

*Análisis de la ecología microbiana del
bacterioplancton de los océanos
mediante secuenciación de genomas e
implicación de Sulfitobacter sp. EE-36*

*Analysis of the microbial ecology of
bacterioplancton of the oceans by
genomes sequencing and implication
of Sulfitobacter sp. EE-36*

Airam Pedro Benítez García

Tutorizado por Dr. José Manuel González Hernández

Trabajo de Fin de Grado en Biología

Junio 2017

Agradecimientos

Al Dr. José M. González, por ayudarme y guiarme en todo momento, además de enseñarme más acerca de una de las áreas más interesantes de la Biología a través de este trabajo. También a Iris Manzano, por colaborar y por su disposición a ayudar en todo aquello que necesitara.

A Martina y Germán, mis padres, y a Delioma, por todo el apoyo recibido, no solo durante la realización de este trabajo, sino durante estos 4 años. Gracias por confiar en mí y haber sido mis pilares para llegar hasta aquí.

Índice

Resumen	1
Abstract	1
1. Introducción	2
1.1. Implicación del bacterioplancton en el ciclo de los elementos	2
1.2. Importancia de la metagenómica en la investigación del bacterioplancton	3
1.3. El plancton bacteriano y el ciclo del azufre.....	4
1.4. Interés de estudio de <i>Sulfitobacter</i> y muestreo de la Expedición Malaspina	8
2. Objetivos.....	9
3. Material y métodos	10
4. Resultados.....	12
4.1. Genomas de referencia	13
4.2. Anotaciones resultantes de alineamientos y comparaciones BLAST	14
4.3. Comparación y homología de contigs con <i>Sulfitobacter</i> sp. EE-36.....	16
5. Discusión.....	18
6. Conclusiones	19
6. Conclusions.....	21
7. Bibliografía	23

Resumen

En los últimos años, los estudios acerca de la ecología del bacterioplancton oceánico han generado gran interés en la comunidad científica debido a su relevante implicación en el ciclo de los elementos de la Tierra. Esta intervención ha comenzado a ser explicada gracias a los avances en las técnicas de secuenciación y análisis de secuencias de ácidos nucleicos, tales como la metagenómica, por la cual se ha podido obtener información sobre rutas metabólicas de vital importancia en miembros del plancton bacteriano cuyo estudio mediante cultivo es inviable. En consecuencia, el interés por conocer cómo se produce la intervención de estas vías metabólicas en los ciclos biogeoquímicos ha aumentado, pudiendo llegar a aplicarse estos conocimientos en algunas áreas de la biotecnología ambiental tales como la biorremediación o la degradación de desechos. *Sulfitobacter* sp. EE-36 despierta este interés debido a la implicación de algunas de sus rutas metabólicas dentro del ciclo del azufre. En el presente trabajo se analiza la presencia de genes así como su anotación funcional, mediante herramientas bioinformáticas, encargados de codificar productos capaces de degradar compuestos orgánicos del azufre, como el DMSP, o inorgánicos, tales como el tiosulfato, en muestras ambientales recogidas en distintos puntos del océano profundo.

Abstract

In recent years, studies on the ecology of oceanic bacterioplankton have aroused great interest in the scientific community due to their important role in the cycle of the elements in Earth. This intervention has started to be explained thanks to the advances in metagenomics, which has allowed to obtain information about metabolism in members of bacterioplankton whose study by means of cultivation is difficult. Consequently, the interest in knowing how the intervention of these metabolic pathways occurs in the biogeochemical cycles has increased, being able to apply this knowledge in environmental biotechnology such as bioremediation or waste degradation. This interest is raised by *Sulfitobacter* sp. EE-36 due to the implication of some of its metabolic pathways within the sulfur cycle. The present project analyzes the presence of genes as well as their functional annotation, using bioinformatic tools, responsible for coding products capable of degrading organic sulfur compounds, such as DMSP, or oxidation of inorganic, such as thiosulfate, in environmental sequences from samples of the deep ocean.

Palabras clave: *Sulfitobacter*, metabolismo, metagenómica, bioinformática, ciclo del azufre

1. Introducción

El bacterioplancton oceánico constituye una gran parte de la biomasa que alberga ecosistemas tan diversos como los existentes en los océanos, y reúne características bioquímicas y ecofisiológicas muy diferentes a los microorganismos terrestres. Este grupo, que constituye especies de microorganismos pertenecientes tanto a *Archaea* como a *Bacteria*, presenta diversos tipos de metabolismo que tienen gran importancia sobre los ciclos de los elementos del planeta, puesto que estas bacterias son capaces de llevar a cabo transformaciones biogeoquímicas exclusivas. El plancton bacteriano oceánico realiza fases fundamentales de los ciclos biogeoquímicos marinos que tienen gran influencia sobre los ambientes marinos y terrestres. Entre los ciclos donde el bacterioplancton lleva una destacada participación, se pueden remarcar el ciclo del carbono, del oxígeno, del nitrógeno y del azufre (González, Pedrós-Alió, & Gasol, 2008).

1.1. Implicación del bacterioplancton en el ciclo de los elementos

Según diversos estudios (González et al., 2008; R.P. Kiene, Linn, & Bruton, 2000), los miembros del bacterioplancton son los responsables de llevar a cabo aproximadamente el 95% de la respiración que se produce en los océanos debido a que éstos oxidan gran parte de la materia orgánica a CO_2 , convirtiéndose ésta en nutrientes disponibles a los demás organismos de la cadena trófica (Teoría del “bucle microbiano” de Pomeroy, 1974). Además, también son capaces de fijar el CO_2 atmosférico, el principal gas causante del efecto invernadero (mediante la realización de la fotosíntesis por parte de las bacterias fotosintéticas y algas unicelulares), y de desprender metano por parte de las bacterias metanógenas (CH_4). Con respecto a su efecto sobre el nitrógeno, este grupo es capaz de metabolizar el N_2 atmosférico (únicos seres vivos capaces de hacerlo), imprescindible para la fotosíntesis, en un producto asimilable por los seres vivos como la forma amino (NH_3). En condiciones anóxicas, las bacterias respiran nitrato (NO_3), y emiten óxido nitroso (NO_2) o bien N_2 . De igual manera se produce la asimilación de fosfatos y hierros, depositados principalmente por el polvo atmosférico sobre la superficie de los océanos, e imprescindibles para el funcionamiento de las bacterias fotosintéticas y nitro fijadoras. También utilizan ciertos gases como fuente de energía, tales como hidrógeno atmosférico (H_2), ácido sulfhídrico (H_2S) y monóxido de carbono (CO), y son capaces de asimilar algunos metales que en altas concentraciones pueden ser contaminantes en el medio (aparte del hierro, algunos como el manganeso, cobre, níquel, etc.). Estos son algunos ejemplos que ilustran que el metabolismo del plancton bacteriano

tiene un papel relevante sobre el ciclo de los elementos. Sin embargo, esto no sucede de igual manera a lo largo de la columna de agua oceánica, pues existen diversos factores a lo largo de ésta que impiden el crecimiento bacteriano óptimo (González et al., 2008; Petri, Podgorsek, & Imhoff, 2001).

En los últimos años se han venido descubriendo nuevas rutas metabólicas llevadas a cabo por miembros del plancton bacteriano, las cuales tienen grandes repercusiones ecológicas en la biosfera, y que incluso, pueden tener importantes aplicaciones biotecnológicas (Friedrich et al., 2005; González et al., 2002; Kiene et al., 2000). Este hecho radica en gran parte a la presión evolutiva a la que se encuentran sometidos los microorganismos marinos, pues los océanos constituyen un medio oligotrófico, sobre todo en las zonas profundas, donde los nutrientes escasean (tales como el hierro, el nitrógeno o el fósforo), y factores como la luz, la temperatura, la salinidad, el pH y la presión hidrostática son críticos a lo largo de la columna de agua oceánica, e influyen en la limitación de la producción marina. Por ello, dado que las bacterias oceánicas se encuentran en abundancia, éstas se ven obligadas a adaptarse a estas estrictas condiciones cambiantes desarrollando distintas estrategias biológicas entre las que se incluyen el aumento de la diversidad metabólica y fisiológica, y consecuentemente, de la diversidad genética (González et al., 2008).

1.2. Importancia de la metagenómica en la investigación del bacterioplancton

Muchas de estas estrategias se han ido investigando gracias a los avances de los análisis genéticos, según los cuáles podemos identificar los péptidos implicados en las rutas que llevan a cabo la transformación de diversos productos. Una de las técnicas más utilizadas en la actualidad para analizar las enzimas del plancton bacteriano oceánico, puesto la gran dificultad que conlleva el cultivo y análisis de éstas especies en el laboratorio, es la metagenómica (Curson et al., 2008; González et al., 2008). Esta técnica lleva a cabo la secuenciación de muestras ambientales, recogiendo los genes de aquellas especies que componen el bacterioplancton a modo de “fotografía” del material genético de esas bacterias, por lo que se analiza en una visión general y no específicamente en una sola bacteria o un solo gen, pudiendo observar la función potencial de toda la comunidad bacteriana, así como detalles que con el uso de un cultivo se podrían perder (Alguacil Martín, 2010; Schloss & Handelsman, 2003; Streit & Schmitz, 2004). Para realizar este tipo de análisis, en primer lugar se realiza una extracción del ADN crudo de la comunidad planctónica bacteriana. La

muestra pasa por un proceso de prefiltración para eliminar los organismos eucariotas que no son de interés así como contaminantes de la muestra, y se realiza la lisis celular así como la extracción del ADN microbiano, el cual es cortado. A continuación, se lleva a cabo la secuenciación mediante distintos sistemas de procesado de alto rendimiento, en los que se obtienen distintas secuencias de ADN las cuáles pasan por un proceso de “scaffolding” o ensamblaje, resultando secuencias ensambladas denominadas “contigs”. El análisis de estos contigs mediante el tratamiento con software bioinformáticos, nos puede dar la idea del mapa físico del genoma de una bacteria específica o del metagenoma de la comunidad (Handelsman, 2005; Riesenfeld et al., 2004; Schloss & Handelsman, 2003; Streit & Schmitz, 2004).

