



universidad  
de león



**FACULTAD DE CIENCIAS BIOLÓGICAS Y AMBIENTALES**

**UTILIZACIÓN DE HERRAMIENTAS  
BIOINFORMÁTICAS PARA EL ANÁLISIS  
EVOLUTIVO DE LA FAMILIA GÉNICA *PMOCAB*  
UTILIZATION OF BIOINFORMATIC TOOLS FOR  
EVOLUTIONARY ANALYSIS OF GENE FAMILY  
*PMOCAB***

**Autor:** Eloy Gómez Ayerbe

Tutor: Luis E. Sáenz de Miera y Carnicer

**GRADO EN BIOTECNOLOGÍA**

**Julio, 2023**

## Índice

Resumen .....	1
Abstract .....	1
1. Introducción .....	2
1.1 Las enzimas monooxigenasas.....	2
1.2 Términos y conceptos importantes .....	4
1.3 Aplicaciones de las monooxigenasas .....	6
1.4 Objetivo .....	6
2. Materiales y métodos .....	7
2.1 Materiales .....	7
2.1.1 Bases de datos .....	7
2.1.2 Programas de búsquedas .....	7
2.1.3 Los genomas y secuencias de las especies con pMMO .....	7
2.1.4 Alineamientos y árboles filogenéticos .....	8
2.2 Métodos .....	8
2.2.1 Búsqueda de secuencias .....	8
2.2.2 Alineamientos Múltiples de Secuencias .....	9
2.2.3 Árboles .....	10
3. Resultados .....	11
3.1. Localización de genes pmo en los genomas .....	11
3.2 ARN ribosomal 16S.....	13
3.3 <i>pmoCAB</i> .....	15
3.4 <i>pmoC</i> .....	17
3.5 <i>pmoA</i> y <i>pmoB</i> .....	19
4. Discusión.....	22
4.1 Genes que acumulan un gran número de mutaciones.....	22
4.2. Semejanza entre genes duplicados de la misma especie .....	23
4.3. Transferencia horizontal en los genes <i>pmoCAB</i> .....	26
4.4 Conclusión .....	27
Referencias .....	28

## **Resumen**

En este trabajo se ha realizado un análisis evolutivo de los genes que codifican para la enzima metano-monooxigenasa presente en bacterias de diversos géneros y especies gracias al empleo de herramientas bioinformáticas y búsquedas en base de datos.

Se ha empleado la base de datos del NCBI para buscar las secuencias de los genes *pmoCAB*, los responsables de codificar las metano-monooxigenasas. Los datos obtenidos se han empleado para generar alineamientos de secuencias y posteriormente filogramas que son representaciones gráficas en forma de árboles que permiten realizar una comparación evolutiva entre las distintas metano-monooxigenasas y el ARN 16S ribosomal que permite conocer la relación filogenética de las especies. Con la comparación de los árboles resultantes se ha podido conocer las distintas relaciones de ortología y paralogía dadas entre los genes codificantes de las proteínas analizadas o las duplicaciones existentes durante el proceso evolutivo. Otra información que podemos extraer del análisis nos permite conocer los puntos donde es posible que se diera una transferencia horizontal y entre qué grupos taxonómicos

## **Abstract**

In this study, an evolutionary analysis of genes encoding the methane monooxygenase enzyme in bacteria from various genera and species was conducted using bioinformatic tools and database searches.

The NCBI database was used to search for the sequences of *pmoCAB* genes, which encode methane monooxygenases. The obtained data were used to generate sequence alignments and subsequently phylogenetic trees, which are graphical representations in the form of trees that allow for evolutionary comparisons among different methane monooxygenases and the 16S ribosomal RNA, providing insight into the phylogenetic relationship of the species. By comparing the resulting trees, it has been possible to determine the different orthology and paralogy relationships among the coding genes of the analyzed proteins or the existing duplications during the evolutionary process. Another piece of information that can be extracted from the analysis allows us to identify the points where horizontal transfer could have occurred and between which taxonomic groups.

**Palabras clave:** Alinamientos, bioinformática, evolución, filogenias, metano-monooxigenasas.

**Keywords:** Alignments, bioinformatics, evolution, methane-monooxygenases, phylogenies.

## 1. Introducción

En este trabajo de fin de grado se presenta un análisis evolutivo de las metano-monooxigenasas, unas enzimas con capacidad para metabolizar metano. Estas enzimas son muy importantes en el proceso de oxidación de metano a metanol en organismos aerobios metanotróficos.

### 1.1 Las enzimas monooxigenasas

Las enzimas monooxigenasas (MMO) en general son consideradas como unas de las enzimas oxidantes más poderosas debido a su amplia especificidad catalítica, que se extiende a la oxigenación de muchos tipos de compuestos, incluyendo monóxido de carbono, varios alcanos, alquenos, éteres, metanos halogenados y compuestos cíclicos y aromáticos (Khider *et al.*, 2021).

Las bacterias metanotróficas representan una vía posible para la utilización del metano y la mitigación de las emisiones de metano. En el primer paso de su ruta metabólica, los metanotrofos aerobios utilizan las monooxigenasas de metano (sMMO) para activar el metano, oxidándolo a metanol. Hay dos formas de sMMO: una forma de la enzima permanece unida a la membrana (pMMO) mientras que la otra es una enzima soluble citoplásmica (sMMO). Las dos sMMO no están relacionadas entre sí, ambas poseen diferentes arquitecturas, cofactores metálicos y mecanismos (Koo y Rosenzweig, 2021).

Aunque las sMMO tienen una gama de sustratos más amplia y una mayor eficiencia catalítica que pMMO, los estudios han indicado que pMMO puede ser más adaptable para la implementación de la biorremediación, debido a la sensibilidad de sMMO al cobre en el medio ambiente. Por lo tanto, la elección de la enzima MMO a utilizar en la biorremediación dependerá de las condiciones específicas del medio (Koo y Rosenzweig, 2021).

De los dos tipos de metano-monooxigenasas mencionados, este trabajo analiza las sintetizadas por tres genes agrupados en el operón *pmoCAB* que son las pertenecientes al grupo de las pMMO.

Este operón se encuentra compuesto por tres genes, *pmoC*, *pmoA* y *pmoB*. Los dos primeros genes dan lugar a dos proteínas hidrofóbicas que poseen un total de seis regiones transmembrana. El último de los genes, *pmoB* da lugar a una proteína que parece insertarse en la membrana a través de solo dos hélices (Figura 1). Según Gilbert *et al.*, (2000) el clúster de

genes completo funciona como un operón que se transcribe entero a partir de un promotor situado aguas arriba del comienzo del gen *pmoC*.

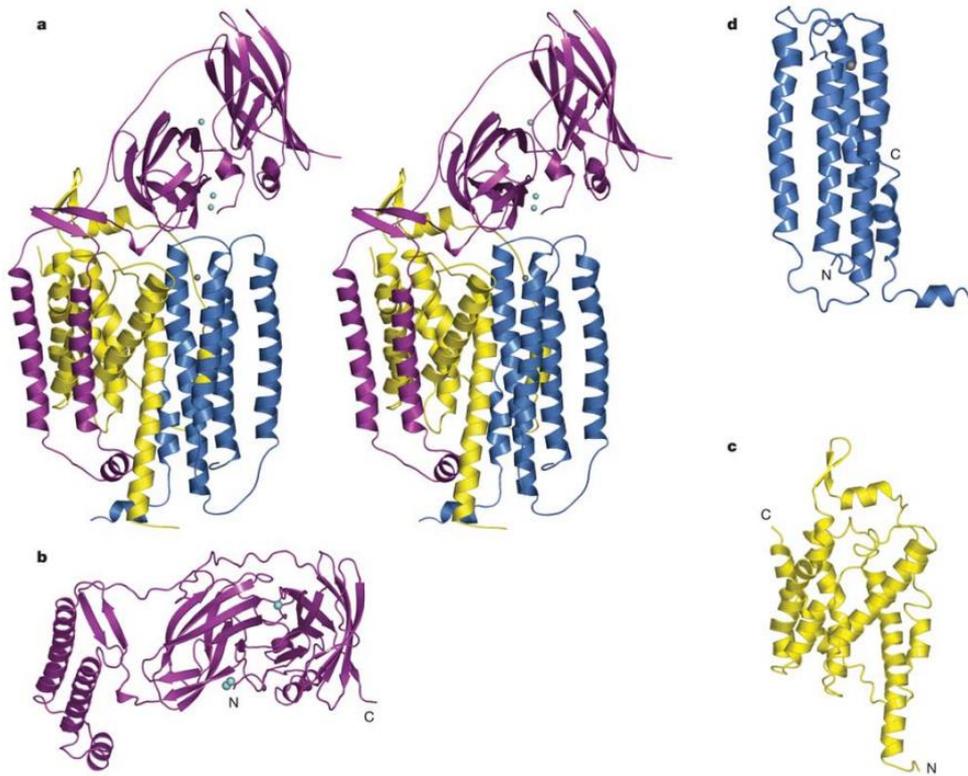


Figura 1: Representación de las subunidades de la pMMO. **a.** Vista de las tres subunidades, en magenta la subunidad B en azul la A y en amarillo la C, las esferas azules representan el cobre y las grises el zinc. **b.** Representación de la subunidad B. **c.** Representación de la subunidad A. **d.** Representación de la subunidad C. Figura extraída de (Lieberman y Rosenzweig, 2005).

Actualmente no se conoce con exactitud la localización del centro activo de la enzima, existen algunos sitios candidatos como el punto en el que se localizan dos átomos de cobre juntos, en la subunidad B, es candidato debido a su similitud con centros activos de otras proteínas, pero la distancia entre los átomos de cobre es mayor que entre los conocidos por lo que podría no serlo. Otra razón para pensar que ese sitio es el centro activo es que cerca de él existe una cavidad con residuos hidrofóbicos conservados que ayudarían en la unión del metano. Otra opción sería que ese sitio funcionase en un proceso de transferencia de electrones debido a que la *pmoB* comparte parecidos estructurales con la subunidad II del citocromo c (Lieberman y Rosenzweig, 2005).

Tampoco se puede descartar la posibilidad de que el centro activo se encuentre en la zona del átomo de zinc y estuviese conformado por residuos de las tres subunidades (Lieberman y Rosenzweig, 2005).

