteria y Bacteroidia (Nayfach et al., 2021) con los mismos criterios que nuestros MAG (Completitud > 75% y Contaminación < 10%). Los módulos de CAZymes y los clústeres de genes de CAZymes (CGC) se anotaron utilizando dbCAN2 (Zhang et al., 2018); (cortes hmmscan: valor E < 1e-15, cobertura > 0.35, cortes DIAMOND: valor E < 1e-102, Hotpep (Frecuencia > 2.6, Hits > 6).

### 6.6 Graficación

El mapa con los puntos situados en las coordenadas geográficas (latitud y longitud) donde fueron extraídas las muestras de sedimento fue realizado por medio de la paquetería de R marmap (Pante & Simon-Bouhet B 2013) utilizando datos topográficos de batimetría y topografía importados de las bases de datos de la Administración Nacional Oceánica y Atmosférica (NOAA por sus siglas en inglés) (<u>https://www.noaa.gov</u>) y la Carta Batimétrica General de los Océanos (GEBCO por sus siglas en inglés) (<u>https://www.gebco.net</u>).

Las matrices integradas obtenidas para el análisis taxonómico de las 37 muestras se escribieron utilizando R, bash, Perl y Python y la manera de integrarlas está disponible en <u>https://github.com/jenniferlu717/KrakenTools</u>. Las matrices de anotación taxonómica se formatearon para la biblioteca R ggplot2 (Wickham 2016) para generar gráficos de barras apiladas en diferentes niveles taxonómicos.

Las matrices integradas para las anotaciones de las CAZymes, CAZymes extracelulares y las CAZymes extracelulares en MAGs se escribieron utilizando programas de Perl que están disponibles en <u>http://github.com/Ales-ibt/Metagenomic-benchmark</u> y que posteriormente fueron modificadas en R. Las matrices completas de anotación de CAZymes, CAZymes extracelulares, beta diversidad y CAZymes en MAGs se encuentran en el material suplementario del artículo "Metagenomic analysis of carbohydrate-active enzymes and their contribution to marine sediment biodiversity" (López-Sanchez et al., 2024).

Las matrices de anotación funcional se formatearon para la biblioteca R ggplot2 (Wickham 2016) para generar gráficos de pie con las familias más abundantes de CAZyme y las gráficas de caja con los conteos de las CAZymes en MAGs de alphaproteobacteria, gammaproteobacteria y bacteroidia en sedimentos y suelos.

La visualización de las PCoAs se realizó utilizando los paquetes R vegan, pragma (Oksanen et al., 2020) y geosphere (Hijmans et al., 2020). La visualización del análisis del tamaño del efecto de la prueba de discriminación lineal (LEfSe) de las matrices taxonómicas de arqueas y bacterias y de CAZymes extracelulares se realizó en la herramienta Galaxy de Hutlab (Segata et al. 2011).

La visualiación de los mapas de calor se realizó por medio de R con la paquetería de ComplexHeatmap (Gu 2022).

La representación de los arboles filogenéticos se realizó con Interactive Tree Of Life (iTOL) Versión 6.8.1. (2023). Recuperado de https://itol.embl.de/ (Letunic et al., 2021).
## 7 Resultados y Discusión

### 7.1 Sedimentos Marinos y su localización

Se hizo una búsqueda exhaustiva de proyectos de sedimentos marinos que contuvieran muestras metagenómicas en la base datos del Centro Nacional para la Información. Una vez que se identificaron los BioProyectos adecuados, se procedió a descargar las secuencias de metagenoma shotgun desde la base de datos del NCBI. Obtuvimos secuencias de 37 muestras metagenómicas que provienen de 12 BioProyectos que representaban muestras de sedimentos marinos y las cuales incluían información valiosa relacionada con cada muestra, como lo son la localización geográfica (coordenadas de latitud/longitud), profundidad en metros por debajo del subsuelo marino (mbsf por sus siglas en inglés) y profundidad en metros bajo el nivel del mar (mbsl por sus siglas en inglés). Esta información adicional ofreció un contexto importante para comprender la distribución espacial y las características ambientales de los ecosistemas marinos muestreados.

Se recopilaron muestras de sedimentos de diversas ubicaciones, con una profundidad que variaba desde 0 hasta 7942 metros bajo el nivel del mar (mbsl) y desde 0 hasta 2.23 metros bajo subsuelo marino sedimentario (mbsf) (ver Figura 6). Con el fin de minimizar cualquier sesgo en la secuenciación, se excluyeron las muestras que no habían sido secuenciadas mediante la tecnología Illumina. Veinte muestras se tomaron por debajo de 1000 mbsl y 17 por encima (Figura 6 y Tabla 2).



Figura 6. Localización de las muestras de sedimento marino. En rojo están las muestras por encima de la termoclina (< 1000 m) y en amarillo las que están por debajo (> 1000 m). El número entre paréntesis dice el número de muestras.

Proyecto	SRA	Etiqueta	# genes	mbsl	mbsf	Longitud	Latitud	Referencia
PRJNA60 9564	SRR1130 8317	CIGOMA4	36537 15	2966	0.10 4	-95	25.83	Raggi et al., 2020
PRJNA60 9564	SRR1130 8316	CIGOMD18	42009 51	1320	0.10 4	-94.41	25.53	Raggi et al., 2020
PRJNA83 5145	SRR1916 0659	KJGOM4	33497 39	1281	0- 0.02	-94.20	19.56	https://www.ncbi.nlm.nih.gov/bioproject/ PRJNA835145
PRJNA83 5146	SRR1916 0658	KJGOM5	36354 96	1281	0.12- 0.14	-94.20	19.56	https://www.ncbi.nlm.nih.gov/bioproject/ PRJNA835145
PRJNA83 5147	SRR1916 0657	KJGOM6	38644 09	1281	0.22- 0.24	-94.20	19.56	https://www.ncbi.nlm.nih.gov/bioproject/ PRJNA835145
PRJEB25 358	ERR2431 945	ART45	11620 120	420	0	-135.34	70.47	https://www.ncbi.nlm.nih.gov/bioproject/ PRJEB25358
PRJEB25 358	ERR2431 946	ART46	79514 92	410	0	-135.34	70.42	https://www.ncbi.nlm.nih.gov/bioproject/ PRJEB25358
PRJEB25 358	ERR2431 947	ART47	58824 31	277	0	-135.56	70.38	https://www.ncbi.nlm.nih.gov/bioproject/ PRJEB25358
PRJEB25 358	ERR2431 948	ART48	12680 711	848	0	-136.12	70.51	https://www.ncbi.nlm.nih.gov/bioproject/ PRJEB25358
PRJEB25 358	ERR2431 949	ART49	14544 454	1534	0	-139.02	70.39	https://www.ncbi.nlm.nih.gov/bioproject/ PRJEB25358
PRJEB25 358	ERR2431 950	ART50	17656 29	246	0	-138.21	70.05	https://www.ncbi.nlm.nih.gov/bioproject/ PRJEB25358
PRJEB25 358	ERR2431 952	KOR52	45191 13	0	0	127.62	34.68	https://www.ncbi.nlm.nih.gov/bioproject/ PRJEB25358
PRJEB25 358	ERR2431 958	KOR58	47724 86	0	0	126.23	35.08	https://www.ncbi.nlm.nih.gov/bioproject/ PRJEB25358
PRJNA29 0197	SRR2133 847	HRSPAC47	12515 55	600	0- 0.12	-125.10	44.67	Kauffman et al., 2018
PRJNA29 7058	SRR2605 792	GYRPAC92	42370 00	5283	0.00- 0.05	-139.8006	-39.3103	Tully et al., 2016
PRJNA32 6769	SRR3715 733	SMMPAC33	12278 742	863	0.06- 0.09	-118.67	33.79	Kauffman et al., 2018
PRJNA36 2212	SRR6301 444	GBGOC44	43719 37	2000	0.03- 0.04	-111.25	27.04	Dombrowski et al., 2017
PRJNA36 2212	SRR6301 445	GBGOC45	38768 59	2000	0.0- 0.01	-111.25	27.04	Dombrowski et al., 2017
PRJNA36 2212	SRR6301 447	GBGOC47	62054 83	2000	0.0- 0.03	-111.25	27.04	Dombrowski et al., 2017
PRJNA43 1796	SRR6660 646	SMMPAC46	42042 30	860. 5	0-0.3	-118.67	33.79	Dombrowski et al., 2017

Tabla 2. Metadatos de muestras metagenómicas de sedimentos marinos alrededor del mundo.

PRJNA48 1090	SRR7614 706	DSANT06	32563 92	0	0- 0.05	77.96	-68.58	Leeming et al., 2015
PRJNA48 1090	SRR7614 708	DSANT08	25591 88	1.8	0- 0.05	77.95	-68.58	Leeming et al., 2015
PRJNA48 1090	SRR7614 709	DSANT09	28070 71	0	0- 0.05	77.92	-68.60	Leeming et al., 2015
PRJNA48 1090	SRR7614 711	DSANT11	26975 62	0.9	0- 0.05	77.86	-68.61	Leeming et al., 2015
PRJNA48 1090	SRR7614 695	DSANT95	27860 59	0	0- 0.05	77.97	-68.57	Leeming et al., 2015
PRJNA50 4765	SRR8451 930	LOKART30	70990 40	3250	1.26	8.46	73.76	Bäckström et al., 2019
PRJNA50 4765	SRR8452 060	LOKART60	11504 903	3236	2.23	8.45	73.75	Bäckström et al., 2019
PRJNA50 4765	SRR8452 062	LOKART62	73690 55	3236	1.63	8.45	73.75	Bäckström et al., 2019
PRJNA50 4765	SRR8482 195	LOKART95	33167 08	3236	1.03	8.45	73.75	Bäckström et al., 2019
PRJNA52 6329	SRR8709 623	MARPAC23	15285 36	7942	0- 0.01	144.93	11.91	https://www.ncbi.nlm.nih.gov/bioproject/ PRJNA526329/
PRJNA55 3005	SRR9649 754	DELGOM54	17543 39	1400	0.10 - 0.15	-94.69	26.94	Zhao R. et al., 2020
PRJNA55 3005	SRR9649 755	DELGOM55	26977 11	1400	0.10 - 0.15	-94.69	26.94	Zhao R. et al., 2020
PRJNA55 3005	SRR9649 756	DELGOM56	27238 78	1400	0.10 - 0.15	-94.69	26.94	Zhao R. et al., 2020
PRJNA55 3005	SRR9649 757	DELGOM57	19719 29	1400	0.10 - 0.15	-94.69	26.94	Zhao R. et al., 2020
PRJNA55 3005	SRR9649 758	DELGOM58	16587 01	1400	0.10 - 0.15	-94.69	26.94	Zhao R. et al., 2020
PRJNA55 3005	SRR9649 759	DELGOM59	15081 47	1400	0.10 - 0.15	-94.69	26.94	Zhao R. et al., 2020
PRJNA55 3005	SRR9649 760	DELGOM60	18268 93	1400	0.10 - 0.15	-94.69	26.94	Zhao R. et al., 2020

#### 7.2 Análisis Gen-Céntrico

#### 7.2.1 Taxonomía de muestras metagenómicas

Las secuencias crudas de las 37 muestras recolectadas se anotaron taxonómicamente por medio del programa Kraken2 (Wood et al., 2019) contra las bases de datos de nucleótidos y la de Secuencias de Referencia (RefSeq por sus siglas en inglés) del NCBI para arqueas y bacterias. Los resultados muestran que en todas las muestras las secuencias más abundantes pertenecientes a bacterias son del filo Proteobacteria (en el rango del 35.14% al 70.99% de las secuencias de cada metagenoma) de las clases Alfa, Beta, Delta, Gamma y Epsilon, siendo las Gammaproteobacteria y las Alfaproteobacterias las más abundantes (13.88 al 47.74% y 6.91 al 33.83% respectivamente) (ver Figura 7).



Figura 7. Abundancia relativa de la bacteria en cada metagenoma analizado a nivel de clase. "OTRO" es cualquier clase que no tiene por lo menos el 1% de abundancia en alguno de los metagenomas. Consulte para el recuento completo del porcentaje de los taxa en todos los metagenomas en https://github.com/RafaelLopez-Sanchez/marine\_sediments.

99.73% de las secuencias de cada metagenoma) de la abundancia relativa y consistente con lo reportado en muestras de sedimento marino. a 8.79% lo cual nos habla de una inmensa diversidad taxonómica en las secuencias el caso de las bacterias la abundancia relativa menor al 1% a nivel de clase va desde 3.96% Thaumarchaeota (0.06 al 91.41% de las secuencias de cada metagenoma) (ver Figura 8) En En el caso de arquea pertenecen a los filos de Euryarchaeota (en el rango del 7.51 al



# 7.2.2 Correlación de diversidad taxonómica de los sedimentos con los metadatos asociados

La mayoría de nuestras muestras de sedimento que se encuentran depositadas en el NCBI no tenían disponibles las variables fisicoquímicas para su comparación. Sin embargo, todas provienen de sedimentos superficiales en la interfaz con la columna de agua, para las cuales se conocen metadatos como parámetros geográficos (latitud y longitud, profundidad en metros por debajo del nivel del mar mbsl y la profundidad del sedimento mbsf. Esto nos permitió probar la correlación entre las variables ambientales y la matriz de diversidad de abundancia a nivel de clase. Empleamos una prueba de Mantel para verificar si la estructura de la comunidad taxonómica estaba correlacionada con los parámetros geográficos y espaciales (Tabla 3).

 Tabla 3. Pruebas de Mantel para ver correlación de los metadatos contra la matriz de disimilitud de Bray-Crutis de diversidad

 taxonómica a nivel de clase realizadas con el paquete de R vegan (Oksanen et al., 2020).

Prueba	R Estadístic	Significanci a (p-value)	Número de permutacione
	a		S
Abundancia vs localización geográfica	0.1076	0.083	9999
(latitud/longitud)			
Abundancia vs mbsl	0.3065	0.00004	9999
Abundancia vs mbsf	0.1343	0.065	9999

Nuestros resultados mostraron una correlación significativa y positiva entre la profundidad (mbsl) y la diversidad taxonómica (matriz de disimilitud de Bray-Curtis), mientras que las distancias geográficas y la profundidad del sedimento (mbsf) no fueron significativas (Tabla 3). Para analizar aún más esta correlación positiva entre la profundidad de la columna de agua y la diversidad taxonómica, realizamos una regresión lineal de ambas matrices de disimilitud (Figura 9). Un valor bajo de R-cuadrado (0.0404) sugiere que la profundidad por debajo del nivel del mar no explica gran parte de la variación en la disimilitud taxonómica. Por eso, decidimos realizar un perfil metagenómico de las muestras basado en la diversidad taxonómica de la comunidad en relación con su potencial metabólico.

#### Regresión de Mantel



Figura 9. Regresión lineal de la prueba de Mantel en abundancia vs. matrices disimilatorias de mbsl.

## 7.2.3 Potencial metabólico para respirar oxígeno en muestras de sedimento

Aunque existe una correlación entre una mayor profundidad de la columna de agua y el grosor del sedimento, debido a que la cantidad de materia orgánica se agota y la penetración de oxígeno se encuentra a lo largo del sedimento, lo que es una característica que haría diferencias sustanciales en las poblaciones microbianas por su tipo de metabolismo, muchas de nuestras muestras se tomaron en los primeros centímetros (de 0 a 2,23 mbsf) donde la comunidad utiliza oxígeno. (D'Hondt et al., 2015). Por ello, decidimos hacer un perfil metagenómico de las muestras basado en la diversidad taxonómica de la comunidad frente a su potencial metabólico. Dado el hecho de que solo una muestra midió oxígeno y entendiendo que todas nuestras muestras superficiales son un gradiente entre las capas óxicas y anóxicas de los sedimentos, decidimos que la mejor manera de obtener una comparación sería ver qué metabolismo respiratorio prevalece en cada muestra. Esto no recrea las condiciones geoquímicas de cada muestra, pero sí hace un buen acercamiento para entender la estructura comunitaria de los sedimentos marinos. Con este fin, asignamos categorías a nuestras muestras (ambientes óxicos/anóxicos) basadas en su contenido genético de reductasas de oxígeno de hemo-cobre (HCO) y reductasas de óxido nítrico (NOR). Las HCO y las NOR son enzimas que se encuentran en los últimos complejos de muchas cadenas respiratorias en microorganismos (Sousa et al., 2011). Como referencia, utilizamos cuatro sedimentos encontrados en el Castillo de Loki etiquetados como anóxicos y uno del Giro del Pacífico Sur etiquetado como óxico y que también tiene mediciones fisicoquímicas de oxígeno. Todo lo que estaba por encima de los recuentos normalizados de HCO y NOR en el control óxico se consideró óxico, y todo lo que estaba por debajo se

consideró anóxico. Nuestros resultados muestran 18 muestras metagenómicas que pueden considerarse óxicas y 13 anóxicas (Figura 10).



#### Oxígeno Reductasas en muestras de sedimento

Conteos normalizados de oxígeno reductasas

Figura 10. Gráfico de dispersión de HCOs y NORs clasificados en cada metagenoma con la base de datos HCO. El eje x muestra los recuentos de genes normalizados de las lecturas, y el eje y es la profundidad en metros bajo el fondo marino de cada muestra de metagenoma (mbsf). Código de colores; muestras de metagenoma menores a 1000 mbsl = naranja; muestras mayores a 1000 mbsl = rosa control óxico = verde; control anóxico = morado.

# 7.2.4 Beta Diversidad de las muestras metagenómicas en cuanto a su disponibilidad de oxígeno

Una vez que establecimos la abundancia de HCO y NOR como una condición en las muestras, un análisis de coordenadas principales (PCoA) basado en la abundancia taxonómica relativa a nivel de clase (matriz de disimilitud de Bray-Curtis) mostró una clara separación de las muestras etiquetadas como óxicas y anóxicas (62.18% de la varianza explicada en CoA1 y CoA2) (Figura 11). Las muestras se agruparon en dos grupos; en el grupo óxico, muestras profundas del Golfo de México (Godoy-Lozano et al., 2018; Zhao R. et al., 2020) que no reportaron filtraciones de hidrocarburos o metano (Zhao R. et al., 2020). El Giro del Pacífico Sur es la única muestra con un nivel óxico y un metabolismo oligotrófico (Tully et al., 2016). Muestras de las playas de Corea y la estación de Davis ubicada en la Antártida presentan perturbaciones antropogénicas; los metagenomas de Corea son muestras de playa, y las de la estación de Davis son muestras superficiales ricas en nutrientes, y el oxígeno se consume en los primeros centímetros del sedimento (Leeming et al., 2015). En el grupo anóxico, muestras del Golfo de México (Universidad de

Delaware), la Cuenca de Guaymas ubicada en el Golfo de California y el Castillo de Loki ubicado en el Ártico, así como la Cresta de Hidrato del Pacífico y los Montículos de Santa Mónica se agruparon juntos. Se ha reportado que estos tienen filtraciones de hidrocarburos o compuestos relacionados (Zhao R. et al., 2020), ventilas hidrotermales con metabolismo anaeróbico (Jaeschke et al., 2012; Kauffman et al., 2018; Bäckström et al., 2019) y volcanes de lodo (Kauffman et al., 2018; Bäckström et al., 2019) (Figura 11).



Figura 11. Beta Diversidad a nivel taxonómico de clase. Análisis de Coordenadas Principales (PCoA) de una matriz de disimilitud de Bray-Curtis de taxones a nivel de clase en muestras de sedimentos. El código de colores indica a qué categoría de metadatos pertenecen (azul = óxico; amarillo = anóxico).

En base a la observación de una clara separación entre las etiquetas, exploramos las diferencias en la composición taxonómica entre las muestras óxicas y anóxicas a través de un análisis LEfSe (Segata et al., 2011) basado en matrices de abundancia de bacterias y arqueas a nivel de clase y filo, respectivamente (Figura 12 y Figura 13). LEfSe

proporciona biomarcadores basados en diferentes categorías de metadatos (en este caso, características óxicas y anóxicas).

Las muestras óxicas mostraron un enriquecimiento en Alphaproteobacteria. Muchas alfaproteobacterias codifican varios tipos diferentes de oxidasas terminales, lo que probablemente refleja la versatilidad ecológica con respecto a los niveles de oxígeno (Cevallos, M. A., y Degli Esposti, M. 2022).

Las muestras anóxicas tuvieron un enriquecimiento en varias clases bacterianas: Epsilonbacteria, Deltaproteobacteria, Bacilli, Clostridia, Fusobacteriia, Dehalococcoidia, Bacteroidia, Sphingobacteriia, Cytophagia y Thermodesulfobacteria (Figura 12).



Figura 12.Análisis Discriminante de Efectos Lineales de Tamaño (LEfSe) para identificar taxones significativos entre muestras con las clases 'anóxicas' y 'óxicas' de bacterias. Los grupos taxonómicos muestran valores de LDA > 3.0 con p < 0.1. El efecto del tamaño y el poder del análisis estadístico se calculó con valores alfa de 0.5 y 0.5 para Kruskal-Wallis (clases) y Wilcoxon (subclases), respectivamente.

Entre los fila de Archaea, Thaumarchaeota está enriquecido en muestras óxicas, mientras que Candidatus Bathyarchaeota, Euryarchaeota y Candidatus Lokiarchaeota están enriquecidos en muestras anóxicas (Figura 13).

Estos resultados son además consistentes con la literatura revisada de sedimentos marinos, donde se sabe que los sedimentos anóxicos están enriquecidos con grupos estrictamente anaerobios como las bacterias sulfato reductoras del filo Chloroflexota (clase Dehalococcoidia) y Deltaproteobacteria y arqueas metanogénicas, como Euryarchaeota, mientras que en los sedimentos óxicos hay prevalencia de la clase Alphaproteobacteria en bacterias y el filo Thaumarchaeota en arqueas (Biddle et al., 2008; Orsi, 2018; Hoshino et al., 2020). Nuestros resultados encontraron que las clases Deltaproteobacteria y Dehalococcoidia y de los fila Chloroflexota y Proteobacteria junto con otras clases anaeróbicas como Clostridia, Thermodesulfobacteria, Fusobacteriia y las arqueas Euryarchaeota son indicativas de un ambiente anóxico, mientras que la clase Alphaproteobacteria de bacterias y Thaumarchaeota arqueas (Tully et al., 2016; Hoshino et al., 2020) fueron indicativas de muestras óxicas (Figuras 12 y 13).



Figura 13.Análisis Discriminante de Efectos Lineales de Tamaño (LEfSe) para identificar taxones significativos entre muestras con los fila 'anóxicas' y 'óxicas' de archaea. Los grupos taxonómicos muestran valores de LDA > 3.0 con p < 0.1. El efecto del tamaño y el poder del análisis estadístico se calculó con valores alfa de 0.5 y 0.5 para Kruskal-Wallis (clases) y Wilcoxon (subclases), respectivamente.