1.3. El plancton bacteriano y el ciclo del azufre

El ciclo del azufre es uno de los ciclos biogeoquímicos más importantes, donde el bacterioplancton toma un papel de especial relevancia en las transformaciones marinas del azufre y su intercambio con la atmósfera. La interacción de las bacterias marinas con el ciclo del azufre se realiza tanto a nivel orgánico como inorgánico, cuyas rutas metabólicas se han ido descubriendo en los últimos años (Friedrich et al., 2005; González et al., 2002)

A nivel orgánico las bacterias llevan a cabo una de las transformaciones más importantes del ciclo tal como es la producción biológica de dimetilo de sulfuro (DMS) a partir de la degradación de dimetilsulfoniopropionato (DMSP), utilizado como fuente de carbono y/o para la obtención de la energía (Fig. 1). El DMS, producto volátil de esta transformación enzimática bacteriana, supone el principal compuesto de intercambio de azufre entre el medio marino y el medio terrestre, pues éste se desprende a la atmósfera una vez que es descompuesto desde su precursor, el DMSP, y es protagonista del olor a mar presente en el aire del océano (Andrew R. J. Curson, Todd, Sullivan, & Johnston, 2011; González et al., 2008; Johnston, Green, & Todd, 2016; R.P. Kiene et al., 2000; Reisch, Moran, & Whitman, 2011). El DMSP se encuentra en gran medida en los ecosistemas marinos y es sintetizado por el fitoplancton marino (dinoflagelados, diatomeas o coccolitóforos), algas verdes y rojas, algunas plantas costeras e incluso algunos tipos de corales. El fitoplancton hace uso de este compuesto como un regulador osmótico para contrarrestar los efectos salinos, y éste se libera al medio externo cuando se produce la lisis celular, siendo susceptible del ataque microbiano (Dickschat, Rabe, & Citron, 2015; Stefels, 2000). El DMS desprendido a la atmósfera tiene un impacto climático importante, pues éste

es oxidado por la radiación solar ultravioleta, formando aerosoles de sulfato que condensan la humedad y provoca la formación de nubes. La consecuencia de este hecho es la disminución del impacto de la radiación solar sobre la superficie terrestre debido a un consecuente aumento del efecto albedo, lo que causa una disminución de las temperaturas. Por tanto, la emisión de DMS tiene un gran impacto climático sobre la biosfera, disminuyendo el efecto invernadero (González et al., 2002, 2008; Ronald P. Kiene & Linn, 2000; Reisch et al., 2011; Simó, 2001).

Existen dos vías para la degradación bacteriana del DMSP: la vía donde se forma DMS debido a que las bacterias utilizan el DMSP como fuente de carbono; y la vía donde las bacterias utilizan el DMSP como fuente de azufre debido a que es energéticamente más eficiente que usar otros compuestos, tales como el sulfato (SO_4) (Curson et al., 2011).

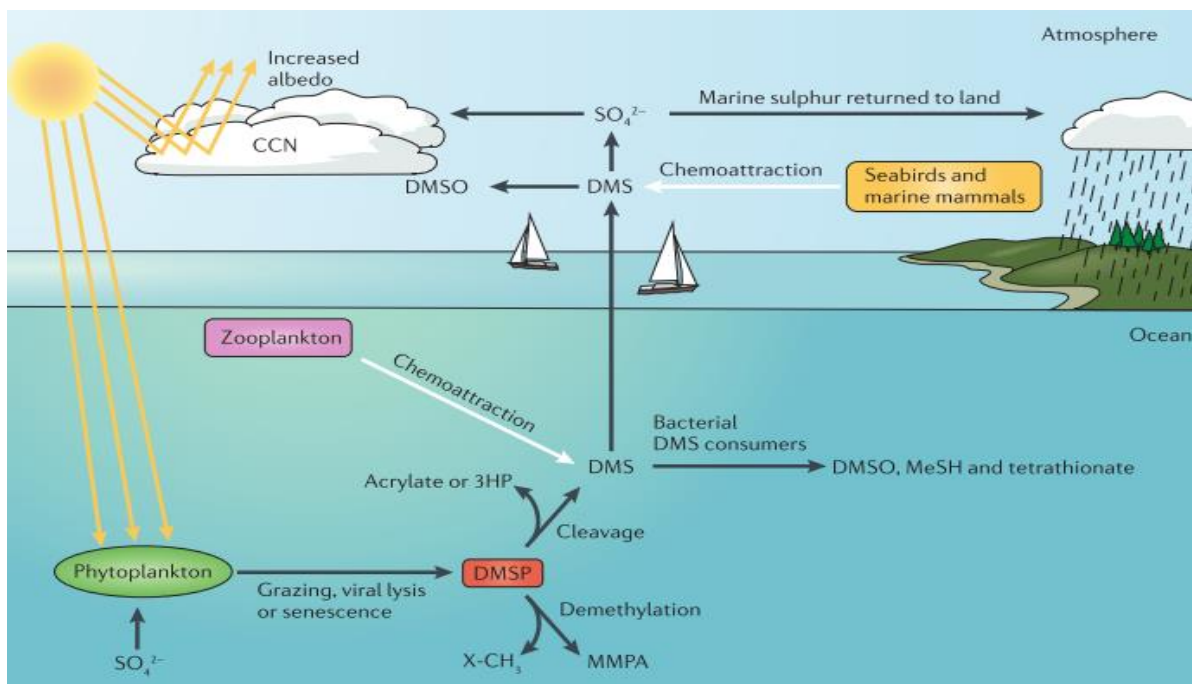


Figura 1. Ciclo del DMSP en los océanos. El DMSP liberado por el fitoplancton es atacado por el plancton bacteriano, cuyos miembros lo catabolizan de dos maneras: mediante la desmetilación o mediante la degradación hasta DMS. Este último puede ser consumido nuevamente por las bacterias y formar compuestos orgánicos e inorgánicos del azufre, o liberado a la atmósfera donde es transformado en dimetilsulfóxido (DMSO) que actúa como núcleo de condensación de nubes. Además, el DMS también funciona como quimioatrayente para el zooplancton así como para mamíferos y aves marinas. Tomado de Curson et al. (2011).

Este mecanismo ha evolucionado y existe gran diversidad con respecto a esta vía metabólica (organismos y enzimas que la llevan a cabo), aunque los últimos estudios destacan 2 rutas principales (Fig. 2): haciendo uso de una liasa de DMSP (degradación del DMSP hasta DMS) o haciendo uso de una demetilasa de DMSP, que lleva a cabo una desmetilación del DMSP para producir metilmercaptopropionato (MMPA). Los genes responsables de codificar

estas enzimas se encuentran distribuidos en el filo Proteobacteria, especialmente en la clase Alphaproteobacteria (González et al., 1999; Kiene et al., 2000; Kiene & Linn, 2000; Sun et al., 2012). Los genes que codifican para las liasas (enzimas Ddd) se denominan genes *ddd* y existen diversos tipos que codifican distintas liasas. La proteína DddD (gen *dddD*) es un tipo de liasa que genera DMS junto a 3-hidroxipropionato (3-HP), la cual posee una disposición distinta, típica de los dominios CoA-transferasa de clase III dispuestos en tándem, probablemente debido a que forma CoA derivado del DMSP. DddL, DddP, DddQ, DddW y DddY (genes *dddL*, *dddP*, *dddq*, *dddW*, y *dddY* respectivamente) son liasas que generan DMS junto a acrilato. DddP pertenece a la familia de las proteasas M24 que se encarga de romper la unión entre el azufre y el carbono, además de no poseer cofactores metálicos; sin embargo DddL, DddQ y DddW poseen en el extremo terminal carboxilo lugares de unión a metales denominados “cupins”; DddY, al contrario que las demás, es una proteína que se encuentra ligada al periplasma bacteriano por el extremo terminal amino, en lugar de encontrarse en el citoplasma. La enzima desmetilasa (DmdA) está codificada por un gen llamado *dmdA*, y pertenece al grupo de proteínas del sistema de la degradación de la glicina T (GcvT) (Curson et al., 2011; Johnston et al., 2016; Ronald P. Kiene & Linn, 2000; Reisch et al., 2011).