Algunas de las especies que poseen estos genes poseen duplicaciones del clúster completo o del gen *pmoC*. En *Methylocystis sp. SC2* se ha demostrado que existen diferencias entre las copias de su operón basadas en la afinidad de la proteína codificada por el metano (Awala, *et al.*, 2021). Las bacterias del filo Verrucomicrobiota poseen varias copias del clúster completo, las del género *Methylacidiphilum* normalmente poseen 3 copias de las cuales las dos primeras poseen una regulación estricta basada en la respuesta a la concentración de oxígeno en el medio, pero la tercera presenta divergencias debido a que produce una expresión muy baja. Al crecer las bacterias en condiciones metanotróficas, se vio un aumento en la expresión al añadir al medio de cultivo otros alcanos (Awala, *et al.*, 2021). Parece indicar que esta duplicación presenta diferencias en su regulación que están haciendo variar su función.

## **1.2 Términos y conceptos importantes**

Para entender este trabajo correctamente hay que conocer una serie de conceptos. El término homólogo fue introducido por Richard Owen en 1843 para designar "el mismo órgano en diferentes animales bajo cualquier variedad de forma y función". Owen claramente distinguía los homólogos de los análogos, los cuales definió como "una parte u órgano en un animal que tiene la misma función que otra parte u órgano en un animal diferente"(Koonin, 2005).

En la década de 1960 ya aparecen los conceptos de genes ortólogos y parálogos, aunque con el uso de diferentes palabras, en el trabajo de Zuckerkandl y Pauling, que sentó las bases de la evolución molecular como disciplina (Koonin, 2005).

Así pues, los genes ortólogos son originados por un proceso de especiación mientras que los genes parálogos aparecen gracias a un proceso de duplicación (Koonin, 2005).

Uno de los conceptos más importantes para entender este estudio es la transferencia horizontal de genes. La transferencia horizontal de genes (HGT, por sus siglas en inglés) fue descrita por primera vez en microorganismos a finales de la década de 1940, y alrededor de 20 años más tarde se especuló que jugaba un papel en la adaptación de eucariotas multicelulares, específicamente plantas. Desde entonces, los métodos para detectar HGT han mejorado, revelando la sorprendente extensión y relevancia de los procesos de HGT que explica el contenido genético de virus, procariotas y eucariotas. El mecanismo se basa en la transferencia de genes entre individuos o células de distintas especies (Soucy *et al.*, 2015).

Se creía que para que un gen transferido sobreviva en la línea celular de recepción durante largos períodos de tiempo, el gen generalmente necesitaba proporcionar una ventaja selectiva.

Sin embargo, ahora se sabe que muchos de los genes identificados como transferidos a través de genómica comparada entre parientes cercanos tienen efectos neutrales o casi neutrales en el receptor tanto en organismos procariontes como eucariotas. Así pues, parece que mientras que el gen transferido no cause algún daño, podrá integrarse con éxito. En el caso de las adquisiciones neutrales, pueden proporcionar más tarde combinaciones novedosas de material genético actuando como fuente de variabilidad genética para los organismos. La HGT ha sido reconocida durante mucho tiempo como una fuerza importante en la evolución de bacterias y arqueas (Soucy *et al.*, 2015). Los tres mecanismos de transferencia genética horizontal más estudiados en procariontes son la conjugación, la transformación y la transducción (Soucy *et al.*, 2015).

La conjugación requiere contacto físico entre una célula donante y una receptora a través de un *pili* de conjugación, mediante el cual se transfiere material genético. La conjugación se restringe canónicamente a células bacterianas como donante y receptora (Soucy *et al.*, 2015).

La transformación es la absorción de ADN exógeno del medio ambiente y se ha observado tanto en arqueas como en bacterias (Soucy *et al.*, 2015).

La transducción es la entrega de material genético a través de la infección de fagos debido a la integración de material genético exógeno del huésped en el genoma de un fago, y este fenómeno se ha observado tanto en bacterias como en arqueas. Hay dos tipos de transducción: generalizada, en la que se incorpora una pieza aleatoria del ADN del huésped durante la lisis celular; y especializada, en la que un profago se excinde de forma imprecisa del genoma del huésped causando la incorporación de parte del ADN flanqueante del huésped a la partícula infectiva. (Soucy *et al.*, 2015).

Finalmente falta hablar del ARN ribosomal 16S y su uso para situar cualquier bacteria dentro de la escala evolutiva y determinar su posición taxonómica. Los ácidos nucleicos presentes en todos los seres vivos cambian con el tiempo debido a la acumulación de mutaciones. Las diferencias en la secuencia de nucleótidos o aminoácidos que conforman las proteínas codificadas por los ácidos nucleicos reflejan la distancia evolutiva existente entre ellas, de esta forma podemos entender que cuanto más cerca se encuentran dos secuencias evolutivamente, más cambios tendrán en común respecto a una tercera secuencia más alejada (Olsen y Woese, 1993).

Para ser capaces de contrastar los cambios entre individuos es necesario fijar una secuencia que tenga su homólogo en todas las especies que se desee analizar. El ARN ribosomal 16S o ARNr 16S es una molécula que se encuentra presente en todos los procariotas. Gracias a esta característica es usada para establecer la relación filogenética de las distintas especies procariotas. Actualmente la taxonomía de procariotas se basa en este gen.

### **1.3 Aplicaciones de las monooxigenasas**

Aunque no son el objetivo de este estudio podemos añadir información sobre las aplicaciones que posee esta enzima. La amplia especificidad catalítica de las MMO las hace útiles para la biorremediación de contaminantes como el tricloroetileno (TCE) (Khider *et al.*, 2021).

Los estudios han demostrado que las bacterias que producen MMO pueden ser utilizadas para la biorremediación tanto *in situ* como *ex situ*. La biorremediación *in situ* se basa en la degradación de compuestos como el TCE por comunidades metanotróficas mixtas enriquecidas procedentes del medio ambiente. La biorremediación *ex situ*, por otro lado, se lleva a cabo en biorreactores de una sola etapa donde los biocatalizadores que producen MMO se cultivan y se usan para la biodegradación dentro del reactor (Khider *et al.*, 2021).

Otra de las aplicaciones biotecnológicas importantes que se le puede dar a este gen es su uso para la predicción de depósitos de petróleo o gas natural en el subsuelo (Qian-Zhi *et al.*, 2020). Si, al analizar una muestra ambiental se detecta un exceso de la presencia de estos genes o de bacterias conocidas que se sabe que poseen alcano-monooxigenasas como son las metano-monooxigenasas, sobre las que se centra este trabajo, se puede llegar a predecir la existencia de petróleo o gas natural.

Para llevar a cabo la predicción se usa como sonda el gen *pmoA* (Qian-Zhi *et al.*, 2020), lo que conlleva a una sobreestimación de resultados debido a que este gen presenta homología con secuencias de un gran número de genomas de distintas especies, muchas de las cuales no poseen en ningún momento el operón de las metano-monooxigenasas completo.

### **1.4 Objetivo**

El objetivo principal de este trabajo es analizar las relaciones filogenéticas entre las distintas metano-monooxigenasas presentes en los principales géneros bacterianos que las poseen y compararlo con las relaciones evolutivas dadas entre las distintas especies bacterianas marcadas por el gen del ARNr 16S.

Un objetivo secundario es tratar de buscar y localizar casos en los que se haya producido transferencia horizontal entre los genes que componen el clúster *pmoCAB* a partir de los resultados obtenidos.

## **2. Materiales y métodos**

### **2.1 Materiales**

Para llevar a cabo el estudio propuesto se utilizaron herramientas bioinformáticas como programas especializados en el alineamiento o en la búsqueda de secuencias similares a una sonda y bases de datos. Gracias a ellos se generaron árboles filogenéticos necesarios para el análisis evolutivo de los genes *pmo*.

#### **2.1.1 Bases de datos**

La base de datos elegida para realizar el estudio fue el GenBank del NCBI (Sayers *et al.*, 2022) debido a su gran cantidad de datos (contiene todas las secuencias publicadas) y la facilidad de búsqueda.

#### **2.1.2 Programas de búsquedas**

El *blast* (Altschul, *et al.*, 1990) es un programa gratuito al que se puede acceder desde la misma base de datos del NCBI el cual permite localizar similitudes de las secuencias elegidas para la consulta (*query* o sonda) con todas las contenidas en la base de datos.

Existen distintos tipos de *blast*. El tipo más común es el llamado *blastn*, este nos permite realizar búsquedas en bases de datos de nucleótidos utilizando como sondas secuencias de nucleótidos. Otro de los tipos, el *tblastx*, permite la búsqueda en bases de datos traducidas de nucleótidos utilizando como *query* secuencias de nucleótidos también traducidas. Este último tipo demostró ser el más eficaz para realizar las búsquedas requeridas para este análisis.

#### **2.1.3 Los genomas y secuencias de las especies con pMMO**

Los genomas completos son necesarios para poder obtener las secuencias génicas implicadas en el análisis, sobre todo para analizar el número de copias del operón *pmoCAB* completo o de cada uno de sus genes y sus semejanzas filogenéticas. A su vez, las secuencias obtenidas son necesarias para la construcción de alineamientos y árboles filogenéticos.

#### 2.1.4 Alineamientos y árboles filogenéticos

Se ha utilizado el programa MEGA en su versión 7.0.26. El programa se puede descargar de forma gratuita desde su página web y cuenta con un número muy elevado de herramientas de análisis filogenéticos altamente útiles para trabajos como éste. El nombre completo es MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets (Kumar et al., 2015).

## **2.2 Métodos**

### 2.2.1 Búsqueda de secuencias

El primer paso que se llevó a cabo en la realización de este trabajo es la búsqueda de genomas de procariontes que tuviesen tanto la secuencia de su metano monooxigenasa como la de su ARN 16S ribosomal correctamente secuenciadas y anotadas.

Inicialmente se intentó realizar este proceso utilizando la barra de búsquedas de la base de datos del NCBI, pero, esta técnica acabó presentando problemas, algunas debido a que se encontraban mal registradas en la propia base de datos y otras probablemente se debieran a malas anotaciones.

Al buscar en los genomas secuenciados la metano monooxigenasa se puede ver como la anotación presentaba distintas formas. En la mayoría de los casos este error se debía a que, como se ha explicado en la introducción, existen dos tipos de metano monooxigenasas las sMMO y las pMMO siendo estas muy diferentes entre sí. Cuando se decidió centrar el estudio en las metano-monooxigenasas particuladas, pMMO, se observó que el tipo de monooxigenasa no siempre estaba anotado. En general los genes que codifican esta familia de metano-monooxigenasas constituyen un clúster formado por tres genes codificantes de las subunidades C, A y B y no siempre estaba marcada la posición de cada gen.