#### 7.2.5 Anotación funcional de CAZymes en muestras de sedimentos

Examinamos la distribución del contenido de CAZyme (Enzimas Activas de Carbohidratos) dentro de los metagenomas. Para lograr esto, realizamos un Análisis de Coordenadas Principales (PCoA) utilizando recuentos normalizados de todas las familias CAZyme identificadas dentro de cada muestra de metagenoma. Como en nuestros hallazgos sobre la beta diversidad, las muestras mostraron una separación entre condiciones óxicas y anóxicas (59.18% de la varianza explicada en CoA1 y CoA2) (Figura 14).





Teniendo en cuenta la suposición de que el reciclaje de carbono en los sedimentos marinos es llevado a cabo por los microorganismos que lo habitan, mismos que usan enzimas secretadas (conocido como "secretoma") para almacenar carbono a largo plazo (Orsi et al., 2018), decidimos buscar CAZymes secretadas. Nos fijamos en las CAZymes ya que estas son enzimas clave para la degradación de carbohidratos que son la mayor parte de macromoléculas en organismos incluidos procariotes que habitan en el detrital marino orgánico que es la nieve marina (Zhao Z. et al., 2020).

Realizamos una anotación funcional de familias CAZymes que tenían un péptido señal contra la base de datos CAZyme (Lombard et al., 2013). Anotamos las secuencias en las seis clases de la base de datos CAZy, que están implicadas en la creación, descomposición e identificación de carbohidratos. Estas clases son Glicosil Transferasas (GTs), Glicosil Hidrolasas (GHs), Esterasas de Carbohidratos (CEs), Módulos de Unión a Carbohidratos (CBMs), Liasas de Polisacáridos (PLs) y Actividades Auxiliares (AAs). Dieciocho familias de CAZyme secretadas se encontraron en más del 1% de todas las anotaciones totales de CAZyme (lo que representa el 55.94% de todas las CAZymes anotadas en nuestras

muestras de metagenoma). De estos las familias, GH109, GH23 y CE1 fueron las más abundantes (Figura 15).



Figura 15. Gráfico de pie con porcentajes de las familias CAZyme secretadas con más anotaciones en los metagenomas de sedimento. "OTRO" se refiere a cualquier módulo que es menor al 1% de abundancia.

Su abundancia fue particularmente alta en los siguientes metagenomas: Cuenca de Guaymas (GBGOC), Estación Davis de la Antártida (DSANT), playas de Corea (KOR), Cresta de Hidrato del Pacífico Sur (HRSPAC47), Castillo de Loki (LOKART) del Ártico, Montículos de Santa Mónica (SMMPAC) y el Golfo de México (CIGOMD18 y KJGOM6) (Figura 16).



Figura 16. Mapa de calor de la abundancia y distribución de familias secretadas CAZyme en muestras de sedimentos. Módulos de Unión a Carbohidratos (CBMs), Esterasas de Carbohidratos (CEs), Hidrolasas de Glicósidos (GHs) y Transferasas de Glicósidos (GTs). El código de color de las familias se refiere al sustrato al que las familias están dirigidas reportados en la literatura (Lombard et al., 2013; Orsi et al., 2018). El color de la barra superior representa la etiqueta de metadatos (amarillo=anóxico; azul=óxico). El tipo de sustrato se encuentra en la barra lateral en círculos de colores. Consulte para el recuento completo de genes que codifican CAZymes en todos los sedimentos en https://github.com/RafaelLopez-Sanchez/marine\_sediments.

Los metagenomas tenían un repertorio de CAZymes secretadas, dirigido principalmente a detritos algales y necromasa (ver Figura 15, Tabla 4 y Tabla 5). Entre las familias prevalentes involucradas en la descomposición de detritos algales estaban los módulos de glicosidasa hidrolasa GH2, GH3 y GH16 3, así como la esterasa de carbohidratos CE1. Los módulos de unión incluían CBM9, CBM44 y CBM67. Estas familias están compuestas de enzimas con actividades de  $\beta$ -galactosidasas,  $\beta$ -glucuronidasas,  $\beta$ -manosidasas, exo- $\beta$ glucosaminidasas en el caso de GH2 y GH3, donde las glicosidasa hidrolasas y fosforilasas realizan una amplia gama de funciones que implican la degradación de biomasa y remodelación de paredes celulares de plantas y bacterias. GH16\_3 rompe laminarasa, un carbohidrato encontrado en algas pardas (Qin et al., 2017) mientras que CE1 tiene actividades de acetil xilano esterasas (EC 3.1.1.72), feruloil esterasas (EC 3.1.1.73) y muchas otras esterasas como las PHB depolimerasas. CBM9 y CBM44 son módulos que se dirigen a dominios de unión a celulosa principalmente xilano y otros carbohidratos y dominios de unión a celulosa y CBM67 se dirige a la unión a L-ramnosa, un carbohidrato producido por microalgas (0 a13.3% de la composición algal) (Brown, 1990) (Figura 15 y Tabla 4 y Tabla 5) (Lombard et al., 2014).

Tabla 4. Familias de los módulos de síntesis de polisacáridos más abundantes extracelulares encontrados en los metagenomas de sedimento con sus posibles sustratos y las actividades enzimáticas más relevantes de cada familia. Entre paréntesis está el número EC en caso de tenerlo. La información de esta tabla está contenida en www.cazy.org, https://www.cazypedia.org y

https://bcb.unl.edu/dbCAN\_sub/data/.

Sustratos	Familia CAZyme	Actividad(es) enzimática(s)
Quitina/Peptidoglicano/Detrito algal	GT2_Glycos_tranf_2_	Biosíntesis de la cápsula bacteriana.
Peptidoglicano	GT51	Peptidoglicano glicosiltransferasa (EC 2.4.1.129)
Detrito algal	GT83	[Undecaprenil fosfato- α-L-Ara4N: 4-amino-4- desoxi-β-L- arabinosiltransferasa (EC 2.4.2.43);
Detrito algal	GT4	Hexosiltransferasas (EC 2.4.1)

Tabla 5. Familias de las familias más abundantes extracelulares de reconocimiento y degradación de polisacáridos encontrados en los metagenomas de sedimento con sus posibles sustratos y las actividades enzimáticas más relevantes de cada familia. Entre paréntesis está el número EC en caso de tenerlo. La información de esta tabla está contenida en www.cazy.org, https://www.cazypedia.org y https://bcb.unl.edu/dbCAN\_sub/data/.

Sustratos	Familia CAZyme	Actividad(es) enzimática(s)
Glicanos del huésped	GH109	alfa-N-acetilgalactosaminidasa (3.2.1.49); beta-N- acetilhexosaminidasa (3.2.1.52)
Quitina/Peptidoglicano	GH23	quitinasa(3.2.1.14); lisozima tipo G (3.2.1.17); peptidoglicano liasa (4.2.2.n1)
Detrito algal	CE1	feruloil esterasa (3.1.1.73); acetil xilano esterasa (3.1.1.72)
Detrito algal	CBM9	Unión a celulosa

Glicanos del huésped	GH29	alfa-1,3/1,4-L-fucosidasa
		(3.2.1.51), alfa-1.2-L-fucosidasa (3.2.1.51): alfa-1.2-L-fucosidasa
		(3.2.1.63)
Quitina/Peptidoglicano/Detrito algal	GH3	exo-1,3-1,4-glucanasa (3.2.1); beta-1,2-glucosidasa (3.2.1); beta-1,3-glucosidasa (3.2.1); beta-glucosidasa (3.2.1.21); xilano 1,4-beta-xilosidasa (3.2.1.37); beta- glucosilceramidasa (3.2.1.45)
Quitina/Peptidoglicano/Detrito algal	CE4	quitina desacetilasa (3.5.1.41);
		quitooligosacárido desacetilasa
		desacetilasa (3.5.1)
Peptidoglicano	GH103	peptidoglicano transglicosilasa
		lítica (3.2.1)
Detrito algal	GH16_3	Endo- $\beta$ -1,3-
		3.2.1.39: endo- $8-1.3(4)$ -
		glucanasa/liquenasa-laminarinasa
		(EC 3.2.1.6)
Detrito algal	GH2	beta-glucosidasa (3.2.1.21); beta-
		galactosidasa $(3.2.1.23)$ ; beta- manosidasa $(3.2.1.25)$ : beta-
		glucuronidasa (3.2.1.31)
Glicanos del huésped	GH33	trans-sialidasa (2.4.1) ; sialidasa
		o neuraminidasa (3.2.1.18)
Detrito algal	CBM67	Unión a ramnosa
Detrito algal	CBM44	Unión a celulosa y xiloglucanos
Quitina/Peptidoglicano	CE14	N-acetil-1-D-mio-inositil-2-
		allino-2-uesoxi-alla-D- gluconiranósido desacetilasa
		(3.5.1.89): quitina disacárido
		desacetilasa (3.5.1.105)

Para la degradación de necromasa, los módulos GH23 y GH103 contienen familias de transglicosilasas líticas de peptidoglicano. GH23 también se ha encontrado que tiene actividad quitinasa. Además, las actividades conocidas de las familias CE4 y CE14 incluyen enzimas como acetilxilan esterasas, deacetilasas de quitina, deacetilasas de oligosacáridos de quitina y deacetilasas de peptidoglicano (CE4) y deacetilasa de diacetilquitobiosa (EC 3.5.1.-) deacetilasas de disacáridos de quitina (CE14). (Lombard et al., 2014). Finalmente, para la degradación de glicanos del huésped, el módulo GH29 contiene  $\alpha$ -L-fucosidasas, y las familias GH109 conforman  $\alpha$ -N-acetilgalactosaminidasa y  $\beta$ -N-acetilhexosaminidasa. La sialidasa GH33 o neuraminidasa (EC 3.2.1.18) se dirige al ácido siálico del glicano del huésped (Figura 15 y Tabla 5).

Está documentado que las comunidades bacterianas dominan los sedimentos someros, que están compuestos principalmente por arcilla, envolturas celulares de organismos planctónicos y materia orgánica (Bienhold et al., 2016). Se espera que los genes relacionados con la degradación de carbono recalcitrante, incluyendo celulosa, quitina o peptidoglicano, jueguen un papel importante en los sedimentos marinos (Tully et al., 2016; Bradley et al., 2018; Orsi et al., 2018). La necromasa contribuye significativamente a satisfacer la demanda energética de hasta el 13% de la comunidad microbiana en sedimentos someros cuando se oxida bajo condiciones óxicas o anóxicas. La oxidación de una célula por año puede proporcionar suficiente energía para sostener la demanda de miles de células en sedimentos con bajos recursos energéticos, posicionando potencialmente la oxidación de necromasa como una fuente primaria de carbono para microorganismos incapaces de sobrevivir en ambientes pobres en energía (Bradley et al., 2018). El hecho de que la mineralización y adsorción de biopolímeros en partículas de sedimento podría reducir la accesibilidad de otros carbohidratos (Orsi et al., 2018), podría hacer que las envolturas celulares, como el peptidoglicano, sean una elección preferida para las familias de CAZyme secretadas que se encontraron más abundantes (Figura 16). La mayoría de estas familias de CAZyme se encuentran en un amplio espectro de formas de vida, pero se concentran en bacterias (Lombard et al., 2014).

Algunas de las familias más abundantes diferían entre las muestras óxicas y anóxicas. Los módulos CAZyme GH23, CBM9, GH16\_3, GT51, CE4 y CE14 fueron significativamente más abundantes en muestras óxicas. Por otro lado, CBM44 y GT83 se encontraron diferentes en relación con muestras anóxicas. Curiosamente, tanto CBM44 como CBM9 pueden unir celulosa (Figura 17). Las familias que están más enriquecidas en los sedimentos óxicos tienen como sustratos principales la necromasa y quitina. De manera adicional la familia de AA7 tiene actividades de oxidación de quitooligosacáridos (www.cazy.org)



Figura 17. Análisis Discriminante de Efectos Lineales de Tamaño (LEfSe) para identificar familias secretadas CAZyme significativos entre muestras con las clases 'óxicas' y 'anóxicas'. Actividades Auxiliares (AAs), Módulos de Unión a Carbohidratos (CBMs), Esterasas de Carbohidratos (CEs), Hidrolasas de Glicósidos (GHs) y Transferasas de Glicósidos (GTs) y Liasas de Polisacáridos (PLs). Los grupos CAZyme muestran valores LDA > 3.5 con p < 0.1. El efecto del tamaño y la potencia del análisis estadístico se calculó con valores alfa de 0.5 y 0.5 para Kruskal-Wallis (clases) y Wilcoxon (subclases), respectivamente.

#### 7.3 Análisis Genoma-Céntrico

#### 7.3.1 Genomas ensamblados a partir de metagenomas (MAG)

Para comprender mejor la comunidad involucrada en el reciclaje de carbohidratos en sedimentos marinos, recuperamos MAGs de cada muestra de metagenoma. Se reconstruyeron 494 genomas ensamblados de metagenomas (MAG) reconstruidos a partir de los 37 metagenomas, cada uno representando una instantánea de las comunidades microbianas muestreadas de diferentes sedimentos. Casi dos tercios de los genomas están sustancialmente completos con una completitud ≥80% y una contaminación <10%, mientras que el resto tienen una completitud ≥75% y contaminación <10%. Los tamaños de los MAG varían de 0.75 a 9.56 Mbps. Los MAG están distribuidos a lo largo del árbol filogenético y se agrupan en 443 MAG bacterianos y 51 MAG arqueales comprendidos en 103 y 3 grupos taxonómicos a nivel de clase respectivamente, con 360 MAG taxonómicamente asignados a nivel de especie basados en una identidad promedio de nucleótidos (ANI por sus siglas en inglés) del 95%. La gran mayoría de ellos pertenecen a las clases de los filos Proteobacteria (Gammaproteobacteria y Alphaproteobacteria) y Bacteroidia. (Figura 18, Tabla Suplementaria 1).

### 7.3.2 Anotación de CAZymes en los MAG

Dado que no es posible reconstruir todos los genomas en muestras de sedimentos, y para establecer una relación entre nuestros resultados metagenómicos y la participación de los MAG recuperados en el reciclaje predominante de carbohidratos en sedimentos marinos, nos enfocamos en las familias de CAZymes más abundantes encontradas en nuestras anotaciones de metagenomas. Esto es especialmente relevante, ya que se ha demostrado que la necromasa y los detritos algales juegan un papel crucial en este proceso (Orsi 2018). Nos centramos en módulos de CAZymes secretadas y familias de CAZymes correspondientes a Clústeres de Genes de CAZymes (CGCs por sus siglas en inglés) (Figura 19, Tabla Suplementaria 2).

El módulo GH23 fue el más abundante en MAG de sedimentos y no se encontraron módulos CBM44 en ninguno de los MAG. MAG de las clases Alphaproteobacteria, Bacteroidia y Gammaproteobacteria tenían más de un módulo de CAZymes extracelulares (Figura 14). Los MAG de Alphaproteobacteria tenían módulos GH103 y GH23. Los MAG de Alphaproteobacteria pertenecen a las familias Rhodobacteraceae y Methyloligellaceae con especies encontradas en ambientes marinos, incluyendo los géneros *Pseudorhodobacter, Sulfitobacter, Roseicyclus* e *Hyphomicrobium* (Uchino et al., 2002; Rathgeber et al., 2005; Yoon et al., 2007; Vuilleumier et al., 2011) y otras especies del género *Methyloceanibacter*, que habían sido previamente reportadas en sedimentos del Mar del Norte (Vekeman et al., 2016, Tabla Suplementaria 2).

Los MAGs de Bacteroidia contenían módulos de CAZymes en al menos un MAG excepto para la familia GT83; la composición de CAZymes de los principales módulos encontrados en nuestras muestras de metagenomas (GH109, GH23 y CBM9) fue mayor en la familia Flavobacteriaceae donde se asignaron MAG al género *Prevotella* (DSANT95\_maxbin.044), *Maribacter* (DSANT06\_maxbin.002, DSANT06\_maxbin.016 y DSANT95\_maxbin.030) *Pricia* (DSANT95\_maxbin.051, DSANT95\_maxbin.024 y DSANT95\_maxbin.016) *Eudoraea* (DSANT11\_maxbin.017), *Aureibaculum* (DSANT06\_maxbin.006 y DSANT08\_maxbin.008) junto con otros módulos abundantes (Figura 18, Figura 19 y Tabla Suplementaria 2 (López-Sanchez et al., 2024).



Figura 18. Árbol filogenético de los MAGs reconstruidos en sedimentos marinos. Las clases taxonómicas de los 10 MAGs más abundantes reconstruidos están codificadas por colores (Consulte el repositorio de Github para obtener anotaciones taxonómicas completas en MAG). Los círculos concéntricos muestran A) La clase taxonómica B) La completitud (%), C) El tamaño del MAG (Mbps). Las clases de taxonomía se anotaron utilizando la base de datos GTDB v-207 (Chaumeil et al., 2020). Los valores de bootstrap en las ramas están coloreados.



Figura 19. Mapa de calor de las familias más abundantes de enzimas de activación de carbohidratos secretadas (CAZymes) encontradas en nuestras muestras de genomas ensamblados de metagenomas (MAG). La etiqueta de color lateral es para la taxonomía a nivel de clase anotada al MAG y se lee de arriba a abajo. Para fines de visualización, los MAG que no tenían un mínimo de los principales Familias de CAZymes fueron descartados de la figura. Consulte para el recuento completo de genes que codifican CAZymes en todos los MAG en <u>https://github.com/RafaelLopez-Sanchez/marine\_sediments</u>.

Las especies de los géneros *Prevotella*, *Maribacter* y *Aureibaculum* se habían recuperado de sedimentos marinos del Océano Pacífico y el Mar Amarillo (Reed et al., 2002; Nedashkovskaya et al., 2004; Zhao et al., 2019). El género *Pricia* se había aislado previamente de una muestra de sedimento intermareal arenoso recogida en la costa antártica (Yu et al., 2012), lo cual es consistente con el lugar de donde se recuperó (Estación Davis). Las especies del género *Eudoraea* fueron aisladas de las aguas costeras del Mar Adriático (Alain et al., 2008).

Finalmente, MAG sin familias GH23 como las clases de Phycisphaerae, UBA2214, Planctomycetia y Bacteroidia contenían GH109, GH2, GH29 y CBM67. UBA2214 también estaba enriquecido con módulos GH3. MAG de UBA2214, Phycisphaerae y Planctomycetia fueron asignados a las familias Zgenome-0027, Anaerohalosphaeraceae y Thermoguttaceae, respectivamente. Las especies de estas familias se encuentran en sedimentos marinos y ambientes acuáticos con bajo oxígeno (Dedysh et al., 2020; Pradel et al., 2020; Chiciudean et al., 2022). Además, cuatro MAG de Bacteroidia asignados al orden Bacteroidales mostraron un inventario de CAZymes similar abundantes (Figura 18, Figura 19 y https://github.com/RafaelLopez-Sanchez/marine\_sediments).

Los MAG pertenecientes a Bacteroidia y Gammaproteobacteria tienen el mayor número de CGC. MAG de Bacteroidia clasificado como *Prevotella* (DSANT95\_maxbin.044) y Gammaproteobacteria *GCA-001735895 sp009937625* (KOR58\_maxbin 012.fasta.contigs.refinado) tuvieron el mayor número de CGC de todos, con cinco incluyendo módulos GH3, GH23 y GH2 en el caso de *Prevotella* y dirigidos a módulos GH23, GH103, GT51 y CE4. (<u>https://github.com/RafaelLopez-Sanchez/marine\_sediments</u>).

El MAG de Alphaproteobacteria contenía CGC dirigidos a módulos GH23, GH103 y GH3. Se encontraron CGC de Gammaproteobacteria dirigidos a módulos CE4, GH103, GT51, CBM9, GH23 y GH3.

Aunque los MAG ensamblados no pueden cubrir toda la diversidad de los sedimentos, encontramos un grupo de MAG anotados a clases que eran abundantes en nuestras muestras, como Bacteroidia, Alphaproteobacteria y Gammaproteobacteria. Encontramos el inventario de CAZymes y CGCs que contenían los módulos más abundantes encontrados en nuestros metagenomas en estas clases de bacterias. Además, los MAG de Bacteroidia, Alphaproteobacteria y Gammaproteobacteria que se encontraron teniendo módulos importantes de CAZymes pertenecen a géneros o familias encontrados o aislados en ambientes marinos, lo que hace que estas clases sean algunos de los principales impulsores de la transformación de carbohidratos en sedimentos marinos. Es bien sabido que el filo Bacteroidota se considera el filo primario para la degradación de carbohidratos (Lapébie et al., 2019). Todos nuestros MAGs de este filo pertenecían a Bacteroidia. El filo Proteobacteria fue el más prevalente en muestras de sedimentos. La mayoría de los taxones que encontramos pertenecen a Gamma y Alphaproteobacteria (47.75-13.88% y 33.83-6.91% de abundancia relativa, respectivamente) (https://github.com/RafaelLopez-Sanchez/marine\_sediments).

Logramos identificar y analizar los MAG de muestras de metagenomas (Figura 18, https://github.com/RafaelLopez-Sanchez/marine\_sediments), de las principales familias en la transformación de carbohidratos en sedimentos marinos. Estas clases mostraron la presencia de los familias más abundantes de CAZymes secretadas y los Clústeres de Genes de CAZymes (CGCs) que corresponden a la degradación de carbohidratos en ambientes marinos. La presencia de estas CAZymes y CGCs en MAG derivados del mar indica su papel crítico en la transformación de carbohidratos en sedimentos marinos. Esto resalta la importancia del filo Bacteroidota en la degradación de carbohidratos, particularmente la clase Bacteroidia, y las contribuciones significativas de ambos, Gamma y Alphaproteobacteria, a los taxones observados en muestras de sedimentos marinos.

### 7.3.3 CAZymes en MAG de sedimento vs CAZymes en MAG de suelo

Dado que Alphaproteobacteria, Gammaproteobacteria y Bacteroidia tenían un inventario tan rico de CAZymes para carbohidratos encontrados en sedimentos marinos, decidimos explorar cuán diferentes eran los inventarios de CAZymes de nuestros MAG en comparación con los de MAG de Alphaproteobacteria, Bacteroidia y Gammaproteobacteria seleccionados de muestras de suelo publicadas por Nayfach et al., (2021) utilizando los mismos criterios de selección (Completitud > 75% Contaminación <10%). Los sedimentos marinos y el suelo son ecosistemas ricos en microorganismos y son componentes cruciales en el ciclo del carbono, ya que pueden secuestrarlo y desempeñar un papel en su reciclaje. (Arndt et al., 2013; Bargett et al., 2014).