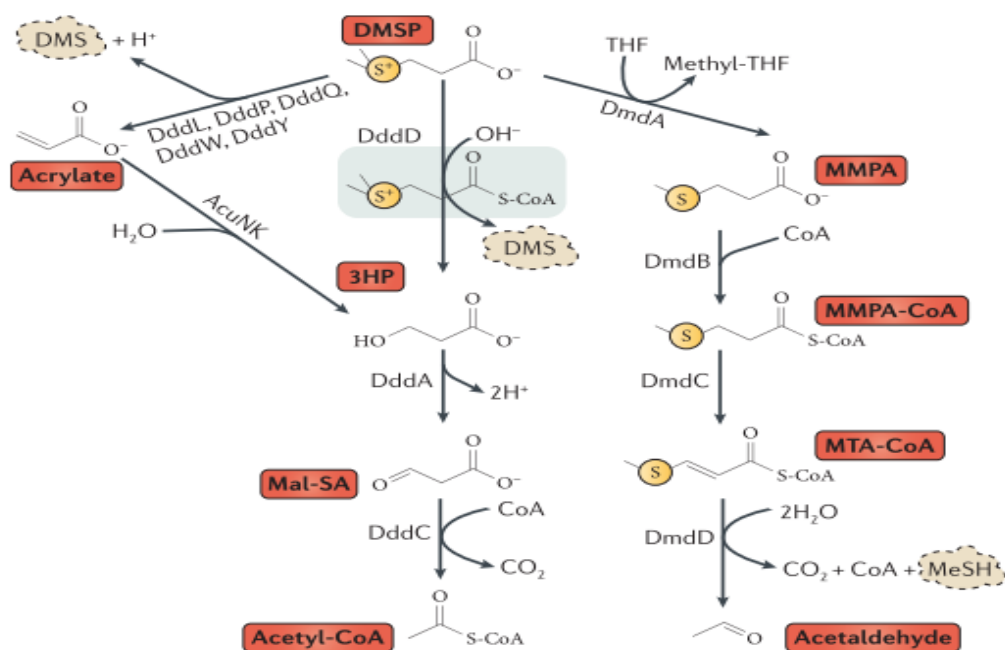


Figura 2. Vías metabólicas del DMSP. Existen 2 mecanismos por los que el DMSP es catabolizado. En una vía, la enzima DmdA se encarga de llevar a cabo la desmetilación del DMSP hasta MMPA, el cual puede ser degradado posteriormente hasta acetaldehído. En la otra vía, las enzimas DddL, DddP, DddQ, DdW o DddY pueden degradar el DMSP hasta acrilato, que posteriormente puede ser degradado hasta 3-HP; o bien pueden degradar el DMSP directamente hasta este último compuesto mediante la enzima DddD. En ambos casos resulta DMS como resultado de las oxidaciones. El 3-HP es luego degradado posteriormente hasta acetil coenzima A (Acetil-CoA). Tomado de Curson et al. (2011).

En relación a la interacción del bacterioplancton con el ciclo del azufre a nivel inorgánico, cabe destacar que las bacterias son capaces de transformar muchos compuestos reducidos inorgánicos del azufre presentes en el medio marino, probablemente procedentes de la degradación del DMSP y también debido a la actividad antropogénica, origen del aumento de la concentración de este tipo de compuestos. Existen dos vías conocidas (Fig. 3) para la oxidación de estos compuestos mediante bacterias heterótrofas (más abundantes que las autótrofas): mediante la producción de tetrionato ($S_4O_6^{2-}$) o mediante la producción de sulfato (SO_4^{2-}) (Friedrich et al., 2005; Friedrich, Rother, Bardischewsky, Ouentmeier, & Fischer, 2001; González et al., 2002).

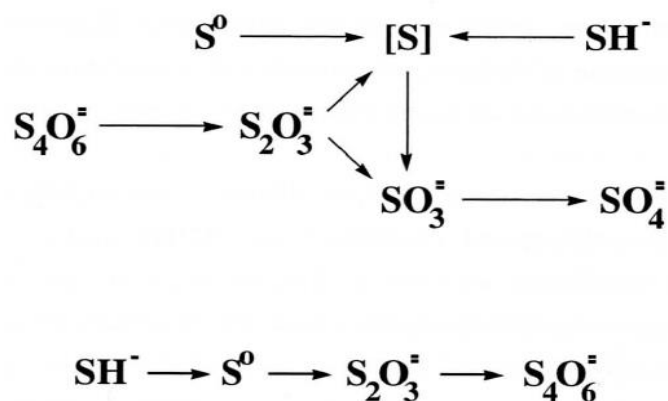


Figura 3. Vías metabólicas de los compuestos inorgánicos del azufre del plancton bacteriano. En la vía productora de sulfato, el tetrionato es reducido a tiosulfato ($S_2O_3^{2-}$) y posteriormente es oxidado para dar lugar a azufre unido a transportadores tales como proteínas (simbolizado como [S]) o bien para ser oxidado a sulfito (SO_3^{2-}) y en último lugar a sulfato. El tiosulfato, el azufre elemental (S^0) o algunos sulfuros pueden originar tetrionato. Estas rutas no siempre siguen una trayectoria lineal, sino que se combinan. Tomado de González et al. (2002).

Los genes relacionados con la oxidación de compuestos reducidos inorgánicos del azufre son diversos. Sin embargo, los últimos estudios (Friedrich et al., 2005, 2001; Lenk et al., 2012) destacan un grupo de genes denominados *sox* encargados de codificar una serie de proteínas que conforman el complejo enzimático periplásmico Sox (Fig. 4), el cual lleva a cabo la oxidación del azufre en distintas especies del bacterioplancton (muy común en Proteobacteria y Chlorobi aunque existe gran diversidad con respecto a la filogenia de las bacterias oxidadoras del azufre). Las proteínas que conforman este complejo y son codificadas por el cluster de genes *soxXYZABCD* se denominan SoxXA (proteína heterodimérica tipo citocromo C), SoxYZ (proteína heterodimérica que une covalentemente el tiosulfato), SoxB (proteína monomérica que contiene dos átomos de manganeso) y SoxCD (proteína heterotetramérica compuesta por un heterodímero que contiene un cofactor de molibdeno y otro que es de tipo Citocromo C). Este complejo enzimático se encarga de oxidar el tiosulfato ($S_2O_3^{2-}$) y el sulfito (SO_3^{2-}) produciendo a su vez sulfato, aunque también puede

ser capaz de oxidar el ácido sulfhídrico (H_2S) (Espinoza, Revah, & Le Borgne, 2010; Friedrich et al., 2005, 2001; Lenk et al., 2012; MetaCyc, 2014; Meyer, Imhoff, & Kuever, 2007; Petri et al., 2001).

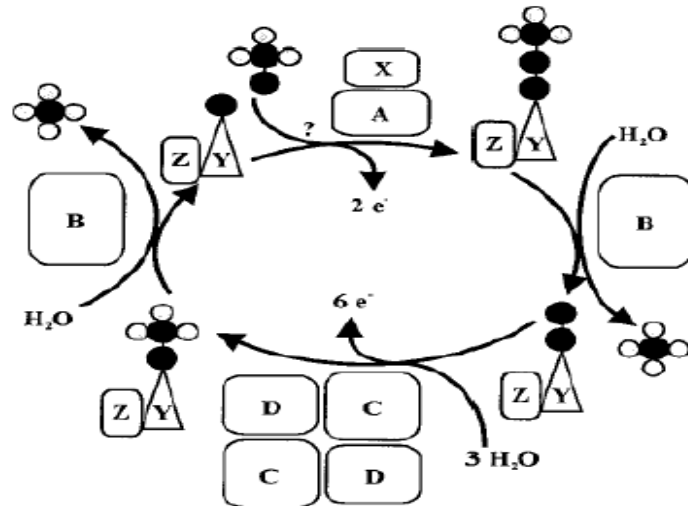


Figura 4. Modelo del funcionamiento del complejo Sox. Los compuestos son representados con círculos negros para los átomos de azufre, y círculos blancos para los de oxígeno. En este caso, se representa la oxidación del tiosulfato por parte de Sox para reducir el citocromo C. El grupo sulfano del tiosulfato se une al grupo tiol del residuo de cisteína de SoxYZ mediante puentes disulfuro a través de la catálisis del complejo SoxXA. A continuación, el grupo sulfona del tiosulfato (R-SO_3^{2-}) es oxidado a sulfato por SoxB. El grupo sulfano es regenerado en grupo sulfona por el complejo Sox(CD)₂ y vuelto a hidrolizar por SoxB a sulfato, regenerándose de nuevo el complejo SoxYZ. De igual manera pueden ser hidrolizados en este complejo el H_2S y SO_3^{2-} . El tioperoximonosulfato, producto de degradación del $\text{S}_4\text{O}_6^{2-}$ por la tetrionato hidrolasa, puede ser catabolizado por este sistema. Tomado de Friedrich et al. (2001).

1.4. Interés de estudio de *Sulfitobacter* y muestreo de la Expedición Malaspina

Según los recientes trabajos (A. R J Curson et al., 2008; González et al., 1999; Meyer et al., 2007; Reisch et al., 2011; Sun et al., 2012) sobre las implicaciones ecológicas de los miembros del bacterioplancton, uno de los géneros que más abundan en el plancton bacteriano oceánico, especialmente en zonas profundas (Kumbhare et al., 2015), y que tiene especial importancia en el ciclo del azufre es *Sulfitobacter*. Las bacterias de este género pertenecen a la clase Alphaproteobacteria, y se encuentran incluidas dentro de la familia *Rhodobacteraceae* en el orden *Rhodobacterales*. Son bacterias aeróbicas, heterótrofas, Gram-negativas y halofílicas moderadas (Roseobase, 2017). Este género, descubierto por primera vez por Sorokin en 1995 en el Mar Negro, está compuesto actualmente por 10 especies reconocidas, que han sido observadas en diversas aguas del planeta. Este género fue caracterizado en primer lugar por llevar a cabo la oxidación del sulfito para la obtención de energía (Sorokin, 1995), y es común encontrar este tipo de especies en zonas donde abundan los compuestos de azufre inorgánico (Roseobase, 2017). Sin embargo, desde entonces se han ido descubriendo nuevas rutas metabólicas en miembros de este género que implican nuevas

funciones ecofisiológicas en las que el azufre es el principal protagonista (A. R J Curson et al., 2008; Lenk et al., 2012). Estas nuevas rutas no solo se han descubierto en las especies caracterizadas, sino en cepas del género que aún no han sido reconocidas como especie. Buen ejemplo de ello es la cepa *Sulfitobacter* sp. EE-36, destacada por su gran metabolismo oxidativo de los compuestos de azufre inorgánico, tales como el tiosulfato y el sulfito, con el fin de obtener energía para el crecimiento; además de ser capaz de oxidar algunos compuestos orgánicos de azufre tales como DMSP y la glicina betaína, utilizados como única fuente de carbono (A. R J Curson et al., 2008; Reisch et al., 2011; Roseobase, 2017; Sun et al., 2012). Esta estirpe ha generado un gran interés debido a las posibles implicaciones ecológicas y biotecnológicas que estas rutas poseen, pues éstas pueden ser aplicadas a la tecnología de degradación de desechos y compuestos contaminantes (González et al., 2002, 1999). Por ello, *Sulfitobacter* sp. EE-36 será la protagonista del presente trabajo, dado que a través del uso de herramientas de análisis bioinformático se investigará si existe en este miembro del plancton bacteriano la presencia de los genes responsables de la síntesis de las enzimas implicadas en las rutas metabólicas nombradas para distintas muestras ambientales de aguas de océano profundo. En el caso de la anotación de los genes de degradación de DMSP, solo se realizó su búsqueda en genomas de aislados de la cepa en el laboratorio presentes en las bases de datos.

