



universidad
de león



FACULTAD DE CIENCIAS BIOLÓGICAS Y AMBIENTALES

**ANÁLISIS FILOGENÉTICO DE ARQUEAS
AMONIO-OXIDANTES ENFOCADO EN LA
EVOLUCIÓN DE LOS GENES DEL OPERÓN *AMO*.**

**PHYLOGENETIC ANALYSIS OF
AMMONIUM-OXIDIZING ARCHAES
FOCUSED ON THE EVOLUTION OF THE
GENES OF THE AMO OPERON.**

Autor: Sara Domínguez Llata

GRADO EN BIOTECNOLOGÍA

Septiembre, 2020

Índice

1. INTRODUCCIÓN	1
2. OBJETIVOS	3
3. MATERIALES Y MÉTODOS	3
3.1. BÚSQUEDA DE SECUENCIAS EN BASES DE DATOS.	3
3.2. PROGRAMACIÓN.	5
3.3: SELECCIÓN DE MÉTODO MSA Y MODELO FILOGENÉTICO.	6
4. RESULTADOS Y DISCUSIÓN	7
4.1 rRNA 16S	8
4.2 <i>AMO_A</i> , GEN CODIFICANTE PARA LA SUBUNIDAD A DE LA AMONIO MONOOXIGENASA.	11
4.3 <i>AMO_B</i> , GEN CODIFICANTE PARA LA SUBUNIDAD B DE LA AMONIO MONOOXIGENASA.	13
4.4 <i>AMO_C</i> , GEN CODIFICANTE PARA LA SUBUNIDAD C DE LA AMONIO MONOOXIGENASA.	15
4.5 <i>ACO_A</i> , GEN CODIFICANTE PARA LA ACETIL-CoA SINTETASA.	18
4.6 <i>ACO</i> , GEN CODIFICANTE PARA LA ENZIMA ACONITASA.	19
5. CONCLUSIONES	24
BIBLIOGRAFÍA	25
ANEXOS	29
ANEXO I: <i>PIPELINE</i> DE BASH	29
ANEXO II: SCRIPT PERL SEPARACIÓN	29
ANEXO III: SCRIPT PERL EXTRACCIÓN	31

Resumen

Las arqueas oxidantes del amonio (AOA) son fundamentales en el desarrollo del ciclo biogeoquímico del nitrógeno. Este proceso metabólico está catalizado por la enzima Amo, la cual está formada por 3 subunidades: AmoA, AmoB y AmoC. El gen *amoA* es el segundo gen más empleado como marcador genético para estudios filogenéticos y como herramienta de clasificación taxonómica. El objetivo del trabajo es analizar la filogenia de distintas especies de AOA del grupo Thaumarchaeota en relación con estos tres genes, junto a dos genes esenciales para el metabolismo, los que codifican para las enzimas Aconitasa y Acetil-CoA sintetasa, y con el gen codificante para el RNA ribosomal 16S rRNA. Otro objetivo es determinar la posición filogenética de las especies estudiadas, con relación a arqueas del Grupo Asgard, como *Promethoarchaeum*. Para este análisis se establece un *pipeline* mediante herramientas bioinformáticas que permita automatizar parte del proceso. Este junto con otras herramientas bioinformáticas facilita la elaboración de análisis filogenéticos que discuten la utilidad de *amoA*, muestran la naturaleza del gen *amoB*, y sugieren la importancia de la transferencia genética horizontal en los procesos evolutivos.

Abstract

Ammonium oxidizing archaea (AOA) are essential in the development of the nitrogen biogeochemical cycle. This metabolic process is catalyzed by the Amo enzyme, which is made up of 3 subunits: AmoA, AmoB and AmoC. The *amoA* gene is the second most widely used gene as a genetic marker for phylogenetic studies and as a taxonomic classification tool. The main aim of the work is to analyze the phylogeny of different species of Thaumarchaeota AOA respecting to these three genes, two genes essential for metabolism (those that encode the enzymes Aconitasa and Acetyl-CoA synthetase) and to the 16S rRNA. Another objective is to determine the phylogenetic position of the species that are studied, with special emphasis on *Promethoarchaeum*, an arch of the Asgard group. For all this, a pipeline is established using bioinformatic tools that automate part of the process. This, along with other bioinformatics tools, allows the development of phylogenetic analysis that corroborate the use of *amoA*, shows the nature of the *amoB* gene, and reveals the importance of horizontal gene transfer in evolutionary processes.

Keywords: AOA, Asgard, Ammonium, Amo, Phylogeny, Thaumarchaeota.

1. Introducción

La nitrificación se conoce como el proceso biológico de oxidación del amonio, presente en el medio por la descomposición de materia orgánica, a nitrato. Dicha oxidación ocurre en dos pasos: el primer paso transforma los iones amonio presentes en el entorno a iones nitrito, y el segundo paso transforma estos en iones nitrato, el cual es el principal aceptor de electrones en procesos de desnitrificación, así como el principal componente a partir del cual obtener N_2O por vías biogénicas (Tolar *et al.*, 2017), y un compuesto químico de gran importancia en suelos agrícolas, pues el crecimiento de plantas depende de su disponibilidad. Además, este proceso metabólico también suele ser aprovechado en plantas de tratamiento de aguas residuales con el fin de eliminar de estas la urea y el amoníaco (Kaynak *et al.*, 2008). Sin embargo, también puede tener consecuencias negativas, como la contaminación de aguas subterráneas por lixiviación. (Wrage *et al.*, 2001)

Por tanto, la nitrificación es una de las reacciones fundamentales dentro del ciclo biogeoquímico del nitrógeno, de gran relevancia a nivel medioambiental y agroalimentario, así como una de las reacciones biológicas más importantes en la obtención de energía en organismos quimiolitautótrofos, los cuales se encuentran en dos de los tres dominios en los que se organizan todos los seres vivos: el dominio más conocido y el primero en el que se suele pensar, el dominio Bacteria, y un segundo dominio, el dominio Archaea, propuesto por Woese *et al.* en 1990 y demostrado por distintos estudios moleculares realizados posteriormente (Spang *et al.*, 2015), siendo el último en incorporarse al Árbol Filogenético de la vida (*Tree of Life* o ToF), y de gran interés debido a su capacidad de crecer en cualquier ambiente, incluso en ambientes extremos (Schleper *et al.*, 2005).

El primer paso de la nitrificación, la oxidación del amonio, conlleva la liberación de protones al medio extracelular y la aportación de electrones a la cadena de transporte de electrones realizada en la membrana plasmática, lo cual contribuye a la obtención de poder reductor ($NADH^+$) y, finalmente, de ATP (Wendeborn, 2020). Este proceso es realizado principalmente por dos consorcios de microorganismos: Las bacterias amonio-oxidantes (AOB), y las arqueas amonio-oxidantes (AOA) (Tolar *et al.*, 2017). Mientras que las AOB fueron descritas en 1890 por Winogradsky en los trabajos donde se describieron por primera vez los procesos de nitrificación, no fue hasta hace algo más de una década cuando se identificaron por primera

vez las AOA (Venter et al., 2004) y comenzaron sus estudios sobre su papel en la oxidación del amonio en distintos entornos.

La oxidación del amonio es catalizada por la Amonio Monooxigenasa (AMO), una enzima integrada en la membrana cuya actividad depende de la presencia de iones de cobre (Ensign *et al.*, 1993) y, quizá, de la presencia de iones hierro no hemo (Zahn et al., 1996) (Gilch *et al.*, 2010). La enzima AMO presenta una elevada semejanza estructural con la enzima Metano Monooxigenasa (MMO), otra enzima ligada a membrana dependiente de iones cobre, lo que parece indicar su homología ancestral (Tolar et al., 2017). Estudios de cristalografía de rayos X han permitido observar que la MMO (y también la AMO) es una enzima formada por tres protómeros, cada protómero conformado por tres subunidades, A, B, y C (Chan and Yu, 2008). Así pues, la enzima AMO estaría compuesta por tres protómeros, cada uno de los cuales tendría tres subunidades: AmoA, AmoB y AmoC. (Sayavedra-Soto and Arp, 2014)

El primer gen detectado mediante análisis metagenómicos perteneciente a las AOA fue el gen que codificada para la AmoA (Venter et al., 2004), por lo que se estableció como marcador molecular para clasificar a las distintas especies de AOA (Francis et al., 2005) debido a que es un gen altamente conservado. La aparición de variabilidad entre las secuencias de una muestra significa que hay cierta diversidad de especies de AOA en la misma, además de que la opinión general es que esta es la subunidad catalítica (Hyman and Wood, 1985). Los genes *amoB* y *amoC*, codificantes para las otras dos subunidades se han obtenido inicialmente mediante el análisis de metagenomas y metatranscriptomas, pero no son usados en estudios de ecología microbiana como *amoA* (Tolar et al., 2017)

En diversas arqueas del filo Thaumarchaeota se encuentran diversas copias del gen *amoC*, sobre todo en aquellas presentes en suelo, como es el caso de *Nitrososphaera viennensis*, la cual además presenta dos copias de *amoB* en su genoma (Kerou et al., 2016). Estas múltiples copias del mismo gen dentro del genoma del propio microorganismo surgen por duplicación, y son conocidos como genes parálogos. Es decir, cada copia del gen es una secuencia paróloga con respecto al resto de copias. *amoA* y *amoC* son los genes que se utilizan como “targets” para estudios de ecología de AOA (Tolar et al., 2017).

Actualmente el gen *amoA* es el marcador genético más secuenciado, después gen del codificante para rRNA 16S, siendo el 56% de las secuencias de arqueas presentes en las bases de datos propias de este grupo taxonómico. La secuencia ancestral de este gen, presente en el

ancestro común de las AOA, ha ido sufriendo eventos de especiación en el tiempo, lo que ha hecho que cada una de estas arqueas posean al menos una copia de este gen, siendo estos genes ortólogos entre sí, pues son genes homólogos, de elevada similitud, presentes en especies distintas. Los estudios realizados a partir de este gen muestran como la abundancia y diversidad de AOA en la naturaleza dependen de muchos factores y varían según el ecosistema estudiado.