Para localizar los genes codificantes para la pMMO sin depender de las anotaciones se decidió emplear los programas de búsqueda de secuencias del NCBI usando los tres genes que componen el operón *pmoCAB* procedentes de *Methylocystis sp. SC2* como referencia. Se usaron las tres secuencias como una sonda triple, es decir se introdujeron las tres secuencias separadas, pero al mismo tiempo lo que permitía analizar los resultados de forma individual y ver que genomas aparecían en los tres casos, confirmando la existencia de al menos un clúster *pmoCAB* completo.

Gracias a una de las opciones del *blast* que permite excluir resultados de los géneros o especies u obtener solo los resultados del género o especie que se desea en cada búsqueda se logró encontrar todas aquellas bacterias que presentaban estos genes. En un primer momento se utilizó para ir buscando genomas completos de especies pertenecientes a géneros que se sabe que poseen la pMMO. Cuando ya se obtuvieron todos los posibles por este método se realizó un *blast* pero ahora excluyendo todos los géneros de los que tenemos datos esperando encontrar algún género nuevo y distinto a los conocidos.

Tras conseguir todas las monooxigenasas de géneros de bacterias distintos utilizando el *blast* con *Methylocystis sp. SC2* se probó a usar otras referencias. Se terminó así la búsqueda de especies para formar parte de los árboles.

Una de las posibilidades que se debe tener en cuenta al tomar las secuencias de la base de datos para su posterior análisis es que las secuencias codificantes fuesen las cadenas mostradas en el archivo del NCBI o sus reversas complementarias. Para realizar el alineamiento es necesario utilizar la misma cadena de todas las secuencias. Utilizando el MEGA se obtenía la reversa complementaria de aquellas que se orientaban al revés en la mitocondria.

### 2.2.2 Alineamientos Múltiples de Secuencias

Para realizar el Alineamiento Múltiple de Secuencias (MSA) se utilizó el programa MEGA7. El programa permite abrir los archivos FASTA donde se guardaron las secuencias y usar diferentes herramientas de análisis.

El alineamiento consiste en situar las distintas secuencias de nucleótidos o aminoácidos una encima de otra y mediante la inserción de gaps o huecos ir tratando de lograr que se formen columnas en las que se sitúen nucleótidos o aminoácidos homólogos en cuanto a sus ancestros, es una búsqueda de homologías posicionales.

El MEGA cuenta con varias herramientas para realizar los MSA como el ClustalW (Higgins, 1994) o el MUSCLE (Edgar, 2004), para llevar a cabo el alineamiento de las metano-monooxigenasas de este análisis se han utilizado ambos algoritmos y además se ha optado por realizar mejoras de forma manual, característica que también permite el propio programa de MEGA que puede realizar los alineamientos traduciendo las secuencias de ADN a proteínas y volver a los nucleótidos fácilmente.

### 2.2.3 Árboles

Una vez alineadas las secuencias se exporta el alineamiento resultante y utilizando el mismo programa MEGA se pueden generar los árboles filogenéticos. El MEGA al igual que para los alineamientos incluye numerosas herramientas que emplean distintos algoritmos y modelos para el análisis filogenético.

Para llevar a cabo el análisis evolutivo de las secuencias se optó por la construcción de árboles usando el método estadístico de *Neighbour-Joining* (Saitou y Nei, 1987). Este método parte de la creación de una matriz de distancias en las que se relaciona cada par de secuencias presentes en el MSA. A cada par se le asigna un valor de distancia. El algoritmo comienza con un árbol en forma de estrella, un único nodo del que parten todas las ramas y mediante iteraciones en las que para cada caso evalúa dos ramas va buscando aquellas dos que posean un menor valor de un parámetro Q, que mide la distancia media de esas dos secuencias al resto. El algoritmo recalcula la longitud de las ramas y genera un nodo que las une y tras calcular las distancias de las distintas ramas con el nuevo nodo realizar una nueva ronda para generar el siguiente nodo. De esta forma iterativa al final se obtiene un árbol que carece de politomías.

La principal ventaja de este método en comparación con otros es su alta velocidad a la hora de generar el árbol lo que facilita su uso junto con otros algoritmos como el Bootstrap (Felsenstein, 1985). Este algoritmo permite generar una serie de árboles réplica que se pueden visualizar como un árbol consenso el cual lleva en cada nodo un valor, dicho valor hace referencia a la proporción de árboles réplica que incluyen ese nodo. A mayor valor de Bootstrap podemos concluir que es más probable que ese nodo sea correcto y viceversa.

Para llevar a cabo el árbol de los genes codificantes para el ARNr 16S se emplea como modelo de sustitución el *Tamura-Nei*, este modelo estima las distancias entre pares de secuencias considerando que la probabilidad de sustituciones nucleotídicas por transversiones y los dos tipos de transiciones son diferentes. Además, considera que las frecuencias de los cuatro nucleótidos no están equilibradas en el conjunto del MSA (Tamura y Nei, 1993). Para el resto de los árboles se empleó otro método llamado *Maximum Composite Likelihood* que consiste en una mejora del anterior desarrollado por los mismos autores (Tamura *et al.*, 2004).

### 3. Resultados

#### 3.1. Localización de genes pmo en los genomas

Al realizar las búsquedas mencionadas con el *tblastx* se obtuvieron una serie de genomas, al ver en que posiciones de cada uno se daba la mayor similitud con la sonda se pudo determinar en qué zona del genoma se encontraba el gen y cuantas copias había. También permitió ver si se trataba de una subunidad aislada o un clúster completo dependiendo de si la región aparecía también en los resultados del alineamiento de las otras dos secuencias.

En la realización de los *tblastx* existieron una serie de problemas que hubo que solventar. Estos problemas se dan con respecto a la anotación de las secuencias. En el clúster, el orden de los genes es C-A-B y se vio como en algunas entradas de la base de datos tenían anotados los tres genes por orden alfabético, A-B-C. Al analizar las secuencias de los tres genes comparándolas con las de otras entradas se apreció que las anotaciones eran incorrectas, el nombre dado a cada gen del operón era erróneo.

Tabla 1: Clasificación taxonómica de las especies estudiadas. Todas las especies presentes en la tabla son bacterias por lo que se ha omitido la categoría de reino.

Filo	Clase	Orden	Familia	Organismo
Pseudomonadota	Alphaproteobacteria	Hyphomicrobiales	Beijerinckiaceae	<i>Methylocapsa acidiphila</i>
			Methylocystaceae	<i>Methylosinus trichosporium</i>
				<i>Methylocystis sp.</i>
				<i>Methylocystis parvus</i>
				<i>Methylocystis rosea</i>
	Gammaproteobacteria	Methylococcales		<i>Methylomicrobium sp.</i>
				<i>Methylobacter sp.</i>
				<i>Methylocaldum marinum</i>
				<i>Methylocaldum szegediense</i>
				<i>Methylococcus geothermalis</i>
				<i>Methylococcus capsulatus</i>
				<i>Methylomonas sp.</i>
				<i>Methylovulum psychrotolerans</i>
				<i>Methylomagnum ishizawai</i>
	candidate division NC10		-	
Verrucomicrobiota				<i>Methylacidimicrobium sp.</i>
	Methylacidiphilae	Methylacidiphilales	Methylacidiphilaceae	<i>Methylacidiphilum kamchatkense</i>

Otro problema que surgió durante la realización de los *tblastx* fue que en ocasiones las secuencias que se encontraron no estaban anotadas. En estos casos lo que se hizo fue fijarse en las posiciones de inicio y fin de la zona en la que se detectaba similitud, incluir 200 nucleótidos extras en cada extremo y usar la secuencia resultante en el análisis. Al alinearlos con el MEGA fue muy fácil eliminar los nucleótidos sobrantes de los extremos y alinear el resto con normalidad. Tras resolver esos problemas se obtuvieron las secuencias de los distintos genes del operón *pmoCAB* y el ARNr 16S de distintos genomas de un conjunto de bacterias de distintos géneros, en la Tabla 1 se puede ver la clasificación taxonómica de cada una de ellas.

Todas ellas cuentan con al menos una copia del clúster de genes *pmoCAB* completo, pero alguna cuenta con duplicaciones de este o con solo duplicaciones de alguno de sus genes. En la Tabla 2 se recoge el número de copias presentes junto con el código del genoma de referencia empleado para la realización del estudio.

Tabla 2: Número de copias del operón y de los genes individuales de las especies ensayadas. En la tabla aparece también una columna extra donde se incluye el código del genoma de referencia empleado para obtener la información recogida.

Organismo	Genoma de Referencia	Número Clúster <i>pmoCAB</i>	Número <i>pmoC</i> individuales
<i>Methylocapsa acidiphila</i>	CT005238	1	0
<i>Methylosinus trichosporium</i>	CP023737	2	1
<i>Methylocystis</i> sp.	AJ584611	1	0
<i>Methylocystis parvus</i>	CP092968	2	5
<i>Methylocystis rosea</i>	CP044328	2	2
<i>Methylomicrobium</i> sp.	CP024202	1	0
<i>Methylobacter</i> sp.	CP091424	1	0
<i>Methylocaldum marinum</i>	AP017928	2	2
<i>Methylocaldum szegediense</i>	OX458333	2	2
<i>Methylocaldum</i> sp.	AB275418	1	0
<i>Methylococcus geothermalis</i>	CP046565	2	1
<i>Methylococcus capsulatus</i>	CP079096	2	1
<i>Methylomonas</i> sp. DH-1	CP014360	1	0
<i>Methylomonas</i> sp. EFPC3	CP116398	1	1
<i>Methylovulum psychrotolerans</i>	CP022129	1	0
<i>Methylomagnum ishizawai</i>	AP019783	2	1
<i>Methylomirabilis oxyfera</i>	FP565575	1	1
<i>Methylacidimicrobium</i> sp.	CP066203	2	1
<i>Methylacidiphilum kamchatkense</i>	CP037899	3	1

Se pudo ver como de las tres subunidades existentes la única que posee duplicaciones en solitario es la C, ni la A ni la B aparecen en ninguna de las especies de forma aislada. Como se indica en la introducción el operón posee un solo promotor situados aguas arriba de *pmoC* por lo que la existencia única de duplicaciones de esta subunidad pueda deberse a que es la única que conserve su expresión pese a perder el resto del operón.

Como se menciona en el apartado de materiales y métodos utilizando el MEGA se generaron una serie de árboles para realizar el análisis filogenético de forma gráfica.