Las muestras analizadas se corresponden con el muestreo realizado en la Expedición Oceanográfica de Circunnavegación Malaspina 2010, una campaña de estudio llevada a cabo alrededor de los océanos del planeta a bordo del Buque Oceanográfico Hespérides, la cual tenía como objetivo obtener una visión en detalle acerca del impacto del cambio global en los ecosistemas oceánicos, así como el estudio de la biodiversidad existente en los mismos, especialmente en las zonas del océano profundo. Entre las distintas disciplinas implicadas en la investigación, el área de Biodiversidad y Procesos Microbianos se encargó de muestrear los microorganismos presentes en el océano profundo (poco estudiado hasta entonces) con el fin de explorar la diversidad filogenética y funcional además de los factores que regulan a los mismos, su implicación en los ciclos biogeoquímicos así como la creación de un repositorio de ADN microbiano para aumentar el conocimiento de los microorganismos presentes en los ecosistemas oceánicos.

2. Objetivos

El objetivo fundamental del presente trabajo es el análisis de la ecología microbiana del bacterioplancton oceánico, especialmente en *Sulfitobacter* sp. EE-36, a través del análisis

de la secuenciación de metagenomas haciendo uso de herramientas bioinformáticas, con el fin de investigar la expresión de distintos genes relacionados con la oxidación de compuestos del azufre orgánicos e inorgánicos. La importancia de llevar a cabo este trabajo radica en obtener un mayor conocimiento acerca del funcionamiento de estas vías metabólicas que tienen una gran implicación ecológica sobre el ciclo global del azufre, y que además, pueden suponer un avance en los conocimientos acerca de las aplicaciones biotecnológicas del bacterioplancton sobre el medio ambiente.

3. Material y métodos

Tal como se describe en los objetivos, el principal interés del trabajo es realizar la búsqueda en las secuencias problema (contigs) de genes cuyos productos se encuentren implicados en la oxidación orgánica e inorgánica de compuestos del azufre como los anteriormente nombrados. Las muestras de bacterioplancton se obtuvieron en distintos puntos de muestreo durante la campaña Malaspina, repartidos a lo largo de todos los océanos y a distintas profundidades de los mismos (Fig. 5). Para la realización de este trabajo, se obtuvieron las muestras secuenciadas y ensambladas procedentes de Malaspina, según las técnicas de metagenómica anteriormente nombradas, con los contigs resultantes (Contig 244 y Contig 120) listos para los análisis genéticos correspondientes cuyo objetivo era llevar a cabo una predicción funcional de los distintos genes, así como la asociación de éstos a distintas especies. Para ello, se realizó la comparación de las secuencias obtenidas en los distintos contigs con las secuencias caracterizadas en las bases de datos públicas (Fig. 6).

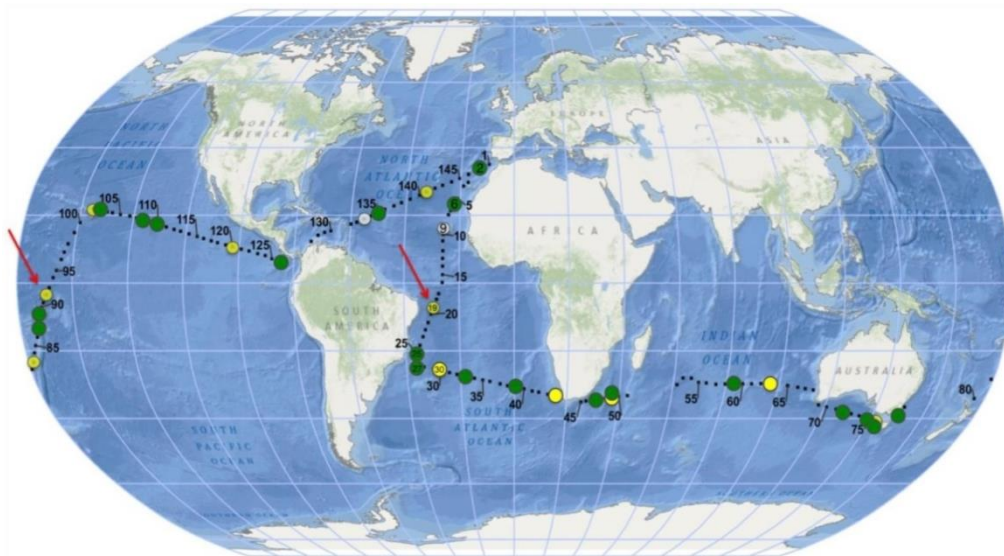


Figura 5. Puntos de muestreo durante la campaña Malaspina. Representación de los puntos donde fueron obtenidas las muestras de metagenomas secuenciados. Los contigs resultantes de la secuenciación de estas muestras suponen la base del estudio en este trabajo. Imagen cedida por el Dr. José M. González.

Para los análisis de secuencias se hizo uso del software BLAST (Basic Local Alignment Search Tool), un programa bioinformático desarrollado por el NCBI (National Center for Biotechnology Information) que realiza el alineamiento de las distintas secuencias problema contra las secuencias caracterizadas existentes en las bases de datos del NCBI y muestra el grado de homología entre las mismas, por lo que fue usado como sistema de identificación y anotación funcional de genes. Para hacer uso de esta herramienta bioinformática de forma local y no en el servidor (usualmente se hace en <https://www.ncbi.nlm.nih.gov/>), se realizó la instalación de BLAST en Linux (Ubuntu 16.04 LTS). Para poder utilizar esta aplicación, se instaló Biopython (bibliotecas de datos en lenguaje Python para el desarrollo de software bioinformáticos) con los datos necesarios para el funcionamiento de la aplicación. También se realizó la búsqueda en bases de datos especializadas alojadas en web tales como PFAM, desarrollada por el EMBL-EBI (European Bioinformatics Institute), y TIGRFAM, desarrollada por el J. Craig Venter Institute, para contrastar los resultados, las cuáles utilizan Modelos Ocultos de Markov en el alineamiento múltiple de secuencias.

Para proceder a la anotación de genes, se decidió realizar el alineamiento comparativo en BLAST de las secuencias de aminoácidos de los péptidos, por lo que se hizo uso de otro software denominado Prodigal (Prokaryotic Dynamic Programming Genefinding Algorithm), una herramienta bioinformática diseñada por la Universidad de Tennessee para la predicción de genes microbianos a través de la utilización de algoritmos simples que aportan resultados adecuados y útiles. Una vez que Prodigal identificó los potenciales genes que codifican proteínas en los contigs, las secuencias de aminoácidos resultantes fueron sometidas a Protein BLAST (BLASTp) con el fin de establecer homologías entre la secuencia problema (“query”) de los contigs y las secuencias caracterizadas de los genomas de las bases de datos del NCBI. Para poder ejecutar BLASTp, se utilizaron como bases de datos de referencia los genomas de distintas bacterias comunes en el bacterioplankton y que pueden albergar los genes de interés para el estudio o similares. Para poder emplear estas bases de datos, una vez descargadas en formato GenBank desde el NCBI, fue necesario utilizar algunos scripts en lenguaje Python para obtener las secuencias de péptidos en formato FASTA, formato utilizado para ejecutar BLASTp

Para realizar un contraste de los resultados obtenidos en la anotación manual en BLAST contra el genoma de *Sulfitobacter* sp. EE-36 de forma gráfica se hizo uso de un software denominado Prokka, desarrollado por Victorian Bioinformatics Consortium. Esta

herramienta bioinformática lleva a cabo la anotación automática de las secuencias de nucleótidos de interés en formato FASTA (en este caso la secuencia del contig y la secuencia del genoma, ambas acotadas en la zona de interés) para producir un archivo GenBank con los resultados, los cuáles puede reproducir en una figura que represente el alineamiento entre las dos secuencias peptídicas analizadas en función del grado de homología y conservación.

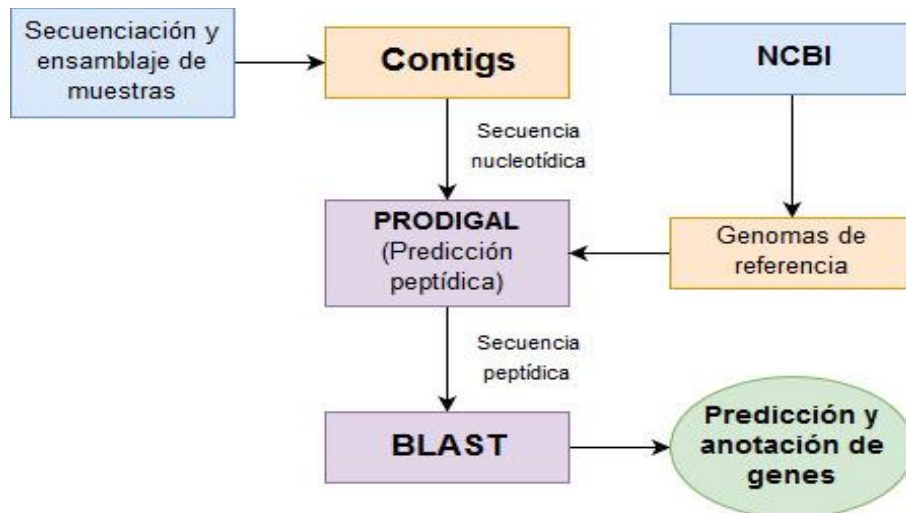


Figura 6. Metodología general usada en la anotación de genes. Las muestras de Malaspina fueron secuenciadas y ensambladas (Metagenómica) resultando los contigs analizados en el presente trabajo. Estos contigs consisten en secuencias de nucleótidos que pasan por Prodigal para predecir las secuencias peptídicas que codifican. Éstas se alinean en Protein BLAST contra las secuencias peptídicas de los genomas de referencia (también tratados con Prodigal para el alineamiento de secuencias peptídicas), resultando en la anotación de genes de la muestra.

