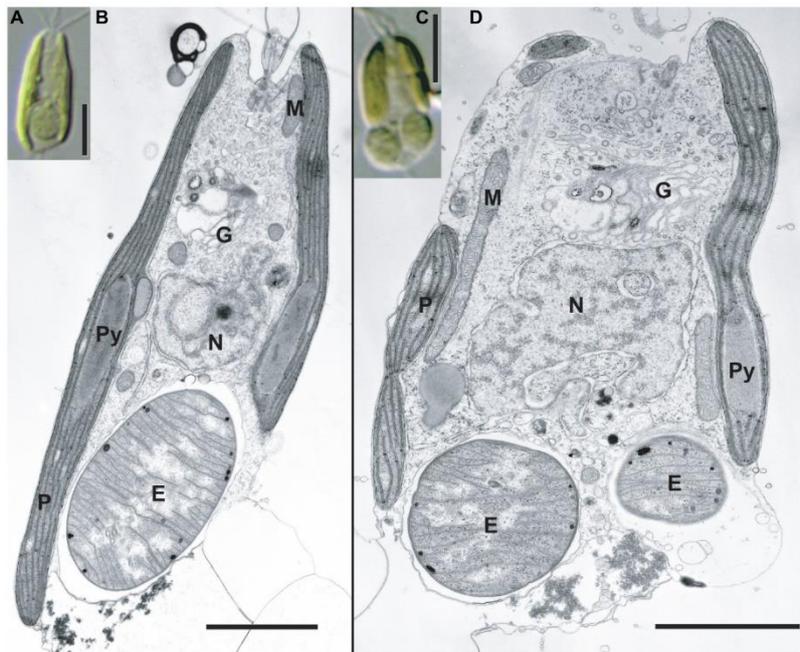




Microorganismos fijadores de N_2 en los océanos

N_2 -fixing microorganisms in the Oceans



Fotografía de MET en la que se muestran dos ejemplares del alga unicelular *Braarudosphaera bigelowii* con la cianobacteria UCYN-A como endosimbionte (E).
Tomada de Suzuki, Kawachi, *et al.* (2021)

Trabajo de Fin de Grado

DANIEL ACOSTA BARRIOS

Tutorizado por José Manuel González Hernández y Semidán Robaina Estévez

Grado en Biología. Julio 2022.

Índice

Resumen	1
Abstract	1
1. Introducción	2
2. Objetivos	6
3. Material y métodos.....	7
4. Resultados y discusión	9
5. Conclusiones	21
6. Conclusions	22
7. Bibliografía.....	23

Resumen

En la actualidad, sigue existiendo un gran desconocimiento acerca de los microorganismos que fijan nitrógeno en los océanos. Con el objetivo de estudiar la diversidad de microorganismos implicados en este proceso, se llevó a cabo un análisis bioinformático de metatranscriptomas, procedentes de muestras recogidas en todo el planeta por la expedición oceanográfica Tara Oceans. Para ello, se creó un árbol filogenético con secuencias ambientales de microorganismos fijadores de N₂, utilizando *nifH* como gen marcador. Analizando los datos obtenidos, se puede afirmar que hay una enorme diversidad de microorganismos involucrados en el complejo proceso de fijación biológica de nitrógeno (FBN), siendo *Gammaproteobacteria* el grupo más abundante. Además, se comprobó la presencia en el Océano Atlántico Sur, de una cianobacteria (UCYN-A) que vive en simbiosis con un alga unicelular, presentando una gran abundancia en aguas superficiales. Sin embargo, no se pudieron clasificar todas las secuencias, por lo que serán necesarios más estudios para conocer en profundidad la diversidad de microorganismos implicados en la fijación de N₂ en los océanos.

Palabras clave: fijación de nitrógeno, metatranscriptoma, bioinformática, *nifH*.

Abstract

Nowadays, there is still a great lack of knowledge about nitrogen-fixing microorganisms in the oceans. In order to study the diversity of microorganisms involved in this process, a bioinformatic analysis of metatranscriptomes was carried out, from samples collected throughout the planet by the Tara Oceans oceanographic expedition. For this reason, a phylogenetic tree was created with environmental sequences of N₂-fixing microorganisms, using *nifH* as a marker gene. Analyzing the obtained data, we can state that there is a huge diversity of microorganisms implicated in the complex process of biological nitrogen fixation (BNF), with *Gammaproteobacteria* being the most abundant group. In addition, the presence of cyanobacteria (UCYN-A), which lives in symbiosis with a unicellular algae, was confirmed in the South Atlantic Ocean, presenting a great abundance in surface water. However, not all sequences could be classified, thus further studies will be necessary to fully understand the diversity of microorganisms involved in N₂-fixation in the oceans.

Key words: nitrogen fixation, metatranscriptome, bioinformatics, *nifH*.

1. Introducción

El nitrógeno es un elemento esencial para la vida, debido a que forma parte de la estructura de proteínas y ácidos nucleicos (ADN y ARN). La forma más abundante de nitrógeno en el agua de mar es su forma molecular (N_2). Sin embargo, el N_2 sólo puede ser utilizado por algunos microorganismos procariotas conocidos como diazotrofos. Estos microorganismos son aquellas bacterias y arqueas con la capacidad de fijar nitrógeno, es decir, son capaces de reducir el dinitrógeno (N_2) a amoníaco (NH_3) mediante la enzima nitrogenasa. Este proceso es conocido como fijación biológica de nitrógeno (FBN), y supone con diferencia la mayor parte de la fijación de nitrógeno del planeta. En la última década se estimó que la FBN marina global tiene valores de entre 140 Tg N año y 177 Tg N año (Großkopf et al., 2012; Voss et al., 2013).

La fijación de nitrógeno que no tiene origen biológico se denomina fijación abiótica, está causada por los rayos en tormentas eléctricas y da lugar a una producción global de 7 Tg N año (Tie et al., 2002), una cantidad mucho menor que la fijada biológicamente. En un primer momento, se creía la FBN de los océanos era un proceso exclusivo de aguas superficiales tropicales y algunos sistemas bentónicos (Blackburn et al., 1988). Pero posteriormente se demostró que también tenía lugar en el fondo del mar y aguas anóxicas (Dekas et al., 2009), e incluso en estuarios (Bentzon-Tilia et al., 2015). Las tasas de fijación de N_2 están correlacionadas significativamente con condiciones ambientales, de manera positiva con la temperatura del agua de mar y la radiación fotosintéticamente activa, y de forma negativa con la profundidad y la salinidad (Bonnet et al., 2018). Además, en muchos hábitats el nitrógeno es limitante, por lo que el proceso de fijación biológica es fundamental para sustentar la red trófica (Vitousek & Howarth, 1991). Las bacterias de los océanos tienen un papel muy importante como fijadores de nitrógeno, ya que solo los procariotas tienen la capacidad de fijar N_2 , por lo que son la base de la cadena trófica y proporcionan nitrógeno biodisponible para el resto de organismos.

La FBN es llevada a cabo por los diazotrofos, gracias a la enzima nitrogenasa que cataliza la siguiente reacción: $N_2 + 16ATP + 8e^- + 8H^+ \rightarrow 2NH_3 + H_2 + 16ADP_i$. La reducción de N_2 a NH_3 tiene lugar en el sitio activo de la enzima, denominado cofactor FeMo (Einsle, 2014). Los diazotrofos pueden presentar distintos tipos de nitrogenasa: dependiente de molibdeno (Mo), dependiente de vanadio (V) y dependiente de hierro (Fe). La mayoría de los microorganismos fijadores de N_2 codifican únicamente la nitrogenasa dependiente de Mo. La información

necesaria para codificar esta nitrogenasa se encuentra en los genes *nif*. Por otra parte, los microorganismos que codifican las nitrogenasas alternativas dependientes de V o Fe, mediante los genes *vnf* o *anf* respectivamente, también codifican la nitrogenasa dependiente de Mo por medio de los genes *nif* (Joerger et al., 1988; Dos Santos et al., 2012). La nitrogenasa de Mo es más eficiente que las nitrogenasas alternativas, pero estas últimas son capaces de expresarse en ausencia de molibdeno (Mus et al., 2019). Hay grandes similitudes estructurales y funcionales entre los tres tipos de nitrogenasa, probablemente debidas a un origen de duplicación genética y divergencia evolutiva (Jasniewski et al., 2020).

La nitrogenasa más estudiada es la dependiente de Mo. Está integrada por dos componentes proteicos: el homodímero dinitrogenasa reductasa γ_2 (proteína NifH) y el heterotetrámero dinitrogenasa $\alpha_2\beta_2$ (proteínas NifD y NifK) (Raymond et al., 2004). Los genes *nif* se encuentran en operones, y son imprescindibles para la FBN ya que codifican las subunidades de la nitrogenasa, así como proteínas involucradas en la activación, regulación y transporte de metales (Kessler et al., 1998; Raymond et al., 2004). El número de genes *nif* que pueden presentar los diazotrofos es variable, siendo NifHDKENB el conjunto mínimo de productos génicos necesarios para codificar una nitrogenasa funcional (Dos Santos et al., 2012; Boyd et al., 2015). Hay organismos que tienen una mayor cantidad de genes *nif*, pudiendo en ocasiones llegar a tener hasta 20 *nif* distintos (Mus et al., 2019). Uno de los genes *nif* más conservados es *nifH*, por ello, es habitual la utilización de secuencias *nifH* como gen marcador para inferir filogenias de organismos fijadores de nitrógeno (Zehr et al., 2003; Raymond et al., 2004; Gaby & Buckley, 2011; Angel et al., 2018). Los genes *nif* están ampliamente distribuidos en bacterias y arqueas, incluso en organismos que carecen de la capacidad de fijar nitrógeno, debido a que existen secuencias parálogas de *nif* que no codifican nitrogenasas. Utilizando *nifH* y *nifD*, se han clasificado las secuencias homólogas de la nitrogenasa en 5 grupos: los grupos I, II y III corresponden a los fijadores de nitrógeno, mientras que los microorganismos que carecen de esta capacidad pertenecen a los grupos IV y V.