2. Objetivos

El objetivo principal de este trabajo es la realización de un análisis filogenético del *cluster* de genes codificantes para la Amonio Monooxigenasa en distintas especies de arqueas amonio-oxidantes presentes en diversos ambientes, tratando de establecer posibles relaciones filogenéticas entre ellos y verificar la idoneidad de su uso como marcadores para estudios evolutivos.

Un objetivo secundario será establecer un *pipe line* mediante código *bash* en el entorno de Linux, que deberá incluir también scripts de Perl, de modo que permita automatizar la búsqueda de nuevos genes en distintas especies y permita establecer relaciones entre ellos de paralogía y ortología con aquellos descritos previamente.

Con el objetivo de identificar posibles casos de transferencia horizontal de los genes del operón *amo* entre grupos de arqueas u organismos procedentes de otros dominios, se estudiarán otras secuencias que permitirán conocer la posición filogenética de las especies estudiadas, prestando especial atención a las arqueas del superfilo Asgard, poco conocido hasta el momento y que ha despertado un gran interés ya que podría ser el ancestro de los organismos del dominio Eukarya (Zaremba-Niedzwiedzka *et al.*, 2017).

3. Materiales y Métodos

3.1. Búsqueda de secuencias en bases de datos.

Se buscaron las secuencias correspondientes a distintos genes para extraerlos y usarlos como sondas, seleccionando como AOA de referencia a partir del cual realizar el estudio a *Nitrososphaera viennensis*, puesto que es un arquea que cuenta con 6 genes parálogos del gen *amoC*, el mayor número de copias halladas hasta el momento en el genoma de una AOA (Tabla 1). Por otra parte, se extrajeron los genomas de referencia de distintas AOA y de las

arqueas de referencia, y con estos se creó una base de datos local sobre las que lanzar las sondas mediante la realización de una búsqueda *BLAST* (*Basic Local Alignment Search Tool*) local utilizando el programa *BLAST+* (Camacho *et al.*, 2009), este método tiene la ventaja de poder localizar distintos genes en un mismo genoma, evita la mezcla de secuencias de individuos diferentes y permite conocer la posición de las secuencias coincidentes en los genomas. El proceso de elaboración de la base de datos local vendrá descrito en el apartado 3.3.

Las AOA cuyos genomas fueron empleados para la creación de la base de datos presentan la siguiente clasificación taxonómica (basada en la base de datos de taxonomía del NCBI), la cual fue utilizada como referencia a la hora de analizar los resultados:

- Filo Thaumarchaeota
 - Orden Cenarchaeales
 - Familia Cenarchaeaceae
 - Género *Cenarchaeum*
 - Orden Nitrosopumilales
 - Familia Nitrosopumilaceae
 - Género *Nitrosopumilus*
 - Género *Nitrosarchaeum*
 - Género *Nitrosomarinus*
 - Clase Nitrososphaeria
 - Orden Nitrosocaldales
 - Familia Nitrosocaldaceae
 - Género *Nitrosocaldus*
 - Orden Nitrososphaerales
 - Familia Nitrososphaeraceae
 - Género *Nitrososphaera*
 - Género *Nitrosocosmicus*
 - Incertae sedis*
 - Género *Nitrosopelagicus*
 - Género *Nitrosotalea*
 - Género *Nitrosotenus*

Junto a los genomas de estas especies, se usaron para la creación de la base de datos el genoma de la arquea de referencia del grupo Asgard (*Promethoarchaeum syntrophicum*) y el genoma de una AOB (*Nitrosomonas europaea*), además del de un organismo eucariota (*Saccharomyces cerevisiae*).

Las secuencias, tanto de las sondas como de los genomas completos, se buscaron en los servidores del NCBI (*National Center for Biotechnology Information*), centrando la búsqueda en las opciones de *Genes* y *Genomes*, donde se almacenan y organizan los accesos a múltiples secuencias obtenidas en distintos trabajos de investigación y compartidas en esta herramienta, por lo que esta base de datos es actualizada de forma frecuente con gran multitud de secuencias obtenidas día a día mediante técnicas de *Next Generation Sequencing*. Mientras que *Genome* contiene información relacionada con secuencias, mapas, cromosomas,

anotaciones y ensamblajes, *Gene* recoge información sobre secuencias de referencia (*RefSeq*, conjunto de secuencias integras, no redundantes, bien anotadas y revisadas) de distintas especies, nomenclatura, variantes, rutas metabólicas en los que participan, y links a entradas relacionadas con la función y posición en el genoma de dichos genes.

Los genes usados como sondas (Tabla 1) fueron los que codificaban para las tres subunidades de la amonio monooxigenasa, el rRNA 16S (usado sistémicamente para la clasificación de procariontes), la Acetil-CoA sintetasa y la aconitasa, estas últimas enzimas esenciales en el metabolismo energético de los procariontes, al participar en el ciclo de los ácidos tricarboxílicos.

Tabla 1: Sondas utilizadas para el estudio. Genes y especie de procedencia.

• Genes codificantes para la Amonio Monooxigenasa	<i>amoA</i>	• <i>Nitrososphaera viennensis</i>
	<i>amoB</i>	• <i>Nitrososphaera viennensis</i>
	<i>amoC</i>	• <i>Nitrososphaera viennensis</i>
• Gen para rRNA 16S		• <i>Nitrososphaera viennensis</i> . • <i>Nitrosomonas europaea</i> . • <i>Promethoarchaeum syntrophicum</i>
• Gen codificante para Acetil-CoA sintetasa		• <i>Nitrososphaera viennensis</i>
• Gen codificante para Aconitasa		• <i>Nitrososphaera viennensis</i> • <i>Promethoarchaeum syntrophicum</i> .

3.2. Programación.

Para la automatización de diferentes análisis bioinformáticos se usaron dos lenguajes de programación distintos:

- Para la preparación de la base de datos y la ejecución del *BLAST* local, se usó el lenguaje de programación *bash*, que es el que se utiliza de forma general para ejecutar órdenes o programas bajo línea de comandos desde el terminal del sistema operativo Linux (Anexo I).
- Para la adaptación del output a los siguientes pasos y la extracción de los fragmentos coincidentes en los genomas obtenidos como resultado de *BLAST*, se escribieron scripts en lenguaje de programación *Perl* (Anexo II).

3.3: Método MSA y modelo filogenético.

Las secuencias de cada uno de los genes analizados se ordenaron mediante la obtención de alineamientos múltiples (MSAs) que fueron obtenidos mediante el alineamiento progresivo que realiza el programa ClustalW (Thompson *et al.*, 1994). El funcionamiento básico de este algoritmo se basa en la comparación de todos los pares de secuencias para la realización de un primer árbol guía. Tras esto, las secuencias se agrupan, de forma individual o en grupos de pares, empezando por las más parecidas, empleando el árbol guía como referencia. Tras esta primera agrupación, se busca, empleando el árbol guía, que secuencia o grupo de secuencias tienen menor distancia con el primer grupo de partida, y se agrupan en un *cluster*. Esto se hace de forma secuencial hasta que todas las secuencias están alineadas y las distancias genéticas entre ellas.

Tras alinear las secuencias y obtener las distancias entre ellas se procedió a realizar el árbol filogenético basado en distancias (Neighbor-Joining) usando el modelo de sustitución LogDet, ya que al tratarse de especies pertenecientes a taxones muy alejados la composición de nucleótidos puede ser muy diferente, y en estos casos el modelo LogDet funciona mejor que los modelos basados en la sustitución de nucleótidos. Mientras otros modelos evolutivos suponen que todas las líneas evolutivas cambian bajo el mismo patrón de sustitución de nucleótidos, el modelo LogDet no lo hace, por lo que es mejor para medir el cambio evolutivo entre secuencias en las que los patrones de sustitución son heterogéneos entre linajes (Tamura and Kumar, 2002).

El output de este proceso es la obtención de un árbol de tipo filograma. Para comprobar la validez de la topología de la filogenia obtenida, se realizó un bootstrapping o test bootstrap, que consiste en un remuestreo aleatorio de los datos que permite generar numerosas réplicas del conjunto de datos (MSAs, distancias, etc.) para la generación de gran cantidad de réplicas de árboles filogenéticos. Cada réplica se analiza mediante el modelo empleado para obtener una topología y al final se obtiene un “árbol consenso” en el cual se muestra la frecuencia con la que aparece cada nodo en el conjunto de las réplicas, mostrando la fiabilidad de este. En este caso el bootstrapping se hizo con 500 réplicas, y en los nodos de cada árbol filogenético se muestra su frecuencia y, por tanto, su fiabilidad.

Tanto para la obtención de los alineamientos como de las filogenias, así como el tratamiento de los árboles se usó el programa MEGA7 en su versión GUI (Gráfica) de Windows (Kumar *et al.*, 2016)

4. Resultados y Discusión

El pipeline elaborado, junto con los scripts desarrollados (Anexos I, II y III), permitieron la identificación y extracción de los genes deseados de las distintas especies que fueron estudiadas en este trabajo.

En la Tabla 2 se esquematizan los genomas utilizados para cada especie, los genes hallados y el número de copias de cada gen.

Tabla 2: Número de copias identificadas de cada gen en cada genoma de las especies estudiadas.