4. Resultados

Los análisis se llevaron a cabo sobre 2 contigs (identificados como 244 y 120) en los que se llevó a cabo la anotación manual de funciones. Antes de llevar a cabo el alineamiento y la comparación de secuencias, se seleccionaron los genomas de referencia donde se encuentran caracterizados estos genes. Tras ello, y posteriormente al tratamiento con Prodigal, se procedió a la comparación mediante BLASTp para realizar la predicción de los genes y la anotación de funciones pertenecientes a dichos contigs. En la comparación BLAST se tienen en cuenta distintos valores para establecer la confianza de la anotación, entre ellos destacan: la Identidad, la cantidad de caracteres idénticos (coincidencias de aminoácidos, en este caso) entre la secuencia de referencia y la secuencia problema; y el E-value, definido como la probabilidad de que el alineamiento realizado haya sido al azar (cuanto más bajo sea este valor, mejor será el resultado de la anotación).

4.1. Genomas de referencia

Los genomas de referencia utilizados para la comparación de secuencias de aminoácidos pertenecían a microorganismos de la clase Alfabroteobacteria, donde los genes relacionados con la oxidación de los compuestos orgánicos e inorgánicos del azufre se encuentran presentes y distribuidos de forma distinta en las especies (Fig. 7). El principal organismo cuyo genoma sirvió de referencia máxima fue el de *Sulfitobacter* sp. EE-36, especie protagonista del presente trabajo, cuyo genoma debía comprender los genes de interés para la anotación en los contigs. Sin embargo, también se utilizaron en los alineamientos secuencias peptídicas pertenecientes a genomas de otras especies relacionadas de géneros como *Roseobacter*, *Rhodobacter*, *Paracoccus*, entre otros.

Especies	Genes de oxidación de azufre orgánico e inorgánico						
	<i>dmdA</i>	<i>dddP</i>	<i>dddL</i>	<i>dddD</i>	<i>dddQ</i>	<i>dddW</i>	<i>soxXYZABCD</i>
<i>Sulfitobacter</i> sp. EE-36			X				X
<i>Paracoccus denitrificans</i>							X
<i>Roseobacter denitrificans</i>	X	X					X
<i>Ruegeria pomeroyi</i>	X	X		X	X	X	X
<i>Dinoroseobacter shibae</i>	X		X	X			X
<i>Rhodobacter sphaeroides</i>			X				X

Figura 7. Genes en genomas de especies utilizados como referencias. En la tabla se presentan los genes presentes en cada especie marcados con una X. Las especies escogidas se hicieron en función de la similitud y cercanía filogenética a los genes de la bacteria estudiada. En los genomas de todas las especies escogidas aparecen los genes *sox* mientras que la distribución de los genes encargados de la degradación del DMSP varía entre ellas, pues *Sulfitobacter* sp. EE-36 únicamente presenta *dddL*. Adaptado de Curson et al. (2011) y Meyer et al. (2007).

A pesar de ejecutar BLAST en los contigs contra los genomas de estas especies, también se llevó a cabo la comparación con las bases en el servidor web del NCBI, PFAM y TIGRFAM, con el fin de contrastar los resultados y obtener información actualizada, reciente o desconocida hasta el momento de la búsqueda, tales como genes filogenéticamente cercanos en determinadas cepas o especies no tenidas en cuenta en la búsqueda local. Esta información podría haberse perdido haciendo la búsqueda únicamente contra los genomas anteriormente nombrados.

4.2. Anotaciones resultantes de alineamientos y comparaciones BLAST

Una vez obtenidas las secuencias peptídicas de los genomas de referencia y de los genes predichos en los contigs haciendo uso de Prodigal, se ejecutó Protein BLAST, dando lugar a la anotación de las funciones de los genes presentes en el contig (es decir, los aminoácidos que constituyen las proteínas codificadas). En el primer contig (ID: 244) analizado resultaron 43 secuencias peptídicas tras ser tratado con Prodigal. Entre los productos anotados al hacer la búsqueda BLAST, se encontraron 11 secuencias (desde la secuencia 20 hasta la 30, ambas inclusive) que pertenecían a la codificación de las proteínas del complejo enzimático Sox (Fig. 8), por lo que estos productos estaban codificados bajo el cluster de genes *soxXYZABCD*. Las secuencias de aminoácidos identificadas correspondían con las proteínas SoxW, SoxX, SoxY, SoxZ, SoxA, SoxB y SoxC, las cuáles fueron confirmadas con altos porcentajes de Identidad y valores correctos de E-value para *Sulfitobacter* sp. EE-36 (máxima Identidad) con respecto a las demás alineaciones. Sin embargo, se encontraron algunas secuencias en que la anotación no era tan exacta, pues se correspondía con resultados no esperados a encontrar en el genoma de referencia, tal como sucedía para las secuencias 20, 21, 22 y 30. En estos casos no se podía confirmar la codificación exacta de un solo producto, pues altos valores de Identidad también se correspondían con distintas funciones en diferentes organismos.

Contig 1 (ID: 244)				
Secuencia peptídica	Producto anotado	Identidad (aa. coincidentes)	E-value	Organismo de alineamiento
20	Proteína reguladora (familia ArsR)	99% (113/114)	5e ⁻⁸³	<i>Sulfitobacter</i> sp. EE-36
21	SoxW	96% (123/128)	5e ⁻⁹²	<i>Sulfitobacter</i> sp. EE-36
22	Citocromo C transmembranal	99% (245/246)	4e ⁻¹⁷⁴	<i>Sulfitobacter</i> sp. EE-36
23	SoxW	98% (187/190)	6e ⁻¹⁴³	<i>Sulfitobacter</i> sp. EE-36
24	SoxX	98% (155/158)	5e ⁻¹¹⁴	<i>Sulfitobacter</i> sp. EE-36
25	SoxY	99% (138/139)	5e ⁻¹⁹⁶	<i>Sulfitobacter</i> sp. EE-36
26	SoxZ	98% (107/109)	5e ⁻⁷⁸	<i>Sulfitobacter</i> sp. EE-36
27	SoxA	99% (280/284)	0.0	<i>Sulfitobacter</i> sp. EE-36
28	SoxB	99% (550/556)	0.0	<i>Sulfitobacter</i> sp. EE-36
29	SoxC	98% (389/397)	0.0	<i>Sulfitobacter</i> sp. EE-36
30	SoxD	94% (321/341)	0.0	<i>Sulfitobacter</i> sp. NAS-14-1

Figura 8. Resultados de anotación para el contig 244 tras alineamiento en BLAST. En la tabla se señala la secuencia peptídica en el contig y su correspondiente anotación (solo aquellas proteínas codificadas de interés en la búsqueda), con el porcentaje de Identidad y el E-value así como el organismo en el que se ha encontrado la mejor puntuación de estos valores, es decir, el genoma de referencia con el que la secuencia peptídica posee mayor correspondencia. Las secuencias 20, 21, 22 y 30 se señalan en rojo debido a que la anotación no es tan exacta debido a la existencia de otros valores de Identidad significativos para otros productos, y es distinta a la esperada para el complejo enzimático Sox.

Con respecto a los resultados obtenidos en el segundo contig analizado (ID: 120), cabe destacar que ofrece anotación similar de genes al contig anterior, salvo que se encuentran en una orientación contraria, pues las secuencias de aminoácidos alineadas se corresponden con las mismas proteínas. En este contig se encontraron 42 secuencias peptídicas, observándose nuevamente productos del complejo enzimático Sox (Fig. 9) en 11 secuencias peptídicas (desde la secuencia 13 a la 23, ambas inclusive). Se confirmaron las anotaciones SoxW, SoxX, SoxY, SoxZ, SoxA, SoxB y SoxC, con alto porcentaje de Identidad y excelentes valores de E-value para *Sulfitobacter* sp. EE-36 (máxima Identidad). Sin embargo, de igual manera que en el contig anterior, quedaron algunas secuencias (13, 21, 22 y 23) sin anotación exacta, pues se producían alineamientos con valores de Identidad altos para distintos productos en diferentes organismos.

Contig 2 (ID: 120)				
Secuencia peptídica	Producto anotado	Identidad (aa. coincidentes)	E-value	Organismo de alineamiento
13	SoxD	94% (321/341)	0.0	<i>Sulfitobacter</i> sp. NAS-14-1
14	Sox C	98% (389/397)	0.0	<i>Sulfitobacter</i> sp. EE-36
15	SoxB	99% (550/556)	0.0	<i>Sulfitobacter</i> sp. EE-36
16	SoxA	99% (280/284)	0.0	<i>Sulfitobacter</i> sp. EE-36
17	SoxZ	98% (107/109)	5e ⁻⁷⁸	<i>Sulfitobacter</i> sp. EE-36
18	SoxY	99% (138/139)	5e ⁻⁹⁶	<i>Sulfitobacter</i> sp. EE-36
18	SoxX	98% (155/158)	5e ⁻¹¹⁴	<i>Sulfitobacter</i> sp. EE-36
20	SoxW	98% (187/190)	6e ⁻¹⁴³	<i>Sulfitobacter</i> sp. EE-36
21	Citocromo C transmembranal (CcdA)	99% (245/246)	4e ⁻¹⁷⁴	<i>Sulfitobacter</i> sp. EE-36
22	SoxW	96% (123/128)	5e ⁻⁹²	<i>Sulfitobacter</i> sp. EE-36
23	Proteína reguladora (familia ArsR)	99% (113/114)	5e ⁻⁹²	<i>Sulfitobacter</i> sp. EE-36

Figura 9. Resultados de anotación para el contig 120 tras alineamiento en BLAST. En la tabla se señala la secuencia peptídica en el contig y su correspondiente anotación (solo aquellas proteínas codificadas de interés en la búsqueda), con el porcentaje de Identidad y el E-value así como el organismo en el que se ha encontrado la mejor puntuación de estos valores. Las secuencias 13, 21, 22 y 23 se señalan en rojo debido a que la anotación no es exacta, y es distinta a la esperada para el complejo enzimático Sox.