En los grupos I, II y III se ha detectado actividad nitrogenasa. El grupo I está constituido por las nitrogenasas de Mo pertenecientes a cianobacterias y proteobacterias (Ghebreamlak & Mansoorabadi, 2020). Estas bacterias están estrechamente relacionadas con el O₂ mediante la respiración aeróbica o la fotosíntesis oxigénica. Por ello, poseen mecanismos espaciales y/o temporales que les permiten mantener separados la nitrogenasa y el O₂. Estos mecanismos reguladores frecuentemente son codificados por genes dentro del propio operón *nif* (Raymond

et al., 2004). El grupo II lo forman principalmente anaerobios obligados como los clostridios, bacterias acetogénicas y arqueas metanógenas del orden methanosarcinales. Poseen nitrogenasas de Mo, muy similares a las del grupo I. Los microorganismos pertenecientes al grupo II tienen un operón *nifHDKEN* muy conservado, teniendo una inserción de aproximadamente 50 aminoácidos en *nifD*, hecho que apoya el origen monofilético de este grupo (Wang et al., 1988; Kim & Woo, 1993). La realización de análisis filogenéticos basados en ARNr 16S indican que se ha producido transferencia horizontal de genes entre los organismos del grupo II, este proceso de intercambio genético puede haber sido favorecido por las frecuentes asociaciones simbióticas en organismos de ambientes anaerobios (Raymond et al., 2004). El grupo III está integrado por las nitrogenasas alternativas que no dependen de Mo, son las nitrogenasas dependientes de V o Fe (Raymond et al., 2004). La aparición de este tipo de nitrogenasas pudo tener lugar bajo condiciones limitantes de Mo en la Tierra Proterozoica, por lo que utilizar otros metales podía resultar beneficioso (Anbar & Knoll, 2002). Sin embargo, la Nitrogenasa de Mo tiene una mayor eficiencia que las dependientes de V y Fe, siendo capaz de fijar una mayor cantidad de nitrógeno con menos nitrogenasa total.

Aquellas secuencias homólogas a la nitrogenasa que no participan en la FBN se han clasificado en los grupos IV y V (Raymond et al., 2004). El grupo IV es un grupo parafilético, formado por metanógenos y bacterias que no fijan nitrógeno (la mayoría de ellas fotosintéticas). También incluye las secuencias similares a *nif* que aún no han sido caracterizadas, conocidas como *nifH* y *nifD* (Staples et al., 2007). Los análisis bioinformáticos proponen que las funciones de estos homólogos podrían ser la adquisición de metales, la fijación de carbono y el metabolismo del azufre (Ghebreamlak & Mansoorabadi, 2020). El grupo V lo constituyen parálogos de *nifH* como los genes *bchX* y *bchL*, que participan de forma esencial en la biosíntesis de bacterioclorofila. El origen de *bchX* y *bchL* está en una duplicación de genes en un ancestro común de todos los fotótrofos. El análisis filogenético de las secuencias de proteínas NifH, BchX y BchL sugiere que la fotosíntesis anoxigénica utilizando bacterioclorofila se originó antes que la fotosíntesis mediada por clorofila de las cianobacterias (Gupta, 2012).

Las campañas oceanográficas han ayudado a la investigación de los microorganismos y sus actividades en los océanos, analizando el papel de arqueas y bacterias en los ciclos biogeoquímicos de elementos como el C, N, P o S. Sin embargo, aún hay mucho desconocimiento acerca del medio marino y los microorganismos presentes en la gran cantidad de ecosistemas que abarcan los océanos. A raíz de esto, han surgido importantes campañas

oceanográficas como el proyecto Tara Oceans, una expedición llevada a cabo en una goleta entre los años 2009 y 2013, cuyos objetivos eran comprender los ciclos biogeoquímicos globales que tienen lugar en el océano y evaluar el impacto del cambio climático en la biodiversidad marina y los procesos oceánicos (Karsenti et al., 2011). Para ello, la expedición Tara Oceans tomó datos de diversos ecosistemas en océanos de todo el mundo, recogiendo 35.000 muestras de 210 ecosistemas en 20 provincias biogeográficas (Sunagawa et al., 2020). Las muestras que obtuvieron consistían en tomas de agua de mar, que incluían la gran variedad de plancton que estaba presente en el agua. El plancton supone una importante parte de la biomasa de los océanos y es clave en el mantenimiento de los ciclos biogeoquímicos globales, por lo que su análisis puede proporcionarnos mucha información (Pesant et al., 2015).

En esta expedición se recolectaron muestras a tres profundidades distintas: la capa de agua superficial (entre 3 y 7 m por debajo la superficie del mar), la capa máxima de clorofila profunda (su profundidad es variable y se determinó a partir de sensores ópticos con un fluorómetro de clorofila) y la zona mesopelágica (entre 200 y 1000 m de profundidad) (Pesant et al., 2015). Para cada muestreo fueron registradas la fecha, hora y ubicación, junto con una serie de propiedades físicas, químicas y biológicas del agua: temperatura, salinidad, pH, concentraciones de O₂, eficiencia fotosintética, puntos calientes de biodiversidad, corrientes o remolinos. También se midieron constantemente factores que afectan a la mezcla de las aguas superficiales: presión, temperatura, humedad, velocidad y dirección del viento. Además, se emplearon varios métodos de muestreo para poder capturar y separar eficazmente según su tamaño, la gran diversidad presente en el plancton. Los organismos muestreados fueron virus, girus (virus gigantes), procariotas (arqueas y bacterias), eucariotas unicelulares y eucariotas multicelulares (Pesant et al., 2015).

El proyecto Tara Oceans proporciona acceso abierto a los datos obtenidos, por lo que se puede acceder libremente a todas sus mediciones ambientales, datos moleculares y listas de nucleótidos. La integración de estos datos con metagenomas oceánicos y genomas de referencia permite obtener un catálogo de genes de referencia, con los que llevar a cabo clasificaciones taxonómicas, filogenéticas y ecológicas. Este proceso permitió distinguir 111,5 millones de secuencias de nucleótidos codificantes de proteínas, las cuales fueron agrupadas con un 95% de identidad nucleotídica en el alineamiento (Sunagawa et al., 2015). De esta manera se redujo el número tan grande de secuencias y se obtuvo un Catálogo de genes de referencia microbiano oceánico con más de 40 millones de genes no redundantes. Además, el 81,4% de los genes

descubiertos en las muestras de Tara Oceans no se habían identificado previamente (Sunagawa et al., 2015). Por tanto, la información resultante del proyecto Tara Oceans es una herramienta muy útil para investigaciones en Biodiversidad Microbiana Marina, Biotecnología y Bioinformática.

En este trabajo se han utilizado muestras de Tara Oceans para estudiar el papel de los microorganismos marinos en el proceso de fijación de nitrógeno. En concreto, analizamos el material genético de metatranscriptomas procedentes de muestras de Tara, lo que nos permitió investigar, mediante la identificación de secuencias del gen *nifH*, la diversidad de microorganismos implicados en la FBN.

2. Objetivos

El objetivo general de este trabajo es estudiar la diversidad existente entre los microorganismos fijadores de N₂ en los océanos, mediante un análisis bioinformático de metatranscriptomas, procedentes de muestras recogidas por todo el planeta durante la expedición Tara Oceans.

Para ello, se siguieron los siguientes objetivos específicos:

1. Utilizar diferentes herramientas computacionales, entre las que está el lenguaje de programación Python, para llevar a cabo un análisis bioinformático de secuencias ambientales de nucleótidos.
2. Construir un árbol filogenético de referencia mediante el método maximum likelihood, utilizando secuencias ya conocidas procedentes de la base de datos MAR.
3. Analizar qué taxones o grupos de bacterias están implicados en el proceso de fijación de N₂ en el medio marino.

3. Material y métodos

Con el fin de estudiar la diversidad de microorganismos fijadores de N₂ en los océanos, analizamos el material genético presente en las muestras recolectadas por la expedición Tara Oceans, entre los años 2009 y 2013, en más de 200 estaciones de muestreo que abarcaban todas las provincias oceánicas. Concretamente analizamos secuencias de ARNm procedentes de metatranscriptomas. Estos suponen el conjunto de transcritos de un sistema, en nuestro caso la comunidad microbiana marina, permitiéndonos observar qué genes se están expresando. Para ello, empleamos *nifH* como gen marcador. La clasificación taxonómica de los microorganismos se llevó a cabo mediante Genome Taxonomy Database Toolkit (GTDB-Tk; Chaumeil et al., 2019), un conjunto de herramientas de software que permiten hacer una asignación taxonómica objetiva para genomas de bacterias y arqueas. GTDB permite trabajar con cientos o miles de genomas ensamblados en un metagenoma y obtenidos directamente de muestras ambientales. Todo el proceso se llevó a cabo con lenguaje de programación Python.