Especie	Codigo RefSeq	16SRNA	amoA	amoB	amoC	acoA	aco
<i>Prometheoarchaeum syntrophicum</i>	NZ_CP042905	3	0	0	0	2	1
<i>Saccharomyces cerevisiae</i>	NC_001144	2	x	x	x	x	x
<i>Nitrosocaldus cavascurensis</i>	NZ_LT981265	3	1	1	1	1	1
<i>Nitrosotalea devanaterrea</i>	NZ_LN890280	3	1	1	1	1	2
<i>Nitrosopelagicus brevis</i>	NZ_CP007026	3	1	1	1	1	2
<i>Nitrosomarinus catalina</i>	NZ_CP021324	3	1	1	1	1	2
<i>Nitrosotenus cloacae</i>	NZ_CP011097	3	1	1	1	1	2
<i>Nitrosocosmicus oleophilus</i>	CP012850	9	1	1	2	1	2
<i>Nitrosocosmicus franklandus</i>	NZ_LR216287	6	1	1	2	1	2
<i>Nitrososphaera viennensis</i>	NZ_CP007536	3	1	1	6	1	2
<i>Nitrososphaera evergladensis</i>	CP007174	3	1	1	5	1	2
<i>Nitrososphaera gargensis</i>	CP002408	3	1	1	2	1	2
<i>Nitrosopumilus piranensis</i>	CP010868	3	1	1	1	1	2
<i>Nitrosopumilus adriaticus</i>	NZ_CP011070	3	1	1	1	1	2
<i>Nitrosopumilus maritimus</i>	CP000866	3	1	1	1	1	2
<i>Nitrosopumilus koreensis</i>	CP003842	3	1	1	0	1	2
<i>Cenarchaeum symbiosium</i>	DP000238	3	1	1	2	0	2
<i>Nitrosopumilus sp.</i>	NZ_CP035425	3	x	x	x	x	x

4.1 16S rRNA

Para la identificación de los genes codificantes para el rRNA 16S en todas las especies se emplearon como sondas los genes de la subunidad ribosomal pequeña de *Nitrososphaera viennensis*, *Nitrosomonas europaea* y *Prometheoarchaeum syntrophicum*. Así, se localizaron las distintas copias presentes en los genomas de la base de datos. El número de copias del gen 16S presentes en las especies van desde 2 en *Saccharomyces cerevisiae* hasta 9 en *Nitrosocosmicus oleophilus*, encontrando en la mayoría de las especies 3 copias, casi siempre situadas en tándem (esta situación puede determinarse por la posición de cada secuencia en su genoma, que queda recogida en la Figura 1). Las secuencias encontradas permitieron generar un MSA de 1775 pb. A partir de este MSA se pudo realizar un árbol filogenético de tipo *Neighbour-Joining*, que es el que se muestra en la Figura 1.

Las copias localizadas dentro del mismo genoma se parecen más entre sí, lo que muestra que pueden haber surgido por fenómenos de duplicación recientes o por una homogeneización de las secuencias dentro del genoma, aunque no estén todas situadas en tándem, como ocurre en *Nitrosocosmicus oleophilus*.

En el árbol se puede ver la agrupación de las especies en sus distintos géneros y la relación filogenética entre estos, siendo prácticamente cada género un claro grupo monofilético. Hay que destacar la presencia de la especie *Nitrosomarinus catalina* entre las arqueas del género *Nitrosopumilus*, por lo que las especies de este género serían las únicas que, en su conjunto, forman un grupo parafilético. El género *Nitrosomarinus* está compuesto exclusivamente por la especie *N. catalina*, y pertenece a la familia Nitrosopumilaceae. De hecho, en artículos recientes algunos autores reclasifican esta especie dentro del género *Nitrosopumilus* (López-Pérez *et al.*, 2019). Por tanto, la clasificación taxonómica de esta especie todavía no es clara.

Se pueden distinguir los tres dominios en los que se clasifican los seres vivos: Bacteria (*Nitrosomonas europaea*), Archaea y Eukarya (*Saccharomyces cerevisiae*). Partiendo de microorganismos del dominio Bacteria como ancestro, cuyo único miembro se ha usado como *outgroup* para poder enraizar el árbol, los microorganismos de los dominios Eukarya y Archaea formarían dos grupos hermanos. Sin embargo, el apoyo del nodo que simboliza el ancestro común de las arqueas por el Bootstrap es muy bajo (64), lo que podría indicar una

incorrecta organización topológica de este dominio, pudiendo ser el ancestro de las arqueas también ancestro de los eucariotas. Además, las especies del dominio Archaea empleadas pertenecen únicamente al grupo Thaumarchaeota, no teniendo representantes de otros grupos, además de estar empleándose para el estudio el genoma de un único organismo eucariota, por lo que se podría estar perdiendo información. Estudios recientes proponen la hipótesis de que el *Tree of Life* de tres dominios independientes que se había propuesto hasta ahora podría ser, en realidad, un árbol en el las arqueas formarían un grupo parafilético y los eucariotas serían un grupo hermano del grupo Asgard de las arqueas, por lo tanto, estarían próximos a *Prometheoarchaeum syntrophicum* en el filograma de la Figura 1. Para algunos autores como Williams *et al.* (2020) los eucariotas descienden de los microorganismos del grupo Asgard. Para poder confirmar el papel de las arqueas del grupo Asgard dentro de la evolución se necesitaría contar con un mayor número de datos de alta calidad, con los cuales no contamos hasta ahora debido a la complejidad en el estudio de estas arqueas, organismos hallados en dorsales oceánicas, difíciles de cultivar y aislar.

Un nuevo fenómeno de especiación dentro del dominio Archaea separó las arqueas del grupo Asgard, y al antecesor del resto de arqueas que se analizan en este trabajo, perteneciendo ellas a un mismo grupo monofilético fuertemente soportado por el Bootstrap (99), formando parte del phylum Thaumarchaeota.

Este árbol permite saber cuál es la posición filogenética de las arqueas AOA para los que se han analizado el resto de los genes, conociendo las especies más alejadas es posible determinar cuál es, en cada caso, la más adecuada para ser empleada como *outgroup* para realizar el enraizamiento de los árboles obtenidos con otros genes.

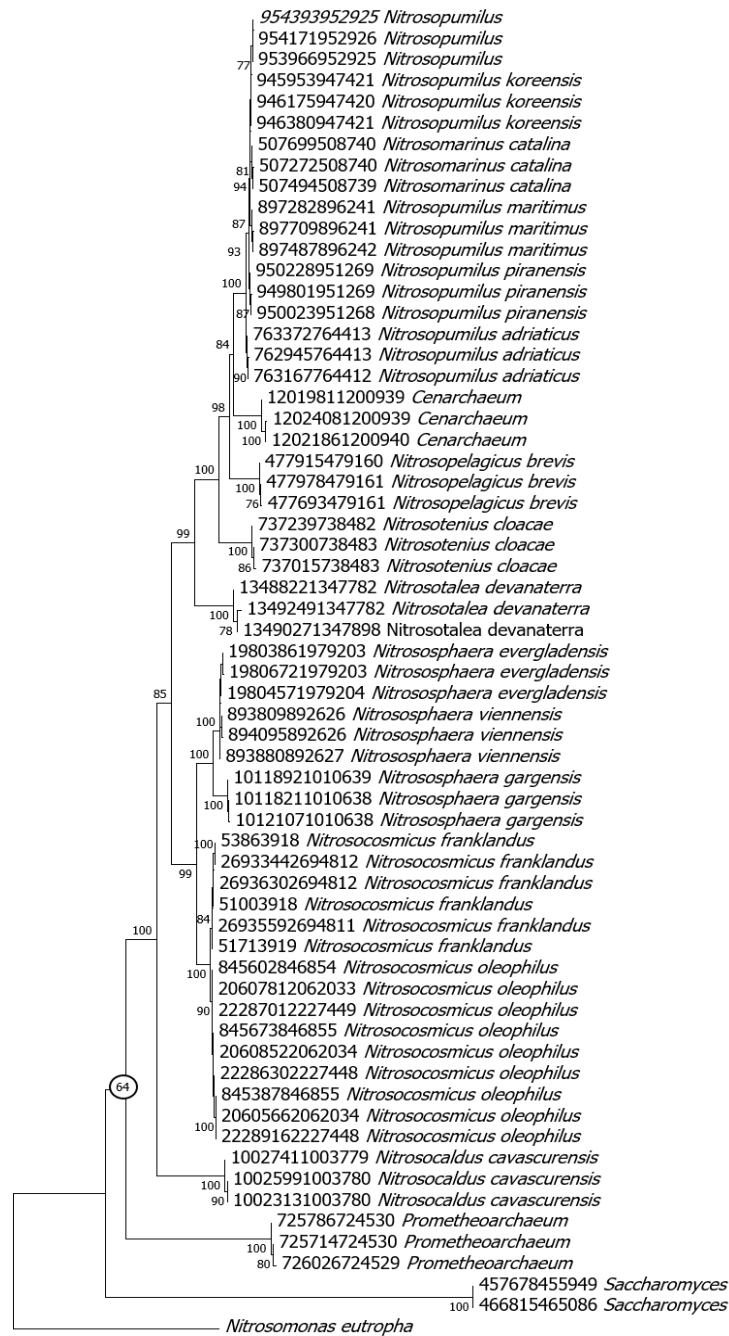


Figura 1: Árbol Filogenético *Neighbour-Joining* construido a partir de las secuencias del gen 16S rRNA de las especies estudiadas, enraizado usando como *outgroup* el gen 16S rRNA de *Nitrosomonas europaea*. Las secuencias fueron identificadas realizando un BLAST que usa como sonda los genes codificantes de *Nitrososphaera viennensis*, *Lokiarchaeum sp.* y *Prometheoarchaeum*. Los nodos contienen puntuación de bootstrap, aunque los valores inferiores al 75% han sido omitidos. El nodo que separa los tres dominios no tiene suficiente apoyo de bootstrap, pero se muestra marcado con un círculo. Los números de las secuencias indican su posición en el genoma.

4.2 amoA, gen codificante para la subunidad A de la Amonio Monooxigenasa.

Se localizaron las secuencias codificantes para la subunidad A de la enzima Amonio Monooxigenasa en las especies que se están estudiando, empleando como sonda la secuencia del gen de *Nitrososphaera viennensis*. En todos los casos, se localizó una única copia en el genoma de la especie.

Nitrosocaldus cavascurensis es el pariente más alejado filogenéticamente, y pertenece a la clase Nitrososphaeria, por lo que el grupo monofilético que conforma este *phylum* fue empleado como *outgroup* para enraizar este árbol. A partir de la única copia de este gen en cada especie se obtuvo un MSA de 666 pb, aunque al igual que en el resto de los genes codificantes para proteínas se utilizó inicialmente la secuencia traducida para obtener el alineamiento más fiable. Partiendo del MSA de nucleótidos se pudo obtener el árbol *Neighbour-Joining* que se puede ver en la Figura 2.

De la secuencia ancestral de *amoA* surgen las secuencias del gen en las arqueas de la clase Nitrososphaeria, y las secuencias ancestrales de este gen en el las arqueas del orden Nitrosopumilales y otras arqueas, todas ellas clasificadas en *incertae sedis*.