### **3.2 ARN ribosomal 16S**

Para llevar a cabo la búsqueda de las distintas secuencias del ARN ribosomal 16S que componen este árbol se emplearon los genomas registrados en la base de datos del NCBI. La gran mayoría de estos se encontraban anotados permitiendo localizar fácilmente las secuencias correspondientes al 16S ya que al tratarse de una secuencia importante para evaluar las filogenias de las especies es común que se encuentre anotado siempre que exista un genoma secuenciado.

Una excepción fue el genoma de *Methylobacterium sp. wino1*, el cual no se encontraba anotado en la base de datos y, además se trataba del único representante del género *Methylobacterium* que tuviera unos genes *pmoCAB* similares al resto de genomas analizados según el *blast*. Para este caso especial lo que se tuvo que hacer fue una búsqueda *blast* del 16S utilizando como sonda una secuencia del 16S anotada de una bacteria del mismo género, en este caso se empleó *Methylobacterium sp. 5B* y se realizó un *blast* especificando que diese como respuesta únicamente los resultados de *Methylobacterium sp. wino1*. A partir del alineamiento resultante se tomó la secuencia de *Methylobacterium sp. wino1* que lo abarcase por completo junto con unos 200 nucleótidos extras en cada extremo.

Esa secuencia se añadió al alineamiento con los otros 16S y gracias al programa MEGA que permite editar las secuencias y el alineamiento se recortaron los extremos sobrantes hasta quedarse con la secuencia correspondiente al ARN ribosomal 16S de *Methylobacterium sp. wino1*.

Para el estudio realizado solo interesa el 16S como reloj molecular por lo que únicamente se ha buscado una copia en cada genoma, es probable que existan más, pero a diferencia de las de los genes *pmoCAB* estas no son relevantes por lo que únicamente se ha tomado una copia de cada uno.

Con las secuencias bacterianas encontradas se realizó un MSA de 1585 pb. A partir de este MSA se pudo realizar el árbol filogenético representado en la Figura 2.

El árbol presentado en la Figura 2 es el basado en las distintas secuencias del ARNr 16S de las bacterias. Como resultados de este árbol esperaríamos obtener las especies analizadas agrupadas según la posición taxonómica en la que están clasificados tal y como se observa en la Tabla 1, las ramas más externas muestran las relaciones filogenéticas de los filos y las clases y las internas las de familias, géneros y especies.

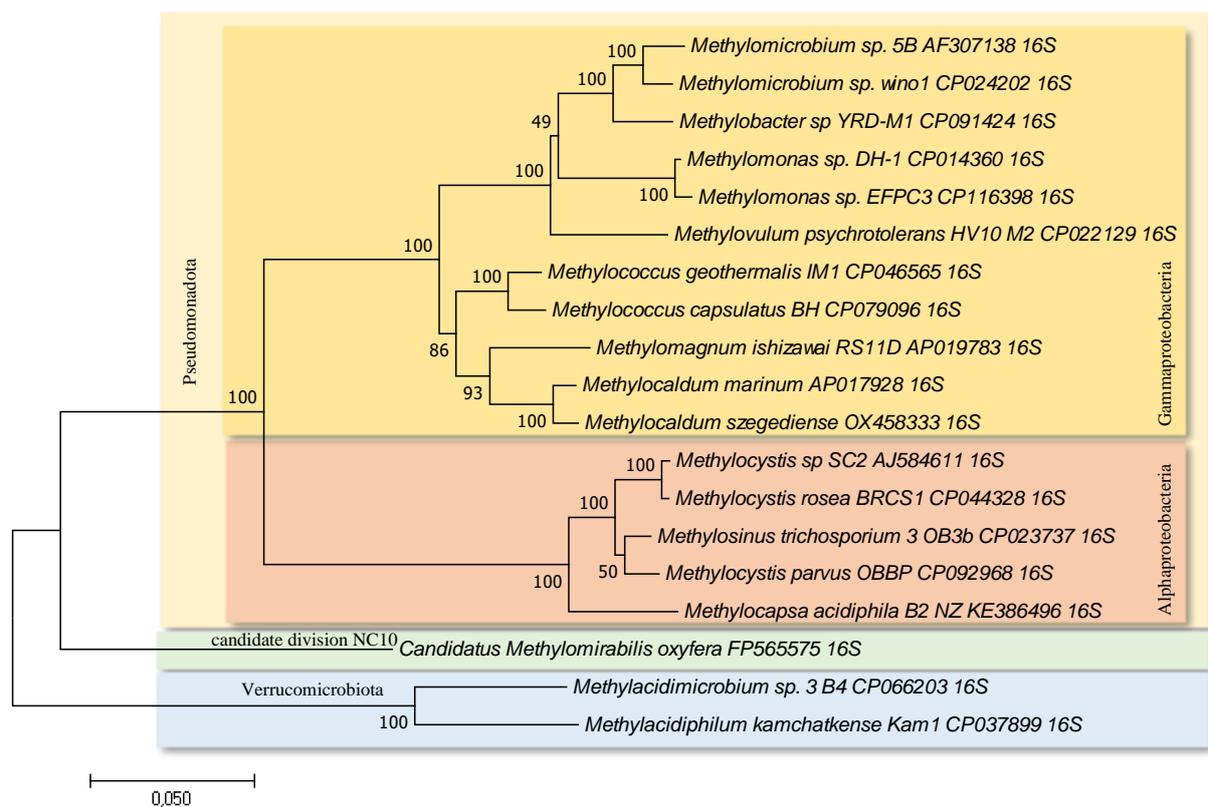


Figura 2: Árbol Filogenético del gen codificante para ARNr 16S. Este árbol se ha realizado a partir de las secuencias del de las especies en estudio siguiendo el método estadístico de *Neighbour-Joining*. Se ha usado el método de *Tamura-Nei* con distribución gamma como modelo de sustitución y 500 réplicas Bootstrap. El amarillo las bacterias pertenecientes al filo Pseudomonadota, en azul las pertenecientes al filo Verrucomicrobiota y en verde el filo candidate division NC10.

Tal y como se puede observar los resultados obtenidos coinciden con lo esperado. Claramente se pueden distinguir tres grandes bloques que coinciden con los filos Pseudomonadota, la división NC10 para la que se propone el nombre de Methylococcoides con categoría de filo según el NCBI y Verrucomicrobiota, siendo los dos primeros los filos más próximos entre sí. Dentro del filo Pseudomonadota se distinguen a su vez dos grandes grupos constituidos por las bacterias de las clases Alphaproteobacteria y Gammaproteobacteria. Al observar los distintos valores de Bootstrap presentes en las ramas se ve como la gran mayoría son de 100 o cercanos

a este valor tratándose por lo tanto de una topología muy fiable. Todos los grupos taxonómicos incluidos en la Tabla 1 serían monofiléticos con un soporte del Bootstrap muy alto (los valores de los nodos que representan a sus ancestros aparecen en el 10% de las réplicas). Destacan únicamente dos nodos con un valor relativamente bajo de Bootstrap, el que une las *Methylomonas* con el grupo de *Methylomicrobium* y *Methylobacter* dentro del grupo de las gammaproteobacterias y el que une *Methylocystis parvus* con *Methylosinus trichosporium* dentro de las alphaproteobacterias, la poca fiabilidad de este último nodo indica que las distintas especies del género *Methylocystis* podrían formar un grupo monofilético.

Al analizar la clase de las gammaproteobacterias podemos ver como el árbol se divide en dos grupos que posteriormente se subdividen hasta llegar a cada especie. La clasificación taxonómica no refleja esta división inicial que genera dos grupos bastante diferenciados y todas las gammaproteobacterias estudiadas se engloban en la familia Methylococcaceae. Por otro lado, en el caso de las alphaproteobacterias se puede ver como la *Methylocapsa acidiphila* aparece relativamente alejada del resto, en este caso, la clasificación taxonómica si recoge a *M. acidiphila* en una familia diferente al resto de Alphaproteobacteria.

### **3.3 *pmoCAB***

Este conjunto de tres genes origina las tres subunidades que dan lugar a las metano-monooxigenasas. El método empleado para localizar las distintas secuencias ya se ha comentado en el apartado de materiales y métodos. Tras importar las secuencias a MEGA se alinearon las secuencias obteniéndose un MSA con una longitud de 3474 pb.

La Figura 3 muestra el árbol filogenético obtenido con el alineamiento de los operones *pmoCAB* completos.