En primer lugar, se construyó un árbol filogenético de referencia mediante el método maximum likelihood, a partir de secuencias de *nifH* presentes en genomas de referencia conocidos, procedentes de la base de datos MAR. Al árbol de referencia se añadieron aquellas secuencias ambientales de Tara que tenían el gen *nifH*, este proceso es denominado placement. Es necesario crear un árbol de referencia, debido a que las secuencias de Tara no se llegan a ensamblar del todo, dado que las secuencias ambientales son de peor calidad que las secuencias procedentes de bacterias aisladas. Por ello, creamos un árbol de referencia fiable con secuencias de *nifH* procedentes de genomas ya conocidos presentes en una base de datos, y posteriormente añadimos las secuencias de Tara sin modificar la topología de referencia. La base de datos utilizada para hacer el árbol de referencia fue MAR, una base de datos de referencia con genomas procarióticos marinos completamente secuenciados y genomas parciales de alta calidad (Klemetsen et al., 2018). Para construir el árbol de referencia utilizamos el software Prodigal, un algoritmo de predicción y reconocimiento de genes procarióticos capaz de identificar el sitio de iniciación de la traducción, permitiéndonos traducir a aminoácidos los nucleótidos de los que partimos.

Para cuantificar la expresión del gen marcador *nifH*, primero se identificó el gen mediante la búsqueda con blastp de todas las secuencias traducidas del metatranscriptoma frente al conjunto representativo de la base de datos, con un bitcore mínimo de 50. Luego, se colocaron los

péptidos en el árbol de referencia utilizando los paquetes PaPaRa, EPA-ng y gappa. El modelo de sustitución de aminoácidos se predijo con IQ-TREE. Por último, se eliminaron las secuencias colocadas con longitudes de rama más largas que la distancia más larga entre todos los pares de secuencias en el árbol de referencia (<10% de las secuencias). Estas secuencias fueron consideradas parálogas. La búsqueda de las secuencias no se llevó a cabo solo con blastp, también se ejecutó mediante el HMM (Modelo Oculto de Markov) específico de *nifH*. Este otro método es ideal para la identificación de organismos desconocidos, y se utiliza cuando la base de datos no recoge toda la diversidad existente, como ocurre en el medio marino. Por otra parte, la mayoría de las secuencias procedentes de muestras ambientales como Tara Oceans están truncadas, por lo que la búsqueda de HMM puede no detectarlas como *nifH* porque no llegan al valor mínimo o necesario que tiene el modelo. Por tal razón, utilizamos dos métodos para la búsqueda de secuencias de *nifH*: blastp y HMM.

Finalmente, clasificamos las secuencias de *nifH* en los grupos: I, II, III y IV. Esto fue posible gracias a una serie de residuos conservados en los péptidos de *nifH* y a un modelo dicotómico (un árbol de decisiones) desarrollado por (Frank et al., 2016). De esta forma, distinguimos qué secuencias correspondían a microorganismos fijadores de nitrógeno (Grupos I, II y III) y cuales no pertenecían realmente a *nifH* (Grupos IV y V).

4. Resultados y discusión

Se construyó un árbol filogenético de referencia a partir de secuencias de *nifH* presentes en genomas de referencia conocidos. A continuación, se alinearon e incorporaron aquellas secuencias de Tara Oceans que presentaban *nifH*. De esta forma, se creó un nuevo árbol que nos permitió analizar, por medio de las secuencias de *nifH*, la diversidad de diazotrofos a distintos niveles taxonómicos. Las secuencias referentes a *nifH* se clasificaron funcional y taxonómicamente en las siguientes categorías: clase, orden, familia y género. Algunas secuencias de *nifH* no se han podido identificar y se indican en las **Figuras 1, 2, 3 y 4** como no especificadas (“Unspecified”). No se pudieron clasificar porque los microorganismos a los que pertenecen estas secuencias todavía no se han aislado, y no hay nada suficientemente parecido en las bases de datos, la única información que tenemos sobre ellos es que son bacterias o arqueas. También puede ocurrir que algunas secuencias no se hayan podido clasificar debido a que la base de datos de referencia no incluya todas las secuencias conocidas, o a que las secuencias cortas que situamos en el árbol (placement) comprendan regiones del gen que están conservadas a un nivel taxonómico superior, por tanto, el algoritmo no puede situarlas bien en niveles taxonómicos inferiores (por ejemplo, si sólo puede situar la secuencia a nivel de familia, entonces el género y la especie aparecerán como “Unspecified”).

A nivel de clase destaca que la mayoría de secuencias se clasifican en *Gammaproteobacteria*. La abundancia de este grupo es tal, que en las muestras analizadas suponen más del 50% de organismos involucrados en la fijación de nitrógeno. Además, podemos observar la presencia de *Cyanobacteria*, grupo al que con frecuencia se atribuye la FBN en los océanos. Sin embargo, podemos afirmar que también hay otros muchos taxones implicados en este proceso, como son *Alphaproteobacteria*, *Phycisphaerae* (filo *Planctomycetota*), *Desulfovibrionia*, *Desulfobacteria*, *Verrucomicrobiae*, *Campylobacteria*, *Bacteroidia* y UBA9042 (**Figura 1**). Este último es una clase perteneciente al filo *Myxococcota*. Se sabe que los organismos incluidos en la clase UBA9042 se diferencian lo suficiente de otros microorganismos como para ser considerados una clase, pero todavía no se han aislado, por lo que aún no se ha asignado un nombre a este grupo. La información conocida sobre la clase UBA9042 se puede encontrar en GTDB (Genome Taxonomy Database). Finalmente, pese a que la mayoría de secuencias han podido ser identificadas a nivel de clase, en aproximadamente un 20% del total esto no ha sido posible, indicándose estas secuencias como no especificadas.

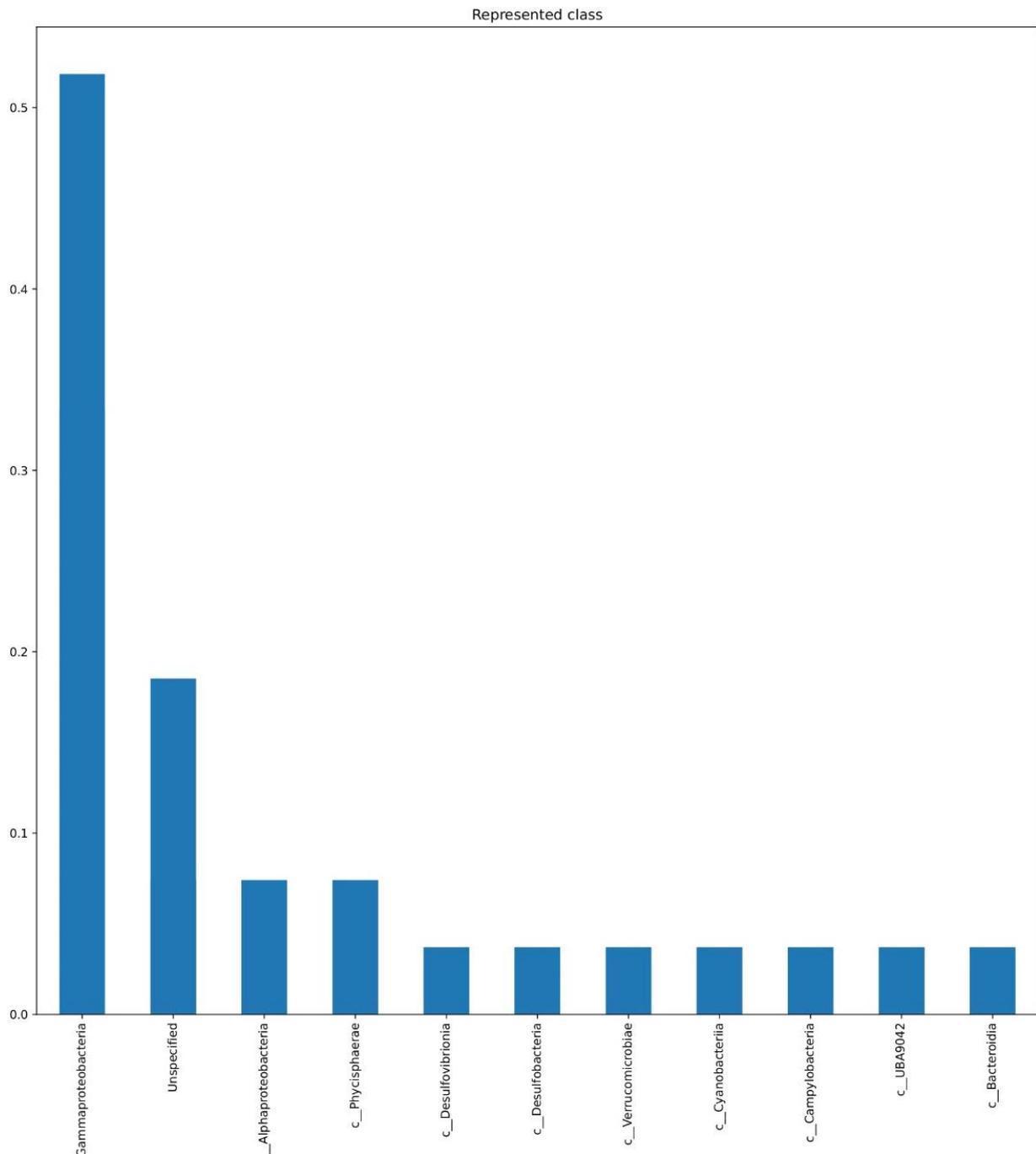


Figura 1. Diversidad de diazotrofos a nivel de clase en las muestras analizadas de Tara Oceans. En el eje x se indica el nombre de las clases más abundantes en las muestras analizadas, mientras que en el eje y se representa de 0 a 1 la fracción de secuencias que se clasifican dentro de cada clase. Más de la mitad de las secuencias se clasifican como *Gammaproteobacteria*. Aproximadamente el 20% de las secuencias no se pudo clasificar.