Dentro de la clase Nitrosphaeria, los tres géneros se muestran como monofiléticos, pero el orden Nitrosphaerales en el que se clasifican *Nitrosocosmicus* y *Nitrosphaera* se muestra como parafilético, ya que junto a este puede verse a la única especie del género *Nitrosocaldus*, aunque el nodo que lo une a *Nitrososphaera* tiene una baja puntuación de bootstrap (69), poniendo en duda la topología de esta región del árbol filogenético.

De nuevo se observa que *Nitrosomarinus catalina* comparte el nodo que corresponde al ancestro de todas las arqueas del género *Nitrosopumilus*, lo que siembra una duda sobre su correcta clasificación taxonómica en cuanto al género en el que está clasificada. Pero en este caso el nodo que agrupa las arqueas de estos dos géneros tiene una baja puntuación de bootstrap. Mencionar también la pertenencia al mismo grupo monofilético de las secuencias codificantes para la AmoA de *Nitrosopelagicus brevis* y *Nitrosotenus cloacae* a pesar de que dichas especies no se encuentran clasificadas en un mismo grupo taxonómico, lo que da indicios de una posible transferencia horizontal. Sin embargo, el apoyo de este fenómeno por el bootstrap es bajo (61). La posición de *Cenarchaeum symbiosium* en el árbol no coincide con la indicada por el gen del rRNA 16S (Figura 1).

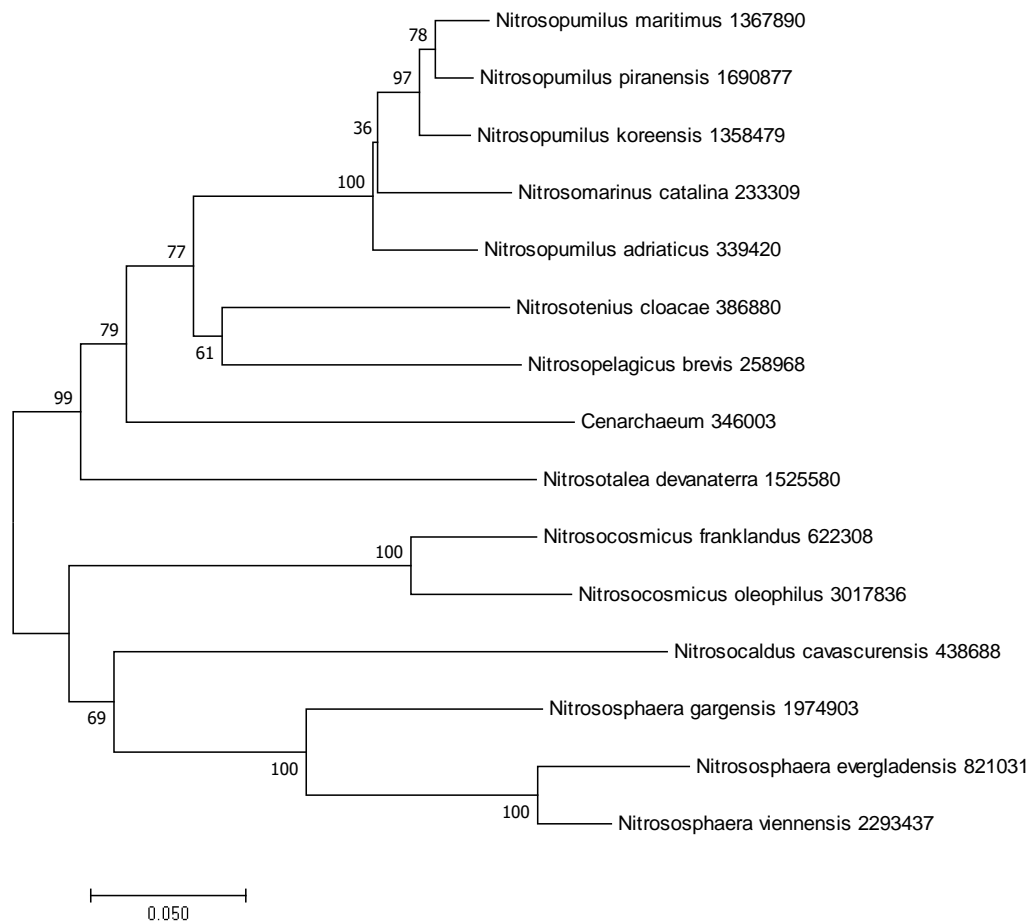


Figura 2: Árbol filogenético *Neighbour-Joining* construido a partir de las secuencias presentes en las especies estudiadas codificantes para la subunidad A de la Amonio Monooxigenasa, identificadas mediante un *BLAST* que usa el gen codificante de *Nitrososphaera viennensis* como sonda. Como *outgroup* se usó el grupo monofilético de la clase Nitrososphaeria. En los nodos se observa la puntuación del bootstrap. Los números junto a los nombres indican la posición de inicio de la secuencia.

amoA es el segundo marcador molecular más empleado en estudios filogenéticos debido a que se suele utilizar para determinar relaciones evolutivas entre arqueas (Tolar *et al.*, 2017), pues permite distinguir en una muestra ambiental las procariontas arqueas y bacterianas, estudiando sólo las primeras. Esto se debe a que, a pesar de encontrar bacterias capaces de oxidar el amonio con la proteína codificada por este gen, la gran mayoría de microorganismos poseedores de *amoA* que se hayan en las muestras pertenecen al dominio Archaea (Xu *et al.*, 2012), por lo que al amplificarlo por PCR para la posterior secuenciación, la cantidad de copias de origen bacteriano pasa desapercibido. Además, los genes de *amoA* de bacterias y arqueas poseen zonas conservadas distintas que permiten distinguirlos al emplear *primers* específicos que hibriden con esas áreas (Morimoto *et al.*, 2011). Por tanto, a la hora de

realizar estudios filogenéticos en arqueas, *amoA* se convierte en una alternativa al gen del rRNA 16S, aunque las posiciones de las especies en los árboles filogenéticos no siempre coinciden y los valores de bootstrap de las secuencias *amoA* aportan una fiabilidad muy inferior a la topología final.

En los árboles obtenidos con el análisis de los MSA del 16S y del *amoA* las arqueas de los principales géneros estudiados, *Nitrososphaera*, *Nitrosocosmicus* y *Nitrosopumilus*, quedan unidos en grupos monofiléticos (o parafiléticos en el caso de *Nitrosopumilus*), lo que apoya tanto la clasificación taxonómica en cuanto a género de estas como la posible utilidad del gen *amoA* como marcador molecular para estudios filogenéticos, si bien los resultados presentarán diferencias con respecto al empleo del rRNA 16S cuando se considera la relación entre los distintos géneros.

4.3 *amoB*, gen codificante para la subunidad B de la Amonio Monooxigenasa.

Las secuencias codificantes para la subunidad B de la Amonio Monooxigenasa de las distintas especies fueron identificadas empleando como sonda la secuencia codificante para esta subunidad presente en el genoma de *Nitrososphaera viennensis*. A partir de estas secuencias se realizó un MSA de 535 pb, el cual se utilizó para la obtención del árbol *Neighbour-Joining* que se encuentra en la Figura 3. Al igual que en *amoA*, solo aparece una única copia en los genomas de cada especie. Las copias de *amoB* están muy próximas a las de *amoA* como revela su posición en el genoma, que puede observarse en las Figuras 2 y 3. Los genes *amo* se encuentran agrupados en operones en bacterias, y todo indica que también lo están en arqueas.

De nuevo en este caso el *outgroup* que permite el enraizamiento del árbol es la clase Nitrososphaeria. Comenzando el análisis desde esta raíz se puede ver cómo, a través de un fenómeno de especiación, se obtienen dos variantes de la secuencia codificante para AmoB, lo que provoca que las secuencias de las especies que aparecen en el árbol se separen en dos grupos monofiléticos. Una de las secuencias ancestrales dio lugar, de nuevo por especiación, a las variantes presentes en *Nitrosocaldus cavascurensis* y en los géneros *Nitrosocosmicus* y *Nitrososphaera*, representando cada género un grupo monofilético independiente, ambos géneros pertenecen al orden Nitrososphaerales de las arqueas.

La otra secuencia ancestral pasó por varios fenómenos de especiación, que dieron lugar a las diferentes secuencias del gen *amoB* en Thaumarchaeotas de los restantes grupos. En este gen las arqueas del género *Nitrosopumilus* forman un grupo monofilético bien soportado por el bootstrap (90%) mientras que *Nitrosomarinus* sería un género hermano muy próximo. Esta topología se ajusta a la clasificación de los géneros a diferencia de lo que ocurría con el rRNA 16S y *amoA*.