En las anotaciones inexactas se seleccionaron los productos con mejores resultados de Identidad y E-value que más se correspondieran con los productos esperados en el genoma de la bacteria analizada, suponiendo los otros resultados de los alineamientos como fallos de la anotación automática en las bases de datos de referencia (NCBI, PFAM o TIGRFAM).

4.3. Comparación y homología de contigs con *Sulfitobacter* sp. EE-36

Una vez anotados los genes de los contigs analizados y comprobado que en su mayoría se corresponden con genes y productos presentes en *Sulfitobacter* sp. EE-36, se llevó a cabo una comparación para contrastar si la secuencia procedente de la muestra se corresponde con parte del genoma de *Sulfitobacter* sp. EE-36. Para ello se hizo uso de Prokka, en el cual se llevó a cabo una alineación en función del grado de homología entre las secuencias peptídicas de los contigs y las correspondientes al genoma de referencia, ambas acotadas en la zona donde se encuentra el complejo *sox* que codifica el complejo multienzimático Sox. Tras introducir las secuencias el programa predice las secuencias peptídicas y anota los genes automáticamente. El archivo resultante es utilizado para representar una figura con los alineamientos de los péptidos resultantes de las anotaciones automáticas. Por tanto, las anotaciones representadas en la figura no corresponden con la anotación manual realizada en la fase anterior con BLASTp debido a que la anotación realizada por Prokka es automática (no contrastada).

En las figuras resultantes (Fig. 10), tanto para el primer contig como para el segundo, se observa una gran relación con las secuencias del genoma de referencia. En el caso de la alineación del primer contig (ID: 244) se observa que todas las secuencias de interés anotadas con BLASTp correspondientes al complejo multienzimático Sox presentan un alto grado de correspondencia con el genoma de *Sulfitobacter* sp. EE-36, pues todas las secuencias se encuentran representadas con mayor o menor grado de conservación o sintenia, exceptuando la anotada manualmente como SoxD, en cuyo caso no encuentra ninguna relación. En el caso del segundo contig (ID: 120) se produce el mismo resultado, salvo que la orientación del mismo es distinta al primero. En este caso el grado de conservación es similar al contig anterior, observándose una alta correspondencia entre las secuencias del contig y las del genoma de *Sulfitobacter* sp. EE-36, exceptuando la proteína SoxD.

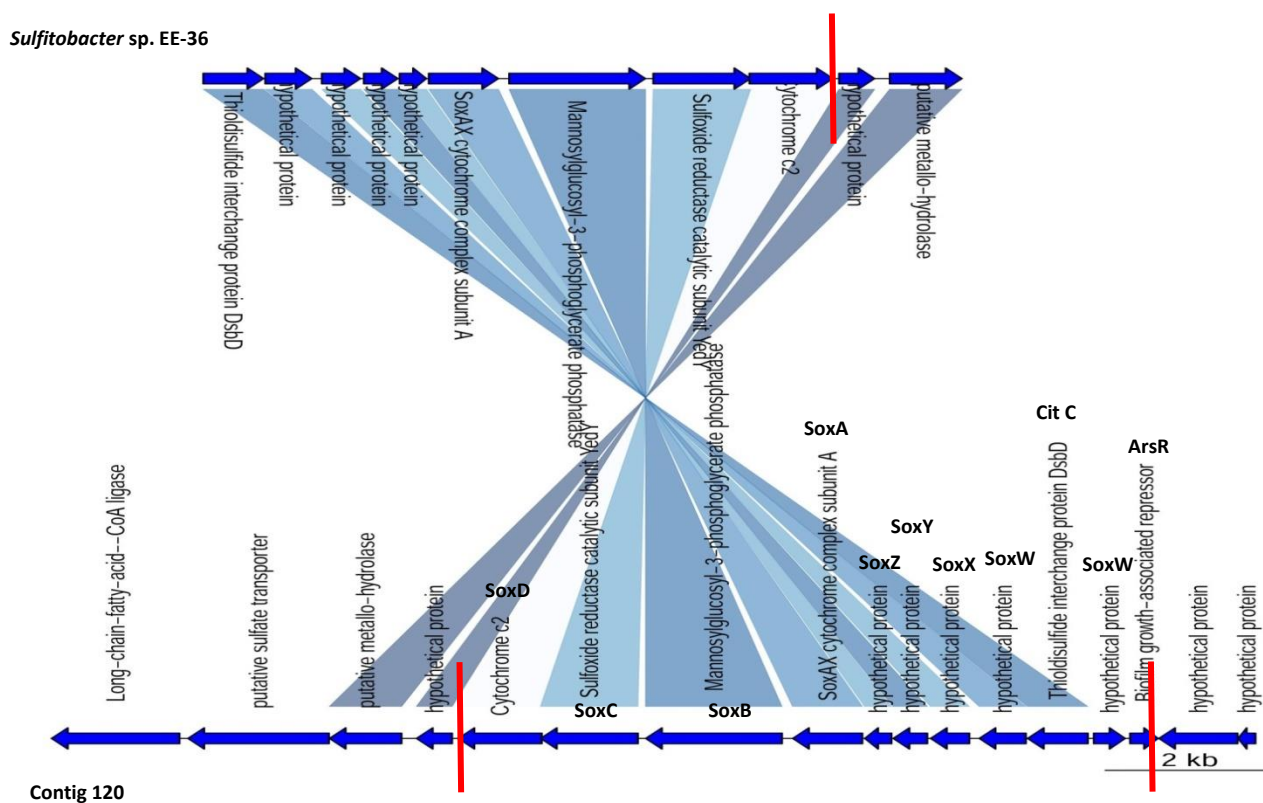
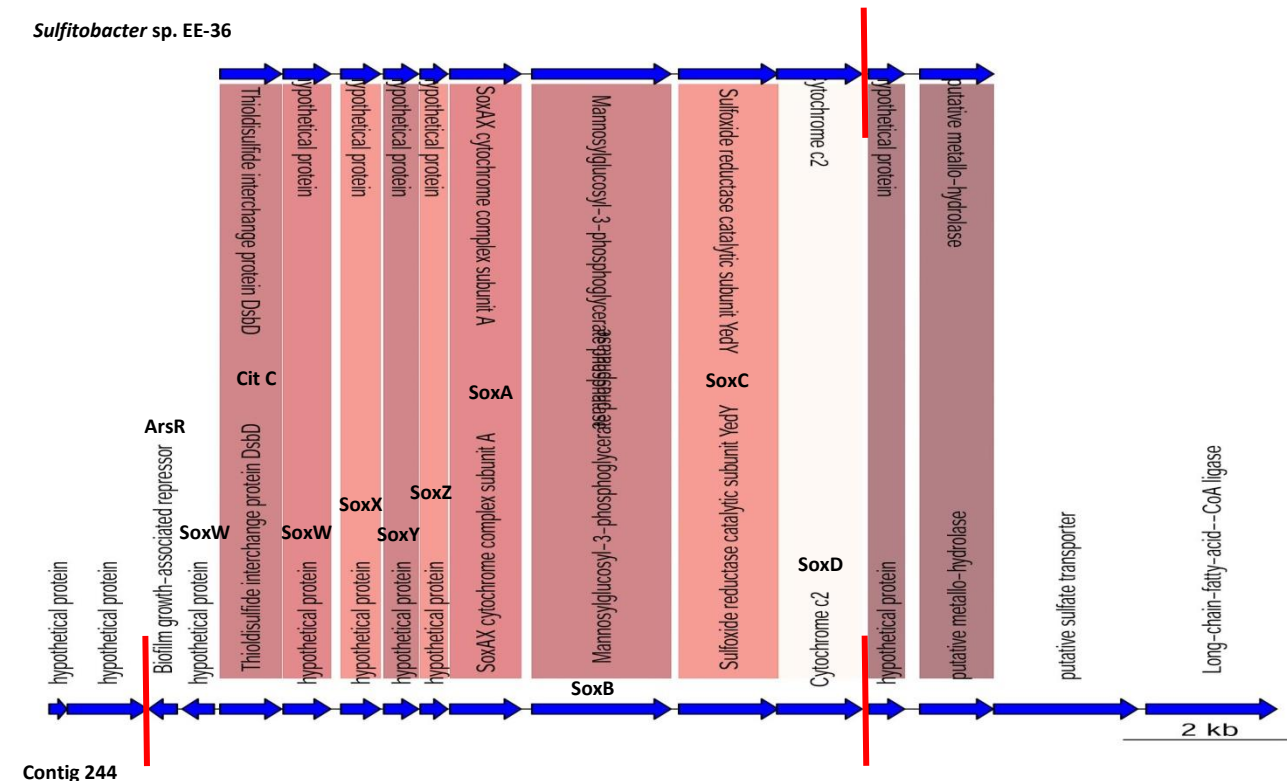


Figura 10. Alineación de genes anotados en el contig y en el genoma de *Sulfitobacter* sp. EE-36. En la parte superior se realiza la comparación del primer contig, mientras que en la figura inferior se representa la comparación del segundo. En las secuencias aparece delimitada entre líneas rojas verticalmente la zona que abarca el cluster *sox* anotado. Encima de las anotaciones automáticas realizadas por Prokka (en vertical) aparecen las anotaciones manuales realizadas con BLAST (en horizontal). Los colores de las bandas de alineamiento representan el grado de similitud o identidad entre las secuencias peptídicas (cuanto más oscuras, mayor homología).