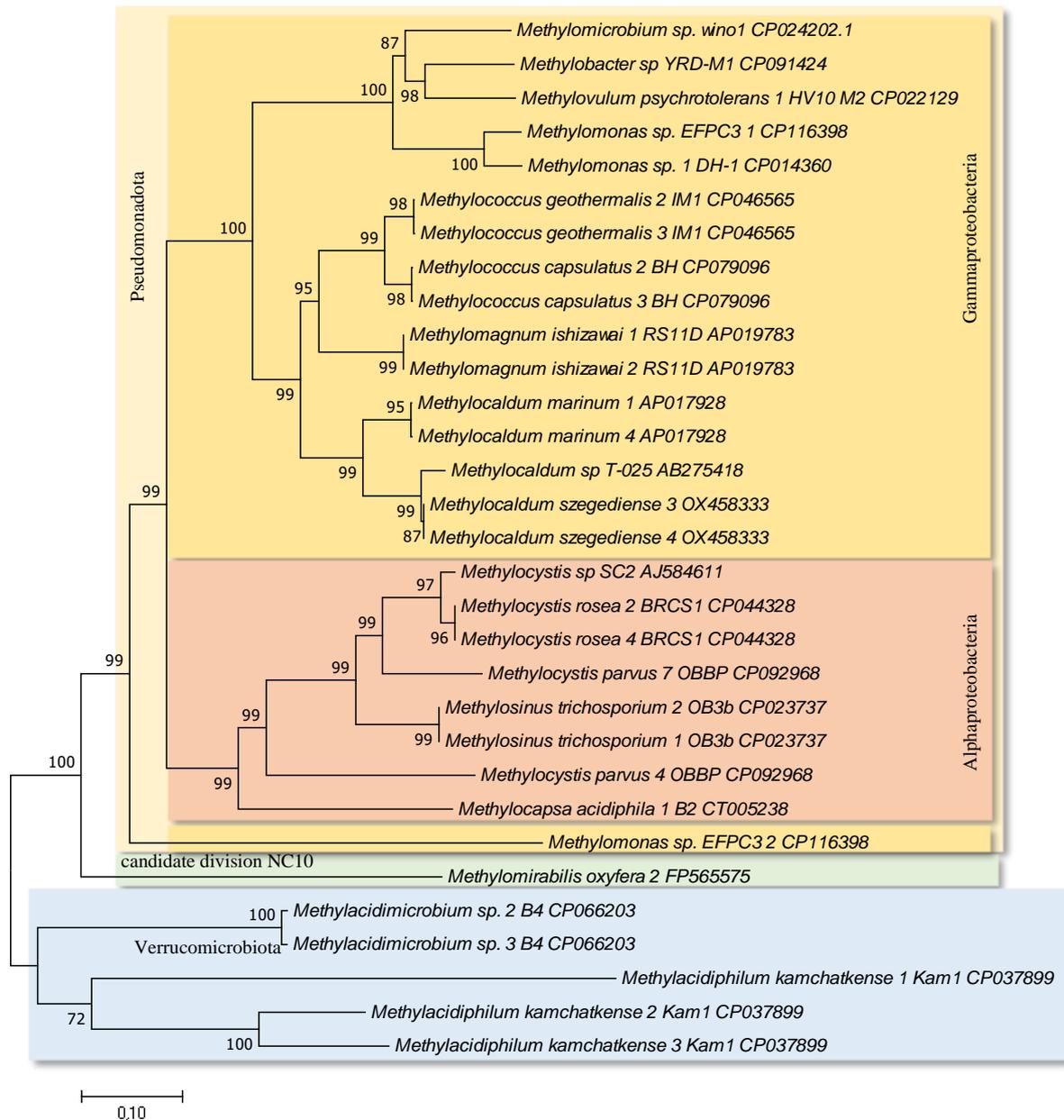


Figura 3: Árbol Filogenético del operón de genes *pmoCAB*. Este árbol se ha realizado a partir de las secuencias de las especies en estudio siguiendo el método de *Neighbour-Joining*. Se ha usado el “*Maximum Composite Likelihood*” como modelo de sustitución y un valor de 500 de Bootstrap. El amarillo las bacterias pertenecientes al filo Pseudomonadota, en azul las pertenecientes al filo Verrucomicrobiota y en verde el filo division NC10.

Al realizar un análisis de la topología del árbol de la Figura 3 puede verse cómo existen diferencias con la Figura 2 las cuales no son solo debidas al aumento de secuencias derivadas de las duplicaciones del operón *pmoCAB*. Lo primero que llama la atención es que, a diferencia de la Figura 2 en la que se veían dos bloques claros en la división de clases del filo

Pseudomonadota ahora vemos como la secuencia número 2 de *Methylomonas sp.* EFPC3 posee tantas diferencias nucleotídicas con respecto al resto que es excluida de los dos bloques principales que representan las dos clases del filo. Posteriormente se analizarán las posibles causas de esta situación.

El soporte de los valores Bootstrap de este árbol indican que todos los nodos son fiables con valores bastante elevados, el menor de todos ellos es de 72 del nodo que une a la secuencia 1 de *Methylacidiphilum kamchatkense* (Verrucomicrobiota) y las otras dos secuencias, 2 y 3 de la misma especie. Al igual que ocurre con *Methylomonas sp.* La secuencia 1 de *M. kamchatkense* posee bastantes diferencias con respecto a las otras dos.

La secuencia 4 de *Methylocystis parvus* (Alphaproteobacteria) es otro de los casos de una posición alejada del resto de las secuencias de su mismo género, que se encuentran formando un bloque.

Existen otros cambios menores como el cambio de posición de *Methylovulum* y *Methylomagnum ishizawai* cambian de posición o la adición de *Methylocaldum sp.* Para la que no ha encontrado el 16S.

### **3.4 *pmoC***

Al analizar las subunidades de forma individual lo primero que destaca es que la subunidad C sea la única que aparece como copia aislada en algunos genomas, incluso en algunas especies como *Methylocystis parvus* OBBP existen varias duplicaciones de la subunidad individual. Los nuevos genes *pmoC* proporcionan información extra por lo que es necesario realizar un tercer alineamiento y su posterior análisis filogenético. El MSA obtenido al alinear solo las *pmoC*, tanto las individuales como las pertenecientes a los clústeres posee una longitud de 993.

Considerando que existían copias extra del gen *pmoC*, se construyó un árbol a partir del alineamiento de estos genes incluyendo tanto las que forman parte de los operones *pmoCAB* completos como las que aparecen en solitario.

En general, la topología del árbol de la Figura 4 es más parecida a la del árbol del operón completo (Figura 3) que a la del gen codificante para el 16S (Figura 2). Pero la posición de los nuevos genes varía en función de la especie a la que pertenezca.

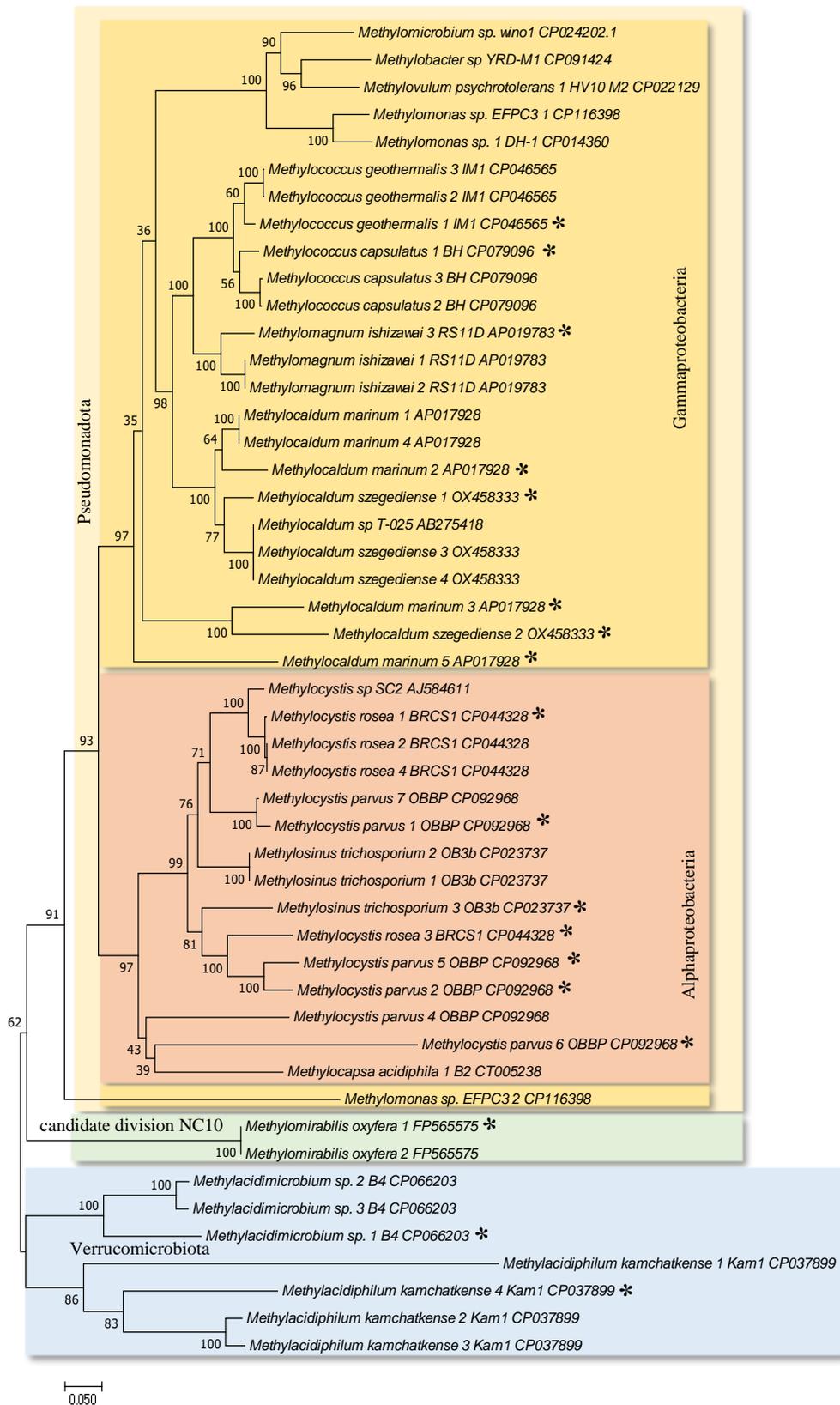


Figura 4: Árbol Filogenético del gen *pmoC*. Este árbol se ha realizado a partir de las secuencias de las especies en estudio siguiendo el método de *Neighbour-Joining*. Se ha usado el “*Maximum Composite Likelihood*” como modelo de sustitución y un valor de 500 de Bootstrap. El amarillo las bacterias pertenecientes al filo Pseudomonadota, en azul las pertenecientes al filo Verrucomicrobiota y en verde el filo candidate division NC10. Las secuencias que presentan un asterisco negro son aquellas que poseen el gen *pmoC* de forma individual.

Al igual que en la Figura 3 vemos como la secuencia 2 de *Methylomonas sp.* EFPC3, la 1 de *M. kamchatkense* y la 4 de *Methylocystis parvus* aparecen más alejadas de donde se esperaría atendiendo a la filogenia del 16S. En el nuevo alineamiento se incluyen 18 nuevas secuencias que no se encuentran en operones completos.

En algunos casos los genes *pmoC* aislados se encuentran próximos filogenéticamente a los genes de la misma especie incluidos en los operones completos, como ocurre en *Methylococcus geothermalis*, *M. capsulatus*, *Methylomagnum ishizawai*, el gen 2 de *Methylocaldum marinum* y el 1 de *M. szegediensie* dentro del segundo grupo de Gammaproteobacteria, pero otros genes aislados se encuentran separados de su especie como ocurre con los genes 3 y 5 de *Methylocaldum marinum* y 2 de *M. szegediensie*. En Alphaproteobacteria también hay copias aisladas junto a copias en operones (*Methylocystis rosea* 1 y *M. parvus* 1) y otras copias mucho más alejadas en la filogenia como *M. trichosporium* 3, *M. rosea* 3, *M. parvus* 5 y 2 que forman un grupo entre ellas y *M. parvus* 6 más alejada. En los filos Verrucomicrobiota y Methylomirabilota cada especie tiene una copia aislada de *pmoC* que se mantienen junto al resto de copias de su especie, pero más alejadas que en Pseudomonadota. Muchos de los nuevos nodos, que sitúan en el árbol a los genes aislados tienen valores de Bootstrap mucho más bajos que el resto. Su posición no es tan fiable como la de otros genes, y de nuevo presentan mayor número de sustituciones de las que existen entre los genes de los operones completos.