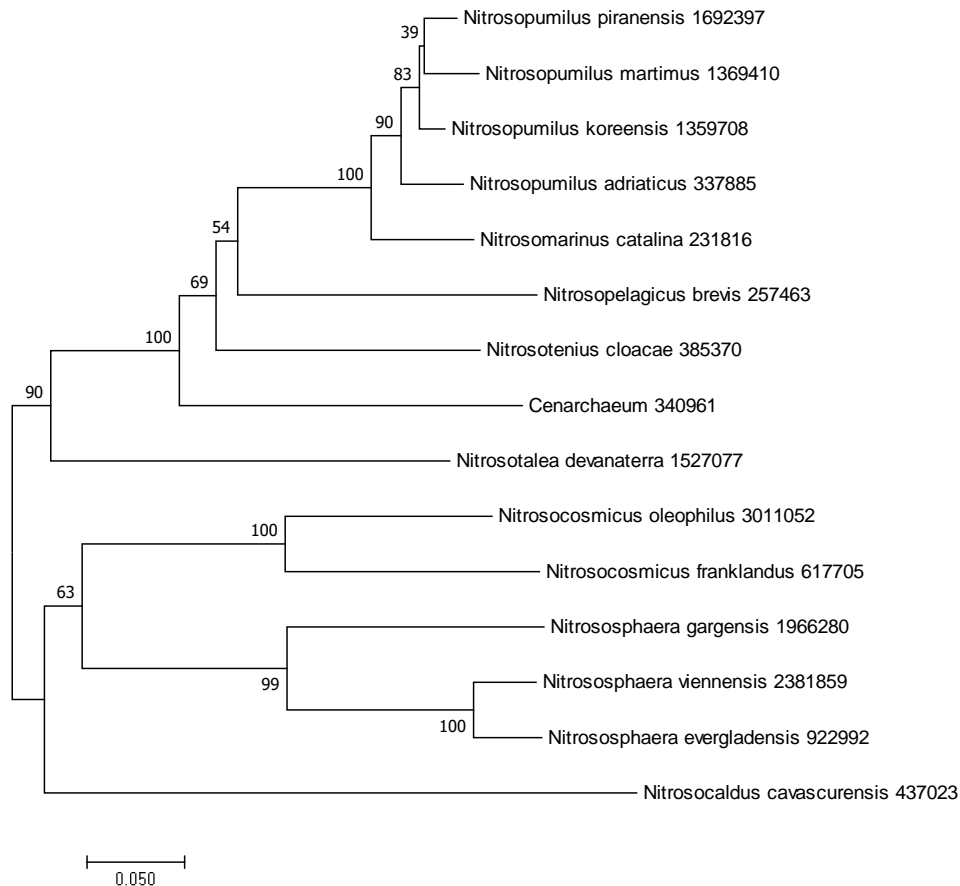


Figura 3: Árbol filogenético *Neighbour-Joining* construido a partir de las secuencias presentes en las especies estudiadas codificantes para la subunidad B de la Amonio Monooxigenasa. Para el enraizamiento se utilizaron las secuencias presentes en las arqueas de clase Nitrososphaeria. Las secuencias fueron identificadas realizando un *BLAST* que usa como sonda el codificante de *Nitrososphaera viennensis*. En los nodos se observa la puntuación del bootstrap. Los números junto a los nombres indican la posición de inicio de la secuencia.

Las secuencias *amoB* muestran distancias menores entre sí que las presentadas por *amoA* (véase la diferencia en las escalas de los árboles de las Figuras 2 y 3), este dato junto a los valores más altos en el bootstrap del árbol de *amoB* indica una evolución más lenta en este último gen. Esto podría indicar que se trata de una secuencia más conservada y, por tanto, podría tratarse de un gen codificante de gran importancia en la realización del metabolismo del nitrógeno. En AOB de los géneros *Nitrosomonas* y *Nitrospira* se ha comprobado que el gen *amoB* es el que más se expresa del operón *amo*. Mientras que se sabe que la subunidad AmoA contiene el centro catalítico de la enzima y AmoC tiene función estabilizadora, la función de AmoB permanece completamente desconocida (Zorz *et al.*, 2018), si bien la base de datos Uniprot (The UniProt Consortium, 2016) revela que el papel de esta subunidad podría relacionarse con la correcta compartimentalización de la enzima, pues este péptido contiene un dominio transmembrana que, junto con la subunidad AmoC, permitiría a la enzima situarse en la membrana plasmática. Es por esto por lo que la subunidad debe estar conservada, pues un cambio en la composición de aminoácidos podría afectar a su localización en la célula.

Al ser una secuencia muy conservada y presente en copia única, este gen puede ser usado como marcador filogenético, permitiendo obtener árboles de calidad. En este caso se puede observar como el árbol filogenético resultante presenta una elevada similitud con la clasificación taxonómica presente en el NCBI que puede observarse en el apartado 3.1. Así, se puede deducir que las especies *incertae sedis* pueden ser incluidas en ordenes propios, fuera de los taxones principales de las Thaumarchaeota descritos.

4.4 *amoC*, gen codificante para la subunidad C de la Amonio Monooxigenasa.

Se alinearon las distintas secuencias halladas en los genomas de las especies con las que se trabajó en un MSA. Dichas secuencias fueron identificadas empleando como sonda el gen codificante para la subunidad C de la Amonio Monooxigenasa en *Nitrososphaera viennensis*. El MSA tiene una longitud de 622 pb. A partir de ese MSA se pudo obtener el árbol *Neighbour-Joining* que aparece en la Figura 4.

Mientras que algunas especies, como las pertenecientes al género *Nitrosopumilus* poseen una única copia de este gen, hay otras, como las pertenecientes al género *Nitrososphaera*, que

poseen varias, siendo el máximo número 6, presentes en *Nitrososphaera viennensis*. Sin embargo, en *Nitrosopumilus koreensis* no se ha detectado ninguna copia.

De nuevo para enraizar el árbol se utilizó el grupo monofilético correspondiente a las arqueas de la clase Nitrososphaeria.

En todas las especies analizadas, incluida *Nitrosocaldus cavascurensis*, la única copia existente o una de las copias se mantiene en el genoma junto a los genes *amoA* y *amoB*, mientras que el resto de las copias aparecen en otras posiciones alejadas del operon *amo*.

Considerando las secuencias situadas en el operón los genes *amoC* muestran en el árbol una topología semejante a la de los genes *amoB*. Aunque nuevamente *Nitrosotenus cloacae*, *Nitrosotalea devanaterre*, *Nitrosopelagicus brevis* y *Cenarchaeum symbiosium* han variado sus posiciones, tal vez influenciados por la duplicación sufrida por el gen en *Cenarchaeum*.

Las dos copias de cada una de las especies del género *Nitrosocosmicus* son más parecidas entre ellas que a la especie opuesta, lo que parece indicar que las duplicaciones que las han originado en las dos especies son independientes, aunque cabe la posibilidad de que sea una duplicación antigua y que las copias de cada especie hayan sufrido algún proceso de homogenización. Lo mismo podría decirse de las dos copias encontradas en *Nitrososphaera gargensis*, se parecen más entre sí que a otras copias de las especies del mismo género, por lo que o bien la duplicación es reciente o se han homogenizado.

En *Nitrososphaera viennensis* y *Nitrososphaera evergladensis* las copias generadas por duplicaciones permiten emparejar ambas especies, cada pareja estaría formada por genes ortólogos y las distintas parejas forman grupos de genes parálogos que se habrían formado por duplicación antes de la divergencia de las dos especies. Algunas de las copias situadas fuera de los operones muestran ramas largas en el árbol, incluso alterando su topología. Posiblemente estas copias externas al operón *amo* han cambiado su función y están sometidas a una menor presión de selección, lo que ha producido que cambien más rápidamente que las copias que en principio mantienen su función original como proteína.

El conjunto de árboles filogenéticos que analizan los eventos evolutivos de las secuencias codificantes para las tres subunidades de la Amonio Monooxigenasa muestran la estrecha relación filogenética existente entre *N. viennensis* y *N. evergladensis* en lo referente al

metabolismo del Nitrógeno y que los procesos de duplicación génica que se han producido antes de su especiación.

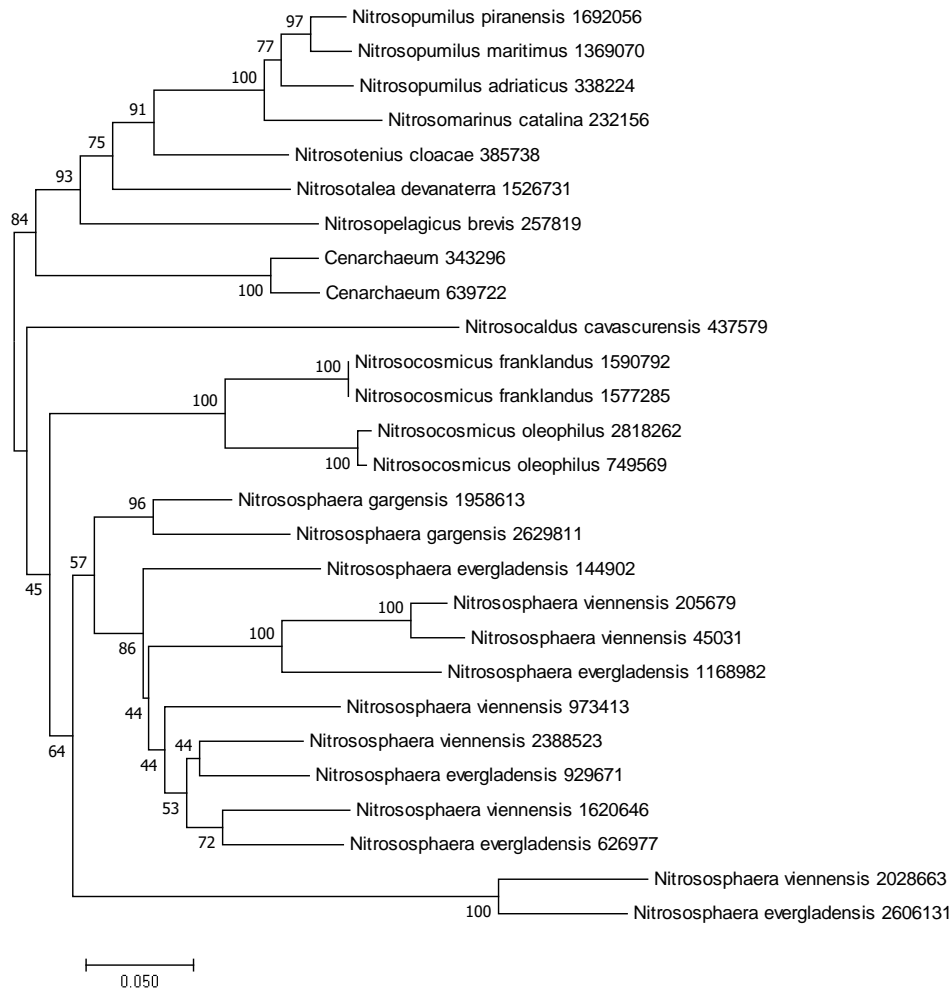


Figura 4: Árbol filogenético de tipo *Neighbour-Joining* construido a partir de las secuencias presentes en las especies estudiadas codificantes para la subunidad C de la Amonio Monooxigenasa. Como *outgroup* se utilizaron el conjunto de secuencias correspondiente al gen codificante de AmoC en arqueas de la clase Nitrososphaeria. Las secuencias fueron identificadas realizando un *BLAST* que usa como sonda el gen codificante de *Nitrososphaera viennensis*. En los nodos se observa la puntuación del bootstrap. Los números junto a los nombres indican la posición de inicio de la secuencia.