Respecto a los órdenes, *Pseudomonadales* fue el más abundante con casi un 35% de las secuencias. Pertenece a la clase *Gammaproteobacteria*, así como el orden *Enterobacterales*, aunque este tiene un porcentaje mucho menor que los *Pseudomonadales*. Es razonable que el orden más abundante pertenezca a *Gammaproteobacteria*, debido a que era la clase dominante

con gran diferencia. Los órdenes que siguen en abundancia a *Pseudomonadales* son *Rhodobacterales* (clase *Alphaproteobacteria*) y *Phycisphaerales* (filo *Planctomycetota*). Otros órdenes con una representación importante son *Desulfovibrionales*, *Desulfobacterales*, *Opitutales*, *Cyanobacterales*, *Enterobacterales*, *Burkholderiales*, *Campylobacterales*, UBA3505 (clase UBA9042) y *Bacteroidales* (**Figura 2**). De la misma manera que ocurría en las clases, en los órdenes casi un 20% de las secuencias no han podido ser identificadas.

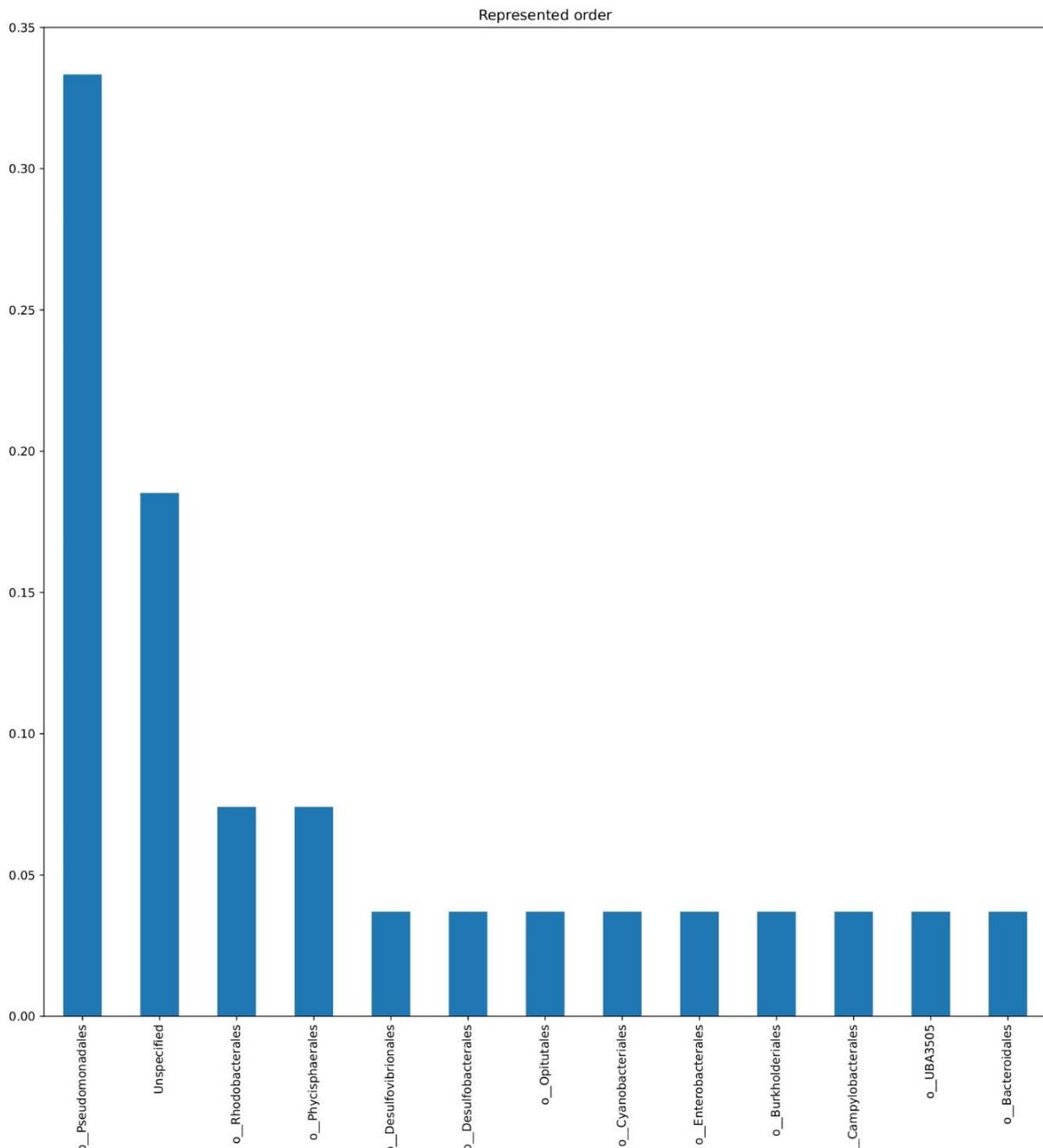


Figura 2. Diversidad de diazotrofos a nivel de orden en las muestras analizadas de Tara Oceans. Se muestran los órdenes más abundantes en el eje x. La fracción de secuencias clasificadas dentro de cada orden (valores de 0 a 1) es mostrada en el eje y. Casi un 35% de las secuencias corresponden al orden Pseudomonadales, perteneciente a la clase Gammaproteobacteria. El 20% de las secuencias no se pudo clasificar a nivel de orden.

A nivel de familia predominan las secuencias no especificadas, debido a que aproximadamente el 50% de las secuencias no se han podido clasificar. Dentro de las familias identificadas, las más representativas cuantitativamente fueron *Pseudomonadaceae*, *Rhodobacteraceae* y *Phycisphaeraceae*, corresponden a familias pertenecientes a los 3 órdenes más abundantes: *Pseudomonadales*, *Rhodobacterales* y *Phycisphaerales*. Las otras familias que se representan en la **Figura 3** son *Desulfovibrionaceae*, UBA3067, *Oleiphilaceae*, *Desulfobacteraceae*, *Opitutaceae*, *Microcystaceae*, *Nitrincolaceae*, *Psychromonadaceae*, *Rhodocyclaceae*, *Arcobacteraceae*, UBA3505 y *Prolixibacteraceae*.