5. Discusión

Una vez llevado a cabo el análisis de las secuencias metagenómicas ensambladas procedentes de muestras ambientales, los resultados arrojaron información sobre los genes presentes en las mismas. Este proceso llevado a cabo mediante la identificación y la anotación funcional, fue crucial para relacionar la presencia de *Sulfitobacter* sp. EE-36 con las muestras analizadas, y tratar de describir el papel que desempeñan las rutas metabólicas de esta estirpe en el ciclo del azufre.

De forma general, en ambos contigs, se observa en los resultados de las anotaciones en BLAST la existencia de genes encargados de codificar enzimas capaces de oxidar compuestos inorgánicos del azufre, es decir, el complejo *sox* principal (conformado por *soxXYZABCD*), pues en las muestras se observan secuencias capaces de codificar estos péptidos. Además de estos, también se encuentran otros genes reguladores existentes en los procesos de downstream de la degradación del azufre inorgánico, reflejado en la anotación doble del gen *soxW* cuyo producto es la proteína SoxW implicada en el crecimiento quimiotrófico con tiosulfato y de la transferencia de agentes reductores (Friedrich et al., 2005). También se observa el citocromo CcdA (probablemente codificado por *soxV*) implicado en la transferencia electrónica y proteínas reguladoras como ArsR (probablemente codificada por el gen *soxR*) encargada de la regulación en los procesos downstream del catabolismo por parte del complejo enzimático Sox (Friedrich et al., 2005, 2001). En todos los casos, los alineamientos de aminoácidos durante la anotación funcional manual (Fig. 8, 9) de las secuencias resultaron con altos valores de Identidad y E-value bajos (sobre todo para SoxABCD, donde éste último fue de 0.0), por lo que existen un alto grado de homología entre las proteínas que codificarían los genes de la muestra y los codificados por los genomas de referencia donde existía mayor Identidad (en su mayoría *Sulfitobacter* sp. EE-36). Esta norma se cumple en todas las proteínas de interés, salvo en SoxD, SoxW, CcdA y ArsR, donde las anotaciones no fueron tan significativas únicamente para *Sulfitobacter* sp. EE-36.

Con respecto al contraste de las secuencias peptídicas que se predicen con Prokka, cabe destacar que existe un alto nivel de homología y conservación en las secuencias de los contigs con respecto a *Sulfitobacter* sp. EE-36. En el caso de algunos genes existe un mayor grado de conservación de sus secuencias peptídicas (Fig. 10) tales como en CcdA, SoxW, SoxY, SoxA y SoxB, donde se observan mayor grado de oscuridad en las bandas que representan la identidad de los alineamientos, por lo que éstas secuencias podrían estar más

relacionadas con *Sulfitobacter* sp. EE-36. En otras como SoxD no existe relación ninguna, probablemente debido a que esta proteína esté codificada por un gen representado en *Sulfitobacter* sp. NAS-14-1 tal como sucedía durante la anotación del mismo, por lo que esta secuencia peptídica de la muestra ambiental no podría relacionarse con la estirpe EE-36.

Con respecto a la aparición de genes o anotación de los mismos relacionados con la degradación de compuestos orgánicos del azufre (DMSP), el gen *dddL* (el cuál codifica para la enzima DddL) se encontró en aislados de *Sulfitobacter* sp. EE-36 en laboratorio, único gen comprobado como responsable de este catabolismo hasta la actualidad (Andrew R. J. Curson et al., 2011). En las secuencias ambientales no se pudo anotar debido a que la parte del ensamblaje del metagenoma analizado solo comprendía los genes *sox* (los genes *ddd* se encuentran alejados).

6. Conclusiones

La primera conclusión general que se obtiene a partir de la realización de este trabajo es la existencia de genes, en los contigs procedentes de las muestras ambientales, que codifican enzimas capaces de degradar compuestos de azufre inorgánico, tales como el tiosulfato, el sulfito o el ácido sulfhídrico para producir sulfato. Esto es debido a la identificación del cluster principal de genes que protagonizan esta ruta metabólica: los genes *soxXYZABCD*. Esta identificación viene dada por la anotación de las proteínas SoxXYZABCD, entre otras, a partir de los péptidos predichos desde las secuencias nucleotídicas de los contigs. Los altos valores de identidad en la anotación de los genes suscita, por tanto, que la actividad de oxidación de compuestos inorgánicos del azufre se está llevando a cabo en las los océanos profundos, lugar de procedencia de las muestras. El gen relacionado con la oxidación del DMSP (gen *dddL*) solo fue detectado en aislados, puesto que no se realizó la búsqueda en las secuencias ambientales. Este hecho, sin embargo, no implica que éste no exista en la muestra ambiental, sino que significa que los contigs analizados no correspondían con la zona del metagenoma ensamblado donde se encuentra este gen que codifica para la enzima DddL encargada de oxidar el DMSP hasta acrilato mediante la producción de DMS.

La segunda conclusión se corresponde con la confirmación del hecho de que los contigs procedentes de las muestras son pertenecientes a *Sulfitobacter* sp. EE-36, debido principalmente al alto grado de similitud de las secuencias de aminoácidos de las proteínas

procedentes de la anotación con aquellas de esta especie, exceptuando SoxD (refiriéndonos a las anotaciones procedentes de los genes de interés). Esta hipótesis se ve justificada en los grandes valores de identidad en las secuencias de péptidos, y que correspondían con *Sulfitobacter* sp. EE-36 en la anotación manual de los genes de los contigs (Fig. 8, 9), así como en los resultados provenientes del análisis en Prokka, donde también el alineamiento señalaba grandes homologías (Fig. 10). Estos valores de identidad, con E-value bastante aceptables, indican que existe la variación de algunos aminoácidos en las secuencias peptídicas de la muestra (contigs) con respecto a las secuencias peptídicas del genoma caracterizado de *Sulfitobacter* sp. EE-36. Este hecho podría verse reflejado en las homologías de las secuencias nucleotídicas, las cuáles tampoco llegarían a ser completas. Esto puede ser debido a una probable variación genética provocada por alguna mutación debida quizás a la presión evolutiva existente en la zona donde se encuentra la bacteria o por otros factores que hayan podido provocar mutagénesis. De esta manera, las muestras indicarían la actividad de *Sulfitobacter* sp. EE-36 en la degradación de compuestos inorgánicos de azufre, tales como el tiosulfato, en las zonas profundas del océano.

Por tanto, como tercera conclusión, en vistas a investigaciones de futuro, sería de vital importancia ahondar en el entendimiento en mayor medida del papel de este importante miembro del plancton bacteriano oceánico para conocer sus consecuencias ecológicas en el ciclo de los elementos. Esto puede abordarse a través de la investigación de las condiciones ambientales bajo las que se encuentran estas muestras (localización de las muestras y los factores físico-químicos, climáticos y bióticos) para identificar la presencia de *Sulfitobacter* sp. EE-36, y para encontrar una relación acerca de la distribución en los océanos de esta maquinaria enzimática así como de los procesos que pueden intervenir en los cambios de los genes que codifican dichas proteínas. A colación de este último punto, también podría investigarse la expresión de estos genes con el fin de conocer si existe algún efecto de los cambios en los aminoácidos en la codificación de las proteínas implicadas en la degradación de los compuestos inorgánicos del azufre, es decir, cambios en las funciones de las enzimas provocados por las variaciones genéticas.

La realización de este tipo de estudios podría aumentar el conocimiento acerca de los genes y enzimas encargadas de llevar a cabo las rutas metabólicas en los océanos, además de conocer la distribución y la ecología (influencia en el ciclo del azufre) de este tipo de bacterias, tales como *Sulfitobacter* sp. EE-36, en la zona donde se haya muestreado (zonas

oceánicas profundas), y de forma general, la distribución y el papel bacterioplancton en el planeta y su implicación en los ciclos biogeoquímicos.

6. Conclusions

The first general conclusion obtained from this project is the existence of genes in contigs from environmental samples that encode enzymes capable of degrading inorganic sulfur compounds such as thiosulfate, sulfite or sulfhydryric acid to produce sulfate. This is due to the identification of the main cluster of genes that lead this metabolic pathway: the genes *soxXYZABCD*. This identification is given by the annotation of SoxXYZABCD proteins from the predicted peptides from the nucleotide sequences of the contigs. The high values of identity in the annotation of the genes arouse that the oxidation activity of inorganic sulfur compounds is taking place in the deep oceans, where the samples come from. The gene related to the oxidation of DMSP (*dddL* gene) was only detected in isolates because we did not search in the environmental sequences. This fact, however, has no implication that there is no success in the environmental sample. It means that the analyzed contigs did not correspond to the area of the assembled metagenome where this gene coding for the enzyme DddL is in charge of oxidizing the acrylate of up to DMSP Through the production of DMS.

The second conclusion corresponds to the fact that the contigs coming from the samples belong to *Sulfitobacter* sp. EE-36, due to the high degree of similarity of the amino acid sequences of the proteins coming from the annotation with those of this species, except SoxD. This hypothesis is justified in the great values of Identity that corresponded to *Sulfitobacter* sp. EE-36 in the manual annotation of the genes of the contigs (Fig. 8, 9), as well as in the results from the Prokka analysis, where the alignment also indicated large homologies (Fig. 10). Identity values, with E-value quite acceptable, indicate that there is a variation of some amino acids in the peptidic sequences of the sample (contigs) with respect to the peptidic sequences of the genome characterized *Sulfitobacter* sp. EE-36. This fact could be reflected in the homologies of the nucleotide sequences, which would not be complete either. This may be due to a probable genetic variation caused by some mutation due perhaps to the existing evolutionary pressure in the area where the bacterium is found or by other factors that may have caused mutagenesis. In this way, the samples would indicate the activity of *Sulfitobacter* sp. EE-36 in the degradation of inorganic sulfur compounds, such as thiosulfate, in the deep ocean.