Tanto en la Figura 3 como en la Figura 4 se ha omitido una secuencia de *Methylocystis parvus*, concretamente la número 3. Esta secuencia pese a estar anotada como un operón *pmoCAB* y tener similitudes con el resto (presenta homología con las secuencias observadas), también presenta un gran número de diferencias. Al ser incluida en el árbol, genera una rama muy larga que dificulta la visualización del resto de secuencias, por tanto, se decidió no incluirla en las figuras.

### **3.5 *pmoA* y *pmoB***

Pese a no existir copias individuales de los genes *pmoA* y *pmoB* estos aparecen dentro de los clústers y pueden ser analizados de forma individual. El alineamiento de los genes *pmoA* presenta una longitud de 975 pb mientras que el alineamiento de *pmoB* es de 1359 pb.

Se construyeron sendas filogenias para los genes *pmoA* y *pmoB* para comprobar si sus relaciones filogenéticas se ajustan a las obtenidas con el operón completo. Los árboles obtenidos se muestran en las Figuras 5 y 6 respectivamente. No se detectaron copias aisladas

de estos genes. Estas figuras permiten analizar si la topología de la Figura 3 se debe únicamente a variaciones en el *pmoC* o también a variaciones en los otros dos genes del operón.

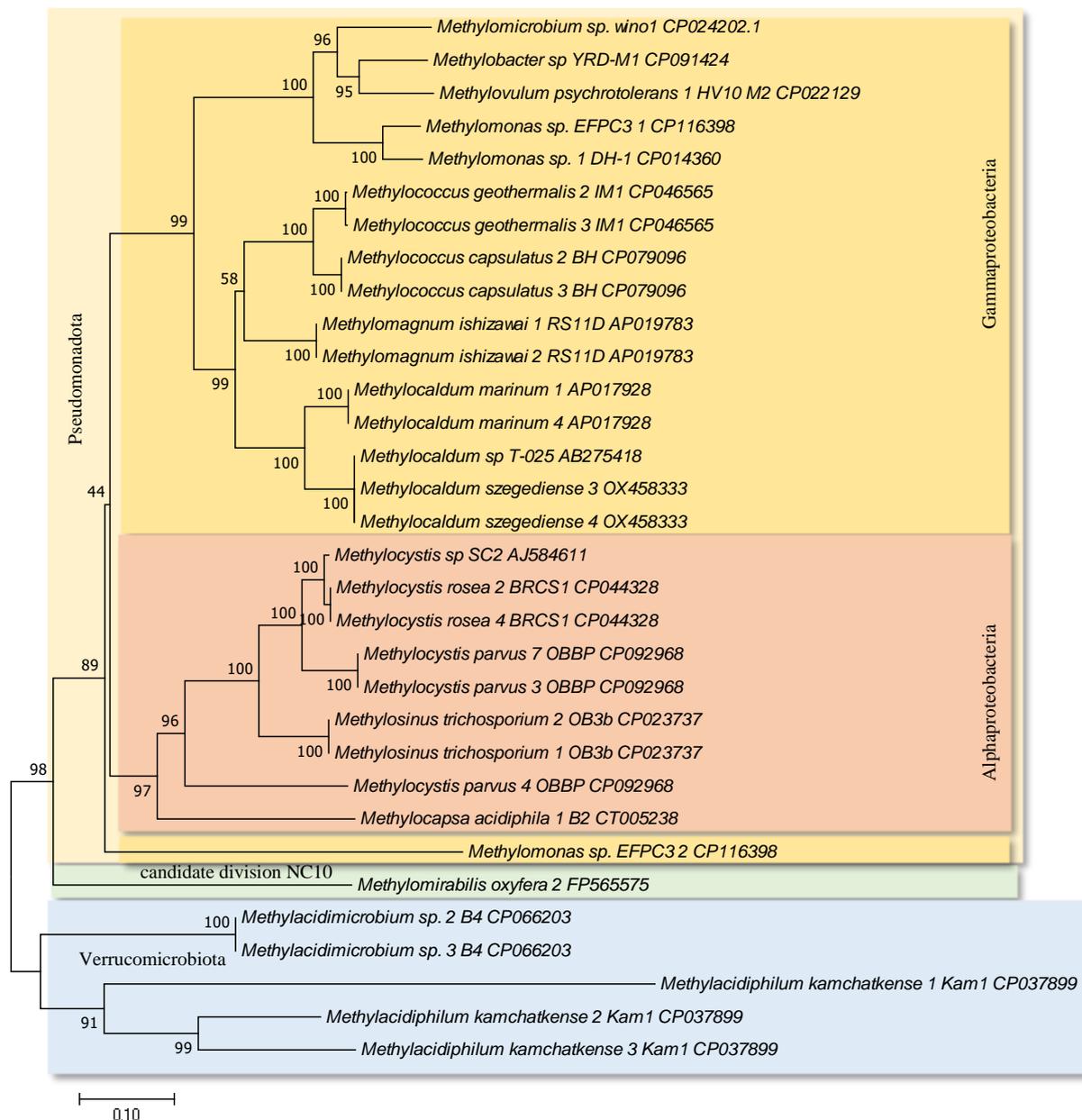


Figura 5: Árbol filogenético del gen *pmoA*. Este árbol se ha realizado a partir de las secuencias de las especies en estudio siguiendo el método de *Neighbour-Joining*. Se ha usado el “*Maximum Composite Likelihood*” como modelo de sustitución y un valor de 500 de Bootstrap. El amarillo las bacterias pertenecientes al filo Pseudomonadota, en azul las pertenecientes al filo Verrucomicrobiota y en verde el filo candidate division NC10.

Las Figuras 5 y 6 mantienen la topología de la Figura 3 como se esperaría. Ambas representan árboles con nodos altamente fiables debido a sus altos valores de Bootstrap a excepción de los nodos de unión entre *Methylobacter sp.* y *Methylovulum psychrotolerans* y entre ambos con *Methylomonas sp.* en la Figura 6, estos genes pertenecen a especies del primer grupo de las

gammaproteobacterias. Debe destacarse que para estos árboles la secuencia de *M. parvus* 3 no presenta tantos cambios como para ser necesario omitirla por lo que si aparece presente en las

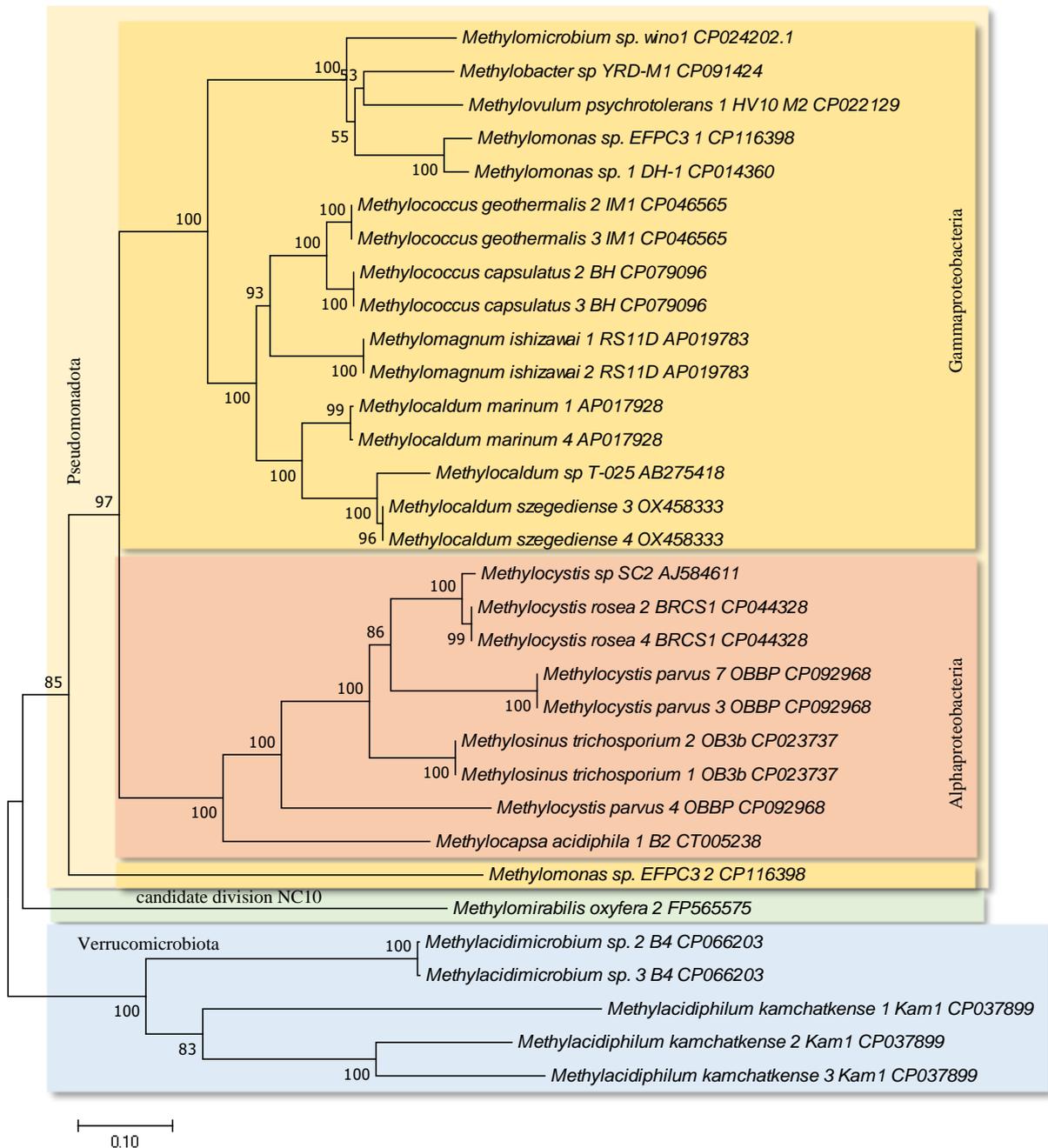


Figura 6: Árbol Filogenético del gen *pmoB*. Este árbol se ha realizado a partir de las secuencias de las especies en estudio siguiendo el método de *Neighbour-Joining*. Se ha usado el “*Maximum Composite Likelihood*” como modelo de sustitución y un valor de 500 de Bootstrap. El amarillo las bacterias pertenecientes al filo Pseudomonadota, en azul las pertenecientes al filo Verrucomicrobiota y en verde el filo candidate division NC10.