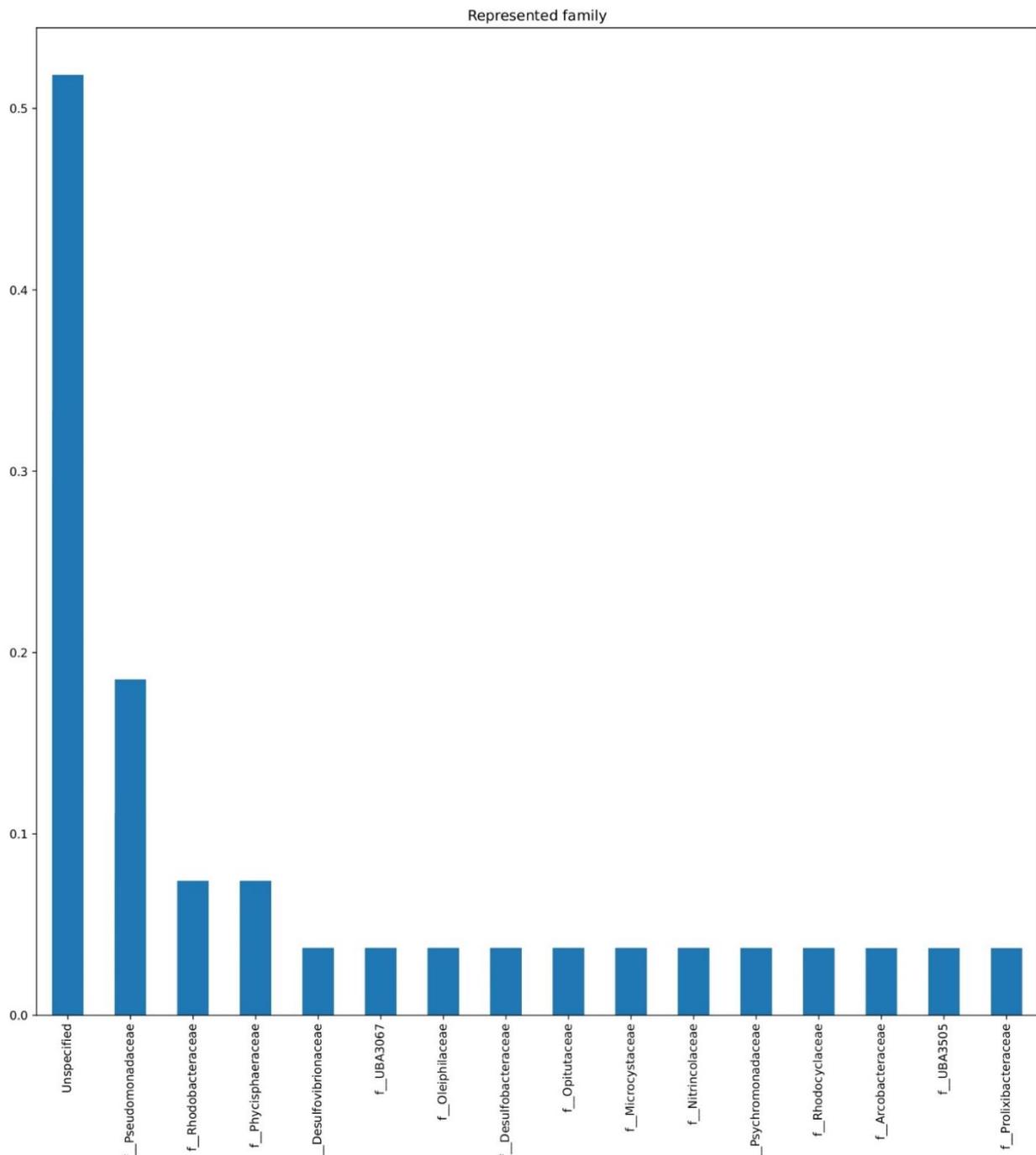


Figura 3. Diversidad de diazotrofos a nivel de familia en las muestras analizadas de Tara Oceans. En el eje x se muestran los nombres de las familias más abundantes, mientras que en el eje y se puede observar la fracción de secuencias que se clasifican dentro de cada familia representadas con valores entre 0 y 1. Las familias más abundantes fueron *Pseudomonadaceae*, *Rhodobacteraceae* y *Phycisphaeraceae*. En más de la mitad de las secuencias no se pudo identificar a qué familia pertenecían.

Por último, el 50% de las secuencias no se han podido clasificar en géneros, hecho que ya sucedió con las familias. Respecto a los géneros que sí se pudieron identificar, los que presentan una mayor abundancia son *Pseudomonas*, UBA7800 y *Marinibacterium*. UBA7800 es un género perteneciente al orden *Phycisphaerales*, pero ha sido descubierto recientemente y las especies que pertenecen a este grupo todavía no han sido descritas. La información conocida sobre el género UBA7800 se puede encontrar en GTDB. Otros géneros que fueron identificados como fijadores de nitrógeno son *Pseudodesulfovibrio*, UBA3067, *Marinobacter*, DG1235, *Atelocyanobacterium*, *Motiliproteus*, *Psychromonas*, *Azoarcus*, *Arcobacter*, UBA3505 y *Sunxiuqinia* (**Figura 4**). Las secuencias de *Azoarcus* se consiguieron identificar especialmente bien, llegando incluso a ser identificadas a nivel de especie, concretamente la especie *Azoarcus communis*. Otro género a destacar es el recientemente descubierto *Atelocyanobacterium*, formado únicamente por la especie UCYN-A, una cianobacteria que vive en simbiosis con un alga primnesiófita unicelular.

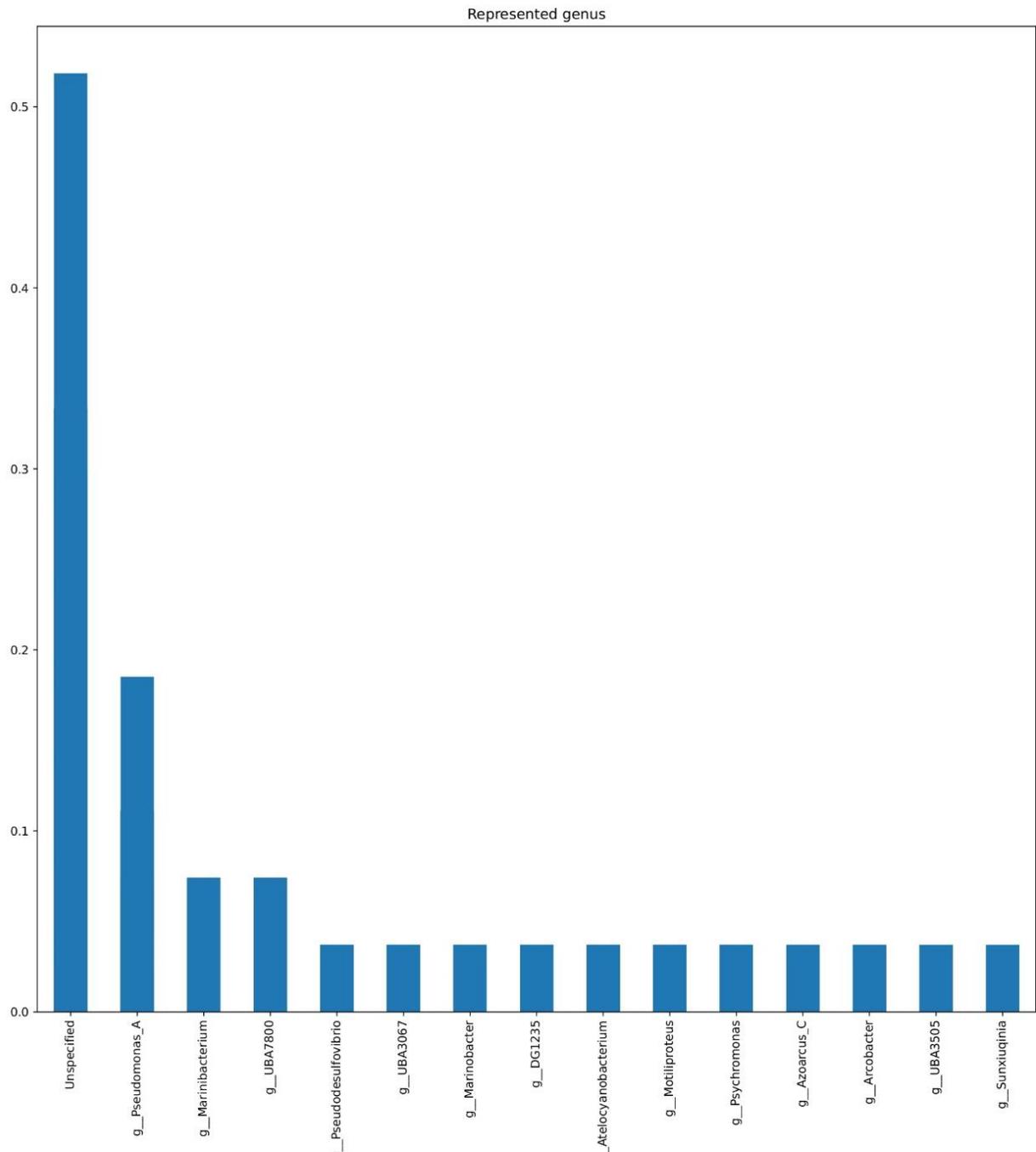


Figura 4. Diversidad de diazotrofos a nivel de género en las muestras analizadas de Tara Oceans. Podemos observar el nombre de los géneros más abundantes en el eje x, mientras que en el eje y se indica de 0 a 1 la fracción de secuencias clasificadas dentro de cada género. Más del 50% de las secuencias no se pudieron clasificar a nivel de género. Respecto a las que sí se pudieron clasificar, el género *Pseudomonas* es el más abundante, seguido por *Marinibacterium* y UBA7800.

Para todos los niveles taxonómicos (**Figuras 1, 2, 3 y 4**) hay microorganismos que se indican como no especificados (“Unspecified”). En algunos casos puede ocurrir que las secuencias no se identificaran porque la base de datos no incluye todas las secuencias conocidas, sin embargo, la base de datos utilizada como referencia contiene una cantidad de secuencias considerable, por lo que es probable que en muchos casos estos microorganismos no se hayan podido clasificar porque todavía no se han aislado. Esto es un ejemplo de que aún hay muchos microorganismos marinos que se desconocen. En las muestras que analizamos, alrededor del 50% de las secuencias no se pudieron clasificar en género y familia. En cambio, en niveles taxonómicos superiores como la clase y el orden, el número de secuencias que no se pudo identificar fue menor, únicamente un poco menos del 20%. En parte, esto puede deberse a que algunas de las secuencias ambientales de Tara incorporadas al árbol comprenden regiones del gen *nifH* conservadas a niveles taxonómicos superiores, por lo que el algoritmo utilizado no puede clasificarlas a nivel de familia y género. Podemos concluir que, como es lógico, hay un mayor desconocimiento de los niveles taxonómicos inferiores.

También es reseñable la gran variedad de microorganismos, pertenecientes a grupos muy diversos, que participan en la fijación de nitrógeno. Esta diversidad de diazotrofos incluye a múltiples clases como *Gammaproteobacteria*, *Cyanobacteria*, *Alphaproteobacteria*, *Desulfovibrionia*, *Desulfobacteria*, *Phycisphaerae*, *Verrucomicrobiae*, *Campylobacteria* y *Bacteroidia*. Por tanto, debido al gran número de organismos implicados, podemos afirmar que la FBN es un proceso muy complejo. El grupo más importante cuantitativamente en la categoría de clase fueron las *Gammaproteobacteria*, incluyendo el 50% de fijadores de nitrógeno en las muestras analizadas. Esto se refleja también en el resto de categorías taxonómicas, puesto que el orden *Pseudomonadales* (perteneciente a la clase *Gammaproteobacteria*) es el orden más abundante y uno de los que presenta mayor diversidad, estando integrado por varias familias con la capacidad de fijar nitrógeno: *Pseudomonadaceae*, UBA3067, *Oleiphilaceae* y *Nitrincolaceae*. El género con mayor abundancia es *Pseudomonas*, que pertenece a los ya mencionados familia *Pseudomonadaceae* y orden *Pseudomonadales*. En consecuencia, podemos concluir que la clase *Gammaproteobacteria* es muy importante para la FBN, a causa de su gran aportación cuantitativa y a la diversidad de organismos dentro del grupo que contribuyen en este proceso.