Los MAG de estas clases eran principalmente de diferentes familias en comparación con los MAG de sedimentos que recuperamos, los cuales se agruparon juntos en un árbol filogenético (Figura 20); los MAG de Gammaproteobacteria encontrados en sedimentos se agrupan en diferentes clados; el primer grupo comprendía bacterias oxidantes de sulfuro de la familia Beggiatoaceae, bacterias que viven en sedimentos superficiales y en interfaces de sedimento-agua (Teske & Salman, 2014); el segundo grupo agrupaba bacterias principalmente del orden Acidiferrobacterales que tiene géneros no cultivados que realizan fijación de carbono oscuro en sedimentos costeros (Dyksma et al., 2016); el tercer clado reune bacterias de seis órdenes diferentes principalmente de dos familias (UBA4575 y SZUA-229) cuyas secuencias se han encontrado principalmente en ambientes marinos (Parks et al., 2022); el cuarto clado tiene bacterias del orden Pseudomonadales que comprende diferentes familias de microorganismos encontrados en ambientes marinos como Moraxellaceae, Halomonadaceae, HTCC2089 y Halieaceae (Park et al., 2012; Matsuyama et al., 2015; Qiu et al., 2021); el quinto clado son bacterias principalmente de la familia Woeseiaceae (orden Woeseiales) que se ha encontrado en sedimentos marinos (Hoffman et al., 2020); y finalmente el clúster 7 tiene bacterias de la familia Nitrosomonadaceae que comprende un grupo de bacterias oxidantes de amoníaco y se ha encontrado en ambientes marinos (Prosser et al., 2014).

En Alphaproteobacteria hay dos clústeres, uno de bacterias que pertenecen a la familia Rhodobacteraceae y otro del orden Rhizobiales (familias Hyphomicrobiaceae y Methyloligellaceae). Los MAG de Bacteroidia de sedimento también se agrupan principalmente en dos grupos, uno del orden Bacteroidales y el otro en la familia Flavobacteriaceae, todos ellos con especies aisladas de ambientes marinos (Nedashkovskaya et al., 2004), como se discutió. También hay un singleton perteneciente al MAG de la especie *Prevotella sp018054505*, que muestra una mayor divergencia genética. Como se mencionó anteriormente, este MAG se encontró que tiene los CGCs más numerosos en cuanto a los módulos de CAZymes más comunes y el hecho de que las especies del género *Prevotella* se han encontrado en ambientes marinos debido a la contaminación antropogénica de alcantarillas cerca de la playa noruega de Bore (Bagi & Skogerbø 2022), al igual que en las muestras de sedimentos marinos de la Estación Davis (donde se reconstruyó este MAG), podría indicar cambios evolutivos más extensos en este MAG en particular. El hecho de que los MAG de sedimentos de la misma clase taxonómica se agrupen juntos sugiere que estos microorganismos a menudo poseen adaptaciones ecológicas compartidas. En la literatura se ha sugerido que las distribuciones de clústeres podrían interpretarse como evidencia de filtrado de hábitat donde un grupo de especies estrechamente relacionadas a menudo comparten un rasgo que les permite persistir en un hábitat dado (Horner-Devine et al., 2006). Estas adaptaciones podrían incluir respuestas a factores de estrés ambiental, como las condiciones de oligotrofia en los sedimentos marinos. Las comunidades de sedimentos marinos se adaptan a condiciones de baja energía y son seleccionadas para sobrevivir en estas condiciones. Aunque las tasas de mutación son bajas, la recombinación podría afectar a los microorganismos de los sedimentos y provocar variaciones en su contenido genético (Orsi, 2018). Con el tiempo, esto puede resultar en el agrupamiento observado en el árbol filogenético. (Figura 20).



Figura 20. Árbol filogenético de los MAG asignados a Alphaproteobacteria, Gammaproteobacteria y Bacteroidia MAGs de suelo y sedimentos. Los círculos concéntricos muestran A) La clase taxonómica con el entorno B) La completitud (%), C) El tamaño del MAG (Mbps). Las clases de taxonomía se anotaron utilizando la base de datos GTDB v-207 (Chaumeil et al., 2020). Las líneas punteadas delimitan los clusters de los MAGs de sedimentos marinos en el Ancestro Común Más Bajo con el que se agruparon (etiquetas de familia u orden). Los valores de bootstrap en las ramas están coloreados.

Para ver cuán diferentes eran estos MAG en términos de repertorio de CAZymes, el análisis de PCoA de los recuentos de todos los módulos de CAZymes encontrados en cada MAG mostró que la composición de CAZymes parecía ser similar entre los filos, donde Alfa y Gamma Proteobacteria se agruparon juntos y lo mismo ocurrió con los MAG de Bacteroidia (explicando el 29.74% de la varianza en CoA1 y CoA2) (Figura 21).

Esto podría explicar por qué los MAG de sedimentos marinos poseen un perfil de CAZymes menos diverso en comparación con los MAG de suelo.



Figura 21. Análisis de coordenadas principales (PCoA) basado en CAZymes identificadas dentro del MAG de suelo y sedimento. La aparición del econsistema y la taxonomía de cada MAG está codificada por colores. El número de MAG está entre paréntesis. Los recuentos de los módulos CAZyme se normalizaron a porcentajes para construir una matriz disimilatoria de Bray-Curtis. b) Diagrama de caja que muestra el recuento total de genes CAZy por MAG (abundancia de módulos CAZyme) dentro del tipo de clase y hábitat (suelo versus sedimento).

La principal diferencia entre las clases recuperadas de los MAG de sedimentos en comparación con los MAG de suelo fue el número de módulos de CAZymes encontrados entre ellos y la diversidad de los módulos de CAZymes: todas las clases de MAG de suelo tenían un número total de módulos mayor (Figura 22) y más diverso (Figura 23) en comparación con los que recuperamos de sedimentos marinos, donde la clase Bacteroidia era la que tenía más recuentos y módulos de CAZymes más diversos. Esto es consistente con estudios de MAGs en ambientes donde los módulos de CAZymes están filogenéticamente conservados entre los filos microbianos, pero hay cierta especificidad hacia el hábitat donde el suelo es un ecosistema en el que se ha encontrado riqueza y diversidad en módulos de CAZymes en contraste con ambientes marinos como los sedimentos marinos (López-Mondéjar et al., 2022). Además, el filo Bacteroidetes al que pertenece la clase Bacteroidia ha sido reportado como la clase principal para la transformación de carbohidratos, ya que posee un gran repertorio de familias CAZymes (Lapébie et al., 2019).



Figura 22. Diagrama de caja que muestra el recuento total de genes CAZyme por MAG (abundancia de módulos CAZyme) dentro del tipo de clase y ecosistema (suelo versus sedimento) Se consideraron todas las clases CAZymes de la clasificación de la base de datos CAZy (Lombard et al., 2013). Los diagramas de caja muestran los valores medianos y los cuartiles inferior y superior. Valores significativos (p -value <0.01)



Figura 23. Diagrama de caja de la diversidad funcional de CAZyme (número de módulos CAZyme por MAG) dentro del tipo de clase y ecosistema (suelo versus sedimento). Se consideraron todas las clases CAZymes de la clasificación de la base de datos CAZy (Lombard et al., 2013). Los diagramas de caja muestran los valores medianos y los cuartiles inferior y superior. Valores significativos (p -value <0.01)

Esta comparación entre los MAGs de sedimentos marinos y los genomas ensamblados de metagenomas de suelo de Alphaproteobacteria, Gammaproteobacteria y Bacteroidia revela diferencias interesantes que resaltan los contrastantes roles ecológicos y presiones ambientales que estas bacterias experimentan en sus respectivos hábitats. El mayor número y diversidad de módulos de CAZymes encontrados en los MAGs de suelo en comparación con los MAGs de sedimentos marinos respaldan la idea de que las comunidades microbianas del suelo están expuestas a una variedad más amplia de sustratos orgánicos, incluyendo biomasa vegetal, detritos animales y materia orgánica del suelo compleja. Esta diversidad de sustratos probablemente impulsa la necesidad de un conjunto más amplio de capacidades enzimáticas en los microorganismos del suelo, como se refleja en su repertorio de CAZymes. Por el contrario, los ambientes de sedimentos marinos pueden ser más homogéneos en términos de disponibilidad de sustrato orgánico, posiblemente debido a la

predominancia de materia orgánica de origen marino, como el fitoplancton y otros organismos marinos.

Otra posible explicación podría ser que el ambiente de sedimentos marinos es más limitado en energía en comparación con el suelo, lo que lleva a una presión selectiva para organismos que pueden degradar eficientemente la materia orgánica disponible con un conjunto más pequeño de enzimas. Esto podría llevar a un perfil de CAZymes más simplificado en las bacterias de sedimentos marinos. Como se describe en la Hipótesis de la Reina Negra, la dependencia de bienes comunes en las poblaciones que viven en limitación de nutrimentos extrema puede llevar a la reducción del genoma. En la biosfera profunda famélica, los microorganismos que usan bienes comunes de manera eficiente pueden tener una ventaja evolutiva ya que estos guardan energía. La necromasa de las células que murieron más rápido que las células que sobreviven en el subsuelo es la mejor fuente de materia orgánica reactiva y bienes comunes. Los genomas reducidos de los microorganismos del subsuelo y la extremada limitación de energía de la biosfera profunda, sugiere que los mecanismos de la Reina Negra contribuyen a la evolución de estos microorganismos, de manera similar a los microorganismos planctónicos que también viven bajo condiciones oligotróficas extremas. Es posible que los grupos cosmopolita del subsuelo experimenten reducción genómica del tipo Reina Negra en ambientes superficiales poco profundos y así pueden tener sus genomas previamente reducidos antes de ser enterrados (Orsi 2018).

A pesar de estas diferencias, el hecho de que Alphaproteobacteria, Gammaproteobacteria y Bacteroidia de suelo y sedimentos marinos se agrupen juntos en el análisis de PCoA sugiere un conjunto central de módulos de CAZymes que se conservan dentro de estos grupos taxonómicos, reflejando probablemente historias evolutivas compartidas y funciones metabólicas centrales.

Este estudio subraya la importancia de considerar el contexto ecológico al estudiar las capacidades funcionales de las comunidades microbianas. Las marcadas diferencias en los perfiles de CAZymes entre las bacterias de suelo y sedimentos marinos resaltan cómo los factores ambientales pueden dar forma al potencial funcional de las comunidades microbianas. Por lo tanto, es esencial tener en cuenta estos factores al estudiar la ecología y función de los microorganismos en diferentes ambientes.

### 8 Conclusiones

En este estudio, clasificamos 37 metagenomas de diferentes lugares del mundo con pocos metadatos fisicoquímicos al comparar la capacidad de la comunidad para utilizar el oxígeno como último aceptor de electrones, usándolo como un marcador para clasificarlos en óxicos o anóxicos según el metabolismo predominante. Encontramos una clara diferencia entre nuestras muestras de sedimentos en términos de taxonomía y contenido de CAZymes en el contexto de esta clasificación.

Para tener una mejor idea de qué polisacáridos se encuentran prevalecientes en el sedimento marino, establecimos un perfil de las CAZymes secretadas más abundantes en nuestras muestras, donde 18 módulos de CAZymes eran abundantes en todas y se encontró que se dirigían principalmente a carbohidratos degradados de necromasa y detrito algal, lo que es coherente con las condiciones ambientales encontradas en los sedimentos. La mayoría de las principales familias de CAZymes que encontramos abundantes eran de origen bacteriano.

Finalmente, logramos recuperar MAGs (Genomas Ensamblados Metagenómicos) de las muestras, que se asignaron a las clases Alphaproteobacteria, Gammaproteobacteria y Bacteroidia. Los MAGs contenían módulos extracelulares de las principales CAZymes anotadas en nuestros metagenomas, así como CGCs (Clústeres de Genes de CAZyme) que apuntaban a esas familias, y se encontró que la familia GH23 estaba presente en casi todos nuestros MAGs y se dirigía a sustratos de peptidoglicano y quitina. Otros MAG que no contenían familias GH23 mostraban otros de los principales módulos que se dirigían a glicanos hospedadores y detritos de plantas. Estos taxones podrían ser las principales bacterias que impulsan la transformación de carbohidratos en sedimentos marinos, pero se necesitan más estudios para respaldar esto. Nuestros hallazgos proporcionan información valiosa sobre la estructura y función de la transformación de carbohidratos en sedimentos marinos, enfatizando los roles clave que desempeñan clases bacterianas específicas y sus inventarios de CAZymes y CGCs asociados.

Es importante señalar que muchos de los MAG que reconstruimos pertenecían a taxones que ya se habían encontrado en ambientes marinos y que las clases Bacteroidia, Alpha y Gammaproteobacteria estaban presentes en nuestros metagenomas en abundancia. Estos taxones se compararon con otros MAGs de muestras de suelo, y observamos un perfil similar, pero con menos módulos de CAZymes totales y diversos en nuestros MAG de sedimentos. Esto podría deberse a las condiciones oligotróficas a las que están sometidas estas especies en los sedimentos marinos, en contraste con las condiciones del suelo. Aunque los MAGs que reconstruimos a partir de nuestras muestras nos dan una idea de la comunidad microbiana, el número de MAG recuperados no es suficiente para abarcar la gran diversidad de la comunidad microbiana de los sedimentos marinos. Además, se necesitan más estudios en sedimentos subsuperficiales profundos para comprender mejor el inventario de CAZymes en comparación con los sedimentos marinos superficiales.

## 9 Perspectivas

- Hacer el análisis de las CAZymes para muestras que tengan mayores profundidades a nivel mbsf.
- Hacer una anotación de genes marcadores de procesos biogeoquímicos para hacer mejores inferencias sobre las muestras.
- Hacer análisis de las CAZymes de metagenomas de muestras de sedimento vs de muestras de columna de agua de las mismas locaciones.
- Realizar análisis de CAZymes en MAGs las Alphaproteobacteria, Gammaproteobacteria y Bacteroidia de ambientes distintos al suelo.

## **10 Referencias bibliográficas**

1. Alain, K., Intertaglia, L., Catala, P., and Lebaron, P. (2008). Eudoraea adriatica gen. nov., sp. nov., a novel marine bacterium of the family Flavobacteriaceae. Int. J. Syst. Evol. Microbiol. 58, 2275–2281. doi:10.1099/ijs.0.65446-0.

2. Albertson NH, Nystrom T, Kjelleberg S. 1990. Exoprotease activity of two marine bacteria during starvation. Appl. Environ. Microbiol. 56:218–23

3. Alderkamp A-C, van Rijssel M, Bolhuis H. 2007. Characterization of marine bacteria and the activity of their enzyme systems involved in degradation of the algal storage glucan laminarin. FEMS Microbiol. Ecol.59:108–17

4. Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37, 420–423. doi:10.1038/s41587-019-0036z.

5. Andrade, A. C. et al. Diversity of Microbial Carbohydrate-Active enZYmes (CAZYmes) Associated with Freshwater and Soil Samples from Caatinga Biome. Microb. Ecol. 74, 89–105 (2017).

6. André, I., Potocki-Véronèse, G., Barbe, S., Moulis, C., and Remaud-Siméon, M. (2014). CAZyme discovery and design for sweet dreams. Curr. Opin. Chem. Biol. 19, 17–24. doi:10.1016/j.cbpa.2013.11.014.

Andrews, S. FastQC: a quality control tool for high throughput sequence data.
 (2010). Available online at: <u>http://www.bioinformatics.babraham.ac.uk/projects/fastqc</u>.

8. Antranikian G, Herzberg C, Gottschalk G. 1987. Production of thermostable  $\alpha$ amylase, pullulanase, and  $\alpha$ -glucosidase in continuous culture by a new Clostridium isolate. Appl. Environ. Microbiol. 53:1668–73

 Arndt, S., Jørgensen, B. B., LaRowe, D. E., Middelburg, J. J., Pancost, R. D., and Regnier, P. (2013). Quantifying the degradation of organic matter in marine sediments: A review and synthesis. Earth-Science Rev. 123, 53–86. doi:10.1016/j.earscirev.2013.02.008.
 Arnosti C. (2000). Substrate specificity in polysaccharide hydrolysis: contrasts between bottom water and sediments. Limnol. Oceanogr. 45:1112–19

11. Arnosti C. (2008). Functional differences between Arctic sedimentary and seawater microbial communities:contrasts in microbial hydrolysis of complex substrates. FEMS Microbiol. Ecol. 66:343–51

12. Arnosti, C. (2011). Microbial extracellular enzymes and the marine carbon cycle. Ann. Rev. Mar. Sci. 3, 401–425. doi:10.1146/annurev-marine-120709-142731.

13. Arnosti, C. (2014). Patterns of Microbially Driven Carbon Cycling in the Ocean: Links between Extracellular Enzymes and Microbial Communities. Adv. Oceanogr. 2014, 1–12. doi:10.1155/2014/706082.

14. Asnicar, F., Thomas, A. M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M., Cumbo, F., May, U., Sanders, J.G., Zolfo, M., Kopylova, E., Pasolli, E., Knight, R., Mirarab, S., Huttenhower, C., & Segata, N. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nature Communications, 11(1), 1–10. <u>https://doi.org/10.1038/s41467-020-16366-7</u>

15. Bäckström D et al., "Virus Genomes from Deep Sea Sediments Expand the Ocean Megavirome and Support Independent Origins of Viral Gigantism.", mBio, 2019 Mar 5;10(2).

16. Bagi, A., and Skogerbø, G. (2022). Tracking bacterial pollution at a marine wastewater outfall site – A case study from Norway. Sci. Total Environ. 829. doi:10.1016/j.scitotenv.2022.154257.

17. Baltar F, Aristegui J, Gasol JM, Sintes E, van Aken HM, Herndl GJ. 2010. High dissolved extracellular enzymeactivity in the deep central Atlantic Ocean. Aquat. Microb. Ecol. 58:287–302

18. Bardgett, R. D., and Van Der Putten, W. H. (2014). Belowground biodiversity and ecosystem functioning. Nature 515, 505–511. doi:10.1038/nature13855.

19. Biddle, J. F., Fitz-gibbon, S., Schuster, S. C., Brenchley, J. E., and House, C. H. (2008). Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. 105, 10583–10588. doi:10.1073/pnas.0709942105.

20. Bienhold, C., Zinger, L., Boetius, A., and Ramette, A. (2016). Diversity and biogeography of bathyal and abyssal seafloor bacteria. PLoS One 11, 1–20. doi:10.1371/journal.pone.0148016.

21. Biersmith, A., and Benner, R. (1998). Carbohydrates in phytoplankton and freshly produced dissolved organic matter. Mar. Chem. 63, 131–144. doi:10.1016/S0304-4203(98)00057-7.

22. Boetius A, Damm E. 1998. Benthic oxygen uptake, hydrolytic potentials and microbial biomas at the Arctic continental slope. Deep-Sea Res. I 45:239–75

23. Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

doi:10.1093/bioinformatics/btu170.

24. Bradley, J. A., Amend, J. P., and Larowe, D. E. (2018). for Microorganisms in Marine Sediments. 577–590.

25. Briggs, N., Dall'Olmo, G., and Claustre, H. (2020). Major role of particle fragmentation in regulating biological sequestration of CO2 by the oceans. Science (80-.). 367, 791–793. doi:10.1126/science.aay1790.

26. Brown MR (1991) The amino-acid and sugar composition of 16 species of microalgae used in mariculture. J Exp Mar Bio Ecol 145, 79–99. https:// doi. org/ 10. 1016/ 0022- 0981(91) 90007-J

27. Buchfink, B., Reuter, K., & Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature Methods, 18(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

Burdige DJ. 2006. Geochemistry of Marine Sediments. Princeton, NJ: Princeton

Univ. Press. 609 pp.

29. Cantarel, B. I., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. 37, 233–238. doi:10.1093/nar/gkn663.

30. Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics, 25(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

31. Cevallos, M. A., and Degli Esposti, M. (2022). New Alphaproteobacteria Thrive in the Depths of the Ocean with Oxygen Gradient. Microorganisms 10. doi:10.3390/microorganisms10020455.

32. Chen, K. & Pachter, L. Bioinformatics for whole-genome shotgun sequencing of microbial communities. PLoS Comput. Biol. 1, 0106–0112 (2005).

33. Chiciudean, I., Russo, G., Bogdan, D. F., Levei, E. A., Faur, L., Hillebrand-Voiculescu, A., et al. (2022). Competition-cooperation in the chemoautotrophic ecosystem of Movile Cave: first metagenomic approach on sediments. Environ. Microbiomes 17, 1–18. doi:10.1186/s40793-022-00438-w.

34. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36, 1925–1927. doi:10.1093/bioinformatics/btz848.

35. De Filippis, F., Parente, E. & Ercolini, D. Metagenomics insights into food fermentations. Microb. Biotechnol. 10, 91–102 (2017).

36. DeVries, T., and Weber, T. (2017). The export and fate of organic matter in the ocean: New constraints from combining satellite and oceanographic tracer observations. Global Biogeochem. Cycles 31, 535–555. doi:10.1002/2016GB005551.

37. D'Hondt, S. D., Jørgensen, B. B., Miller, D. J., Batzke, A., Blake, R., Cragg, B. A., et al. (2004). Esearch rticle. 2216–2222.

38. D'Hondt S. Global distribution of microbial abundance and biomass in subseafloor sediment. Proc Natl Acad Sci U S A. 2012;109(40):16213-16216. doi:10.1073/pnas.1203849109.

39. D'Hondt, S., Inagaki, F., Zarikian, C. A., Abrams, L. J., Dubois, N., Engelhardt, T., et al. (2015). Presence of oxygen and aerobic communities from sea floor to basement in deep-sea sediments. Nat. Geosci. 8, 299–304. doi:10.1038/ngeo2387.

40. Dedysh, S. N., Kulichevskaya, I. S., Beletsky, A. V., Ivanova, A. A., Rijpstra, W. I. C., Damsté, J. S. S., et al. (2020). Lacipirellula parvula gen. nov., sp. nov., representing a lineage of planctomycetes widespread in low-oxygen habitats, description of the family Lacipirellulaceae fam. nov. and proposal of the orders Pirellulales ord. nov., Gemmatales ord. nov. and Isosphaerales ord. nov. Syst. Appl. Microbiol. 43, 126050.

doi:10.1016/j.syapm.2019.126050.

41. Dombrowski, N., Seitz, K. W., Teske, A. P., and Baker, B. J. (2017). Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. Microbiome 5, 106. doi:10.1186/s40168-017-0322-2.

42. Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022). The carbohydrate-active enzyme database: Functions and literature. Nucleic Acids Res. 50, D571–D577. doi:10.1093/nar/gkab1045.