## 4. Discusión

### 4.1 Genes que acumulan un gran número de mutaciones

Al comparar los resultados de las Figuras 2, 3 y 4 se puede ver la existencia de diferencias en las topologías. Las diferencias se deben sobre todo a la presencia de duplicaciones en los genes codificantes para monooxigenasas particuladas. Las duplicaciones que se han producido son frecuentes tanto para el operón completo como para el gen *pmoC*. Las bacterias del filo Verrucomicrobiota destacan por sus duplicaciones del operón completo, mientras que para el gen *pmoC* es *Methylocystis parvus* la que más duplicaciones ha sufrido, llegando a tener 5 copias individuales.

Las duplicaciones de las dos especies de Verrucomicrobiota se mantienen en grupos monofiléticos sin diferencias notorias entre sí, sin embargo, esto no ocurre al revisar otros organismos como el mencionado *Methylocystis parvus*, en este caso se ve claramente en la Figura 4 como las secuencias 7 (*pmoCAB*) y 1 (*pmoC*) se encuentran en la posición que marca el 16S, mientras que las secuencias 4 (*pmoCAB*) y 2, 5 y 6 (*pmoC*) se encuentra alejadas filogenéticamente de las primeras. Pese a las diferencias existentes entre las duplicaciones, todavía se mantienen en la zona del árbol correspondiente a alphaproteobacterias, indicando que pese a sus diferencias es más probable que todas partan de una secuencia inicial que sufrió varias duplicaciones a que de alguna forma las copias provengan de un proceso de transferencia horizontal entre bacterias, en este caso hubieran sido bacterias del mismo grupo taxonómico. Las claras diferencias que presentan entre si probablemente sean debidas a mutaciones que podrían acabar dando una pérdida de función o un cambio de función. La pérdida de función podría estar asociada a una pérdida de presión de selección depuradora, en esta selección las mutaciones que forman alelos deletéreos son eliminadas. El cambio de función estaría asociado o a un aumento de selección positiva, en la que las mutaciones que producen cambios en las proteínas son fijadas con mayor frecuencia de lo esperado por azar. En ambos casos se produce un aumento de cambios en sus nucleótidos y como consecuencia una mayor distancia en los árboles filogenéticos. Estos aumentos en los cambios nucleotídicos suelen producir nodos incorrectos que tienen poco respaldo Bootstrap. En el caso de la secuencia 3 (*pmoCAB*) de *Methylocystis parvus* debido a que presenta un gran número de diferencias con todas las demás, incluyendo las de otros filos, ha tenido que ser eliminada de las Figuras 3 y 4.

Aunque la secuencia 3 (*pmoCAB*) de *M. parvus* no se haya podido incluir en los árboles de los operones completos (Figura 3) y de los genes *pmoC* (Figura 4), los otros dos genes del operón

son muy semejantes a los de la copia 7 de la misma especie como puede apreciarse en los árboles de *pmoA* (Figura 5) y *pmoC* (Figura 6). Este comportamiento anómalo del gen *pmoC* en una copia concreta del operón corrobora la pérdida o el cambio de función de alguno de los genes que se ha generado por duplicación.

En resumen, la secuencia 3 de *Methylocystis parvus* se originó por duplicación del operón completo, pero que a diferencia de las demás, sufrió un gran número de cambios en el gen que codificaba la subunidad C de la pMMO. Esta subunidad participa principalmente en el anclaje de la proteína a la membrana, aunque como indican Lieberman y Rosenzweig (2005) puede ser que en ella se encuentre parte del centro activo por lo que esta secuencia podría dar lugar a una enzima que tuviese una función o especificidad algo distinta al resto de secuencias de *Methylocystis parvus*, como ocurre en la cepa *Methylocystis sp.* SC2 estudiada por Awala *et al.*, (2021).

Otro de los casos de secuencias pertenecientes a una misma especie que aparecen altamente separadas en las figuras es la que se da con la secuencia 2 de *Methylomonas sp.* EFPC3 correspondiente aun operón completo. Este caso es sumamente llamativo debido a que en los cuatro árboles filogenéticos de genes *pmo* Figuras 3, 4, 5 y 6 representa la única copia de Gammaproteobacteria que aparece lo suficientemente alejada como para encontrarse fuera de la rama de la que surgen las demás. A diferencia de lo que ocurre con la copia 3 de *Methylocystis parvus* donde los genes *pmoA* y *pmoB* están muy conservados, todos los genes de la secuencia 2 de *Methylomonas sp.* EFPC3 han acumulado gran número de mutaciones, lo que parece indicar que todo el operón ha perdido su función.

#### **4.2. Semejanza entre genes duplicados de la misma especie**

Cuando se encuentran en el mismo genoma diversos genes homólogos originados por duplicaciones se considera que están formando familias génicas. Estos genes son parálogos y es importante conocer a qué genes se corresponden en el genoma de otras especies; un mismo gen en dos especies se correspondería con dos genes ortólogos. Cuando las duplicaciones son antiguas, es decir anteriores a la diversificación de las especies, es fácil identificar ortólogos ya que los genes parálogos son más distantes y muestran mayor número de cambios entre ellos que los genes ortólogos.

En ocasiones cuando hay familias génicas, como ocurre en la mayoría de las especies analizadas con los genes *pmoCAB*, las copias dentro de la misma especie se parecen más entre sí que a los

genes de otras especies, y esto es así en prácticamente todas las copias de los operones completos *pmoCAB* (Figura 3) y en muchos de los genes aislados de *pmoC* (Figura 4). La explicación más sencilla sería que estos genes proceden de duplicaciones muy recientes, posteriores a la especiación e independientes en cada especie. Sin embargo, el patrón de dos copias completas del operón y una copia aislada del gen de *pmoC* se repite en todas las especies del segundo grupo de gammaproteobacterias, en *Methylocystis rosea* y en las dos especies de Verrucomicrobiota. Lo que hace pensar que el ancestro de todas ellas también pudo tener esas tres copias en su genoma. Nei y Rooney (2005) propusieron dos modelos para explicar la mayor semejanza entre genes parálogos dentro de cada especie, el modelo de *birth-and-death* y la evolución concertada.

El primero de los modelos, *birth-and-death*, propone que a lo largo de todo el periodo evolutivo se generan nuevos genes mediante procesos de duplicación génica (*birth*) y que estos permanecen en el genoma durante mucho tiempo o se inactivan o se eliminan (Nei y Rooney, 2005). Gracias a este modelo se podrían explicar las múltiples copias individuales del gen *pmoC* como casos en los que, tras duplicarse el operón completo se perdieran las otras dos subunidades. En este caso también podrían detectarse copias más antiguas, o copias modernas que están acumulando cambios por una pérdida de función como la copia 3 de *Methylocystis parvus* u otros genes señalados en el apartado anterior, en este caso serían copias en proceso de muerte (*death*). El principal argumento en contra de esta hipótesis sería la aparición de dos operones completos y uno parcial de forma repetida en las especies antes mencionadas. El paralelismo observado entre especies y grupos taxonómicos supondría un proceso poco probable que se repite simultáneamente y de forma independiente en distintos genomas.

El segundo modelo, la evolución concertada, es el más antiguo de los dos. El modelo propone que durante el proceso evolutivo todos los genes miembros de una misma familia evolucionan a la vez y por medio de recombinaciones internas los distintos parálogos de un genoma comparten los cambios nucleotídicos que se van fijando (Nei y Rooney, 2005). El proceso de convergencia por recombinación génica está bien estudiado en las diferentes copias de los genes ribosomales de todas las especies bacterianas que guardan gran similitud en sus secuencias, aunque existe, rara vez se observa diversidad intra genómica (Espejo y Plaza en 2018 publicaron una revisión). Un proceso de conversión génica semejante al de los genes ribosomales podría ocurrir con las copias de *pmoCAB*.

Bajo la hipótesis de la evolución concertada, parece que las tres copias, dos completas y un *pmoC* aislado sería el estándar y podría ser la estructura de los genes en el ancestro de todas las especies estudiadas. Considerando las distancias genéticas del árbol de la Figura 4, en Verrucomicrobiota la conversión génica sería menos frecuente o eficiente. La presencia de una única copia de los genes codificantes para la metanomonoxigenasa en el primer grupo de gammaproteobacterias formado por *Methylomicrobium*, *Methylobacter*, *Methylomonas* y *Methylovulum* sería una pérdida secundaria o una adquisición tardía en estos taxones de los genes *pmoCAB* por transferencia horizontal.

Este modelo no explicaría el caso de la secuencia 3 de *Methylocystis parvus* que como se ha comprobado con las Figuras de los genes *pmoC*, principalmente en la Figura 5, ha evolucionado de forma distinta a los otros dos miembros de la familia. Este gen y junto con otros entre los que se incluyen las copias 2, 4, 5 y 6 de *Methylocystis parvus*, 3 de *M. rosea*, 3 y 5 de *Methylocaldum marinum* y la 2 de *M. szegediense* habrían dejado de sufrir la conversión génica del resto de las copias y habrían evolucionado de forma independiente al resto.

La secuencia 5 de *Methylocaldum marinum* pudo surgir después de la especiación y acabar perdiendo su función causando que, por un proceso de selección positiva, se fijasen muchos cambios

Hay que destacar la rama de la secuencia 1 de *Methylacidiphilum kamchatkense*, la cual posee una longitud mayor que las demás de su misma especie indicando que posee una mayor capacidad para acumular mutaciones evolucionando de forma independiente a las otras dos secuencias de la misma especie. Esta característica es típica de genes que han acabado perdiendo su función provocando que la selección depuradora que actúa sobre ellos deje de funcionar y cause este aumento de cambios que se ve reflejado gráficamente como un aumento en la longitud de la rama. Las ramas largas como las de la secuencia 1 de *Methylacidiphilum kamchatkense*, la 2 de *Methylomonas sp.* EFPC3, la 6 y la 4 de *Methylocystis parvus* o el gen *pmoC* de la secuencia 3 de *Methylocystis parvus* deben tener un origen reciente ya que todas ellas parecen ser pseudogenes y estos suelen durar poco en bacterias debido a que los procariotas tienden a reducir su genoma por selección.

El hecho de que un gen de *pmoC* haya evolucionado de forma altamente coordinada con los clústeres completos del operón hace pensar que este ha de poseer alguna función especial que lo haga necesario o al menos que su presencia se vea favorecida por la selección. Tal vez el gen *pmoC* del operón se expresa en cantidades más bajas de las requeridas y es necesario para estas

bacterias mantener los niveles de esta subunidad proteica para mejorar el proceso de oxidación del metano que llevan a cabo o puede ser que ayude fijando metano para que posteriormente lo procesen las pMMO producidas al transcribir y traducir los operones completos.