Therefore, as a third conclusion, in a future research view, it would be important to delve more deeply into the understanding of this important member of oceanic bacterioplankton to know its ecological consequences in the cycle of elements. This can be approached through the investigation of the environmental conditions under which these samples are located (location of samples and physical-chemical, climatic and biotic factors) to identify the presence of *Sulfitobacter* sp. EE-36, and to find a relation about the distribution in the oceans of this enzymatic machinery as well as the processes that can intervene in the changes of the genes that codify this proteins. In relation with this latter point could be also investigate the expression of these genes in order to know if there is any effect of changes in amino acids in the coding of proteins involved in the degradation of inorganic sulfur compounds, therefore, changes in the functions of enzymes caused by genetic variations.

The realization of this type of studies could increase the knowledge about the genes and enzymes responsible for the metabolic routes in the oceans, besides knowing the distribution and the ecology (influence in the sulfur cycle) of this type of bacteria, such as *Sulfitobacter* sp. EE-36, in the area where it has been sampled (deep ocean areas), and in general, the distribution and role of bacterioplankton in the planet and its involvement in biogeochemical cycles.

7. Bibliografía

- Alguacil Martín, S. (2010). Metagenómica: la microbiología del futuro. Consultado el 5 de abril de 2017 en <http://www.agenciasinc.es/Reportajes/Metagenomica-la-microbiologia-del-futuro>
- Curson, A. R. J., Rogers, R., Todd, J. D., Brearley, C. A., & Johnston, A. W. B. (2008). Molecular genetic analysis of a dimethylsulfoniopropionate lyase that liberates the climate-changing gas dimethylsulfide in several marine ??-proteobacteria and *Rhodobacter sphaeroides*. *Environmental Microbiology*, 10(3), 757–767. <https://doi.org/10.1111/j.1462-2920.2007.01499.x>
- Curson, A. R. J., Todd, J. D., Sullivan, M. J., & Johnston, A. W. B. (2011). Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes. *Nature Reviews Microbiology*, 9(12), 849–859. <https://doi.org/10.1038/nrmicro2653>
- Dickschat, J. S., Rabe, P., & Citron, C. A. (2015). The chemical biology of dimethylsulfoniopropionate. *Org. Biomol. Chem.*, 13, 1954–1968. <https://doi.org/10.1039/C4OB02407A>
- Espinoza, J., Revah, S., & Le Borgne, S. (2010). Rutas Metabólicas De Oxidación Del Azufre En Bacterias Quimiolitotróficas , Relevancia Ambiental Y Biotecnología. *Mensaje Bioquímico*, XXXIV, 101–120.
- Friedrich, C. G., Bardischewsky, F., Rother, D., Quentmeier, A., & Fischer, J. (2005). Prokaryotic sulfur oxidation. *Current Opinion in Microbiology*, 8(3), 253–259. <https://doi.org/10.1016/j.mib.2005.04.005>
- Friedrich, C. G., Rother, D., Bardischewsky, F., Ouentmeier, A., & Fischer, J. (2001). Oxidation of Reduced Inorganic Sulfur Compounds by Bacteria: Emergence of a Common Mechanism? *Applied and Environmental Microbiology*, 67(7), 2873–2882. <https://doi.org/10.1128/AEM.67.7.2873-2882.2001>
- Gonzalez, J. M., Kiene, R. P., Joye, S. B., Sorokin, D. Y., & Moran, M. A. (2002). Oxidation of organic and inorganic sulfur compounds by aerobic heterotrophic marine bacteria. *Biotransformations: Bioremediation Technology for Health and Environmental Protection*, 291–310.
- Gonzalez, J. M., Kiene, R. P., & Moran, M. A. (1999). Transformation of sulfur compounds by an abundant lineage of marine bacteria in the alpha-subclass of the class Proteobacteria. *Applied and Environmental Microbiology*, 65(9), 3810–3819.
- Gonzalez, J. M., Pedrós-Alió, C., & Gasol, J. M. (2008). Plancton bacteriano de los océanos. *Investigación Y Ciencia*, 387, 76–84. Consultado el 5 de abril de 2017 en http://www.investigacionyciencia.es/archivos/12-08_gonzalez.pdf
- Handelsman, J. (2005). Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, 69(1), 195–195. <https://doi.org/10.1128/MMBR.69.1.195.2005>
- Johnston, A. W. B., Green, R. T., & Todd, J. D. (2016). Enzymatic breakage of dimethylsulfoniopropionate - a signature molecule for life at sea. *Current Opinion in Chemical Biology*, 31(Figure 1), 58–65. <https://doi.org/10.1016/j.cbpa.2016.01.011>
- Kiene, R. P., & Linn, L. J. (2000). The fate of dissolved dimethylsulfoniopropionate (DMSP) in seawater: Tracer studies using ³⁵S-DMSP. *Geochimica et Cosmochimica Acta*, 64(16), 2797–2810. [https://doi.org/10.1016/S0016-7037\(00\)00399-9](https://doi.org/10.1016/S0016-7037(00)00399-9)

- Kiene, R. P., Linn, L. J., & Bruton, J. A. (2000). New and important roles for DMSP in marine microbial communities. *Journal of Sea Research*, 43(3), 209–224. [https://doi.org/10.1016/S1385-1101\(00\)00023-X](https://doi.org/10.1016/S1385-1101(00)00023-X)
- Kumbhare, S. V., Dhotre, D. P., Dhar, S. K., Jani, K., Apte, D. A., Shouche, Y. S., & Sharma, A. (2015). Insights into diversity and imputed metabolic potential of bacterial communities in the continental shelf of Agatti Island. *PLoS ONE*, 10(6), 1–14. <https://doi.org/10.1371/journal.pone.0129864>
- Lenk, S., Moraru, C., Hahnke, S., Arnds, J., Richter, M., Kube, M., ... Mußmann, M. (2012). Roseobacter clade bacteria are abundant in coastal sediments and encode a novel combination of sulfur oxidation genes. *The ISME Journal*, 2178–2187. <https://doi.org/10.1038/ismej.2012.66>
- MetaCyc. (2014). MetaCyc Enzyme: Sox enzyme system. Consultado el 2 de mayo de 2017 en <https://biocyc.org/META/NEW-IMAGE?type=ENZYME&object=CPLX-882>
- Meyer, B., Imhoff, J. F., & Kuever, J. (2007). Molecular analysis of the distribution and phylogeny of the soxB gene among sulfur-oxidizing bacteria - Evolution of the Sox sulfur oxidation enzyme system. *Environmental Microbiology*, 9(12), 2957–2977. <https://doi.org/10.1111/j.1462-2920.2007.01407.x>
- Petri, R., Podgorsek, L., & Imhoff, J. F. (2001). Phylogeny and distribution of the soxB gene among thiosulfate-oxidizing bacteria. *FEMS Microbiology Letters*, 197(2), 171–178. [https://doi.org/10.1016/S0378-1097\(01\)00111-2](https://doi.org/10.1016/S0378-1097(01)00111-2)
- Reisch, C. R., Moran, M. A., & Whitman, W. B. (2011). Bacterial catabolism of dimethylsulfoniopropionate (DMSP). *Frontiers in Microbiology*, 2(AUG), 1–12. <https://doi.org/10.3389/fmicb.2011.00172>
- Riesenfeld, C. S., Schloss, P. D., & Handelsman, J. (2004). Metagenomics: Genomic Analysis of Microbial Communities. *Annual Review of Genetics*, 38(1), 525–552. <https://doi.org/10.1146/annurev.genet.38.072902.091216>
- Roseobase. (2017). Sulfitobacter sp EE-36 (General features). Consultado el 2 de mayo de 2017 en http://www.roseobase.org/Species/sulf_ee36.html
- Schloss, P. D., & Handelsman, J. (2003). Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology*, 14(3), 303–310. [https://doi.org/10.1016/S0958-1669\(03\)00067-3](https://doi.org/10.1016/S0958-1669(03)00067-3)
- Simó, R. (2001). Production of atmospheric sulfur by oceanic plankton: Biogeochemical, ecological and evolutionary links. *Trends in Ecology and Evolution*, 16(6), 287–294. [https://doi.org/10.1016/S0169-5347\(01\)02152-8](https://doi.org/10.1016/S0169-5347(01)02152-8)
- Sorokin, D. Y. (1995). Sulfitobacter pontiacus gen. nov., sp. nov. - A New Heterotrophic Bacterium from the Black Sea , Specialized on Sulfite Oxidation. *Mikrobiologiya*, 64(3), 354–365.
- Stefels, J. (2000). Physiological aspects of the production and conversion of DMSP in marine algae and higher plants. *Journal of Sea Research*, 43(3–4), 183–197. [https://doi.org/10.1016/S1385-1101\(00\)00030-7](https://doi.org/10.1016/S1385-1101(00)00030-7)
- Streit, W. R., & Schmitz, R. A. (2004). Metagenomics - The key to the uncultured microbes. *Current Opinion in Microbiology*, 7(5), 492–498. <https://doi.org/10.1016/j.mib.2004.08.002>
- Sun, L., Curson, A. R. J., Todd, J. D., & Johnston, A. W. B. (2012). Diversity of DMSP transport in marine bacteria, revealed by genetic analyses. *Biogeochemistry*, 110(1–3), 121–130. <https://doi.org/10.1007/s10533-011-9666-z>