43. Dyksma, S., Bischof, K., Fuchs, B. M., Hoffmann, K., Meier, D., Meyerdierks, A., et al. (2016). Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments. ISME J. 10, 1939–1953. doi:10.1038/ismej.2015.257.

44. Eddy, S. R. (2011). Accelerated Profile HMM Searches. PLOS Computational Biology, 7(10), e1002195. doi:10.1371/journal.pcbi.1002195

45. Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics, 5, 1–19. https://doi.org/10.1186/1471-2105-5-113

46. Escobar-Briones, E., and García-Villalobos, F. J. (2009). Distribution of total organic carbon and total nitrogen in deep-sea sediments from the southwestern Gulf of Mexico. Bol. la Soc. Geol. Mex. 61, 73–86. doi:10.18268/BSGM2009v61n1a7.

47. Escobar-Zepeda, A., De León, A. V. P., and Sanchez-Flores, A. (2015). The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. Front. Genet. 6, 1–15. doi:10.3389/fgene.2015.00348.

48. Flemming, H. C., and Wuertz, S. (2019). Bacteria and archaea on Earth and their abundance in biofilms. Nat. Rev. Microbiol. 17, 247–260. doi:10.1038/s41579-019-0158-9.
49. Frieden, Earl. "The Chemical Elements of Life." Scientific American, July 1972, pp. 52-60.

50. Froelich, P. N., Klinkhammer, G. P., Bender, M. L., Luedtke, N. A., Heath, G. R., Cullen, D., et al. (1979). Early oxidation of organic matter in pelagic sediments of the eastern equatorial Atlantic: suboxic diagenesis. Geochim. Cosmochim. Acta 43, 1075–1090. doi:10.1016/0016-7037(79)90095-4.

51. Gu, Z. (2022). Complex heatmap visualization. iMeta 1, 1–15. doi:10.1002/imt2.43.
52. Godoy-Lozano, E. E. et al. Bacterial diversity and the geochemical landscape in the southwestern Gulf of Mexico. Front. Microbiol. 9, 1–15 (2018).

53. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W., and Brumer, H. (2017). Polysaccharide utilization loci: Fueling microbial communities. J. Bacteriol. 199. doi:10.1128/JB.00860-16.

54. Handelsman, J., Tiedje, J., National Research Council (US) Committee on Metagenomics: Challenges and Functional & Applications. THE NEW SCIENCE OF METAGENOMICS: Revealing the Secrets of Our Microbial Planet. National Academies Press (2007). doi:NBK54006 [bookaccession].

55. Hijmans, R. J. (2020). Geosphere: Spherical Trigonometry [R package]. Available at: https://CRAN.R-project.org/package=geosphere.

56. Hoffmann, K., Bienhold, C., Buttigieg, P. L., Knittel, K., Laso-Pérez, R., Rapp, J. Z., et al. (2020). Diversity and metabolism of Woeseiales bacteria, global members of marine sediment communities. ISME J. 14, 1042–1056. doi:10.1038/s41396-020-0588-4.

57. Horner-Devine, M. C., and Bohannan, B. J. M. (2006). Phylogenetic clustering and overdispersion in bacterial communities. Ecology 87, 100–108. doi:10.1890/0012-9658(2006)87[100:pcaoib]2.0.co;2.

58. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11. doi:10.1186/1471-2105-11-119.

59. Hoshino, T., Doi, H., Uramoto, G. I., Wörmer, L., Adhikari, R. R., Xiao, N., et al. (2020). Global diversity of microbial communities in marine sediment. Proc. Natl. Acad. Sci. U. S. A. 117, 27587–27597. doi:10.1073/pnas.1919139117.

60. Huang, L., Zhang, H., Wu, P., Entwistle, S., Li, X., Yohe, T., et al. (2018). DbCANseq: A database of carbohydrate-active enzyme (CAZyme) sequence and annotation. Nucleic Acids Res. 46, D516–D521. doi:10.1093/nar/gkx894.

61. Jaeschke, A., Jørgensen, S. L., Bernasconi, S. M., Pedersen, R. B., Thorseth, I. H., and Früh-Green, G. L. (2012). Microbial diversity of Loki's Castle black smokers at the Arctic Mid-Ocean Ridge. Geobiology 10, 548–561. doi:10.1111/gbi.12009.

62. Jiao, N., Herndl, G. J., Hansell, D. A., Benner, R., Kattner, G., Wilhelm, S. W., et al. (2010). Microbial production of recalcitrant dissolved organic matter: Long-term carbon storage in the global ocean. Nat. Rev. Microbiol. 8, 593–599.

doi:10.1038/nrmicro2386.Kallmeyer J, Pockalny R, Adhikari RR, Smith DC

63. Kappelmann, L., Krüger, K., Hehemann, J. H., Harder, J., Markert, S., Unfried, F., et al. (2019). Polysaccharide utilization loci of North Sea Flavobacteriia as basis for using

SusC/D-protein expression for predicting major phytoplankton glycans. ISME J. 13, 76–91. doi:10.1038/s41396-018-0242-6.

64. Kauffman, K. M., Hussain, F. A., Yang, J., Arevalo, P., Brown, J. M., Chang, W. K., et al. (2018). A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. Nature 554, 118–122. doi:10.1038/nature25474.

65. Kerkhof Lee J., Goodman Robert M.,Ocean microbial metagenomics, Deep Sea Research Part II: Topical Studies in Oceanography, Volume 56, Issues 19–20, 2009, Pages 1824-1829, ISSN 0967-0645, https://doi.org/10.1016/j.dsr2.2009.05.005.

66. Kultima, J. R., Sunagawa, S., Li, J., Chen, W., Chen, H., Mende, D. R., et al. (2012). MOCAT: A Metagenomics Assembly and Gene Prediction Toolkit. PLoS One 7, 1–6. doi:10.1371/journal.pone.0047656.

67. Kunath, B.J., Bremges, A., Weimann, A., McHardy, A.C., Pope, P.B. (2017). Metagenomics and CAZyme Discovery. In: Abbott, D., Lammerts van Bueren, A. (eds) Protein-Carbohydrate Interactions. Methods in Molecular Biology, vol 1588. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-6899-2\_20

68. Lapébie, P., Lombard, V., Drula, E., Terrapon, N., and Henrissat, B. (2019). Bacteroidetes use thousands of enzyme combinations to break down glycans. Nat. Commun. 10, 2043. doi:10.1038/s41467-019-10068-5.

69. Lee, C., Hedges, J. I., Baldock, J. A., Ge, Y., Gelinas, Y., Peterson, M., et al. (2001). Evidence for non-selective preservation of organic matter in sinking marine particles. Nature 409, 801–804.

70. Leeming, R., Stark, J., & Smith, J. (2015). Novel use of faecal sterols to assess human faecal contamination in Antarctica: A likelihood assessment matrix for environmental monitoring. Antarctic Science, 27(1), 31-43. doi:10.1017/S0954102014000273.

71. Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49, W293–W296. doi:10.1093/nar/gkab301.

72. Li, D., Liu, C-M., Luo, R., Sadakane, K., and Lam, T-W., (2015) MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics, doi: 10.1093/bioinformatics/btv033.

73. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(D1):490-495. doi:10.1093/nar/gkt1178. CAZy Database. (2023). Carbohydrate-Active enZYmes Database. Retrieved from http://www.cazy.org.

74. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. In Cazypedia, the Carbohydrate-Active enZYmes Encyclopedia(2022). Retrieved from http://www.cazypedia.org

75. López-Mondéjar, R., Tláskal, V., da Rocha, U. N., and Baldrian, P. (2022). Global Distribution of Carbohydrate Utilization Potential in the Prokaryotic Tree of Life. mSystems 7. doi:10.1128/msystems.00829-22.

76. López-Sánchez, R., Rebollar, E. A., Gutiérrez-Ríos, R. M., Garciarrubio, A., Juarez, K., and Segovia, L. (2024). Metagenomic analysis of carbohydrate-active enzymes and their contribution to marine sediment biodiversity. World J. Microbiol. Biotechnol. 40, 1–15. doi:10.1007/s11274-024-03884-5.

77. Madigan et al. "Brock: Biología de los microorganismos." 13¬™ Edición. Editorial Pearson/Prentice Hall. 2010.

78. Matsuyama, H., Minami, H., Sakaki, T., Kasahara, H., Watanabe, A., Onoda, T., Hirota, K., & Yumoto, I. (2015). Psychrobacter oceani sp. nov., isolated from marine sediment. International journal of systematic and evolutionary microbiology, 65(Pt 5), 1450–1455. https://doi.org/10.1099/ijs.0.000118

79. Meyer-Reil L-A. 1986. Measurement of hydrolytic activity and incorporation of dissolved organic substrates by microorganisms in marine sediments. Mar. Ecol. Prog. Ser. 31:143–49

80. Moran MA, Kujawinski EB, Stubbins A, et al. Deciphering Ocean carbon in a changing world. Proc Natl Acad Sci U S A. 2016;113(12):3143-3151. doi:10.1073/pnas.1514645113

81. Murali, R., Pace, L. A., Sanford, R. A., Ward, L. M., Lynes, M. M., Hatzenpichler, R., et al. (2024). Diversity and evolution of nitric oxide reduction in bacteria and archaea. Proc. Natl. Acad. Sci. 121, e2316422121. doi:10.1073/pnas.2316422121.

82. Murali, R., Hemp, J., and Gennis, R. B. (2022). Evolution of quinol oxidation within the heme-copper oxidoreductase superfamily. Biochim. Biophys. Acta - Bioenerg. 1863, 148907. doi:10.1016/j.bbabio.2022.148907.

83. National Center for Biotechnology Information (2019). PubMed. In National Center for Biotechnology Information. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/
84. Nayfach, S., Rodriguez-Mueller, B., Garud, N., and Pollard, K. S. (2016). An

Nayrach, S., Rodriguez-Mideler, B., Garud, N., and Pohard, K. S. (2010). An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res. 26, 1612–1625. doi:10.1101/gr.201863.115.
Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., et al. (2021). A genomic catalog of Earth's microbiomes. Nat. Biotechnol. 39, 499–509.

doi:10.1038/s41587-020-0718-6.

86. Nealson KH. Geomicrobiology: Sediment reactions defy dogma. Nature. 2010 Feb 25;463(7284):1033-4. doi: 10.1038/4631033a. PMID: 20182504.

87. Nedashkovskaya, O. I., Kim, S. B., Han, S. K., Lysenko, A. M., Rohde, M., Rhee, M. S., et al. (2004). Maribacter gen. nov., a new member of the family Flavobacteriaceae, isolated from marine habitats, containing the species Maribacter sedimenticola sp. nov., Maribacter aquivirus sp. nov., Maribacter orientalis sp. nov. and Maribacter ulvicola sp. nov. Int. J. Syst. Evol. Microbiol. 54, 1017–1023. doi:10.1099/ijs.0.02849-0.

88. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution, 32(1), 268–274. https://doi.org/10.1093/molbev/msu300

89. Ohnuma, T., Onaga, S., Murata, K., Taira, T., and Katoh, E. (2008). LysM domains from Pteris ryukyuensis chitinase-A: A stability study and characterization of the chitinbinding site. J. Biol. Chem. 283, 5178–5187. doi:10.1074/jbc.M707156200.

90. Orsi, W. D. (2018). Ecology and evolution of seafloor and subseafloor microbial communities. Nat. Rev. Microbiol. 16, 671–683. doi:10.1038/s41579-018-0046-8.

91. Orsi, W. D., Richards, T. A., and Francis, W. R. (2018). Predicted microbial secretomes and their target substrates in marine sediment. Nat. Microbiol. 3, 32–37. doi:10.1038/s41564-017-0047-9.

92. Oksanen, F.J., et al. (2017) Vegan: Community Ecology Package. R package Version 2.4-3. https://CRAN.R-project.org/package=vegan
93. Orcutt, B. N., Sylvan, J. B., Knab, N. J. & Edwards, K. J.Microbial ecology of the dark ocean above, at, and below the seafloor. Microbiol. Mol. Biol. Rev. 75, 361–422 (2011).

94. Pante, E., and Simon-Bouhet, B. (2013). marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R. PLoS One 8, e73051. Available at: https://doi.org/10.1371/journal.pone.0073051.

95. Park, S., Yoshizawa, S., Inomata, K., Kogure, K., & Yokota, A. (2012). Halioglobus japonicus gen. nov., sp. nov. and Halioglobus pacificus sp. nov., members of the class Gammaproteobacteria isolated from seawater. International journal of systematic and evolutionary microbiology, 62(Pt 8), 1784–1789. https://doi.org/10.1099/ijs.0.031443-0

96. Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., and Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 50, D785–D794. doi:10.1093/nar/gkab776.

97. Parkes, R. J., Cragg, B., Roussel, E., Webster, G., Weightman, A., and Sass, H. (2014). A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere: Geosphere interactions. Mar. Geol. 352, 409–425. doi:10.1016/j.margeo.2014.02.009.

98. Poremba K, Hoppe H-G. 1995. Spatial variation of benthic microbial production and hydrolytic enzymatic activity down the continental slope of the Celtic Sea. Mar. Ecol. Prog. Ser. 118:237–45

99. Pereira, M. M., Santana, M., and Teixeira, M. (2001). A novel scenario for the evolution of haem-copper oxygen reductases. Biochim. Biophys. Acta - Bioenerg. 1505, 185–208. doi:10.1016/S0005-2728(01)00169-4.

100. Pradel, N., Fardeau, M. L., Tindall, B. J., and Spring, S. (2020). Anaerohalosphaera lusitana gen. nov., sp. nov., and limihaloglobus sulfuriphilus gen. nov., sp. nov., isolated from solar saltern sediments, and proposal of anaerohalosphaeraceae fam. nov. within the order sedimentisphaerales. Int. J. Syst. Evol. Microbiol. 70, 1321–1330. doi:10.1099/ijsem.0.003919.

101. Prosser, J. I., Head, I. M., and Stein, L. Y. (2014). "The Family Nitrosomonadaceae," in The Prokaryotes: Alphaproteobacteria and Betaproteobacteria, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 901–918. doi:10.1007/978-3-642-30197-1\_372.

102. Qin, H. M., Miyakawa, T., Inoue, A., Nakamura, A., Nishiyama, R., Ojima, T., et al. (2017). Laminarinase from Flavobacterium sp. reveals the structural basis of thermostability and substrate specificity. Sci. Rep. 7, 1–9. doi:10.1038/s41598-017-11542-0.

103. Qiu, X., Yu, L., Cao, X., Wu, H., Xu, G., & Tang, X. (2021). Halomonas sedimenti sp. nov., a Halotolerant Bacterium Isolated from Deep-Sea Sediment of the Southwest Indian Ocean. Current microbiology, 78(4), 1662–1669. https://doi.org/10.1007/s00284-021-02425-9

104. R Core Team (2023). R: A language and environment for statistical computing. R
Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
105. Raggi, L., García-Guevara, F., Godoy-Lozano, E. E., Martínez-Santana, A.,
Escobar-Zepeda, A., Gutierrez-Rios, R. M., et al. (2020). Metagenomic Profiling and
Microbial Metabolic Potential of Perdido Fold Belt (NW) and Campeche Knolls (SE) in the
Gulf of Mexico. Front. Microbiol. 11, 1–18. doi:10.3389/fmicb.2020.01825.

106. Rathgeber, C., Yurkova, N., Stackebrandt, E., Schumann, P., Beatty, J. T., and Yurkov, V. (2005). Roseicyclus mahoneyensis gen. nov., sp. nov., an aerobic phototrophic bacterium isolated from a meromictic lake. Int. J. Syst. Evol. Microbiol. 55, 1597–1603. doi:10.1099/ijs.0.63195-0.

107. Reed, D. W., Fujita, Y., Delwiche, M. E., Blackwelder, D. B., Sheridan, P. P., Uchida, T., et al. (2002). Microbial communities from methane hydrate-bearing deep marine sediments in a forearc basin. Appl. Environ. Microbiol. 68, 3759–3770. doi:10.1128/AEM.68.8.3759-3770.2002.

108. Seeberger, P. H. (2017). Monosaccharide Diversity. In A. Varki (Eds.) et. al.,
Essentials of Glycobiology. (3rd ed., pp. 19–30). Cold Spring Harbor Laboratory Press.
109. Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., &
Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. Genome
Biology, 12(6), R60. Doi: 10.1186/gb-2011-12-6-r60.

110. Segata N, Börnigen D, Morgan XC, Huttenhower C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun.

2013;4:2304. doi: 10.1038/ncomms3304. PMID: 23942190; PMCID: PMC3760377.111. Segata, N. et al. Computational meta'omics for microbial community studies. Mol.

Syst. Biol. 9, 666–666 (2014).

112. Sousa, F. L., Alves, R. J., Pereira-Leal, J. B., Teixeira, M., and Pereira, M. M. (2011). A bioinformatics classifier and database for Heme-Copper oxygen reductases. PLoS One 6, 1–9. doi:10.1371/journal.pone.0019117.

113. Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics, 22(21), 2688–2690. https://doi.org/10.1093/bioinformatics/btl446

114. Stryer, L., Berg, J. M., Tymoczko, J. L., & Gatto, G. J. (2012). Biochemistry (7th ed.). W.H. Freeman and Company.

115. Tamames, J., Puente-Sánchez, F., SqueezeMeta Development Team, et al. (2019). SqueezeMeta: a highly portable, fully automatic metagenomic analysis pipeline. PeerJ, 7, e7559. Doi: 10.7717/peerj.7559.

116. Taş, N., de Jong, A. E., Li, Y., Trubl, G., Xue, Y., and Dove, N. C. (2021). Metagenomic tools in microbial ecology research. Curr. Opin. Biotechnol. 67, 184–191. doi:10.1016/j.copbio.2021.01.019.

117. Teeling, H., Fuchs, B. M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.
M., et al. (2012). Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science (80-.). 336, 608–611.
Doi:10.1126/science.1218344.

118. Teeling, H., Fuchs, B. M., Bennke, C. M., Krüger, K., Chafee, M., Kappelmann, L., et al. (2016). Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. Elife 5, 1–31. doi:10.7554/eLife.11888.

119. Teske, A., and Salman, V. (2014). "The Family Beggiatoaceae," in The Prokaryotes: Gammaproteobacteria, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 93–134. doi:10.1007/978-3-642-38922-1\_290.

120. Tully, B. J., and Heidelberg, J. F. (2016). Potential mechanisms for microbial energy acquisition in oxic deep- sea sediments. Appl. Environ. Microbiol. 82, 4232–4243. doi:10.1128/AEM.01023-16.

121. Uchino, Y., Hamada, T., and Yokota, A. (2002). Proposal of Pseudorhodobacter ferrugineus gen. nov., comb. nov., for a non-photosynthetic marine bacterium, Agrobacterium ferrugineum, related to the genus Rhodobacter. J. Gen. Appl. Microbiol. 48, 309–319. doi:10.2323/jgam.48.309.

122. Vekeman, B., Kerckhof, F. M., Cremers, G., de Vos, P., Vandamme, P., Boon, N., et al. (2016). New Methyloceanibacter diversity from North Sea sediments includes methanotroph containing solely the soluble methane monooxygenase. Environ. Microbiol. 18, 4523–4536. doi:10.1111/1462-2920.13485.

123. Vetter YA, Deming JW, Jumars PA, Krieger-Brockett BB. 1998. A predictive model of bacterial foraging by means of freely released extracellular enzymes. Microb. Ecol. 36:75–92

124. Vuilleumier, S., Nadalig, T., Ul Haque, M. F., Magdelenat, G., Lajus, A., Roselli, S., et al. (2011). Complete genome sequence of the chloromethane-degrading

Hyphomicrobium sp. Strain MC1. J. Bacteriol. 193, 5035–5036. doi:10.1128/JB.05627-11. 125. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

126. Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol. 20, 1–13. doi:10.1186/s13059-019-1891-0.

127. Pricia antarctica gen. nov., sp. nov., a member of the family flavobacteriaceae, isolated from antarctic intertidal sediment. Int. J. Syst. Evol. Microbiol. 62, 2218–2223. doi:10.1099/ijs.0.037515-0.

128. Yang, C., Chowdhury, D., Zhang, Z., Cheung, W. K., Lu, A., Bian, Z., et al. (2021). A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. Comput. Struct. Biotechnol. J. 19, 6301–6314. doi:10.1016/j.csbj.2021.11.028.

129. Yoon, J. H., Kang, S. J., Lee, M. H., and Oh, T. K. (2007). Description of Sulfitobacter donghicola sp. nov., isolated from seawater of the East Sea in Korea, transfer of Staleya guttiformis Labrenz et al. 2000 to the genus Sulfitobacter as Sulfitobacter guttiformis comb. nov. and emended description of the genus Sulfitobacter. Int. J. Syst. Evol. Microbiol. 57, 1788–1792. doi:10.1099/ijs.0.65071-0.

130. Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46, W95–W101. doi:10.1093/nar/gky418.

131. Zhao, H., Wu, Y., Zhang, C., Feng, J., Xu, Z., Ding, Y., et al. (2019). Aureibaculum marinum gen. nov., sp. nov., a Novel Bacterium of the Family Flavobacteriaceae Isolated from the Bohai Gulf. Curr. Microbiol. 76, 975–981. doi:10.1007/s00284-019-01691-y.

132. Zhao, R., Summers, Z.M., Christman, G.D. et al. Metagenomic views of microbial dynamics influenced by hydrocarbon seepage in sediments of the Gulf of Mexico. Sci Rep 10, 5772 (2020). https://doi.org/10.1038/s41598-020-62840-z

133. Zhao, Z., Baltar, F., and Herndl, G. J. (2020). Linking extracellular enzymes to phylogeny indicates a predominantly particle-associated lifestyle of deep-sea prokaryotes. Sci. Adv. 6, 1–10. doi:10.1126/sciadv.aaz4354.

134. Zheng, J., Hu, B., Zhang, X., Ge, Q., Yan, Y., Akresi, J., et al. (2023). dbCAN-seq update: CAZyme gene clusters and substrates in microbiomes. Nucleic Acids Res. 51, D557–D563. doi:10.1093/nar/gkac1068.