Dentro de la diversidad existente entre los fijadores de nitrógeno, cabe destacar también la presencia de *Atelocyanobacterium thalassa*, comúnmente denominada UCYN-A, ya que es una cianobacteria unicelular que vive en simbiosis con un alga unicelular de la clase *Primnesiophyceae*. UCYN-A carece de genes que codifican el fotosistema II y otras vías esenciales para la fijación de carbono mediante el ciclo de Calvin. La relación que mantiene con esta alga se basa en un intercambio de carbono y nitrógeno fijos, lo que permite a la cianobacteria UCYN-A habitar en ambientes oligotróficos a pesar de tener un genoma reducido. Esta cianobacteria tiene una amplia distribución y contribuye de manera importante a la fijación de nitrógeno en los océanos.

Existen 4 linajes de UCYN-A que coexisten en tiempo y espacio frecuentemente, teniendo una distribución global: UCYN-A1, UCYN-A2, UCYN-A3, y UCYN-A4. Cada linaje está asociado a un alga primnesiófita específica, que actúa como anfitrión. Se sabe que la especie con la que se asocia UCYN-A2 es *Braarudosphaera bigelowii*, mientras que UCYN-A1 tiene como anfitrión a un alga primnesiófita de otra especie, pero estrechamente relacionada filogenéticamente con *B. bigelowii*. Los anfitriones de UCYN-A3 y UCYN-A4 todavía se desconocen. El número de células de UCYN-A implicadas en la asociación varía en función del linaje y el huésped. El linaje UCYN-A1 presenta únicamente una o dos células por cada primnesiófita, mientras que en el linaje UCYN-A2 cada huésped puede contener hasta 10 individuos de UCYN-A. Además, el primnesiófita con el que se relaciona cada linaje tiene diferente tamaño, lo que permite que UCYN-A participe en la fijación de nitrógeno en diversas fracciones de tamaño de plancton.

La bacteria UCYN-A puede ser importante en el proceso de fijación de nitrógeno en algunos puntos del océano. Con el objetivo de comprobar esta importancia en el Océano Atlántico Sur, se analizaron un total de cuatro metatranscriptomas, pertenecientes a las estaciones de Tara 076 y 078 (**Figura 5**). Para cada estación utilizamos los datos de dos muestras de metatranscriptomas tomadas a diferente profundidad, una en la capa superficial (SRF) y otra en la zona de máxima clorofila profunda (DCM). El tamaño de la fracción de plancton utilizado en la toma de todas las muestras estuvo entre 0.22-0.3 μm . Esta información se recoge en la **Tabla 1**.

Tabla 1. Muestras utilizadas para el reclutamiento de fragmentos de linajes de UCYN-A. Se indica el nombre de identificación de la muestra (ID), la estación en la que fue tomada con sus coordenadas geográficas y la respectiva región, el tamaño en μm de la fracción tomada y la profundidad en m a la que se recogió la muestra junto con la capa correspondiente (SRF: Capa superficial; DCM: Capa máxima de clorofila profunda).

Muestra (ID)	Estación	Región	Latitud	Longitud	Capa	Profundidad (m)	Fracción (μm)
ERS490542	076	Océano Atlántico Sur	-20,9354	-35,1803	SRF	5	0.22-3
ERS490597	076	Océano Atlántico Sur	-21,0292	-35,3498	DCM	150	0.22-3
ERS490659	078	Océano Atlántico Sur	-30,1367	-43,2899	SRF	5	0.22-3
ERS490691	078	Océano Atlántico Sur	-30,1484	-43,2705	DCM	120	0.22-3



Figura 5. Mapa en el que se representan las estaciones de Tara Oceans 076 y 078. Se indica la ubicación de las estaciones de Tara 076 y 078, localizadas en la región occidental del Océano Atlántico Sur. En estas ubicaciones se recogieron las muestras que utilizamos para llevar a cabo un reclutamiento de fragmentos de linajes de UCYN-A, con el fin de comprobar la presencia de UCYN-A en esta región.

El análisis realizado consistió en el reclutamiento de fragmentos de cuatro metatranscriptomas pertenecientes a las estaciones 076 y 078. Además, se utilizó como genoma de referencia un genoma de UCYN-A1 que ya había sido secuenciado. Los fragmentos se clasificaron según su % de identidad respecto a las secuencias del genoma de referencia de UCYN-A, estableciendo el 95% de identidad como el límite para considerarlos la misma especie (**Figura 6**).