4.5 acoA, gen codificante para la Acetil-CoA sintetasa.

Se utilizó la secuencia codificante para la Acetil-CoA sintetasa de *Nitrososphaera viennensis* como sonda para identificar los genes presentes en las especies con las que se trabajó mediante una búsqueda *BLAST*. Tras identificar los genes codificantes de cada genoma, se alinearon realizando un MSA de 2023 pb. Se encontró una copia por especie, a excepción de *Cenarchaeum* que no posee ninguna y *Promethoarchaeum* que contaba con dos copias en su genoma. A partir de este MSA se realizó el árbol filogenético *Neighbour-Joining* que se muestra en la Figura 5. Como *outgroup* se utilizó la secuencia ancestral de los genes identificados en *Promethoarchaeum*.

Las copias de *acoA* en *Promethoarchaeum* proceden de una única secuencia ancestral, que actúan de raíz en este árbol. Dicha secuencia ancestral sufre una duplicación, dando lugar a las dos copias de este gen.

Las arqueas del orden Nitrososphaerales no comparten un mismo nodo ancestral y forman un grupo parafilético. Sin embargo, la puntuación de bootstrap del nodo que une la secuencia ancestral de los genes de *Nitrosocosmicus spp.* con el resto de los genes es baja (29), por lo que la posición topológica de las arqueas del género *Nitrososphaera* podría no ser correcta. La separación de las arqueas del orden Nitrososphaerales puede deberse a fenómenos de transferencia horizontal entre la secuencia ancestral de *Nitrososphaera spp* y *Nitrosocaldus cavascurensis*, si bien el nodo que relaciona las secuencias de estas especies tiene un soporte de bootstrap muy bajo (38), por lo que esta hipótesis es poco probable. Otra posibilidad es que la agrupación de *N.cavascurensis* y *Nitrososphaera spp* se deba a que ambas especies pertenecen a la misma clase. Sin embargo, el análisis filogenético del gen *16S rRNA* indicaba una gran distancia evolutiva entre estas especies.

A pesar de tratarse de no pertenecer a un mismo *phylum* conocido, *Nitrosotalea devanaterre* y *Nitrosotenus cloacae* aparecen en el mismo grupo monofilético del árbol, lo que hace sospechar de un fenómeno de transferencia horizontal del gen *acoA* entre estas especies. Sin embargo, en lo análisis filogenéticos anteriores se ve como estas especies van cambiando de posición en las distintas topologías y dichas posiciones no tiene un soporte de bootstrap alto (60), por lo que el apoyo de esta hipótesis en este trabajo es bajo.

Los genes codificantes de las arqueas del género *Nitrosopumilus* provienen de un mismo ancestro que sufre eventos de especiación, dando las distintas secuencias ancestrales a partir de las que ha evolucionado este gen en cada especie. Hay que destacar que, en este caso, las arqueas del género *Nitrosopumilus* si forman un grupo monofilético, siendo *Nitrosomarinus catalina* una especie hermana.

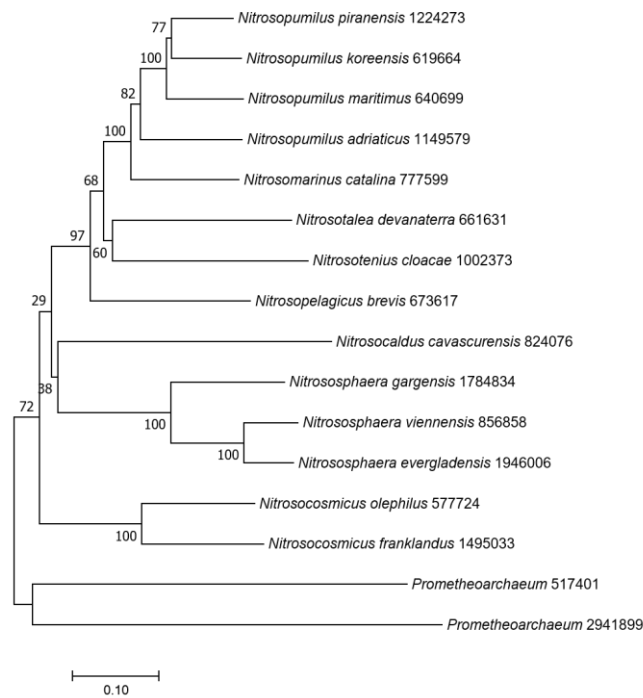


Figura 5: Árbol filogenético de tipo *Neighbour-Joining* construido a partir de las secuencias presentes en las especies estudiadas codificantes para la Acetil-CoA sintetasa identificadas mediante *BLAST*, usando como sonda el gen codificante para dicha enzima de *Nitrososphaera viennensis*. Para el enraizamiento se usó como *outgroup* la secuencia identificada en el genoma de *Prometheoarchaeum*. En los nodos se observa la puntuación del bootstrap. Los números junto a los nombres indican la posición de inicio de la secuencia.

4.6 *aco*, gen codificante para la enzima Aconitasa.

En este caso, para la identificación de los genes codificantes para la aconitasa en los distintos genomas se emplearon dos secuencias, la secuencia codificante de *Nitrososphaera viennensis* y la de *Prometheoarchaeum*. Estas permitieron identificar un total de 31 secuencias, detectándose, en la mayoría de los casos, dos copias en cada especie (a excepción de *Nitrosocaldus cavascurensis* y *Prometheoarchaeum*, con una sola copia). Sin embargo, al intentar alinear todas las copias el MSA no fue posible. Esto se debe a que las sondas empleadas no son homólogas entre sí, por tanto, tampoco las secuencias resultantes. Fue por

esto por lo que las secuencias fueron alineadas en dos MSA distintos según su homología con una u otra sonda.

El primer MSA se hizo con las secuencias homólogas al gen de *Promethoarchaeum*. Este MSA contaba con 15 secuencias, una copia por especie, y tenía una longitud de 1946 pb. El segundo MSA se obtuvo al alinear las secuencias homólogas al gen de *N.viennensis*, contiene un total de 15 secuencias y tiene una longitud de 1279 pb. Las especies que poseían dos copias para el gen de la Aconitasa, realmente poseían una secuencia homóloga al gen codificante de *Promethoarchaeum* y otra homóloga al gen de *N.viennensis*.

A partir de estos MSAs se realizaron los árboles filogenéticos *Neighbour-Joining* que se exponen en la Figura 6. En la parte alta de la figura se observa el árbol obtenido a partir del primer MSA descrito, mientras que el árbol obtenido a partir del segundo MSA se muestra en la parte de abajo.

Empezando por el árbol de arriba, se usó como *outgroup* las secuencias codificantes de *Promethoarchaeum* y *Nitrososphaera gargensis*, al ser las secuencias más alejadas filogenéticamente. Aunque *N. gargensis* es filogenéticamente más próxima al resto de arqueas, su secuencia codificante diverge mucho de estas, lo que se puede comprobar con la longitud de la rama (indica la distancia evolutiva), que muestra que la secuencia ha sufrido diversos cambios, algunos de ellos recientes. Dicha secuencia parece estar más emparentada filogenéticamente con la de *Promethoarchaeum*, lo que podría estar indicando un posible fenómeno de transferencia horizontal de las secuencias y una posterior acumulación de mutaciones. La raíz la compone, por tanto, la secuencia ancestral de los genes codificantes de estas dos especies.

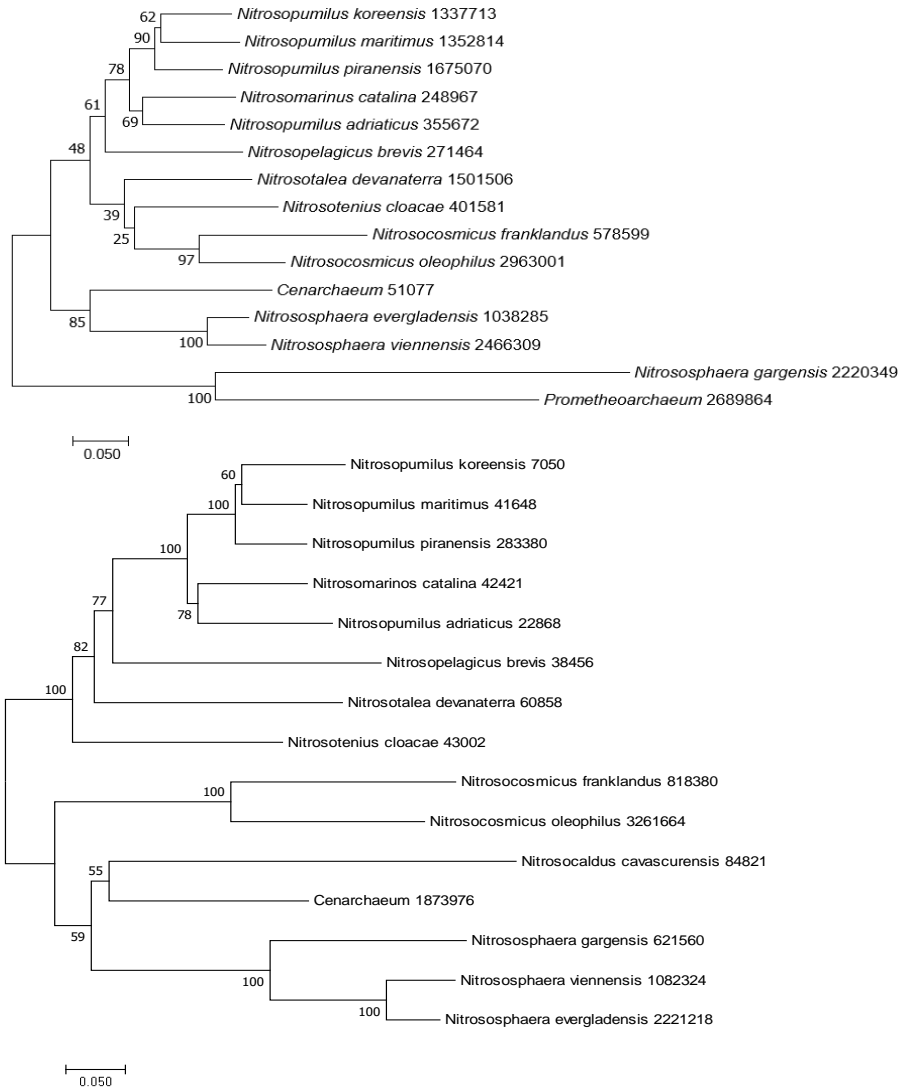


Figura 6: Árboles filogenéticos de tipo *Neighbour-Joining* contruidos a partir de las secuencias codificantes para la Aconitasa presentes en las especies estudiadas. Arriba: Secuencias homólogas a la codificante para la enzima Aconitasa en *Prometheoarchaeum*, usando la secuencia codificante de esta especie como *outgroup*. Abajo: Secuencias homólogas a la codificante para la Aconitasa en *Nitrososphaera viennensis*, usando como *outgroup* la clase *Nitrososphaeria*. Secuencias identificadas mediante la realización de un *BLAST*. En los nodos se observa la puntuación de bootstrap. Los números junto a los nombres indican la posición de inicio de la secuencia.