135. www.cazy.org,

136. https://www.cazypedia.org

- 137. https://bcb.unl.edu/dbCAN\_sub/data/.
- 138. https://cran.r-project.org/web/packages/marmap/index.html
- 139. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA835145
- 140. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA526329/
- 141. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB25358
- 142. https://es.slideshare.net/slideshow/presentacin-sobre-importancia-de-la-diversidad-

biolgica-del-suelo/61714850#16

# 11 Bibliografía

## Portada del artículo producto del doctorado.

Work Journal of Microbiology and Biotechnology (2024) 40:95 https://doi.org/10.1007/s11274-024-03884-5

RESEARCH



## Metagenomic analysis of carbohydrate-active enzymes and their contribution to marine sediment biodiversity

Received: 30 June 2023 / Accepted: 2 January 2024 / Published online: 13 February 2024 @The Author(s) 2024

#### Abstract

Mar.ne sediments consitute the world's most substantial long-term carbon repository. The microorganisms dwelling in these sediments mediate the transformation of fixed oceanic carbon, but their contribution to the carbon cycle is not fully understood. Previous culture-independent investigations into sedimentary microorganisms have underscored the significance of carbohydrates in the carbon cycle. In this study, we employ a metagenomic methodology to investigate the distribution and abundance of carbohydrate-active enzymes (CAZymes) in 37 marine sediments sites. These sediments exhibit varying oxygen availability and were isolated in diverse regions worldwide. Our comparative analysis is based on the metabolic potential for oxygen utilisation, derived from genes present in both oxic and anoxic environments. We found that extracellular CAZyme modules targeting the degradation of plant and algal detritus, necromass, and host glycans were abundant across all metagenomic samples. The analysis of these results indicates that the exic/anoxic conditions not only influence the taxonomic composition of the microbial communities, but also affect the occurrence of CAZyme modules involved in the transformation of necromass, algae and plant detritus. To gain insight into the sediment microbial axa, we reconstructed metagenome assembled genomes (MAG) and examined the presence of primary extracellular carbohydrate active enzyme (CAZyme) modules. Our findings reveal that the primary CAZyme modules and the CAZyme gene clusters discovered in our metagenomes were prevalent in the Bacteroidia, Gammaproteobacteria, and Alphaproteobacteria classes. We compared those MAGs to organisms from the same taxonomic classes found in soil, and we found that they were similar in its CAZyme repertoire, but the soil MAG contained a more abundant and diverse CAZyme content. Furthermore, the data indicate that aburdant classes in our metagenomic samples, namely Alphaproteobacteria, Bacteroidia and Gammaproteobacteria, play a pivo:al role in carbohydrate transformation within the initial few metres of the sediments.

GT

CGC

MAG

CAZymes

Keywords Anoxic · Bioinformatics · CAZymes · Oxic · Marine sediments · Metagenomics

#### Abbreviations

AA	Auxiliary activities
CE	Carbohydrate esterase
CBM	Carbohydrate binding module
GH	Glycoside hydrolase

🖾 Lorenzo Segovia

lerenzo.segovia@ibt.unam.mx

Departamento de Ingeniería Celular y Biocatáisis, Instituto de Biotecnología, Universidad Nacional Autóroma de México, Cuernavaca, Morelos, Mexico

<sup>2</sup> Centro de Ciencias Genômicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

<sup>3</sup> Departamento de Microsiología Molecular, Instituto de Eintecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico Mbsl Meters below sea level Mbsf Meters below sea floor PL Polysaccharide lyase

Glycosyltransferase

CAZ/me gene cluster

Carbohydrate-active enzymes

Metagenomic assembly genomes

### Introduction

The ocean floor is the recipient of all the organic matter coming from the water column and is considered the major carbon repository on the planet. Therefore, microorganisms that live in marine sediments control the storage of massive amounts of carbon (Orcutt et al.

2 Springer

Tabla Suplementaria 1. Taxonomía asignada a los MAGs encontrados en sedimentos marinos.

Identificador	Taxonomía
ART45_max bin.004	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Arcobacteraceae;g_NORP36;s_
ART45_max bin.008	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Arcobacteraceae;g_NORP36;s_
ART45_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Burkholderia
bin.009	les;f_Nitrosomonadaceae;g_GCA-2721545;s_
ART45_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_UBA4486;f_
bin.010	_UBA4486;g_SMWN01;s_
ART45_max	d_Bacteria;p_Desulfobacterota;c_Desulfobulbia;o_Desulfobulbales;f
bin.050	_BM004;g_BM004;s_BM004 sp012974315
ART45_met	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_UBA4575;f_
abat2.83	_UBA4575;g_JABDMD01;s_
ART46_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_UBA4486;f_
bin.016	_UBA4486;g_SMWN01;s_
ART46_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Burkholderia
bin.023	les;f_Nitrosomonadaceae;g_GCA-2721545;s_
ART46_max bin.025	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomona dales;f_Halieaceae;g_Halioglobus;s_
ART46_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_HK1;f_HK
bin.029	1;g_JAABYT01;s_
ART46_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_UBA6522;f_
bin.038	_UBA6522;g_SMWC01;s_
ART46_max bin.044	d_Bacteria;p_Acidobacteriota;c_UBA6911;o_RPQK01;f_;g_;s_
ART46_max	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA5794;f_UB
bin.050	A5794;g_;s_
ART46_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_UBA9214;f_
bin.051	_UBA9214;g_UBA9214;s_
ART46_max	d_Bacteria;p_Myxococcota_A;c_UBA9160;o_UBA9160;f_UBA44
bin.055	27;g_UBA4427;s_
ART46_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_UBA4575;f_
bin.071	_UBA4575;g_JABDMD01;s_

ART46_max	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_Acidimicrobiales;
bin.076	f_Ilumatobacteraceae;g_;s_
ART46_met abat2.52	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomona dales;f_HTCC2089;g_;s_
ART47_max bin.011	d_Bacteria;p_Acidobacteriota;c_UBA6911;o_RPQK01;f_;g_;s_
ART47_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_HK1;f_HK
bin.012	1;g_JAABYT01;s_
ART47_max bin.017	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Burkholderia les;f_Nitrosomonadaceae;g_GCA-2721545;s_GCA-2721545 sp003230615
ART47_max	d_Bacteria;p_Nitrospirota;c_Nitrospiria;o_Nitrospirales;f_UBA863
bin.029	9;g_UBA8639;s
ART47_max bin.033	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomona dales;f_HTCC2089;g_;s_
ART47_max	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA5794;f_ZC
bin.046	4RG35;g_SZUA-217;s_
ART47_max	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_Anaerolineales;f_E4
bin.052	4-bin32;g_W1033;s_
ART47_max bin.062	d_Bacteria;p_;c_;o_;f_;g_;s_
ART47_met	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_UBA4575;f_
abat2.6	_UBA4575;g_JABDMD01;s_
ART47_met	d_Bacteria;p_Desulfobacterota_D;c_UBA1144;o_UBA2774;f_UBA
abat2.85	2774;g_CR02bin9;s_CR02bin9 sp004356555
ART48_max bin.021	d_Bacteria;p_Acidobacteriota;c_UBA6911;o_RPQK01;f_;g_;s_
ART48_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacter
bin.026	ales;f_Alteromonadaceae;g_Thalassomonas;s_
ART49_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Burkholderia
bin.008	les;f_Nitrosomonadaceae;g_GCA-2721545;s_
ART49_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Arenicellales
bin.019	;f_RKSH01;g_;s_
ART50_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Burkholderia
bin.006	les;f_Nitrosomonadaceae;g_GCA-2721545;s_
ART50_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomona
bin.015	dales;f_Halieaceae;g_Halioglobus;s_

ART50_max bin.018	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_GCA-2729495;f_GCA-2729495;g_;s_
ART50_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Coxiellales;f
bin.109	_Coxiellaceae;g_2-12-FULL-42-15;s_
ART50_met	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_UBA4575;f_
abat2.39	_UBA4575;g_JABDMD01;s_
CIGOMA4_	d_Bacteria;p_Methylomirabilota;c_Methylomirabilia;o_Methylomira
maxbin.004	bilales;f_CSP1-5;g_CSP1-5;s_CSP1-5 sp012974305
CIGOMA4_ maxbin.009	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_GCA-2729495;f_GCA-2729495;g_;s_
CIGOMD18	d_Bacteria;p_Desulfobacterota;c_Syntrophobacteria;o_BM002;f_B
_maxbin.001	M002;g_BM002;s_
CIGOMD18	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_AKS1;f_A
_maxbin.003	KS1;g_Sulfuriflexus;s_
CIGOMD18	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Woeseiales;f
_maxbin.031	Woeseiaceae;g;s
DELGOM54	d_Bacteria;p_Methylomirabilota;c_Methylomirabilia;o_Methylomira
_maxbin.001	bilales;f_CSP1-5;g_CSP1-5;s_CSP1-5 sp012974305
DELGOM57	d_Bacteria;p_Desulfobacterota;c_Syntrophobacteria;o_BM002;f_B
_maxbin.001	M002;g_BM002;s_
DELGOM57 _maxbin.002	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_PS1;f_Thio globaceae;g_Ruthia;s_
DELGOM57	d_Bacteria;p_Methylomirabilota;c_Methylomirabilia;o_Methylomira
_maxbin.003	bilales;f_CSP1-5;g_CSP1-5;s_CSP1-5 sp012974305
DELGOM59 _maxbin.007	d_Bacteria;p_VGIX01;c_VGIX01;o_VGIX01;f_VGIX01;g_;s_
DSANT06_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.002	obacteriaceae;g_Maribacter_A;s
DSANT06_ maxbin.003	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Sulfurovaceae;g_Sulfurovum;s_
DSANT06_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.006	obacteriaceae;g_Aureibaculum;s_
DSANT06_ maxbin.014	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Sulfurovaceae;g_Sulfurovum;s_
DSANT06_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.016	obacteriaceae;g_Maribacter;s_

DSANT06_	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA5794;f_UB
maxbin.019	A5794;g_;s_
DSANT06_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Woeseiales;f
maxbin.020	_Woeseiaceae;g_JAACFB01;s_
DSANT06_ maxbin.024	d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhodobacteral es;f_Rhodobacteraceae;g_Sulfitobacter;s_
DSANT06_	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA5794;f_UB
maxbin.029	A5794;g_;s_
DSANT06_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
maxbin.037	s;f_UBA11574;g_;s_
DSANT06_	d_Bacteria;p_Gemmatimonadota;c_Gemmatimonadetes;o_Longimicr
maxbin.040	obiales;f_UBA6960;g_JABDNZ01;s_
DSANT06_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Xanthomona
maxbin.045	dales;f_SZUA-36;g_SZUA-36;s_
DSANT06_	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_LZORAL
maxbin.071	124-64-63;f_LZORAL124-64-63;g_LZORAL124-64-63;s_
DSANT06_	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_Krumholz
maxbin.074	ibacteriales;f_Krumholzibacteriaceae;g_;s_
DSANT08_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.002	obacteriaceae;g_Maribacter_A;s_
DSANT08_	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_Anaerolineales;f_E4
maxbin.006	4-bin32;g_E44-bin32;s_
DSANT08_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Woeseiales;f
maxbin.007	_Woeseiaceae;g_UBA1847;s_
DSANT08_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.008	obacteriaceae;g_Aureibaculum;s_
DSANT08_ maxbin.010	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_SZUA-229;f_SZUA-229;g_GCA-2746365;s_
DSANT08_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Woeseiales;f
maxbin.017	Woeseiaceae;g_SZUA-117;s_
DSANT08_	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA5794;f_UB
maxbin.018	A5794;g_;s_
DSANT08_	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_LZORAL
maxbin.019	124-64-63;f_LZORAL124-64-63;g_LZORAL124-64-63;s_
DSANT08_ maxbin.020	d_Bacteria;p_Myxococcota;c_Polyangia;o_Polyangiales;f_SG8-38;g_SG8-38;s_

DSANT08_	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_Krumholz
maxbin.030	ibacteriales;f_Krumholzibacteriaceae;g_;s_
DSANT08_ maxbin.032	d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhodobacteral es;f_Rhodobacteraceae;g_JL08;s_
DSANT08_ maxbin.035	d_Bacteria;p_Fermentibacterota;c_Fermentibacteria;o_Fermentibacterateratera;s_Fermentibacteraceae;g_Aegiribacteria;s_
DSANT08_ maxbin.037	d_Bacteria;p_Atribacterota;c_JS1;o_SB-45;f_34-128;g_34-128;s_
DSANT08_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.038	obacteriaceae;g_QNYL01;s_
DSANT08_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.067	obacteriaceae;g_;s_
DSANT08_	d_Bacteria;p_Desulfobacterota;c_DSWW01;o_DSWW01;f_DSWW
maxbin.069	01;g_S015-6;s_
DSANT09_	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA5794;f_UB
maxbin.013	A5794;g_;s_
DSANT09_	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA5794;f_UB
maxbin.028	A5794;g_;s_
DSANT09_ maxbin.031	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Xanthomona dales;f_SZUA-36;g_SZUA-36;s_
DSANT09_	d_Bacteria;p_Desulfobacterota;c_Desulfobulbia;o_Desulfobulbales;f
maxbin.042	_Desulfocapsaceae;g_;s_
DSANT09_	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_LZORAL
maxbin.069	124-64-63;f_LZORAL124-64-63;g_LZORAL124-64-63;s_
DSANT09_	d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhodobacteral
maxbin.070	es;f_Rhodobacteraceae;g_Roseicyclus;s_
DSANT11_ maxbin.001	d_Bacteria;p_Myxococcota;c_Polyangia;o_Polyangiales;f_SG8-38;g_SG8-38;s_
DSANT11_	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA5794;f_UB
maxbin.005	A5794;g_;s_
DSANT11_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Woeseiales;f
maxbin.012	Woeseiaceae;g_UBA1847;s
DSANT11_	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_LZORAL
maxbin.015	124-64-63;f_LZORAL124-64-63;g_LZORAL124-64-63;s_
DSANT11_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.017	obacteriaceae;g_Eudoraea;s_

DSANT11_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Chromatiales
maxbin.018	;f_Chromatiaceae;g_;s_
DSANT11_	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_Krumholz
maxbin.019	ibacteriales;f_Krumholzibacteriaceae;g_;s_
DSANT11_	d_Bacteria;p_Desulfobacterota;c_DSWW01;o_DSWW01;f_DSWW
maxbin.021	01;g_S015-6;s_
DSANT11_	d_Bacteria;p_Acidobacteriota;c_Aminicenantia;o_Aminicenantales;f
maxbin.026	Aminicenantaceae;g;s
DSANT11_	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_LZORAL
maxbin.029	124-64-63;f_LZORAL124-64-63;g_LZORAL124-64-63;s_
DSANT11_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Woeseiales;f
maxbin.030	Woeseiaceae;g_UBA1847;s
DSANT11_ maxbin.033	d_Bacteria;p_Acidobacteriota;c_;o_;f_;g_;s_
DSANT95_ maxbin.002	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomona dales;f_Moraxellaceae;g_Psychrobacter;s_
DSANT95_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Granulosicoc
maxbin.006	cales;f_Granulosicoccaceae;g_Granulosicoccus;s_
DSANT95_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.007	obacteriaceae;g_QNYL01;s_
DSANT95_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Burkholderia
maxbin.009	les;f_Neisseriaceae;g_Neisseria;s_Neisseria suis
DSANT95_	d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhodobacteral
maxbin.011	es;f_Rhodobacteraceae;g_Roseovarius;s_
DSANT95_	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_Acidimicrobiales;
maxbin.013	f_SZUA-35;g_SZUA-35;s_
DSANT95_ maxbin.014	d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Sphingomonad ales;f_Sphingomonadaceae;g_Parasphingorhabdus;s_
DSANT95_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.016	obacteriaceae;g_Pricia;s_
DSANT95_	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA5794;f_JA
maxbin.023	ENVV01;g_JAENVV01;s_
DSANT95_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.024	obacteriaceae;g_Pricia;s_

DSANT95_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.030	obacteriaceae;g_Maribacter_A;s_
DSANT95_	d_Bacteria;p_Firmicutes_A;c_Clostridia;o_Oscillospirales;f_Acutali
maxbin.032	bacteraceae;g_Ruminococcus_E;s_Ruminococcus_E bromii_B
DSANT95_	d_Bacteria;p_Firmicutes_A;c_Clostridia;o_Lachnospirales;f_Lachn
maxbin.034	ospiraceae;g_Agathobacter;s_Agathobacter rectalis
DSANT95_	d_Bacteria;p_Fusobacteriota;c_Fusobacteriia;o_Fusobacteriales;f_L
maxbin.035	eptotrichiaceae;g_Sebaldella;s_Sebaldella termitidis
DSANT95_	d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhodobacteral
maxbin.038	es;f_Rhodobacteraceae;g_Pseudorhodobacter;s_
DSANT95_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_Bactero
maxbin.044	idaceae;g_Prevotella;s_Prevotella sp018054505
DSANT95_	d_Bacteria;p_Firmicutes_A;c_Clostridia;o_Clostridiales;f_Clostridi
maxbin.050	aceae;g_Proteiniclasticum;s_Proteiniclasticum ruminis
DSANT95_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flav
maxbin.051	obacteriaceae;g_Pricia;s_
DSANT95_	d_Bacteria;p_Desulfobacterota_I;c_Desulfovibrionia;o_Desulfovibrio
maxbin.067	nales;f_Desulfovibrionaceae;g_Desulfovibrio;s_
DSANT95_	d_Bacteria;p_Actinobacteriota;c_Actinomycetia;o_Propionibacteriale
maxbin.073	s;f_Propionibacteriaceae;g_Arachnia;s_
DSANT95_	d_Bacteria;p_Firmicutes_C;c_Negativicutes;o_Veillonellales;f_Meg
maxbin.075	asphaeraceae;g_Megasphaera;s_Megasphaera sp000417505
DSANT95_	d_Bacteria;p_Firmicutes_C;c_Negativicutes;o_Selenomonadales;f_
maxbin.083	Selenomonadaceae;g;s_
DSANT95_ maxbin.085	d_Bacteria;p_Verrucomicrobiota;c_Verrucomicrobiae;o_Verrucomicr obiales;f_Akkermansiaceae;g_Akkermansia;s_Akkermansia muciniphila
DSANT95_	d_Bacteria;p_Firmicutes_A;c_Clostridia;o_Lachnospirales;f_Lachn
maxbin.118	ospiraceae;g_Fusicatenibacter;s_Fusicatenibacter saccharivorans
DSANT95_	d_Bacteria;p_Planctomycetota;c_Planctomycetia;o_Pirellulales;f_Pi
maxbin.126	rellulaceae;g_Stieleria;s_
DSANT95_	d_Bacteria;p_Verrucomicrobiota;c_Verrucomicrobiae;o_Verrucomicr
maxbin.134	obiales;f_DEV007;g_Arctic95D-9;s_
DSANT95_	d_Bacteria;p_Actinobacteriota;c_Actinomycetia;o_Nitriliruptorales;f
metabat2.56	;g;s
GBGOC44_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Beggiatoales
maxbin.014	;f_Beggiatoaceae;g_UBA10656;s_UBA10656 sp002085445

GBGOC44_ maxbin.033	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_B4-G1;f_B4-G1;g_B30-G15;s_B30-G15 sp003648085
GBGOC44_ maxbin.035	d_Bacteria;p_UBP14;c_UBA6098;o_UBA6098;f_B30-G16;g_;s_
GBGOC44_	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B26-
maxbin.042	1;f_UBA233;g_PIYA01;s_PIYA01 sp003601785
GBGOC44_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_C00003060;f_
maxbin.043	C00003060;g_B23-G16;s_B23-G16 sp003647435
GBGOC44_ maxbin.047	d_Bacteria;p_Desulfobacterota_C;c_Anaeroferrophillalia;o_Anaeroferrophillales;f_Anaeroferrophillaceae;g_DQWO01;s_
GBGOC44_	d_Archaea;p_Thermoproteota;c_Thermoproteia;o_EX4484-217-
maxbin.048	1;f_EX4484-217-1;g_EX4484-217-1;s_EX4484-217-1 sp002254745
GBGOC44_	d_Archaea;p_Methanobacteriota_B;c_Thermococci;o_Methanofastidi
maxbin.049	osales;f_B48-G16;g_B48-G16;s_B48-G16 sp003660605
GBGOC44_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_F082;g
maxbin.050	_UBA4459;s_UBA4459 sp003648625
GBGOC44_ maxbin.052	d_Bacteria;p_Desulfobacterota;c_DSM- 4660;o_Desulfatiglandales;f_B25-G16;g_B33-G16;s_B33-G16 sp003646875
GBGOC44_ maxbin.063	d_Bacteria;p_Desulfobacterota_C;c_Anaeroferrophillalia;o_Anaerofe rrophillales;f_Anaeroferrophillaceae;g_DQWO01;s_DQWO01 sp011042595
GBGOC44_	d_Bacteria;p_Bipolaricaulota;c_Bipolaricaulia;o_UBA7950;f_UBA
maxbin.064	9294;g_S016-54;s_
GBGOC44_	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_JACIWR01;f_JACI
maxbin.067	WR01;g_JACIWR01;s_
GBGOC44_ maxbin.069	d_Bacteria;p_WOR-3;c_UBA3072;o_UBA3072;f_B12-G15;g_B12-G15;s_
GBGOC44_	d_Archaea;p_Asgardarchaeota;c_Baldrarchaeia;o_Baldrarchaeales;f_
maxbin.070	_DXJG01;g_DXJG01;s_
GBGOC44_	d_Archaea;p_Altiarchaeota;c_Altiarchaeia;o_IMC4;f_QMZM01;g_
maxbin.072	_QMZJ01;s_
GBGOC44_	d_Bacteria;p_Bipolaricaulota;c_Bipolaricaulia;o_UBA7950;f_UBA
maxbin.077	9294;g_M55B144;s_
GBGOC44_	d_Archaea;p_Thermoplasmatota;c_E2;o_UBA202;f_DSCA01;g_V
maxbin.081	BQP01;s_
GBGOC44_ maxbin.083	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Hydrogenimonadaceae;g_Hydrogenimonas;s_