#### **4.3. Transferencia horizontal en los genes *pmoCAB***

Respecto a los datos que apoyan la existencia de algún caso posible de transferencia horizontal destaca la aparición de estos genes en *Methylomirabilis oxyfera* (candidate división NC10). La idea de que el operón de *M. oxyfera* se encuentre en la bacteria debido a un evento de transferencia horizontal gana fuerza si se ve que, a diferencia de los otros dos filos en estudio, que poseen varias especies representantes de cada uno, la especie *Methylomirabilis oxyfera* es la única de su filo, seguramente debido a que la metano-monooxigenasa no estaba presente en sus ancestros, sino que pasó a la especie en concreto por un proceso de transferencia horizontal.

Según Khadka *et al.* (2019), que realizaron un análisis de los genes tanto de las metano monooxigenasas como de las amonio monooxigenasas, proponen que los genes del operón *amo* (este operón contiene tres genes con los mismos nombres que *pmoCAB* y son altamente parecidos a los *pmo*) se transfirieron horizontalmente desde un ancestro del género *Nitrosococcus*, perteneciente a las gammaproteobacterias, que desarrolló capacidad metanotrófica a partir de sus genes *amo*, al ancestro de las actuales alphaproteobacterias metanotróficas y después el operón se transfirió al ancestro de las gammaproteobacterias. Por otro lado, Kang *et al.*, (2019) defienden que puesto que para el correcto funcionamiento de esta enzima se necesita cobre es necesario que las bacterias no solo reciban los genes sino también un mecanismo de captación de cobre. En su artículo discuten acerca de la existencia previa de genes de captación de cobre en algún ancestro de las methylococcales, familia a la que pertenecen todas las gammaproteobacterias metanotróficas que habrían obtenido el operón *pmoCAB* por un fenómeno de transferencia horizontal. Destaca el filo Verrucomicrobiota debido a que este no posee ningún mecanismo de captación de cobre, parece ser que sus genes *pmoCAB* también provienen de un fenómeno de transferencia horizontal y no necesitan mecanismos de captación de cobre, esta capacidad podría estar asociada a la abundancia de metales existente en los medios en los que habitan.

En el trabajo de Knief (2015), se realizan filogenias basándose solo en el gen *pmoA*, en sus árboles uede verse como varias de las bacterias que se han analizado aquí aparecen en una posición similar. Así *Methylomirabilis oxyfera* aparece más relacionada a las Pseudomonadota que con las Verrucomicrobiota o la división en dos ramas, por un lado, las alphaproteobacterias

y por otro las gammaproteobacterias. En el caso del trabajo de Knief (2015) existen muchas más especies debido a que un fragmento del gen *pmoA* se ha amplificado en un gran número de bacterias de muestras ambientales. La amplificación de este fragmento de 510 pb se ha utilizado para detectar la presencia de la actividad metano-monooxigenasa en suelos o charcas contaminadas.

#### 4.4 Conclusión

En conclusión, el análisis de los genes reveló la existencia de secuencias que acumulan un gran número de cambios ya sea a lo largo de todo el operón, como es el caso de *Methylomonas sp.* EFPC3 o en el gen *pmoC* del mismo genoma, como en *Methylocystis parvus*. La existencia de estos cambios se debe a cambios de función de los genes afectados. Al igual que se han visto diferencias también se han observado semejanzas como el patrón de dos copias completas del operón y una copia aislada del gen de *pmoC* presente en todas las especies del segundo grupo de gammaproteobacterias, en *Methylocystis rosea* y en las dos especies de Verrucomicrobiota.

Existen procesos a favor del modelo evolutivo de *birth-and-death* como la existencia de genes *pmoC* individuales que pueden explicarse como copias completas del operón en los que se han perdido los otros dos genes, pero, la existencia de dos operones completos y uno parcial de forma repetida en un gran número de especies parece rechazar esta primera hipótesis y decantarse por el segundo modelo, la evolución concertada que propone que durante el proceso evolutivo todos los genes miembros de una misma familia evolucionan a la vez por conversión génica.

El abundante número de secuencias *pmoC* repetidas lleva a la conclusión de que estas han de poseer alguna función especial ya que por economía del genoma en bacterias se perderían rápidamente y verlas tan conservadas en varios grupos distintos reduce drásticamente la probabilidad de que sean copias de función similar que aún no han desaparecido.

No se han detectado casos evidentes de transferencia horizontal, sin embargo, la presencia de *Methylomirabilis oxyfera* como único representante de su filo permite concluir la existencia con una alta probabilidad de que el gen haya llegado a dicha especie por un proceso de HGT. También, la presencia de genes *pmoCAB* en algunos grupos concretos de dos de las clases de Pseudomonadota y un grupo de Verrucomicrobiota apoyan los procesos de HGT.

## **Referencias**

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., *et al.* (1990) “Basic local alignment search tool”, *J. Mol. Biol.*, 215(3), pp. 403-10. doi: 10.1016/S0022-2836(05)80360-2.
- Awala, S. I., Gwak, JH., Kim, YM., Kim, SJ., *et al.* (2021) “Verrucomicrobial methanotrophs grow on diverse C3 compounds and use a homolog of particulate methane monooxygenase to oxidize acetone”, *ISME J.*, 15(12), pp. 3636-3647. doi: 10.1038/s41396-021-01037-2.
- Edgar, R. C. (2004). “MUSCLE: multiple sequence alignment with high accuracy and high throughput”, *Nucleic Acids Research*, 32(5), pp. 1792-97. doi: 10.1093/nar/gkh340.
- Espejo R. T. y Plaza N. (2018) “Multiple Ribosomal RNA Operons in Bacteria; Their Concerted Evolution and Potential Consequences on the Rate of Evolution of Their 16S rRNA”, *Frontiers in Microbiology*, 9: 1232. doi: 10.3389/fmicb.2018.01232.
- Felsenstein, J., (1985) “Confidence limits on phylogenies: An approach using the Bootstrap”, *Evolution*, 39(4), pp. 783-791. doi: 10.1111/j.1558-5646.1985.tb00420.x.
- Gilbert, B., McDonald, IR., Finch, R., Stafford, GP., *et al.* (2000) “Molecular analysis of the pmo (particulate methane monooxygenase) operons from two type II methanotrophs”, *Appl Environ Microbiol*, 66(3), pp. 966-75. doi: 10.1128/AEM.66.3.966-975.2000.
- Higgins D., Thompson J., Gibson T., Thompson J. D., *et al.* (1994) “CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice”, *Nucleic Acids Res.*, 22(22), pp. 4673-80. doi: 10.1093/nar/22.22.4673.
- Kang, C. S., Dunfield, P. F. y Semrau, J. D. (2019) “The origin of aerobic methanotrophy within the Proteobacteria”, *FEMS Microbiology Letters*, 366(9): fnz096. doi: 10.1093/femsle/fnz096.
- Khadka, R., Clothier, L., Wang L., Lim, CK., *et al.* (2018) “Evolutionary History of Copper Membrane Monooxygenases”, *Front. Microbiol.*, 9: 2493. doi: 10.3389/fmicb.2018.02493.
- Khider, M. L. K., Brautaset, T. e Irla, M. (2021) “Methane monooxygenases: central enzymes in methanotrophy with promising biotechnological applications”, *World J Microbiol Biotechnol*, 37(4), pp. 72. doi: 10.1007/s11274-021-03038-x.
- Knief, C., (2015) “Diversity and Habitat Preferences of Cultivated and Uncultivated Aerobic Methanotrophic Bacteria Evaluated Based on *pmoA* as Molecular Marker”, *Frontiers in Microbiology*, 6: 1346. doi: 10.3389/fmicb.2015.01346.
- Koo, CW. y Rosenzweig AC. (2021) “Biochemistry of aerobic biological methane oxidation”, *Chem Soc Rev*, 50(5), pp. 3424-3436. doi: 10.1039/d0cs01291b.
- Koonin, E. V. (2005) “Orthologs, Paralogs, and Evolutionary Genomics”, *Annual Review of Genetics*, 39(1), pp. 309-338. doi: 10.1146/annurev.genet.39.073003.114725.
- Kumar, S., Stecher, G. y Tamura, K. (2015) “MEGA7: Molecular Evolutionary Genetics Analysis version 7.0.”, *Molecular Biology and Evolution*. doi: 10.1093/molbev/msw054.
- Lieberman, R. L., y Rosenzweig, A. C. (2005). “Crystal structure of a membrane-bound metalloenzyme that catalyses the biological oxidation of methane”, *Nature*, 434(7030), pp. 177–182. doi :10.1038/nature03311.
- Nei, M. y Rooney, AP., (2005) “Concerted and birth-and-death evolution of multigene families”, *Annu Rev Genet* 39, pp. 121-52. doi: 10.1146/annurev.genet.39.073003.112240.
- Olsen, G. J. y Woese, C. R. (1993) “Ribosomal RNA: a key to phylogeny”, *The FASEB Journal*, 7(1), pp. 113-23. doi:10.1096/fasebj.7.1.8422957.

- Zhou, Q., Xu, X., Xu, H., Zheng, G., *et al.* (2020) "Surface microbial geochemistry of the Beihanzhuang Oilfield, northern Jiangsu, China" *Journal of Petroleum Science and Engineering*, 191: 107140. doi: 10.1016/j.petrol.2020.107140.
- Rodicio, M. R. y Mendoza, M. C. (2004) "Identificación bacteriana mediante secuenciación del ARNr 16S: fundamento, metodología y aplicaciones en microbiología clínica", *Enfermedades Infecciosas y Microbiología Clínica*, 22(4), pp. 238-245.
- Saitou, N. y Nei, M., (1987) "The neighbor-joining method: A new method for reconstructing phylogenetic trees", *Molecular Biology and Evolution* 4(4), pp. 406-25. doi: 10.1093/oxfordjournals.molbev.a040454.
- Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., *et al.* (2022) "Database resources of the national center for biotechnology information", *Nucleic Acids Research*, 50(D1), pp. D20-D26. doi: 10.1093/nar/gkab1112.
- Soucy, S., Huang, J. y Gogarten, J. (2015) "Horizontal gene transfer: building the web of life", *Nat Rev Genet*, 16(8), pp. 472-82. doi: 10.1038/nrg396.
- Tamura, K. y Nei, M., (1993) "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees", *Molecular Biology and Evolution*, 10(3), pp. 512-26. doi: 10.1093/oxfordjournals.molbev.a040023.
- Tamura, K., Nei, M. y Kumar, S., (2004) "Prospects for inferring very large phylogenies by using the neighbor-joining method", *Proceedings of the National Academy of Sciences*, 101(30), pp. 11030-11035. doi: 10.1073/pnas.0404206101.