Esta secuencia ancestral sufre un fenómeno de especiación que da lugar, por un lado, a la secuencia codificante de *Nitrososphaera gargensis*, y por el otro, a la secuencia ancestral del resto de arqueas, la cual sufre un nuevo fenómeno de especiación, apareciendo la secuencia ancestral del resto de arqueas del género *Nitrososphaera* y de *Cenarchaeum symbiosium*. Esta especie forma un grupo monofilético con algunas especies del género *Nitrososphaera*, que a

su vez forman un grupo polifilético con las arqueas del género *Nitrosocosmicus*, a pesar de pertenecer al mismo orden. Esto podría indicar una posible transferencia horizontal del gen *aco* de *C. symbiosium* al ancestro de *N. viennensis* y *N. evergladensis*.

El último fenómeno de especiación mencionado provoca da lugar también a una secuencia ancestral que, por otro fenómeno de especiación, da por una parte la secuencia ancestro de las secuencias codificantes de las arqueas del género *Nitrosocosmicus*. Sin embargo, la puntuación de bootstrap del nodo que da lugar a su secuencia ancestral es bastante baja (25), por lo que la relación filogenética de estas secuencias con respecto al resto no es clara. Tampoco está soportada por el bootstrap la posición topológica de *Nitrosotalea devanaterre* (39).

La otra secuencia producto de la especiación da a la secuencia ancestral de los genes de *Nitrosopelagicus brevis*, *Nitrosomarinus catalina* y el género *Nitrosopumilus*. Sin embargo, esta divergencia está poco apoyada por el bootstrap (48).

En esta ocasión, de nuevo el género *Nitrosopumilus* forma un grupo parafilético, con *Nitrosomarinus catalina* compartiendo ancestro ya no sólo lejano con todas las arqueas *Nitrosopumilus*, sino una última secuencia ancestral con *Nitrosopumilus adriaticus*, formando estas dos un grupo monofilético. En este caso, la estrecha relación entre *N.catalina* y *N.adriaticus* podría deberse a un fenómeno de transferencia horizontal, probablemente de *N.adriaticus* a *N.catalina*, lo que podría justificar su posición en el árbol.

Con relación al árbol filogenético de abajo, como *outgroup* se volvió a emplear las secuencias de las arqueas de clase Nitrososphaeria.

En este árbol se puede ver como *Cenarchaeum symbiosium* aparece emparejada junto a *Nitrosocaldus cavascurensis* y dentro del grupo de las arqueas de clase Nitrososphaeria, a pesar de no pertenecer a un mismo grupo taxonómico. Esto indica una transferencia horizontal del gen *aco* entre estas especies.

Las arqueas del orden Nitrososphaerales no están posicionadas formando un grupo monofilético. De nuevo se puede ver como el nodo del cual emerge la rama correspondiente a la secuencia ancestral de las arqueas del género *Nitrososphaera* tiene una baja puntuación de bootstrap (59), por lo que se puede deducir que su posición topológica en este árbol no es

adecuada. Sin embargo, a nivel género, todas las arqueas están posicionadas en el árbol de forma adecuada, con un buen respaldo del bootstrap.

Comentar también de nuevo la posición de *Nitrosomarinus catalina*, de nuevo dentro del grupo parafilético de las arqueas *Nitrosopumilus spp.*, aunque con un apoyo del bootstrap ligeramente bajo (78), lo que nos podría hacer dudar sobre una posible transferencia horizontal.

Las secuencias de las arqueas del género *Nitrosopumilus* forman un grupo parafilético entre sí, puesto que las secuencias de *Nitrosopumilus adriaticus* y *Nitrosomarinus catalina* están emparejadas. Esto, de nuevo, podría deberse a un fenómeno de transferencia horizontal entre especies.

La topología de este árbol parece ser similar a la del árbol de *amoA*, aunque en este caso las puntuaciones de bootstrap respaldan más la topología. Este hecho, junto con los demás hallazgos en otros árboles filogenéticos, indica que la clasificación taxonómica de *Nitrosomarinus catalina* fuera del género *Nitrosopumilus* es incorrecta.

Se analizó la identidad de las dos secuencias homólogas empleadas como sondas mediante la realización de una búsqueda *BLAST*, y se comprobó que ambas sondas tenían dos orígenes distintos. Así, el gen de *Promethoarchaeum* presenta homología con secuencias del género *Clostridium* e *Hydrogenobaculum*, lo que muestra un origen bacteriano muy antiguo transmitido horizontalmente. Sin embargo, el gen presente en el genoma de *Nitrososphaera viennensis* tiene como origen una secuencia ancestral de una Thaumarchaeota. Esto plantea la posibilidad de que, mientras que el gen de *N. viennensis* procede de evolución y transferencia vertical, *Promethoarchaeum* adquirió la copia empleada como sonda, y aparentemente la única presente en el genoma, mediante transferencia horizontal desde el dominio Bacteria.

5. Conclusiones

El pipeline y los scripts desarrollados funcionaron adecuadamente y permitieron la automatización de parte del proceso de búsqueda de los genes de interés, facilitando el estudio realizado en este trabajo.

La hipótesis de que las arqueas del grupo Asgard, representado por *Prometheoarchaeum*, fuesen el ancestro de los organismos eucariotas y el nexo entre el dominio Archaea y Eukarya, y que por tanto las arqueas constituyen un grupo parafilético junto a las eucariotas, no puede confirmarse ni rechazarse por los análisis realizados en este trabajo.

amoA podría ser una alternativa eficiente a la hora de realizar un estudio filogenético que permitiera analizar la diversidad de arqueas dentro de una muestra ambiental, así como realizar las clasificaciones taxonómicas. Sin embargo, los resultados filogenéticos obtenidos no son idénticos a los obtenidos a partir de las secuencias de rRNA 16S.

amoB es el único gen del operón que no se usa para clasificación taxonómica, aunque se le puede considerar un gen idóneo para la clasificación de las AOA, siendo posible determinar la posición filogenética de los géneros *incerta sedis* según el NCBI, viendo como estos pueden ser clasificados en grupos diferentes a los existentes.

El gen *amoC* es el gen que más fenómenos de duplicación y especiación sufre a lo largo de la evolución. En especies en las que se han producido las duplicaciones se generan nuevas copias, que ocupan posiciones en el genoma alejados del operón. Estas copias alejadas han sufrido mayor número de modificaciones en el tiempo, lo que convierte este gen en un marcador filogenético inadecuado. El caso más evidente es el de las múltiples copias de *Nitrososphaera viennensis* y *N. evergladensis*.

La transferencia horizontal tiene un importante papel en la evolución de las arqueas, que sufren este fenómeno con mucha frecuencia. Así, la transferencia horizontal puede relacionar filogenéticamente especies que taxonómicamente no están clasificadas en un mismo taxón, permitiendo así explicar las diferencias entre la clasificación taxonómica y la filogenia asociada a estos genes.

Bibliografía

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009) "BLAST+: Architecture and applications", *BMC bioinformatics*. BMC Bioinformatics, 10.
- Chan, S. I. and Yu, S. S. F. (2008) "Controlled oxidation of hydrocarbons by the membrane-bound methane monooxygenase: The case for a tricopper cluster", *Accounts of chemical research*, 41(8), pp. 969–979.
- Ensign, S. A., Hyman, M. R. and Arp, D. J. (1993) "In vitro activation of ammonia monooxygenase from *Nitrosomonas europaea* by copper", *Journal of bacteriology*, 175(7), pp. 1971–1980.
- Francis, C. A., Roberts, K. J., Beman, J. M., Santoro, A. E. and Oakley, B. B. (2005) "Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean", *Proceedings of the National Academy of Sciences of the United States of America*, 102(41), pp. 14683–14688.
- Gilch, S., Meyer, O. and Schmidt, I. (2010) "Electron paramagnetic studies of the copper and iron containing soluble ammonia monooxygenase from *Nitrosomonas europaea*", *BioMetals*, 23(4), pp. 613–622.
- Hyman, M. R. and Wood, P. M. (1985) "Suicidal inactivation and labelling of ammonia mono-oxygenase by acetylene.", *The biochemical journal*, 227(3), pp. 719–725.
- Kaynak, B., Hu, Y., Martin, R. V., Russell, A. G., Choi, Y. and Wang, Y. (2008) "The effect of lightning NO_x production on surface ozone in the continental United States", *Atmospheric chemistry and physics*. European Geosciences Union, 8(17), pp. 5151–5159.
- Kerou, M., Offre, P., Valledor, L., Abby, S. S., Melcher, M., Nagler, M., Weckwerth, W. and Schleper, C. (2016) "Proteomics and comparative genomics of *Nitrososphaera viennensis* reveal the core genome and adaptations of archaeal ammonia oxidizers", *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(49), pp. E7937–E7946.