GBGOC44_	d_Bacteria;p_Cloacimonadota;c_Cloacimonadia;o_JGIOTU-
maxbin.087	2;f_B8-G9;g_B8-G9;s_B8-G9 sp003647765
GBGOC44_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
maxbin.089	s;f_Desulfosudaceae;g_Desulfosudis;s_Desulfosudis sp002085115
GBGOC44_	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B26-
maxbin.091	1;f_UBA233;g_JdFR-07;s_JdFR-07 sp003601605
GBGOC44_	d_Bacteria;p_WOR-3;c_32-
maxbin.092	111;o_QNBC01;f_QNBC01;g_QNBC01;s_QNBC01 sp003641665
GBGOC44_	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_Anaerolineales;f_UB
maxbin.093	A11858;g_B10-G9;s_B10-G9 sp003648195
GBGOC44_ maxbin.094	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B26-1;f_B26-1;g_B26-1;s_B26-1 sp001593925
GBGOC44_ maxbin.096	d_Archaea;p_Thermoproteota;c_Thermoproteia;o_EX4484-217-1;f_EX4484-217-1;g_DRYM01;s_
GBGOC44_ maxbin.099	d_Bacteria;p_WOR-3;c_UBA3072;o_UBA3072;f_B12-G15;g_B12-G15;s_
GBGOC44_ maxbin.101	d_Bacteria;p_Desulfobacterota;c_DSM- 4660;o_Desulfatiglandales;f_Desulfatiglandaceae;g_4484-190- 2;s_4484-190-2 sp002050025
GBGOC44_	d_Bacteria;p_WOR-
maxbin.102	3;c_Hydrothermia;o_UBA1063;f_DRBW01;g_DRBW01;s_
GBGOC44_	d_Bacteria;p_Ratteibacteria;c_UBA8468;o_B48-
maxbin.107	G9;f_JAFGKM01;g_;s_
GBGOC44_ maxbin.110	d_Bacteria;p_Desulfobacterota;c_DSM- 4660;o_Desulfatiglandales;f_Desulfatiglandaceae;g_B46-G9;s_B46- G9 sp003646815
GBGOC44_ maxbin.112	d_Bacteria;p_Omnitrophota;c_Koll11;o_4484-171;f_B32-G15;g_;s_
GBGOC44_	d_Bacteria;p_Verrucomicrobiota;c_Kiritimatiellae;o_Kiritimatiellales
maxbin.115	;f_Pontiellaceae;g_UBA5540;s_
GBGOC44_ maxbin.117	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Sulfurovaceae;g_Sulfurovum;s_
GBGOC44_	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_TCS64;f_TCS64;
maxbin.119	g_B23;s_
GBGOC44_ maxbin.122	d_Bacteria;p_Acidobacteriota;c_Holophagae;o_Thermotomaculales;f Thermotomaculaceae;g_UM-FILTER-49-7;s_UM-FILTER-49-7 sp003650445

GBGOC44_ maxbin.124	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_QNZT01;g_SZUA-1433;s_
GBGOC44_	d_Archaea;p_Halobacteriota;c_Methanosarcinia;o_Methanosarcinale
maxbin.125	s;f_EX4572-44;g_Ethanoperedens;s_
GBGOC44_ maxbin.127	d_Bacteria;p_Desulfobacterota;c_BSN033;o_B13-G15;f_;g_;s_
GBGOC44_ maxbin.128	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_4484-276;g_4484-276;s_
GBGOC44_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
maxbin.129	s;f_;g_;s_
GBGOC44_	d_Bacteria;p_WOR-
maxbin.137	3;c_Hydrothermia;o_UBA1063;f_DRBW01;g_DRBW01;s_
GBGOC44_ maxbin.142	d_Bacteria;p_Desulfobacterota_C;c_S145-22;o_S145-22;f_S145-22;g_S145-22;s_
GBGOC44_	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacte
maxbin.144	rales;f_Hydrogenimonadaceae;g_Hydrogenimonas;s_
GBGOC44_	d_Bacteria;p_Desulfobacterota;c_Desulfuromonadia;o_Desulfuromo
maxbin.151	nadales;f_Geopsychrobacteraceae;g_Desulfuromusa;s_
GBGOC44_	d_Bacteria;p_KSB1;c_UBA2214;o_UBA2214;f_Zgenome-
maxbin.152	0027;g_;s_
GBGOC44_	d_Bacteria;p_Patescibacteria;c_ABY1;o_BM507;f_UBA12075;g_;
maxbin.158	s_
GBGOC44_ maxbin.159	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_4484-276;g_4484-276;s_4484-276 sp002254335
GBGOC44_	d_Archaea;p_Thermoproteota;c_Thermoproteia;o_Sulfolobales;f_N
maxbin.161	BVN01;g_EX4484-204;s_
GBGOC44_ maxbin.162	d_Bacteria;p_B130-G9;c_B130-G9;o_B130-G9;f_B130-G9;g_;s_
GBGOC44_ maxbin.166	d_Bacteria;p_WOR-3;c_WOR-3;o_SM23-42;f_SM23-42;g_4484-100;s_4484-100 sp002049785
GBGOC44_ maxbin.167	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_QNZT01;g_;s_
GBGOC44_ maxbin.172	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_;f <u>;g</u> ;s
GBGOC44_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_F082;g
maxbin.177	_JAADID01;s_

GBGOC44_	d_Bacteria;p_Desulfobacterota;c_DSM-
maxbin.181	4660;o_Desulfatiglandales;f_Desulfatiglandaceae;g_;s_
GBGOC44_ maxbin.194	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Arcobacteraceae;g_B24-G6;s_B24-G6 sp003643835
GBGOC44_	d_Bacteria;p_Patescibacteria;c_Paceibacteria;o_Moranbacterales;f_
maxbin.198	M55B143;g_M55B143;s_
GBGOC44_	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B26-
maxbin.209	1;f_UBA233;g_PIYF01;s_PIYF01 sp003601875
GBGOC44_ maxbin.210	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Sulfurovaceae;g_SZUA-451;s_
GBGOC44_ maxbin.212	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B26-1;f_B26-1;g_PIYB01;s_PIYB01 sp003601835
GBGOC44_	d_Archaea;p_Thermoproteota;c_Thermoproteia;o_Thermofilales;f_
maxbin.216	Thermofilaceae;g_B17-G15;s_B17-G15 sp003649495
GBGOC44_	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacte
maxbin.218	rales;f_Sulfurovaceae;g_Sulfurovum;s_Sulfurovum sp002085305
GBGOC44_	d_Bacteria;p_Omnitrophota;c_Koll11;o_4484-49;f_4484-
maxbin.229	49;g_4484-49;s_4484-49 sp002085265
GBGOC44_ maxbin.236	d_Bacteria;p_UBP14;c_UBA6098;o_UBA6098;f_B30-G16;g_;s_
GBGOC44_ maxbin.240	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_;f_;g_;s
GBGOC44_	d_Archaea;p_Thermoproteota;c_Thermoproteia;o_Thermoproteales;f
maxbin.244	Thermocladiaceae;g;s
GBGOC44_	d_Bacteria;p_Latescibacterota;c_4484-107;o_4484-107;f_4484-
maxbin.245	107;g_4484-107;s_4484-107 sp002059205
GBGOC44_ maxbin.259	d_Bacteria;p_UBP14;c_UBA6098;o_UBA6098;f_B30-G16;g_;s_
GBGOC44_	d_Archaea;p_Thermoproteota;c_Thermoproteia;o_Sulfolobales;f_N
maxbin.264	BVN01;g_EX4484-204;s_
GBGOC44_	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B26-
maxbin.266	1;f_UBA233;g_UBA233;s_UBA233 sp003661945
GBGOC44_ maxbin.301	d_Bacteria;p_Poribacteria;c_WGA-4E;o_B28-G17;f_B28-G17;g_;s_
GBGOC44_	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B25;f_B25;g_E
maxbin.338	X4484-218;s_EX4484-218 sp002254975

GBGOC44_ maxbin.354	dArchaea;pThermoproteota;cBathyarchaeia;oEX4484- 135;fEX4484-135;gEX4484-135;sEX4484-135 sp011040545
GBGOC44_ metabat2.19 0	d_Bacteria;p_RBG-13-66-14;c_B26-G2;o_B26-G2;f_B26-G2;g_B26-G2;s_B26-G2 sp003647715
GBGOC44_ metabat2.23 0	d_Bacteria;p_JdFR-76;c_JdFR-76;o_JdFR-76;f_4484-219;g_4484-219;s_4484-219;s_4484-219 sp002085035
GBGOC44_ metabat2.23 7	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_JAADJ A01;g_JAADJA01;s_
GBGOC44_ metabat2.25 0	d_Bacteria;p_Campylobacterota;c_Desulfurellia;o_;f_;g_;s_
GBGOC44_ metabat2.27 9	d_Bacteria;p_Desulfobacterota;c_Desulfobaccia;o_Desulfobaccales;f _Desulfobaccaceae;g_Desulfobacca_B;s_Desulfobacca_B sp002049795
GBGOC44_ metabat2.29 3	d_Bacteria;p_Acidobacteriota;c_Aminicenantia;o_Aminicenantales;f _Aminicenantaceae_A;g_JdFR-80;s_JdFR-80 sp011042725
GBGOC44_ metabat2.4	d_Archaea;p_Thermoproteota;c_Thermoproteia;o_Sulfolobales;f_D esulfurococcaceae;g_EX4484-58;s_EX4484-58 sp002254665
GBGOC44_ metabat2.47	d_Archaea;p_Thermoproteota;c_Thermoproteia;o_Sulfolobales;f_;g _;s_
GBGOC44_ metabat2.72	d_Bacteria;p_Desulfobacterota;c_Syntrophobacteria;o_B119-G9;f_B119-G9;g_B119-G9;s_B119-G9 sp003646715
GBGOC44_ metabat2.80	d_Archaea;p_Thermoproteota;c_Thermoproteia;o_Thermofilales;f_; g_;s_
GBGOC45_ maxbin.024	d_Archaea;p_Halobacteriota;c_Methanosarcinia;o_Methanosarcinale s;f_Methanogasteraceae;g_Methanogaster;s_Methanogaster sp003601535
GBGOC45_ maxbin.025	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Beggiatoales ;f_Beggiatoaceae;g_UBA10656;s_UBA10656 sp002085445
GBGOC45_ maxbin.029	d_Bacteria;p_Desulfobacterota;c_Thermodesulfobacteria;o_Thermod esulfobacteriales;f_Thermodesulfobacteriaceae;g_Thermodesulfobacteri um;s_
GBGOC45_ maxbin.030	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_C00003060;f_ C00003060;g_B23-G16;s_B23-G16 sp003647435

GBGOC45_ maxbin.031	d_Bacteria;p_UBP14;c_UBA6098;o_UBA6098;f_B30-G16;g_;s_
GBGOC45_ maxbin.032	d_Bacteria;p_Desulfobacterota;c_DSM- 4660;o_Desulfatiglandales;f_B25-G16;g_B33-G16;s_B33-G16 sp003646875
GBGOC45_	d_Archaea;p_Methanobacteriota_B;c_Thermococci;o_Methanofastidi
maxbin.034	osales;f_B48-G16;g_B48-G16;s_B48-G16 sp003660605
GBGOC45_	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B26-
maxbin.035	1;f_UBA233;g_PIYA01;s_PIYA01 sp003601785
GBGOC45_	d_Archaea;p_Altiarchaeota;c_Altiarchaeia;o_IMC4;f_QMZM01;g_
maxbin.036	_QMZJ01;s_
GBGOC45_	d_Archaea;p_Halobacteriota;c_Methanosarcinia;o_Methanosarcinale
maxbin.038	s;f_EX4572-44;g_Ethanoperedens;s_
GBGOC45_	d_Bacteria;p_Cloacimonadota;c_Cloacimonadia;o_JGIOTU-
maxbin.046	2;f_B8-G9;g_B8-G9;s_B8-G9 sp003647765
GBGOC45_ maxbin.052	d_Bacteria;p_Omnitrophota;c_4484-213;o_4484-213;f_4484-213;g_4484-213;s_4484-213 sp002085025
GBGOC45_	d_Bacteria;p_Bipolaricaulota;c_Bipolaricaulia;o_UBA7950;f_UBA
maxbin.058	9294;g_S016-54;s_
GBGOC45_	d_Archaea;p_Thermoproteota;c_Thermoproteia;o_Thermofilales;f_
maxbin.059	B20-G17;g_B20-G17;s_
GBGOC45_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_F082;g
maxbin.060	_UBA4459;s_UBA4459 sp003648625
GBGOC45_	d_Bacteria;p_Desulfobacterota_C;c_Anaeroferrophillalia;o_Anaerofe
maxbin.071	rrophillales;f_Anaeroferrophillaceae;g_DQWO01;s_
GBGOC45_ maxbin.073	d_Bacteria;p_Omnitrophota;c_Koll11;o_4484-171;f_B32-G15;g_;s_
GBGOC45_ maxbin.074	d_Bacteria;p_Desulfobacterota_C;c_Anaeroferrophillalia;o_Anaerofe rrophillales;f_Anaeroferrophillaceae;g_DQWO01;s_DQWO01 sp011042595
GBGOC45_ maxbin.089	d_Bacteria;p_Desulfobacterota;c_DSM- 4660;o_Desulfatiglandales;f_Desulfatiglandaceae;g_B46-G9;s_B46- G9 sp003646815
GBGOC45_	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_Anaerolineales;f_UB
maxbin.092	A11858;g_B10-G9;s_B10-G9 sp003648195
GBGOC45_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
maxbin.099	s;f_UBA11574;g_S5133MH16;s_

GBGOC45_ maxbin.102	d_Archaea;p_Methanobacteriota_B;c_Thermococci;o_Thermococcal es;f_Thermococcaceae;g_Thermococcus_A;s_Thermococcus_A sp003663525
GBGOC45_ maxbin.103	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B26-1;f_B26-1;g_B63;s_B63 sp003601695
GBGOC45_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
maxbin.105	s;f_4be13;g_;s_
GBGOC45_	d_Bacteria;p_Bipolaricaulota;c_Bipolaricaulia;o_UBA7950;f_UBA
maxbin.106	9294;g_M55B144;s_
GBGOC45_	d_Bacteria;p_Bipolaricaulota;c_Bipolaricaulia;o_Bipolaricaulales;f_
maxbin.107	Bipolaricaulaceae;g_UBA3571;s_
GBGOC45_ maxbin.115	d_Bacteria;p_Desulfobacterota;c_Desulfobaccia;o_Desulfobaccales;f _Desulfobaccaceae;g_Desulfobacca_B;s_Desulfobacca_B sp002049795
GBGOC45_ maxbin.117	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B26-1;f_B26-1;g_B26-1;s_B26-1 sp001593925
GBGOC45_	d_Bacteria;p_Acidobacteriota;c_Aminicenantia;o_Aminicenantales;f
maxbin.118	_Aminicenantaceae_A;g_JdFR-80;s_JdFR-80 sp011042725
GBGOC45_	d_Bacteria;p_WOR-
maxbin.123	3;c_Hydrothermia;o_UBA1063;f_DRBW01;g_DRBW01;s_
GBGOC45_	d_Archaea;p_Thermoproteota;c_Thermoproteia;o_EX4484-217-
maxbin.129	1;f_EX4484-217-1;g_EX4484-217-1;s_EX4484-217-1 sp002254745
GBGOC45_	d_Archaea;p_Methanobacteriota_B;c_Thermococci;o_Thermococcal
maxbin.130	es;f_Thermococcaceae;g_Palaeococcus;s_Palaeococcus sp003663695
GBGOC45_	d_Bacteria;p_Cloacimonadota;c_Cloacimonadia;o_Cloacimonadales;
maxbin.139	f_TCS61;g_4484-275;s_4484-275 sp002085205
GBGOC45_ maxbin.140	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Hydrogenimonadaceae;g_Hydrogenimonas;s_
GBGOC45_	d_Bacteria;p_Verrucomicrobiota;c_Kiritimatiellae;o_Kiritimatiellales
maxbin.147	;f_Pontiellaceae;g_UBA5540;s_
GBGOC45_	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_Anaerolineales;f_UB
maxbin.149	A11858;g_G519;s_G519 sp011176615
GBGOC45_ maxbin.151	d_Bacteria;p_JdFR-76;c_JdFR-76;o_JdFR-76;f_4484-219;g_4484-219;s_4484-219;s_4484-219 sp002085035
GBGOC45_ maxbin.153	d_Bacteria;p_B130-G9;c_B130-G9;o_B130-G9;f_B130-G9;g_B130-G9;s_B130-G9 sp003647755
GBGOC45_	d_Bacteria;p_Desulfobacterota;c_Desulfuromonadia;o_Desulfuromo
maxbin.155	nadales;f_Geopsychrobacteraceae;g_Desulfuromusa;s_

GBGOC45_ maxbin.158	d_Bacteria;p_WOR-3;c_WOR-3;o_SM23-42;f_SM23-42;g_4484-100;s_4484-100 sp002049785
GBGOC45_	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B25;f_B25;g_E
maxbin.161	X4484-218;s_EX4484-218 sp002254975
GBGOC45_	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B26-1;f_B26-
maxbin.166	1;g_PIYB01;s_PIYB01 sp003601835
GBGOC45_	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_TCS64;f_TCS64;
maxbin.174	g_B23;s_
GBGOC45_	d_Bacteria;p_Bipolaricaulota;c_Bipolaricaulia;o_Bipolaricaulales;f_
maxbin.193	Bipolaricaulaceae;g_UBA3574;s_
GBGOC45_	d_Bacteria;p_UBP14;c_UBA6098;o_4484-93;f_4484-93;g_4484-
maxbin.199	93;s_4484-93 sp002085385
GBGOC45_ maxbin.201	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Arcobacteraceae;g_B24-G6;s_
GBGOC45_	d_Bacteria;p_WOR-
maxbin.238	3;c_Hydrothermia;o_UBA1063;f_DRBW01;g_DRBW01;s_
GBGOC45_ maxbin.245	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Sulfurovaceae;g_SZUA-451;s_
GBGOC45_ maxbin.254	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_;f_;g_;s_
GBGOC45_	d_Archaea;p_Thermoproteota;c_Thermoproteia;o_EX4484-217-
maxbin.298	1;f_EX4484-217-1;g_EX4484-217-1;s_
GBGOC45_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_S143-
metabat2.15	33;g_;s_
GBGOC45_ metabat2.15 4	d_Bacteria;p_KSB1;c_UBA2214;o_UBA2214;f_Zgenome- 0027;g_;s_
GBGOC45_ metabat2.17 8	d_Bacteria;p_Spirochaetota;c_B62-G9;o_B62-G9;f_B62-G9;g_B62-G9;s_
GBGOC45_ metabat2.19 0	d_Bacteria;p_Omnitrophota;c_Koll11;o_4484-49;f_4484- 49;g_4484-49;s_4484-49 sp002085265
GBGOC45_ metabat2.21 5	d_Bacteria;p_Acidobacteriota;c_Aminicenantia;o_Aminicenantales;f _Aminicenantaceae_A;g_JdFR-80;s_JdFR-80 sp003648765

GBGOC45_ metabat2.22 7	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_F082;g _JAADID01;s_
GBGOC45_ metabat2.23	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale s;f_SURF-3;g_;s_
GBGOC45_ metabat2.23 5	d_Bacteria;p_Omnitrophota;c_Koll11;o_;f_;g_;s_
GBGOC45_ metabat2.25 7	d_Bacteria;p_Desulfobacterota;c_DSM- 4660;o_Desulfatiglandales;f_Desulfatiglandaceae;g_;s_
GBGOC45_ metabat2.28 1	d_Bacteria;p_B130-G9;c_B130-G9;o_B130-G9;f_B130-G9;g_;s_
GBGOC45_ metabat2.28 4	d_Archaea;p_Asgardarchaeota;c_Heimdallarchaeia;o_JABLTI01;f_ JABLTI01;g_JABLTI01;s_JABLTI01 sp019056785
GBGOC45_ metabat2.28 7	d_Bacteria;p_Desulfobacterota;c_DSM- 4660;o_Desulfatiglandales;f_Desulfatiglandaceae;g_4484-190- 2;s_4484-190-2 sp002050025
GBGOC45_ metabat2.29 8	d_Bacteria;p_RBG-13-66-14;c_B26-G2;o_B26-G2;f_B26-G2;g_B26-G2;s_B26-G2 sp003647715
GBGOC45_ metabat2.47	d_Archaea;p_Asgardarchaeota;c_Heimdallarchaeia;o_UBA460;f_U BA460;g;s_
GBGOC45_ metabat2.54	d_Bacteria;p_Zixibacteria;c_MSB-5A5;o_UBA10806;f_4484- 95;g_4484-95;s_4484-95 sp002085375
GBGOC45_ metabat2.82	d_Bacteria;p_KSB1;c_UBA2214;o_AABM5-25-91;f_AABM5-25-91;g_;s_
GBGOC45_ metabat2.9	d_Bacteria;p_Desulfobacterota;c_DSM- 4660;o_Desulfatiglandales;f_Desulfatiglandaceae;g_;s_
GBGOC47_ maxbin.001	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Beggiatoales ;f_Beggiatoaceae;g_QNES01;s_QNES01 sp003645255
GBGOC47_ maxbin.016	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_4572-128;g_4572-128;s_4572-128 sp002084355
GBGOC47_ maxbin.017	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_UBA12 170;g_;s_
GBGOC47_ maxbin_032	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_4572-78;f_4572-78;g_4572-78;s_4572-78 sp002084875

GBGOC47_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Beggiatoales
maxbin.036	;f_Beggiatoaceae;g_UBA10656;s_UBA10656 sp002085445
GBGOC47_	d_Bacteria;p_Desulfobacterota;c_DSM-
maxbin.037	4660;o_Desulfatiglandales;f_Desulfatiglandaceae;g_NaphS2;s_
GBGOC47_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_F082;g
maxbin.041	_UBA4459;s_UBA4459 sp003648625
GBGOC47_	d_Bacteria;p_Spirochaetota;c_Spirochaetia;o_Spirochaetales_E;f_D
maxbin.043	SM-19205;g;s
GBGOC47_ maxbin.045	d_Bacteria;p_Desulfobacterota;c_Desulfuromonadia;o_Desulfuromo nadales;f_Desulfuromonadaceae;g_Desulfuromonas;s_Desulfuromona s sp002084665
GBGOC47_ maxbin.046	dArchaea;pHalobacteriota;cMethanosarcinia;oMethanosarcinale s;fMethanogasteraceae;gMethanogaster;sMethanogaster sp003601535
GBGOC47_	d_Bacteria;p_Desulfobacterota;c_Dissulfuribacteria;o_Dissulfuribact
maxbin.053	erales;f_UBA3076;g_UBA3076;s_UBA3076 sp003194485
GBGOC47_	d_Bacteria;p_Fusobacteriota;c_Fusobacteriia;o_Fusobacteriales;f_F
maxbin.055	usobacteriaceae;g_Psychrilyobacter;s_Psychrilyobacter sp002084825
GBGOC47_ maxbin.058	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Sulfurovaceae;g_Sulfurovum;s_
GBGOC47_	d_Bacteria;p_Cloacimonadota;c_Cloacimonadia;o_Cloacimonadales;
maxbin.059	f;g;s
GBGOC47_ maxbin.062	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Sulfurospirillaceae;g_Sulfurospirillum_A;s_
GBGOC47_	d_Bacteria;p_Firmicutes_A;c_Clostridia;o_Tissierellales;f_Caldisali
maxbin.065	nibacteraceae;g;s
GBGOC47_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_UBA12
maxbin.066	170;g_UBA12170;s_
GBGOC47_	d_Bacteria;p_Cloacimonadota;c_Cloacimonadia;o_Cloacimonadales;
maxbin.070	f_;g_;s_
GBGOC47_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
maxbin.071	s;f_Desulfobacteraceae_A;g_NBLS01;s_NBLS01 sp002084385
GBGOC47_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
maxbin.076	s;f_Desulfobacteraceae;g_4572-130;s_4572-130 sp002084425
GBGOC47_ maxbin.077	d_Bacteria;p_Desulfobacterota_C;c_Anaeroferrophillalia;o_Anaerofe rrophillales;f_Anaeroferrophillaceae;g_DQWO01;s_DQWO01 sp011042595