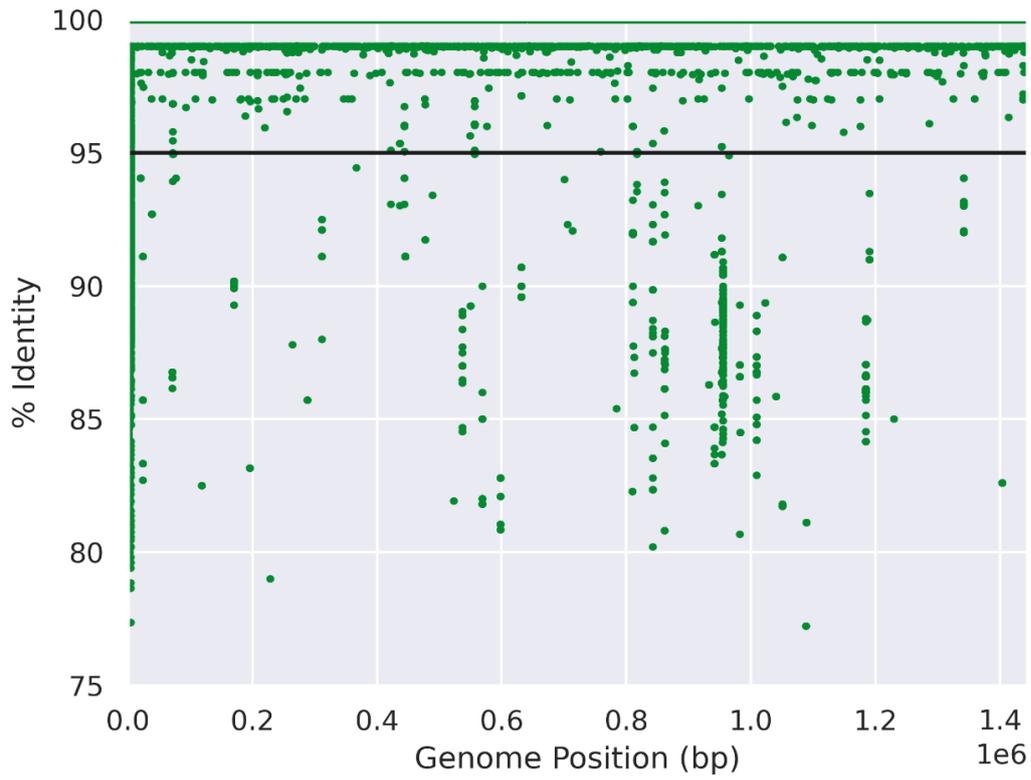


Figura 6a. Muestra 490542 (Estación 076). Profundidad: 5m

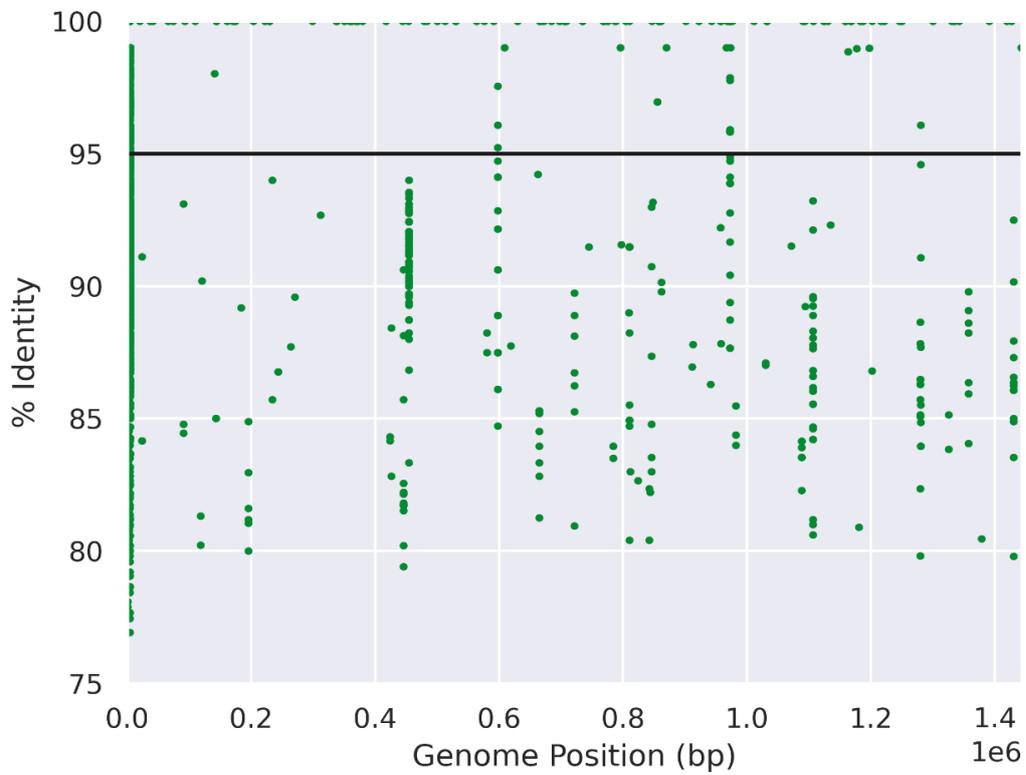


Figura 6b. Muestra 490597 (Estación 076). Profundidad: 150m

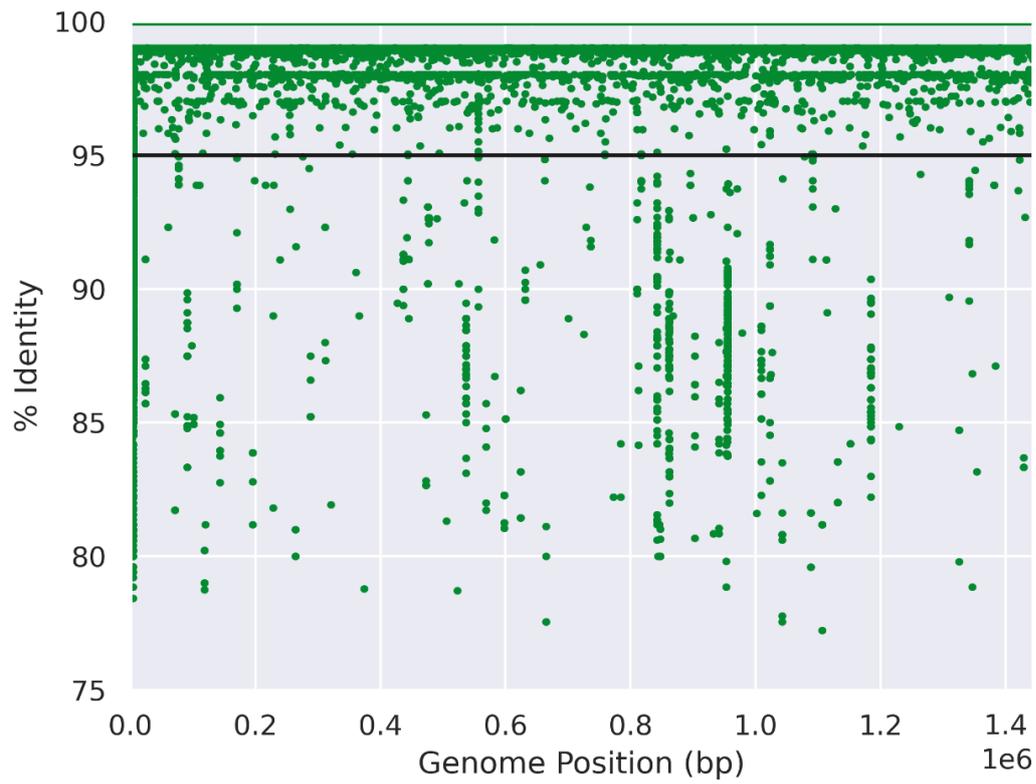


Figura 6c. Muestra 490659 (Estación 078). Profundidad: 5m

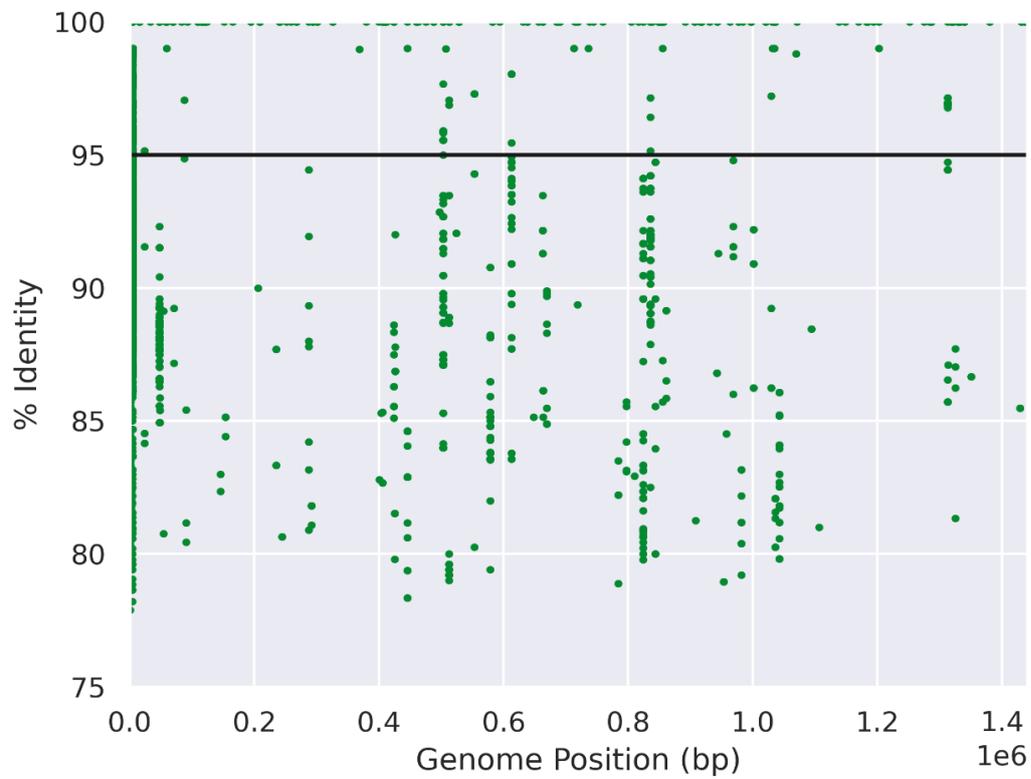


Figura 6d. Muestra 490691 (Estación 078). Profundidad: 120m

Figura 6. Representación de lecturas metatranscriptómicas utilizando el genoma de UCYN-A1 como referencia. Cada lectura se representa con un punto, situado en la gráfica según su posición en el genoma en pares de base (eje x) y el % de identidad respecto al genoma de referencia (eje y). Con una línea negra se resalta el 95% de identidad, umbral establecido para considerar qué lecturas pertenecen a la misma especie que el genoma de referencia. Se analizaron un total de 1.400.000 posiciones en el genoma.

Analizando los resultados obtenidos tras el reclutamiento de fragmentos de cuatro metatranscriptomas (**Figura 6**), podemos afirmar que UCYN-A se distribuye principalmente en aguas someras y poco profundas, ya que se observa una cantidad de lecturas mucho mayor en las **Figuras 6a y 6c**. Existe una enorme abundancia en las muestras tomadas en la capa superficial a 5m de profundidad (**Figuras 6a y 6c**), mientras que en las muestras tomadas en la capa máxima de clorofila profunda a 120m y 125m de profundidad, la abundancia es mucho menor (**Figuras 6b y 6d**). A la izquierda de las gráficas (**Figura 6**) hay una gran cantidad de secuencias, incluso llegando a formar una línea, esto se debe a que corresponde al principio del genoma del ARN ribosómico, que está muy conservado, por lo que coincide en muchas bacterias aunque estén muy distantes filogenéticamente. Finalmente, estos análisis parecen confirmar la presencia de UCYN-A en el Océano Atlántico Sur. Además, posiblemente esta bacteria tenga una gran importancia en el proceso de fijación de nitrógeno, debido al gran número de secuencias correspondientes a UCYN-A, que implican que es muy numerosa.

5. Conclusiones

1. La bioinformática es una herramienta muy útil para el estudio de los microorganismos, permitiendo clasificarlos eficazmente y analizar una enorme cantidad de muestras. Para la mayoría de los microorganismos marinos, los métodos bioinformáticos suponen la única forma de estudio, ya que muy pocos son cultivables.
2. El proceso de FBN es muy complejo, existiendo una gran diversidad taxónomica de bacterias implicadas. En las muestras analizadas el taxón más abundante fue *Gammaproteobacteria*, grupo al que pertenecían más del 50% de las secuencias de *nifH*. Además, comprobamos que UCYN-A, una *Cyanobacteria* que vive en simbiosis con un alga primnesiófita unicelular, está presente y tiene una gran importancia en el Océano Atlántico Sur, distribuyéndose principalmente en aguas superficiales.
3. Pese a que el análisis de secuencias de *nifH* es un método eficaz para clasificar a los microorganismos fijadores de N₂ en los océanos, no se pudieron identificar todas las secuencias. En algunos casos se puede deber a que el gen *nifH* tiene regiones conservadas en niveles taxonómicos superiores, pero una gran parte de las secuencias no pudieron ser clasificadas porque hay muchos microorganismos marinos desconocidos o que todavía no han sido aislados.
4. En la actualidad aún queda mucho por investigar acerca de los microorganismos que fijan N₂ en los océanos. En futuros estudios se podrá determinar con mayor precisión qué microorganismos, y en qué proporción, participan en la FBN.