- Kumar, S., Stecher, G. and Tamura, K. (2016) "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets", *Molecular biology and evolution*, 33(7), pp. 1870–1874.
- López-Pérez, M., Haro-Moreno, J. M., de la Torre, J. R. and Rodríguez-Valera, F. (2019) "Novel *Caudovirales* associated with Marine Group I Thaumarchaeota assembled from metagenomes", *Environmental microbiology*. Blackwell Publishing Ltd, 21(6), pp. 1980–1988.
- Morimoto, S., Hayatsu, M., Takada Hoshino, Y., Nagaoka, K., Yamazaki, M., Karasawa, T., Takenaka, M. and Akiyama, H. (2011) "Quantitative Analyses of Ammonia-oxidizing Archaea (AOA) and Ammonia-oxidizing Bacteria (AOB) in Fields with Different Soil Types", *Microbes and environments*. Japanese Society of Microbial Ecology / Japanese Society of Soil Microbiology / Taiwan Society of Microbial Ecology / Japanese Society of Plant Microbe Interactions / Japanese Society for Extremophiles, 26(3), pp. 248–253.
- Sayavedra-Soto, L. A. and Arp, D. J. (2014) "Ammonia-Oxidizing Bacteria: Their Biochemistry and Molecular Biology", *Nitrification*. American Society of Microbiology, pp. 11–37. doi:10.1128/9781555817145.ch2.
- Schleper, C., Jurgens, G. and Jonscheit, M. (2005) "Genomic studies of uncultivated archaea", *Nature reviews microbiology*. Nature Publishing Group, pp. 479–488. doi:10.1038/nrmicro1159.
- Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., Van Eijk, R., Schleper, C., Guy, L. and Ettema, T. J. G. (2015) "Complex archaea that bridge the gap between prokaryotes and eukaryotes", *Nature*. Nature Publishing Group, 521(7551), pp. 173–179.
- Tamura, K. and Kumar, S. (2002) "Evolutionary Distance Estimation Under Heterogeneous Substitution Pattern Among Lineages", *Molecular biology and evolution*, 19(10), pp. 1727–1736.
- The UniProt Consortium (2016) "UniProt: the universal protein knowledgebase", *Nucleic acids research*, 45(D1), pp. D158–D169.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice", *Nucleic acids research*, 22(22), pp. 4673–4680.

Tolar, B. B., Herrmann, J., Bargar, J. R., van den Bedem, H., Wakatsuki, S. and Francis, C. A. (2017) "Integrated structural biology and molecular ecology of N-cycling enzymes from ammonia-oxidizing archaea", *Environmental microbiology reports*, 9(5), pp. 484–491.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H. and Smith, H. O. (2004) "Environmental Genome Shotgun Sequencing of the Sargasso Sea", *Science*, 304(5667), pp. 66–74.

Wendeborn, S. (2020) "The Chemistry, Biology, and Modulation of Ammonium Nitrification in Soil", *Angewandte chemie international edition*. Wiley-VCH Verlag, 59(6), pp. 2182–2202.

Williams, T. A., Cox, C. J., Foster, P. G., Szöllősi, G. J. and Embley, T. M. (2020) "Phylogenomics provides robust support for a two-domains tree of life", *Nature ecology and evolution*. Nature Research, 4(1), pp. 138–147.

Wrage, N., Velthof, G. L., Van Beusichem, M. L. and Oenema, O. (2001) "Role of nitrifier denitrification in the production of nitrous oxide", *Soil biology and biochemistry*. Elsevier Ltd, 33(12–13), pp. 1723–1732.

Xu, M., Schnorr, J., Keibler, B. and Simon, H. M. (2012) "Comparative analysis of 16S rRNA and amoA Genes from archaea selected with organic and inorganic amendments in enrichment culture", *Applied and environmental microbiology*. American Society for Microbiology (ASM), 78(7), pp. 2137–2146.

Zahn, J. A., Arciero, D. M., Hooper, A. B. and DiSpirito, A. A. (1996) "Evidence for an iron center in the ammonia monooxygenase from *Nitrosomonas europaea*.", *FEBS letters*, 397(1), pp. 35–8.

Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, Di., Juzokaite, L.,

Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., Stott, M. B., Nunoura, T., Banfield, J. F., Schramm, A., Baker, B. J., Spang, A. and Ettema, T. J. G. (2017) "Asgard archaea illuminate the origin of eukaryotic cellular complexity", *Nature*. Nature Publishing Group, 541(7637), pp. 353–358.

Zorz, J. K., Kozlowski, J. A., Stein, L. Y., Strous, M. and Kleiner, M. (2018) "Comparative proteomics of three species of ammonia-oxidizing bacteria", *Frontiers in microbiology*. Frontiers Media S.A., 9(MAY).

Anexos

Anexo I: Pipeline de bash

PipeLine empleado para el estudio evolutivo de familias génicas empleando código BASH.

1° paso: Elaboracion de base de datos *BLAST* local a partir de los genomas obtenidos:

```
makeblastdb -in
genomas/__.fasta -dbtype
nucl -out
blast_databases/__db
```

2° paso: Concatenacion de sondas a utilizar en un documento unico:

```
cat sondas/*.fasta
>mis_sondas.fasta
```

3° paso: Realizacion de *BLASTN* local y concatenación de resultados

```
blastn -db
blast_databases/__db -query
sondas/mis_16S.fasta -out
blast_results/blast__db.txt -
outfmt "6 qseqid sseqid sstart
send sseq"
tblastx -db
blast_databases/__db -query
sondas/mis_sondas.fasta -out
blast_results/__.txt -outfmt
"6 qseqid qstart qend length
sseqid sstart send"
Cat blast_results/*.txt >
blast_results/resultados.fasta
```

4° Separación según sonda, eliminación del qseqid:

```
perl perl/separacion.pl
```

Anexo II: Script perl separación

```
#!/usr/bin/perl
open (RESULTADOS, "<", "blast_results/resultados.fasta");
open (GENERAL16S, ">", "msas/general16S.fasta");
open (ACoA, ">", "msas/AcoA.fasta");
open (Acon, ">", "msas/Acon.fasta");
open (AmoA, ">", "msas/AmoA.fasta");
open (AmoB, ">", "msas/AmoB.fasta");
open (AmoC, ">", "msas/AmoC.fasta");
```



```

while ($RESULTADOS= <RESULTADOS>) {

    if ($RESULTADOS =~ /_16S/) {
        &eliminacion; print GENERAL16S "$RESULTADOS"
    }
    if ($RESULTADOS =~ /_18S/) {
        &eliminacion; print GENERAL18S "$RESULTADOS"
    }
    if ($RESULTADOS =~ /Nvi_amoC/) {
        &eliminacion; print AmoC "$RESULTADOS";
    }
    if ($RESULTADOS =~ /Nvi_amoA/) {
        &eliminacion; print AmoA "$RESULTADOS";
    }
    if ($RESULTADOS =~ /Nvi_amoB/) {
        &eliminacion; print AmoB "$RESULTADOS";
    }
    if ($RESULTADOS =~ /aconase/){
        &eliminacion; print Acon "$RESULTADOS";
    }
    if ($RESULTADOS =~ /ACoA/) {
        &eliminacion; print ACoA "$RESULTADOS";
    }
};

close(RESULTADOS);
close(GENERAL16S);
close(Acon);
close (ACoA);
close(AmoA);
close(AmoB);
close(AmoC);

sub eliminacion {
    $RESULTADOS =~ s/Nvienn_16S//;
    $RESULTADOS =~ s/loki_16S//;
    $RESULTADOS =~ s/Promarch_16S//;
    $RESULTADOS =~ s/Nmonas_16S//;
    $RESULTADOS =~ s/Scer_18S//;
    $RESULTADOS =~ s/Nvi_aconase//;
    $RESULTADOS =~ s/Parch_aconase//;
    $RESULTADOS =~ s/Nvi_ACoA_s//;
    $RESULTADOS =~ s/Nvi_amoA//;
    $RESULTADOS =~ s/Nvi_amoB//;
    $RESULTADOS =~ s/Nvi_amoC1//;
    $RESULTADOS =~ s/Nvi_amoC2//;
    $RESULTADOS =~ s/Nvi_amoC3//;
    $RESULTADOS =~ s/Nvi_amoC4//;
    $RESULTADOS =~ s/Nvi_amoC5//;
    $RESULTADOS =~ s/Nvi_amoC6//;
return $RESULTADOS;
}

```

Anexo III: Script perl extracción

```
#!/usr/bin/perl
$especie=;

$via='C:\Users\Sarit\Genomas\.fasta';
open (INFILE, '<', $via) or die;
$sec="";
while ($a = <INFILE>) {chomp($a); $sec .= $a;}
$ini=; $fin=; $len=$fin-$ini+1; $cadena= "";
$gen=substr($sec,$ini,$len);
if ($cadena eq "menos") {
    $gen= reverse($gen);
    $gen =~ tr/ATGCNatgcn/TACGNTacgn/;
    print ">$especie $fin\n$gen\n\n";}
else {print ">$especie $ini\n$gen\n\n"}

close(INFILE);
```

Agradecimientos y dedicatorias.

A los profesores de la Universidad de León y la Universidad Autónoma de Barcelona que se han esforzado en enriquecerme a mí y otros de forma personal y profesional.

Al equipo de investigación de Microbiología del IBBTEC dirigido por Juan M. García-Lobo, y en especial a Candela, por descubrirme la Microbiología Ambiental e inspirar la temática de este trabajo.

A Luis, por tantísima paciencia, ayuda, y comprensión, por confiar en mí y exigirme ser mejor cuando tocaba. Mi correo siempre estará abierto para recibir más separatas.

A mi familia, por poner los medios, la intención, el cariño, la confianza y los valores de esfuerzo y superación sin los que no estaría aquí. Gracias por recordarme siempre de qué era capaz.

A mis amigos de León y Cantabria. Por haberme apoyado tanto a nivel personal como emocional. Por tantísimos recuerdos memorables, y haber sido mi familia adoptiva fuera de casa. Gracias por haber hecho de la experiencia universitaria una maravilla.

A Andreu, que me cogiste de la mano al inicio de esta aventura y no la has soltado ni cuando tuviste que tirar por los dos. Por ser “casa” durante 4 años. Ojalá podamos compartir también el convertirnos en nuestros sueños y cumplir nuestras aspiraciones profesionales y personales.

Fdo.: Sara Domínguez Llata