GBGOC47_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
maxbin.080	s;f_UBA2174_A;g;s
GBGOC47_	d_Bacteria;p_Desulfobacterota;c_DSM-
maxbin.081	4660;o_Desulfatiglandales;f_Desulfatiglandaceae;g_NaphS2;s_
GBGOC47_	d_Bacteria;p_Verrucomicrobiota;c_Kiritimatiellae;o_Kiritimatiellales
maxbin.082	;f_Pontiellaceae;g_UBA5540;s_
GBGOC47_	d_Bacteria;p_Atribacterota;c_JS1;o_SB-45;f_34-128;g_34-
maxbin.083	128;s_34-128 sp002084865
GBGOC47_ maxbin.089	d_Bacteria;p_Spirochaetota;c_Spirochaetia;o_Spirochaetales_E;f_S pirochaetaceae_B;g_Oceanispirochaeta;s_Oceanispirochaeta sp002084805
GBGOC47_ maxbin.091	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_4572-78;f_4572-78;g_DUEH01;s_DUEH01 sp011192195
GBGOC47_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
maxbin.092	s;f_Desulfobacteraceae;g_Desulfobacula;s_
GBGOC47_	d_Bacteria;p_4572-55;c_4572-55;o_4572-55;f_4572-55;g_4572-
maxbin.096	55;s_4572-55 sp002084765
GBGOC47_ maxbin.098	d_Bacteria;p_Desulfobacterota;c_DSM- 4660;o_Desulfatiglandales;f_B25-G16;g_B33-G16;s_B33-G16 sp003646875
GBGOC47_ maxbin.101	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_4484-276;g_4484-276;s_4484-276 sp002254335
GBGOC47_	d_Bacteria;p_Cloacimonadota;c_Cloacimonadia;o_Cloacimonadales;
maxbin.102	f_TCS61;g_B137-G9;s_B137-G9 sp003647825
GBGOC47_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_NBLH0
maxbin.107	1;g_NBLH01;s_NBLH01 sp002084525
GBGOC47_	d_Bacteria;p_Firmicutes;c_Bacilli;o_Izemoplasmatales;f_UBA5603
maxbin.108	;g_4572-104;s_
GBGOC47_ maxbin.115	d_Bacteria;p_Planctomycetota;c_Phycisphaerae;o_Sedimentisphaeral es;f_Anaerohalosphaeraceae;g_4572-13;s_4572-13 sp002084585
GBGOC47_ maxbin.120	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_B4-G1;f_B4-G1;g_B4-G1;s_B4-G1 sp003648025
GBGOC47_	d_Archaea;p_Halobacteriota;c_Syntropharchaeia;o_ANME-
maxbin.125	1;f_ANME-1;g_ANME1a;s_ANME1a sp003194425
GBGOC47_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
maxbin.126	s;f_4572-123;g_4572-123;s_4572-123 sp002084545
GBGOC47_	d_Bacteria;p_Desulfobacterota;c_Desulfuromonadia;o_Desulfuromo
maxbin.131	nadales;f_Geopsychrobacteraceae;g_Desulfuromusa;s_

GBGOC47_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
maxbin.160	s;f_Desulfobacteraceae;g_NBML01;s_NBML01 sp002085465
GBGOC47_ maxbin.167	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Sulfurimonadaceae;g_Sulfurimonas;s_
GBGOC47_	d_Bacteria;p_Verrucomicrobiota;c_Kiritimatiellae;o_Kiritimatiellales
maxbin.172	;f_Pontiellaceae;g_UBA5540;s_
GBGOC47_ maxbin.182	d_Bacteria;p_Desulfobacterota;c_Dissulfuribacteria;o_Dissulfuribacteria;s_QOAL01;g_S144-17;s_
GBGOC47_ maxbin.192	d_Bacteria;p_Campylobacterota;c_Campylobacteria;o_Campylobacterales;f_Sulfurovaceae;g_Sulfurovum;s_
GBGOC47_	d_Bacteria;p_Planctomycetota;c_Phycisphaerae;o_Sedimentisphaeral
maxbin.193	es;f_;g_;s_
GBGOC47_	d_Bacteria;p_Desulfobacterota;c_DSM-
maxbin.198	4660;o_Desulfatiglandales;f_Desulfatiglandaceae;g_Desulfacyla;s_
GBGOC47_	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_B26-
maxbin.200	1;f_UBA233;g_PIYA01;s_PIYA01 sp003601785
GBGOC47_	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA5794;f_UB
maxbin.204	A5794;g_;s_
GBGOC47_	d_Bacteria;p_Desulfobacterota;c_Desulfofervidia;o_Desulfofervidale
maxbin.208	s;f_DG-60;g_;s_
GBGOC47_	d_Bacteria;p_Desulfobacterota;c_Desulfuromonadia;o_Desulfuromo
maxbin.217	nadales;f_Geopsychrobacteraceae;g_Desulfuromusa;s_
GBGOC47_	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_LZORAL
maxbin.219	124-64-63;f_LZORAL124-64-63;g_LZORAL124-64-63;s_
GBGOC47_	d_Bacteria;p_Myxococcota;c_Bradymonadia;o_REDH01;f_;g_;s_
maxbin.224	-
GBGOC47_ maxbin.246	d_Bacteria;p_Patescibacteria;c_ABY1;o_UBA2591;f_;g_;s_
GBGOC47_ maxbin.260	d_Bacteria;p_Planctomycetota;c_SZUA-567;o_;f_;g_;s_
GBGOC47_	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
maxbin.274	s;f_JAABTS01;g_;s_
GBGOC47_	d_Bacteria;p_Cloacimonadota;c_Cloacimonadia;o_Cloacimonadales;
maxbin.299	f_TCS61;g_DRTC01;s_DRTC01 sp011372345
GBGOC47_	d_Bacteria;p_Thermotogota;c_Thermotogae;o_Thermotogales;f_DS
maxbin.311	M-5069;g_B6-G13;s_B6-G13 sp003641785

GBGOC47_	d_Bacteria;p_Desulfobacterota;c_Desulfobulbia;o_Desulfobulbales;f
maxbin.325	_Desulfurivibrionaceae;g_;s_
GBGOC47_ maxbin.338	d_Bacteria;p_KSB1;c_UBA2214;o_AABM5-25-91;f_;g_;s_
GBGOC47_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacter
maxbin.345	ales;f_Moritellaceae;g_;s_
GBGOC47_	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_LZORAL
maxbin.373	124-64-63;f_LZORAL124-64-63;g_LZORAL124-64-63;s_
GBGOC47_ maxbin.377	d_Bacteria;p_Fermentibacterota;c_Fermentibacteria;o_Fermentibacterateratera;s_Fermentibacteraceae;g_;s_
GBGOC47_	d_Bacteria;p_Bipolaricaulota;c_Bipolaricaulia;o_UBA7950;f_UBA
maxbin.393	9294;g_M55B144;s_
GBGOC47_	d_Bacteria;p_Zixibacteria;c_MSB-
maxbin.413	5A5;o_GN15;f_PGXB01;g_PGXB01;s_PGXB01 sp003641425
GBGOC47_ metabat2.40 3	d_Bacteria;p_KSB1;c_UBA2214;o_UBA2214;f_Zgenome- 0027;g_;s_
GBGOC47_ metabat2.41 3	d_Bacteria;p_KSB1;c_UBA2214;o_UBA2214;f_Zgenome- 0027;g_;s_
GBGOC47_	d_Bacteria;p_Firmicutes_A;c_Clostridia;o_Peptostreptococcales;f_
metabat2.7	Acidaminobacteraceae;g;s_
GBGOC47_	d_Bacteria;p_Acidobacteriota;c_Aminicenantia;o_Aminicenantales;f
metabat2.96	Aminicenantaceae_A;g_JdFR-80;s_JdFR-80 sp011042725
HRSPAC47_ maxbin.011	d_Archaea;p_Halobacteriota;c_Methanosarcinia;o_Methanosarcinale s;f_Methanocomedenaceae;g_Methanomarinus;s_Methanomarinus sp002926195
HRSPAC47_	d_Bacteria;p_Firmicutes_A;c_Clostridia;o_Eubacteriales;f_Eubacter
maxbin.033	iaceae;g_Acetobacterium;s_
HRSPAC47_ maxbin.034	d_Bacteria;p_Desulfobacterota;c_Desulfuromonadia;o_Desulfuromo nadales;f_Desulfuromonadaceae;g_Desulfuromonas;s_Desulfuromona s sp001751155
HRSPAC47_ maxbin.071	d_Bacteria;p_Desulfobacterota;c_Desulfuromonadia;o_Desulfuromo nadales;f_Desulfuromonadaceae;g_Desulfuromonas;s_Desulfuromona s sp002084665
HRSPAC47_ maxbin.076	d_Bacteria;p_Firmicutes_A;c_Clostridia;o_Tissierellales;f_Dethiosulfatibacteraceae;g_Dethiosulfatibacter;s_

HRSPAC47_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_UBA6429;f_
maxbin.084	_UBA6429;g;s
HRSPAC47_	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_UBA
metabat2.19	1820;g_JADGDT01;s_
HRSPAC47_	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomona
metabat2.53	dales;f_Halomonadaceae;g_Halomonas;s_Halomonas sedimenti
KJ4_maxbin	d_Bacteria;p_Methylomirabilota;c_Methylomirabilia;o_Methylomira
.001	bilales;f_CSP1-5;g_CSP1-5;s_CSP1-5 sp012974305
KJ5_maxbin	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Woeseiales;f
.003	_Woeseiaceae;g_;s_
KJ5_maxbin	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Acidiferroba
.008	cterales;f_SPGG2;g_SPGG2;s_
KJ5_maxbin .009	d_Bacteria;p_Nitrospinota;c_UBA9942;o_;f_;g_;s_
KJ5_maxbin	d_Bacteria;p_Zixibacteria;c_MSB-
.010	5A5;o_UBA10806;f_UBA10806;g_;s_
KJ5_maxbin	d_Bacteria;p_Planctomycetota;c_Brocadiae;o_Brocadiales;f_Scalin
.032	duaceae;g;s
KJ6_maxbin	d_Bacteria;p_Zixibacteria;c_MSB-
.007	5A5;o_UBA10806;f_UBA10806;g_;s_
KJ6_maxbin	d_Bacteria;p_Desulfobacterota;c_Syntrophobacteria;o_BM002;f_B
.015	M002;g_BM002;s_
KJ6_metaba t2.15	d_Bacteria;p_Planctomycetota;c_JACQVX01;o_;f_;g_;s_
KOR52_max bin.001	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_SZUA-229;f_SZUA-229;g_JABDQW01;s_
KOR52_max bin.002	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_GCA-001735895;f_GCA-001735895;g_GCA-001735895;s_
VOD51	
bin.004	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_SZUA-229;f_SZUA-229;g_JABDQW01;s_
KOR52_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_SZUA-
bin.004	229;f_SZUA-229;g_JABDQW01;s_
KOR52_max	d_Bacteria;p_Desulfobacterota;c_Desulfobulbia;o_Desulfobulbales;f
bin.011	_Desulfocapsaceae;g_JABDQA01;s_JABDQA01 sp905479735
KOR52_max bin.004 KOR52_max bin.011 KOR52_max bin.020	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_SZUA-229;f_SZUA-229;g_JABDQW01;s_         d_Bacteria;p_Desulfobacterota;c_Desulfobulbia;o_Desulfobulbales;f         _Desulfocapsaceae;g_JABDQA01;s_JABDQA01 sp905479735         d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Cytophagales;f_Cyclob acteriaceae;g_JAGTVG01;s_

KOR52_max bin.116	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomona dales;f_HTCC2089;g_UBA4421;s_
KOR52_met	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Woeseiales;f
abat2.201	_Woeseiaceae;g_;s_
KOR58_max	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Woeseiales;f
bin.006	_Woeseiaceae;g_JAACFE01;s_JAACFE01 sp009937505
KOR58_max	d_Archaea;p_Thermoproteota;c_Nitrososphaeria;o_Nitrososphaerales
bin.010	;f_Nitrosopumilaceae;g_Nitrosopumilus;s_
KOR58_max bin.012	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_GCA-001735895;f_GCA-001735895;g_GCA-001735895;s_GCA-001735895 sp009937625
KOR58_max	d_Bacteria;p_Desulfobacterota;c_Syntrophobacteria;o_BM002;f_B
bin.016	M002;g_BM002;s_
KOR58_max	d_Bacteria;p_Desulfobacterota;c_Desulfuromonadia;o_Desulfuromo
bin.017	nadales;f_BM103;g_M0040;s_
KOR58_met	d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhodobacteral
abat2.14	es;f_Rhodobacteraceae;g_Roseobacter;s_
LOKART30	d_Bacteria;p_Patescibacteria;c_Paceibacteria;o_Paceibacterales;f_R
_maxbin.038	BG-13-36-15;g_RBG-13-36-15;s_RBG-13-36-15 sp011050465
LOKART30	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Acidiferroba
_maxbin.040	cterales;f_SZUA-150;g_;s_
LOKART30 _maxbin.045	d_Bacteria;p_Planctomycetota;c_Phycisphaerae;o_Sedimentisphaeral es;f_Anaerohalosphaeraceae;g_B28-G16;s_
LOKART30	d_Bacteria;p_Dependentiae;c_Babeliae;o_Babeliales;f_Vermiphilac
_maxbin.061	eae;g;s_
LOKART30	d_Bacteria;p_Myxococcota_A;c_UBA9160;o_UBA9160;f_SMWR
_maxbin.065	01;g_;s_
LOKART30	d_Bacteria;p_Chloroflexota;c_Dehalococcoidia;o_SM23-28-
_maxbin.066	2;f_RBG-16-64-32;g_RBG-16-64-32;s_
LOKART30	d_Bacteria;p_Spirochaetota;c_B62-G9;o_B62-G9;f_B62-
_maxbin.069	G9;g_DRHN01;s_DRHN01 sp011049595
LOKART30	d_Bacteria;p_Planctomycetota;c_Phycisphaerae;o_SM23-
_maxbin.072	33;f_SM23-33;g_;s_
LOKART30	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA2766;f_;g_
_maxbin.077	_;s_
LOKART30	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
_maxbin.079	s;f_UBA2156;g_UBA2156;s_UBA2156 sp011049715

LOKART30	d_Bacteria;p_BMS3Abin14;c_BMS3Abin14;o_BMS3Abin14;f_JA
_maxbin.081	FGWO01;g_;s_
LOKART30	d_Bacteria;p_AABM5-125-
_maxbin.083	24;c_BMS3BBIN04;o_BMS3BBIN04;f_BMS3BBIN04;g_;s_
LOKART30 _maxbin.084	d_Bacteria;p_Abyssubacteria;c_SURF-5;o_;f_;g_;s_
LOKART30 _maxbin.085	d_Bacteria;p_Nitrospinota;c_UBA7883;o_JACRGQ01;f_;g_;s_
LOKART30 _maxbin.087	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_SM23-62;g_SM23-62;s_SM23-62 sp011053095
LOKART30 _maxbin.092	d_Bacteria;p_Gemmatimonadota;c_Glassbacteria;o_GWA2-58-10;f_;g_;s_
LOKART30 _maxbin.097	d_Bacteria;p_Planctomycetota;c_Phycisphaerae;o_Sedimentisphaeral es;f_Anaerohalosphaeraceae;g_;s_
LOKART30	d_Bacteria;p_Planctomycetota;c_Planctomycetia;o_Pirellulales;f_T
_maxbin.099	hermoguttaceae;g;s
LOKART30 _maxbin.107	d_Bacteria;p_Myxococcota;c_Polyangia;o_JAAYKL01;f_;g_;s_
LOKART30 _maxbin.109	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_GCA-2748055;g_GCA-2748055;s_
LOKART30	d_Bacteria;p_Hydrogenedentota;c_Hydrogenedentia;o_Hydrogenede
_maxbin.114	ntiales;f_SLHB01;g_;s_
LOKART30	d_Bacteria;p_Actinobacteriota;c_Thermoleophilia;o_Solirubrobactera
_maxbin.122	les;f_;g_;s_
LOKART30 _maxbin.146	d_Bacteria;p_Actinobacteriota;c_;o_;f_;g_;s_
LOKART30 _maxbin.153	d_Bacteria;p_Actinobacteriota;c_Aquicultoria;o_;f_;g_;s_
LOKART30	d_Bacteria;p_Planctomycetota;c_Planctomycetia;o_Pirellulales;f_T
_maxbin.158	hermoguttaceae;g_JAFGRD01;s_
LOKART30	d_Bacteria;p_Planctomycetota;c_Planctomycetia;o_Pirellulales;f_P
_maxbin.175	ALSA-1355;g;s_
LOKART30	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
_maxbin.177	s;f_UBA2156;g_;s_
LOKART30	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA5794;f_UB
_maxbin.190	A5794;g_;s_

LOKART30 _metabat2.1 39	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Cytophagales;f_Cyclob acteriaceae;g_SZUA-230;s_
LOKART60	d_Bacteria;p_Chlamydiota;c_Chlamydiia;o_Chlamydiales;f_SM23-
_maxbin.005	39;g_JAAKFX01;s_JAAKFX01 sp011064875
LOKART60	d_Bacteria;p_Chloroflexota;c_Dehalococcoidia;o_Dehalococcoidales
_maxbin.008	;f_W498;g_;s_
LOKART60 _maxbin.012	d_Bacteria;p_Planctomycetota;c_Phycisphaerae;o_Sedimentisphaeral es;f_SG8-4;g_CS2-K091;s_
LOKART60	d_Bacteria;p_Chloroflexota;c_Dehalococcoidia;o_Dehalococcoidales
_maxbin.018	;f_RBG-16-60-22;g_E44-bin89;s_E44-bin89 sp004375725
LOKART60 _maxbin.019	d_Bacteria;p_Zixibacteria;c_MSB-5A5;o_GN15;f_FEB-12;g_;s_
LOKART60	d_Bacteria;p_Chlamydiota;c_Chlamydiia;o_Chlamydiales;f_SM23-
_maxbin.022	39;g_JAAKFX01;s_JAAKFX01 sp011064905
LOKART60	d_Bacteria;p_Desulfobacterota;c_DSM-
_maxbin.025	4660;o_Desulfatiglandales;f_B25-G16;g_B33-G16;s_
LOKART60	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_C00003060;f_
_maxbin.026	C00003060;g;s_
LOKART60	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_VGOG01;f_VGOG0
_maxbin.034	1;g_VGOG01;s_
LOKART60 _maxbin.042	d_Bacteria;p_Atribacterota;c_JS1;o_SB-45;f_34-128;g_;s_
LOKART60	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_Anaerolineales;f_E4
_maxbin.043	4-bin32;g_E44-bin32;s_
LOKART60	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Acidiferroba
_maxbin.045	cterales;f_SZUA-150;g_;s_
LOKART60	d_Bacteria;p_Actinobacteriota;c_Humimicrobiia;o_Humimicrobiales;
_maxbin.049	f_Humimicrobiaceae;g_M55B120;s_
LOKART60	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_C00003060;f_
_maxbin.052	C00003060;g_C00003060;s_
LOKART60 _maxbin.054	d_Bacteria;p_Chlamydiota;c_Chlamydiia;o_Chlamydiales;f_Rhabd ochlamydiaceae;g_JAAKFS01;s_JAAKFS01 sp011064965
LOKART60	d_Bacteria;p_Armatimonadota;c_UBA5377;o_UBA5377;f_JAAYI
_maxbin.062	R01;g_;s_
LOKART60	d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Bacteroidales;f_VadinH
_maxbin.064	A17;g_JADFUQ01;s_

LOKART60	d_Bacteria;p_Planctomycetota;c_Phycisphaerae;o_UBA1845;f_UB
_maxbin.066	A1845;g_;s_
LOKART60 _maxbin.071	d_Bacteria;p_Omnitrophota;c_Koll11;o_B26-G9;f_B26-G9;g_;s_
LOKART60	d_Bacteria;p_Zixibacteria;c_MSB-
_maxbin.080	5A5;o_UBA10806;f_UBA10806;g_;s_
LOKART60 _maxbin.140	d_Bacteria;p_Planctomycetota;c_DG-23;o_;f_;g_;s_
LOKART60 _metabat2.1 78	d_Bacteria;p_BMS3Abin14;c_BMS3Abin14;o_BMS3Abin14;f_BM S3Abin14;g_SZUA-178;s_
LOKART62	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Acidiferroba
_maxbin.021	cterales;f_SZUA-150;g_;s_
LOKART62	d_Bacteria;p_RBG-13-66-14;c_RBG-13-66-14;o_RBG-13-66-
_maxbin.022	14;f_RBG-13-66-14;g_RBG-13-66-14;s_
LOKART62	d_Bacteria;p_Armatimonadota;c_CAIYQO01;o_WVXJ01;f_DTIZ0
_maxbin.027	1;g_;s_
LOKART62	d_Bacteria;p_RBG-13-66-14;c_B26-
_maxbin.041	G2;o_WVWN01;f_WVWN01;g_WVWN01;s_
LOKART62	d_Bacteria;p_Planctomycetota;c_Planctomycetia;o_Pirellulales;f_T
_maxbin.042	hermoguttaceae;g;s
LOKART62	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_Desulfobacterale
_maxbin.047	s;f_UBA11574;g_;s_
LOKART62	d_Bacteria;p_BMS3Abin14;c_BMS3Abin14;o_BMS3Abin14;f_JA
_maxbin.048	FGWO01;g;s_
LOKART62	d_Bacteria;p_Chlamydiota;c_Chlamydiia;o_Chlamydiales;f_Rhabd
_maxbin.053	ochlamydiaceae;g_SZUA-160;s_SZUA-160 sp011065225
LOKART62	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_C00003060;f_
_maxbin.059	C00003060;g_;s_
LOKART62 _maxbin.065	d_Bacteria;p_Actinobacteriota;c_;o_;f_;g_;s_
LOKART62	d_Bacteria;p_Armatimonadota;c_UBA5377;o_UBA5377;f_JAAYI
_maxbin.070	R01;g_;s_
LOKART62 _maxbin.073	d_Bacteria;p_Aerophobota;c_Aerophobia;o_Aerophobales;f_AE-B3A;g_SOJT01;s_
LOKART62	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_Anaerolineales;f_E4
_maxbin.077	4-bin32;g_E44-bin32;s_