6. Conclusions

1. Bioinformatics is a very useful tool for the study of microorganisms, it allows to make an effective classification and analyze a large number of samples. For most marine microorganisms, bioinformatic methods are the only way to study, since very few are cultivable.
2. The BNF process is very complex, there is a great taxonomic diversity of bacteria involved. In the analyzed samples, the most abundant taxon was *Gammaproteobacteria*, which have more than 50% of *nifH* sequences. In addition, we verify UCYN-A, a *Cyanobacteria* that lives in symbiosis with an unicellular prymnesiophyte algae, is present and of great importance in the South Atlantic Ocean, mainly found in surface waters.
3. Although *nifH* sequences analysis is an effective method for classifying N₂-fixing microorganisms in the oceans, not all sequences could be identified. In some cases, it may be because *nifH* gene has conserved regions at higher taxonomic levels, but a large part of the sequences could not be classified because there are many unknown marine microorganisms, which have not been isolated yet.
4. At present, there is still much to investigate about N₂-fixing microorganisms in the oceans. In future studies, it will be possible to determine with higher precision which microorganisms, and in what proportion, participate in BNF.

7. Bibliografía

- Anbar, A. D., & Knoll, A. H.** (2002). Proterozoic Ocean Chemistry and Evolution: A Bioinorganic Bridge? *Science*, 297(5584), 1137-1142.
- Angel, R., Nepel, M., Panhölzl, C., Schmidt, H., Herbold, C. W., Eichorst, S. A., & Wobken, D.** (2018). Evaluation of Primers Targeting the Diazotroph Functional Gene and Development of NifMAP – A Bioinformatics Pipeline for Analyzing nifH Amplicon Data. *Frontiers in Microbiology*, 9.
- Bentzon-Tilia, M., Traving, S. J., Mantikci, M., Knudsen-Leerbeck, H., Hansen, J. L., Markager, S., & Riemann, L.** (2015). Significant N₂ fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. *The ISME Journal*, 9(2), 273-285.
- Blackburn, T. H., Sørensen, J., & International Council of Scientific Unions (Eds.).** (1988). *Nitrogen cycling in coastal marine environments*. Published on behalf of the Scientific Committee on Problems of the Environment (SCOPE) of the International Council of Scientific Unions (ICSU) by Wiley.
- Bonnet, S., Caffin, M., Berthelot, H., Grosso, O., Benavides, M., Helias-Nunige, S., Guieu, C., Stenegren, M., & Foster, R. A.** (2018). In-depth characterization of diazotroph activity across the western tropical South Pacific hotspot of N₂ fixation (OUTPACE cruise). *Biogeosciences*, 15(13), 4215-4232.
- Boyd, E. S., Costas, A. M. G., Hamilton, T. L., Mus, F., & Peters, J. W.** (2015). Evolution of Molybdenum Nitrogenase during the Transition from Anaerobic to Aerobic Metabolism. *Journal of Bacteriology*, 197(9), 1690-1699.
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H.** (2019). GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics (Oxford, England)*, btz848.
- Dekas, A. E., Poretsky, R. S., & Orphan, V. J.** (2009). Deep-Sea Archaea Fix and Share Nitrogen in Methane-Consuming Microbial Consortia. *Science*, 326(5951), 422-426.
- Dos Santos, P. C., Fang, Z., Mason, S. W., Setubal, J. C., & Dixon, R.** (2012). Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics*, 13(1), 162.
- Einsle, O.** (2014). Nitrogenase FeMo cofactor: An atomic structure in three simple steps. *Journal of Biological Inorganic Chemistry: JBIC: A Publication of the Society of Biological Inorganic Chemistry*, 19(6), 737-745.

- Frank, I. E., Turk-Kubo, K. A., & Zehr, J. P. (2016).** Rapid annotation of *nifH* gene sequences using classification and regression trees facilitates environmental functional gene analysis. *Environmental Microbiology Reports*, 8(5), 905-916.
- Gaby, J. C., & Buckley, D. H. (2011).** A global census of nitrogenase diversity. *Environmental Microbiology*, 13(7), 1790-1799.
- Ghebreamlak, S. M., & Mansoorabadi, S. O. (2020).** Divergent Members of the Nitrogenase Superfamily: Tetrapyrrole Biosynthesis and Beyond. *Chembiochem: A European Journal of Chemical Biology*, 21(12), 1723-1728.
- Großkopf, T., Mohr, W., Baustian, T., Schunck, H., Gill, D., Kuypers, M. M. M., Lavik, G., Schmitz, R. A., Wallace, D. W. R., & LaRoche, J. (2012).** Doubling of marine dinitrogen-fixation rates based on direct measurements. *Nature*, 488(7411), 361-364.
- Jasniewski, A. J., Lee, C. C., Ribbe, M. W., & Hu, Y. (2020).** Reactivity, Mechanism, and Assembly of the Alternative Nitrogenases. *Chemical Reviews*, 120(12), 5107-5157.
- Joerger, R. D., Bishop, P. E., & Evans, H. J. (1988).** Bacterial Alternative Nitrogen Fixation Systems. *CRC Critical Reviews in Microbiology*, 16(1), 1-14.
- Karsenti, E., Acinas, S. G., Bork, P., Bowler, C., Vargas, C. D., Raes, J., Sullivan, M., Arendt, D., Benzoni, F., Claverie, J.-M., Follows, M., Gorsky, G., Hingamp, P., Iudicone, D., Jaillon, O., Kandels-Lewis, S., Krzic, U., Not, F., Ogata, H., ... Consortium, the T. O. (2011).** A Holistic Approach to Marine Eco-Systems Biology. *PLOS Biology*, 9(10), e1001177.
- Kessler, P. S., Blank, C., & Leigh, J. A. (1998).** The *nif* gene operon of the methanogenic archaeon *Methanococcus maripaludis*. *Journal of Bacteriology*, 180(6), 1504-1511.
- Kim, J., & Woo, D. (1993).** X-ray Crystal Structure of the Nitrogenase Molybdenum-Iron Protein from *Clostridium pasteurianum* at 3.0-Å Resolution. *Protein Structure*, 12.
- Klemetsen, T., Raknes, I. A., Fu, J., Agafonov, A., Balasundaram, S. V., Tartari, G., Robertsen, E., & Willassen, N. P. (2018).** The MAR databases: Development and implementation of databases specific for marine metagenomics. *Nucleic Acids Research*, 46(D1), D692-D699.
- Mus, F., Colman, D. R., Peters, J. W., & Boyd, E. S. (2019).** Geobiological feedbacks, oxygen, and the evolution of nitrogenase. *Free Radical Biology & Medicine*, 140, 250-259.
- Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., Iudicone, D., Karsenti, E., Speich, S., Troublé, R., Dimier, C., & Searson, S. (2015).** Open science resources for the discovery and analysis of Tara Oceans data. *Scientific Data*, 2(1), 150023.
- Raymond, J., Siefert, J. L., Staples, C. R., & Blankenship, R. E. (2004).** The Natural History of Nitrogen Fixation. *Molecular Biology and Evolution*, 21(3), 541-554.

- Staples, C. R., Lahiri, S., Raymond, J., Von Herbulis, L., Mukhophadhyay, B., & Blankenship, R. E.** (2007). Expression and Association of Group IV Nitrogenase NifD and NifH Homologs in the Non-Nitrogen-Fixing Archaeon *Methanocaldococcus jannaschii*. *Journal of Bacteriology*, *189*(20), 7392-7398.
- Sunagawa, S., Acinas, S. G., Bork, P., Bowler, C., Eveillard, D., Gorsky, G., Guidi, L., Iudicone, D., Karsenti, E., Lombard, F., Ogata, H., Pesant, S., Sullivan, M. B., Wincker, P., & de Vargas, C.** (2020). Tara Oceans: Towards global ocean ecosystems biology. *Nature Reviews Microbiology*, *18*(8), 428-445.
- Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., Bork, P.** (2015). Structure and function of the global ocean microbiome. *Science*, *348*(6237), 1261359.
- Suzuki, S., Kawachi, M., Tsukakoshi, C., Nakamura, A., Hagino, K., Inouye, I., & Ishida, K.** (2021). Unstable Relationship Between *Braarudosphaera bigelowii* (= *Chrysochromulina parkeae*) and Its Nitrogen-Fixing Endosymbiont. *Frontiers in Plant Science*, *12*.
- Tie, X., Zhang, R., Brasseur, G., & Lei, W.** (2002). Global NO_x Production by Lightning. *Journal of Atmospheric Chemistry*, *43*(1), 61-74.
- Vitousek, P., & Howarth, R.** (1991). Nitrogen limitation on land and in the sea: How can it occur? *Biogeochemistry* 13:87-115. *Biogeochemistry*, *13*, 87-115.
- Voss, M., Bange, H. W., Dippner, J. W., Middelburg, J. J., Montoya, J. P., & Ward, B.** (2013). The marine nitrogen cycle: Recent discoveries, uncertainties and the potential relevance of climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1621), 20130121.
- Wang, S. Z., Chen, J. S., & Johnson, J. L.** (1988). Distinct structural features of the .alpha. and .beta. subunits of nitrogenase molybdenum-iron protein of *Clostridium pasteurianum*: An analysis of amino acid sequences. *Biochemistry*, *27*(8), 2800-2810.
- Zehr, J. P., Jenkins, B. D., Short, S. M., & Steward, G. F.** (2003). Nitrogenase gene diversity and microbial community structure: A cross-system comparison. *Environmental Microbiology*, *5*(7), 539-554.