LOKART62	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_Anaerolineales;f_E4
_maxbin.088	4-bin32;g_;s_
LOKART62	d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_
_maxbin.095	_Hyphomicrobiaceae;g_Hyphomicrobium_A;s_
LOKART62	d_Bacteria;p_Acidobacteriota;c_Aminicenantia;o_Aminicenantales;f
_maxbin.108	_Aminicenantaceae;g_;s_
LOKART62	d_Bacteria;p_Chloroflexota;c_Dehalococcoidia;o_SM23-28-
_maxbin.111	2;f_RBG-16-64-32;g_CF-162;s_
LOKART62	d_Bacteria;p_Krumholzibacteriota;c_Krumholzibacteria;o_Krumholz
_maxbin.119	ibacteriales;f_Krumholzibacteriaceae;g_;s_
LOKART62 _maxbin.134	d_Bacteria;p_Zixibacteria;c_MSB-5A5;o_DSPP01;f_;g_;s_
LOKART62	d_Bacteria;p_Actinobacteriota;c_Humimicrobiia;o_Humimicrobiales;
_maxbin.136	f_Humimicrobiaceae;g_;s_
LOKART62	d_Bacteria;p_Dependentiae;c_Babeliae;o_Babeliales;f_CAIQNR01;
_maxbin.143	g_;s_
LOKART62 _maxbin.156	d_Bacteria;p_Fermentibacterota;c_Fermentibacteria;o_Fermentibacterales;f_Fermentibacteraceae;g_;s_
LOKART62 _maxbin.158	d_Bacteria;p_Chlamydiota;c_Chlamydiia;o_Chlamydiales;f_Rhabd ochlamydiaceae;g_SZUA-160;s_SZUA-160 sp011065235
LOKART62	d_Archaea;p_Thermoproteota;c_Bathyarchaeia;o_TCS64;f_TCS64;
_maxbin.164	g_RBG-16-57-9;s_
LOKART62	d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_
_maxbin.166	_Methyloligellaceae;g;s
LOKART62 _maxbin.188	d_Bacteria;p_Acidobacteriota;c_Vicinamibacteria;o_Bin61;f_;g_;s
LOKART62	d_Bacteria;p_Planctomycetota;c_Phycisphaerae;o_SM23-
_maxbin.191	33;f_SM23-33;g_;s_
LOKART62	d_Bacteria;p_Actinobacteriota;c_Acidimicrobiia;o_UBA5794;f_ZC
_maxbin.222	4RG35;g_SZUA-217;s_
LOKART62	d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_
_maxbin.226	_Methyloligellaceae;g_Methyloceanibacter;s_
LOKART62	d_Bacteria;p_Zixibacteria;c_MSB-
_maxbin.231	5A5;o_UBA10806;f_UBA10806;g_;s_
LOKART62 _metabat2.2 30	d_Bacteria;p_Dependentiae;c_Babeliae;o_Babeliales;f_Vermiphilac eae;g_;s_

LOKART62 _metabat2.6 4	d_Bacteria;p_Chloroflexota;c_Dehalococcoidia;o_Dehalococcoidales ;f_JAHJRU01;g_;s_
LOKART95	d_Bacteria;p_Chloroflexota;c_Dehalococcoidia;o_Dehalococcoidales
_maxbin.015	;f_UBA5760;g_UBA5760;s_
LOKART95 _maxbin.017	d_Bacteria;p_Patescibacteria;c_GCA-2792135;o_;f <u>;g</u> ;s_
LOKART95	d_Bacteria;p_Desulfobacterota;c_DSM-
_maxbin.019	4660;o_Desulfatiglandales;f_B25-G16;g_B33-G16;s_
LOKART95 _maxbin.030	d_Bacteria;p_Patescibacteria;c_Kazan-3B-28;o_Kazan-3B-28;f_UBA10110;g_;s_
LOKART95	d_Bacteria;p_Desulfobacterota;c_DSM-
_maxbin.039	4660;o_Desulfatiglandales;f_B25-G16;g_B33-G16;s_
LOKART95	d_Bacteria;p_Acidobacteriota;c_Thermoanaerobaculia;o_Thermoana
_maxbin.042	erobaculales;f_FEB-10;g_JACXWA01;s_
LOKART95	d_Bacteria;p_Chloroflexota;c_Dehalococcoidia;o_Dehalococcoidales
_maxbin.043	;f_E44-bin46;g_E44-bin46;s_
LOKART95	d_Bacteria;p_Actinobacteriota;c_Humimicrobiia;o_Humimicrobiales;
_maxbin.047	f_Humimicrobiaceae;g_M55B120;s_
LOKART95	d_Bacteria;p_Gemmatimonadota;c_Gemmatimonadetes;o_Longimicr
_maxbin.054	obiales;f_UBA6960;g;s_
LOKART95	d_Bacteria;p_Dependentiae;c_Babeliae;o_Babeliales;f_Vermiphilac
_maxbin.057	eae;g;s
LOKART95	d_Bacteria;p_BMS3Abin14;c_BMS3Abin14;o_BMS3Abin14;f_;g_
_maxbin.059	_;s_
LOKART95	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_C00003060;f_
_maxbin.061	C00003060;g;s_
LOKART95	d_Archaea;p_Asgardarchaeota;c_Thorarchaeia;o_Thorarchaeales;f_
_maxbin.080	Thorarchaeaceae;g_MP8T-1;s_
LOKART95	d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_
_maxbin.092	_Hyphomicrobiaceae;g_Hyphomicrobium_A;s
LOKART95 _maxbin.126	d_Bacteria;p_UBP14;c;o;f;g;s
LOKART95 _maxbin.129	d_Bacteria;p_4484-113;c_;o_;f_;g_;s_
LOKART95 _metabat2.1 54	d_Bacteria;p_Planctomycetota;c_Phycisphaerae;o_SM23- 33;f_SM23-33;g_;s_

LOKART95 _metabat2.1 69	d_Bacteria;p_Patescibacteria;c_UBA1384;o_XYA2-FULL-43- 10;f_XYA2-FULL-43-10;g_;s_
MARPAC23 _maxbin.015	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_GCA-001735895;f_GCA-001735895;g_GCA-001735895;s_
MARPAC23 _maxbin.026	d_Bacteria;p_Acidobacteriota;c_Vicinamibacteria;o_Vicinamibactera les;f_UBA8438;g_NP936;s_
MARPAC23 _maxbin.030	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_UBA6522;f_ _UBA6522;g_SMWC01;s_
MARPAC23 _maxbin.032	d_Bacteria;p_Acidobacteriota;c_UBA6911;o_RPQK01;f_;g_;s_
MARPAC23 _metabat2.4 6	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_UBA6522;f_ _UBA6522;g_SMWC01;s_SMWC01 sp004357505
MARPAC23 _metabat2.5 6	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomona dales;f_HTCC2089;g_SMWK01;s_SMWK01 sp004356895
SMMPAC33 _maxbin.009	d_Archaea;p_Halobacteriota;c_Methanosarcinia;o_Methanosarcinale s;f_Methanocomedenaceae;g_Methanocomedens;s_
SMMPAC33 _maxbin.028	d_Bacteria;p_Desulfobacterota;c_Desulfobacteria;o_C00003060;f_ C00003060;g_S7086C20;s_S7086C20 sp001751075
SMMPAC33 _maxbin.031	d_Archaea;p_Asgardarchaeota;c_Heimdallarchaeia;o_JABLTI01;f_ JABLTI01;g_JABLTI01;s_
SMMPAC33 _maxbin.049	d_Bacteria;p_Spirochaetota;c_Spirochaetia;o_DSM- 27196;f_SLST01;g_JAHOYX01;s_
SMMPAC33 _maxbin.077	d_Bacteria;p_Acidobacteriota;c_Aminicenantia;o_Aminicenantales;f _Aminicenantaceae;g_;s_
SMMPAC33 _maxbin.131	d_Bacteria;p_Planctomycetota;c_;o_;f_;g_;s_
SMMPAC33 _metabat2.1 04	d_Archaea;p_Asgardarchaeota;c_Heimdallarchaeia;o_JABLTI01;f_ JABLTI01;g_JABLTI01;s_
SMMPAC33 _metabat2.1 5	d_Bacteria;p_Desulfobacterota;c_DSM- 4660;o_Desulfatiglandales;f_Desulfatiglandaceae;g_;s_
SMMPAC33 _metabat2.1 90	d_Bacteria;p_QNDG01;c_QNDG01;o_QNDG01;f_DSAB01;g_;s_ -

SMMPAC33 _metabat2.5 6	d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_ _Methyloligellaceae;g_Methyloceanibacter;s_
SMMPAC46	d_Bacteria;p_WOR-3;c_WOR-3;o_SM23-42;f_SM23-
_maxbin.019	42;g_SOIY01;s_
SMMPAC46 _maxbin.024	d_Archaea;p_Halobacteriota;c_Methanosarcinia;o_Methanosarcinale s;f_Methanocomedenaceae;g_Methanomarinus;s_Methanomarinus sp002926195
SMMPAC46	d_Bacteria;p_Chloroflexota;c_Anaerolineae;o_Promineofilales;f_U
_maxbin.041	BA4811;g_GLR113;s_GLR113 sp013139935
SMMPAC46 _maxbin.052	d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_SZUA-229;f_SZUA-229;g_JABDQW01;s_
SMMPAC46	d_Bacteria;p_RBG-13-66-14;c_B26-
_maxbin.061	G2;o_WVWN01;f_WVWN01;g_WVWN01;s_

Tabla Suplementaria 2. Clústers de genes CAZyme (CGC) de los MAGs encontrados en sedimentos marinos.

ID	CGC (Módulos CAZyme)	Taxonomía( Class)
ART45_maxbin .050	GH23+CBM50	Desulfobulbi a
ART46_maxbin .023	GH103	Gammaprote obacteria
ART46_maxbin .029	CE4	Gammaprote obacteria
ART46_maxbin .029	GH23+CE11	Gammaprote obacteria
ART46_maxbin .051	GH23	Gammaprote obacteria
ART46_maxbin .055	GT4	UBA9160
ART46_maxbin .071	GH23+CBM50	Gammaprote obacteria
ART48_maxbin .026	GH23	Gammaprote obacteria
CIGOMA4_ma xbin.004	GH23	Methylomira bilia

CIGOMA4_ma xbin.004	GT4	Methylomira bilia
DELGOM57_ maxbin.002	GH23	Gammaprote obacteria
DSANT06_max bin.002	CBM51+GH2	Bacteroidia
DSANT06_max bin.002	CE14	Bacteroidia
DSANT06_max bin.016	GH3	Bacteroidia
DSANT06_max bin.024	GH23	Alphaproteo bacteria
DSANT06_max bin.029	GH3+CBM9	Acidimicrob iia
DSANT06_max bin.045	GH23	Gammaprote obacteria
DSANT08_max bin.002	GH29+GH163+CBM14+GH18+CBM54+GH92+CBM9 +CBM0+GH130+GH92+CBM47	Bacteroidia
DSANT08_max bin.006	GT4+GT4	Anaerolinea e
DSANT08_max bin.010	GT4	Gammaprote obacteria
DSANT08_max bin.020	GT4	Polyangia
DSANT08_max bin.030	GT51	Krumholzib acteria
DSANT09_max bin.031	GH3	Gammaprote obacteria
DSANT09_max bin.042	GH23+CBM50	Desulfobulbi a
DSANT11_max bin.001	GT51	Polyangia
DSANT11_max bin.005	GT4	Acidimicrob iia
DSANT11_max bin.021	GH29	DSWW01

DSANT11_max bin.021	GT2_Glycos_transf_2+GT2	DSWW01
DSANT11_max bin.026	CBM9	Aminicenant ia
DSANT95_max bin.006	CBM9	Gammaprote obacteria
DSANT95_max bin.007	CBM54+GH16_3+CBM4+GH16+CBM54+GH16_3+C BM4+GH16+GH2+GH97+CBM6+GH3	Bacteroidia
DSANT95_max bin.007	GH3	Bacteroidia
DSANT95_max bin.011	GH103	Alphaproteo bacteria
DSANT95_max bin.014	GH23	Alphaproteo bacteria
DSANT95_max bin.014	GT4	Alphaproteo bacteria
DSANT95_max bin.014	GT51	Alphaproteo bacteria
DSANT95_max bin.016	CE14	Bacteroidia
DSANT95_max bin.016	GH109	Bacteroidia
DSANT95_max bin.016	GH16_3+GH16+GH13_38+GH65	Bacteroidia
DSANT95_max bin.016	GT2_Glycos_transf_2+GT2	Bacteroidia
DSANT95_max bin.024	GH0+GH109	Bacteroidia
DSANT95_max bin.024	GH117+CBM13+GH29+GH2	Bacteroidia
DSANT95_max bin.024	GH29+GH110	Bacteroidia
DSANT95_max bin.024	GT4	Bacteroidia
DSANT95_max bin.026	GH67+GH2+CBM6+CBM7+CBM6+GH3+CE17+GH4 3+GH43_12	Clostridia
DSANT95_max bin.035	CE4	Fusobacterii a
-------------------------	-----------------------------------	----------------------
DSANT95_max bin.035	GH2+CBM67	Fusobacterii a
DSANT95_max bin.035	GH23	Fusobacterii a
DSANT95_max bin.035	GT83	Fusobacterii a
DSANT95_max bin.044	CBM32+GH2+CBM71	Bacteroidia
DSANT95_max bin.044	CBM32+GH43+GH43_28+GH144+CBM6+GH3	Bacteroidia
DSANT95_max bin.044	CBM50+GH23+GT19+GH5_2+GH5+CE8	Bacteroidia
DSANT95_max bin.044	GH2	Bacteroidia
DSANT95_max bin.044	GH2	Bacteroidia
DSANT95_max bin.050	GT51	Clostridia
DSANT95_max bin.051	GH16_3+GH13_38+GH65	Bacteroidia
DSANT95_max bin.051	GT4	Bacteroidia
DSANT95_max bin.073	GH23+CBM50	Actinomycet ia
DSANT95_max bin.083	CE4	Negativicute s
DSANT95_max bin.083	CE4+GH153	Negativicute s
DSANT95_max bin.083	GT2_Glycos_transf_2+GT2	Negativicute s
DSANT95_max bin.085	CBM32+GH2	Verrucomicr obiae
DSANT95_met abat2.56	GH3	Actinomycet ia

DSANT95_met abat2.56	GT4	Actinomycet ia
GBGOC44_ma xbin.033	GH3	Anaerolinea e
GBGOC44_ma xbin.033	GT4	Anaerolinea e
GBGOC44_ma xbin.050	GT4	Bacteroidia
GBGOC44_ma xbin.067	GT4	Anaerolinea e
GBGOC44_ma xbin.067	GT51	Anaerolinea e
GBGOC44_ma xbin.089	GH103	Desulfobact eria
GBGOC44_ma xbin.089	GH23	Desulfobact eria
GBGOC44_ma xbin.101	GH23+CBM50	DSM-4660
GBGOC44_ma xbin.110	GH23	DSM-4660
GBGOC44_ma xbin.110	GT4+GT4	DSM-4660
GBGOC44_ma xbin.112	GT4+GT4	Koll11
GBGOC44_ma xbin.122	GT4+CE14	Holophagae
GBGOC44_ma xbin.125	GT4	Methanosarc inia
GBGOC44_ma xbin.128	GH29	Bacteroidia
GBGOC44_ma xbin.137	GT4	Hydrothermi a
GBGOC44_ma xbin.142	GT51	S145-22
GBGOC44_ma xbin.142	GT83	S145-22

GBGOC44_ma xbin.152	CE4	UBA2214
GBGOC44_ma xbin.152	GH2	UBA2214
GBGOC44_ma xbin.152	GH3+GH28	UBA2214
GBGOC44_ma xbin.152	GH92+GH3+GH3+CBM6	UBA2214
GBGOC44_ma xbin.167	GT4	Campylobac teria
GBGOC44_ma xbin.172	GH23	Krumholzib acteria
GBGOC44_ma xbin.177	GH8+GH2	Bacteroidia
GBGOC44_ma xbin.177	GT51	Bacteroidia
GBGOC44_ma xbin.198	GT4	Paceibacteri a
GBGOC44_ma xbin.212	GT4	Bathyarchaei a
GBGOC44_ma xbin.218	GT2_Glycos_transf_2+GT2	Campylobac teria
GBGOC44_met abat2.230	CBM9	JdFR-76
GBGOC44_met abat2.230	CE4	JdFR-76
GBGOC44_met abat2.279	GH23	Desulfobacc ia
GBGOC44_met abat2.293	CBM67+GH2	Aminicenant ia
GBGOC44_met abat2.293	GH3+CBM6	Aminicenant ia
GBGOC44_met abat2.72	GT51	Syntrophoba
ubut2172		cteria

GBGOC45_ma xbin.089	GH23+GH23	DSM-4660
GBGOC45_ma xbin.118	GH33+GH0+CBM6+GH3	Aminicenant ia
GBGOC45_ma xbin.140	GT51	Campylobac teria
GBGOC45_ma xbin.149	GH3	Anaerolinea e
GBGOC45_ma xbin.151	CBM9	JdFR-76
GBGOC45_ma xbin.151	CE4	JdFR-76
GBGOC45_ma xbin.166	GH3+GH3	Bathyarchaei a
GBGOC45_ma xbin.199	GT51	UBA6098
GBGOC45_met abat2.227	GH2+GH8	Bacteroidia
GBGOC45_met abat2.287	CBM50+GH23	DSM-4660
GBGOC45_met abat2.287	GT51	DSM-4660
GBGOC45_met abat2.54	CBM50+GH23	MSB-5A5
GBGOC47_ma xbin.017	GH2	Bacteroidia
GBGOC47_ma xbin.043	GH23	Spirochaetia
GBGOC47_ma xbin.058	GH103	Campylobac teria
GBGOC47_ma xbin.081	CBM50+GH23	DSM-4660
GBGOC47_ma xbin.083	GH130+GH2+GH38	JS1
GBGOC47_ma xbin.089	GH2	Spirochaetia

GBGOC47_ma xbin.089	GH3	Spirochaetia
GBGOC47_ma xbin.089	GT51	Spirochaetia
GBGOC47_ma xbin.091	GH3	Anaerolinea e
GBGOC47_ma xbin.098	GH2	DSM-4660
GBGOC47_ma xbin.107	GH23+CBM50	Bacteroidia
GBGOC47_ma xbin.107	GH92+GH109	Bacteroidia
GBGOC47_ma xbin.108	CE4	Bacilli
GBGOC47_ma xbin.108	GH2	Bacilli
GBGOC47_ma xbin.120	GT51	Anaerolinea e
GBGOC47_ma xbin.126	GH23	Desulfobact eria
GBGOC47_ma xbin.131	GH103	Desulfurom onadia
GBGOC47_ma xbin.167	GH23	Campylobac teria
GBGOC47_ma xbin.217	GH23+CBM50	Desulfurom onadia
GBGOC47_ma xbin.224	GT83	Bradymonad ia
GBGOC47_ma xbin.325	GH103	Desulfobulbi a
GBGOC47_met abat2.403	GH2	UBA2214
GBGOC47_met abat2.413	GH3	UBA2214
GBGOC47_met abat2.96	CBM67+GH2	Aminicenant ia

HRSPAC47_m axbin.033	GT51	Clostridia
HRSPAC47_m axbin.034	CBM50+GH23	Desulfurom onadia
HRSPAC47_m etabat2.19	GH2	Bacteroidia
HRSPAC47_m etabat2.53	CE4	Gammaprote obacteria
HRSPAC47_m etabat2.53	GH23	Gammaprote obacteria
KJ6_maxbin.00 7	GH23+CBM50	MSB-5A5
KJ6_metabat2. 15	GH23+CBM50	JACQVX01
KOR58_maxbi n.012	CE4	Gammaprote obacteria
KOR58_maxbi n.012	GH102+GH23	Gammaprote obacteria
KOR58_maxbi n.012	GH103	Gammaprote obacteria
KOR58_maxbi n.012	GH23	Gammaprote obacteria
KOR58_maxbi n.012	GT51	Gammaprote obacteria
KOR58_metab at2.14	GH103	Alphaproteo bacteria
LOKART30_m axbin.040	GH95+GH29	Gammaprote obacteria
LOKART30_m axbin.040	GT51	Gammaprote obacteria
LOKART30_m axbin.045	GH3+GH163+GH95	Phycisphaer ae
LOKART30_m axbin.065	GT51	UBA9160
LOKART30_m axbin.069	GT51	B62-G9

LOKART30_m axbin.081	GH23	BMS3Abin1 4
LOKART30_m axbin.084	GH23	SURF-5
LOKART30_m axbin.085	GH23	UBA7883
LOKART30_m axbin.085	GT51	UBA7883
LOKART30_m axbin.087	GH2	Bacteroidia
LOKART30_m axbin.087	GH23	Bacteroidia
LOKART30_m axbin.092	GT51	Glassbacteri a
LOKART30_m axbin.099	CBM67+CBM42+GH2	Planctomyce tia
LOKART30_m axbin.099	GH2	Planctomyce tia
LOKART30_m axbin.109	GH2+GH123+GH123+GH123	Bacteroidia
LOKART30_m axbin.109	GH3	Bacteroidia
LOKART30_m axbin.109	GT51	Bacteroidia
LOKART30_m axbin.114	GH29	Hydrogened entia
LOKART30_m axbin.114	GT51	Hydrogened entia
LOKART30_m axbin.146	GT51	Unknown
LOKART30_m axbin.177	GH23+CBM50	Desulfobact eria
LOKART30_m etabat2.139	GH109+GH130	Bacteroidia
LOKART30_m etabat2.139	GH29+CBM13	Bacteroidia

LOKART60_m axbin.034	GH3	Anaerolinea e
LOKART60_m axbin.034	GH3+CBM6+GH2+GH4+GH130	Anaerolinea e
LOKART60_m axbin.062	CE4+GH148	UBA5377
LOKART60_m axbin.062	GH23	UBA5377
LOKART62_m axbin.070	CE4+GH148	UBA5377
LOKART62_m axbin.077	GH3	Anaerolinea e
LOKART62_m axbin.108	GH23	Aminicenant ia
LOKART95_m axbin.042	GH20+GH3+CBM2+CBM6	Thermoanae robaculia
MARPAC23_m axbin.015	GH103	Gammaprote obacteria
MARPAC23_m axbin.026	GT51	Vicinamibac teria
MARPAC23_m axbin.026	GT83	Vicinamibac teria