Universitat Oberta de Catalunya

uoc.edu

Estudio de genómica comparada de genes bacterianos relacionados con la ruta catabólica inferior del naftaleno: metabolismo del salicilato

Irene Galli Medina
Máster en Bioinformática y Bioestadística
Microbiología, biotecnología y biología molecular

Paloma Pizarro Tobías
Antoni Pérez Navarro
Fecha de entrega: 06/2021

Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada 3.0 España de Creative Commons

FICHA DEL TRABAJO FINAL

Título del trabajo:	Estudio de genómica comparada de genes bacterianos relacionados con la ruta catabólica inferior del naftaleno: metabolismo del salicilato
Nombre del autor:	*Irene Galli Medina*
Nombre del consultor/a:	*Paloma Pizarro Tobías*
Nombre del PRA:	*Antoni Pérez Navarro*
Fecha de entrega (mm/aaaa):	06/2021
Titulación:	*Máster de Bioinformática y Bioestadística*
Área del Trabajo Final:	*Área 4, subárea 6. Microbiología, biotecnología y biología molecular.*
Idioma del trabajo:	Castellano
Número de créditos:	15 créditos
Palabras clave	*Biorremediación, catecol, gentisato*

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.*

El aumento de la presencia de xenobióticos en el medio ambiente ha incrementado el interés por estudiar la degradación de estos compuestos, incluyendo la biorremediación. Dentro de estos xenobióticos se encuentra el grupo de los hidrocarburos aromáticos policíclicos, donde destaca el naftaleno por ser el compuesto más simple y, por ende, más estudiado. La degradación del naftaleno por bacterias se lleva a cabo en dos fases, aunque en este caso, el estudio se centrará en la vía inferior que se puede realizar mediante dos vías: la vía del catecol y la vía del ácido gentísico. En este trabajo se propone un estudio de genómica comparativa de genes seleccionados de la vía del catecol (nahG, nahH y nahI), usando a *Pseudomonas putida* G7 como modelo, y de la vía del gentisato (nagG, nagH, nagI y nagAa) a partir de *Ralstonia* sp. U2. Como resultado se obtiene un árbol filogenético para cada gen mencionado y, además, la predicción y comparación de algunos operones seleccionados. Se ha observado que en las bases de datos para la ruta del catecol destacan cepas pertenecientes a la clase *Gammaproteobacteria*, mientras que para la ruta del ácido gentísico destaca la clase *Betaproteobacteria*. Además, se ha observado una alta identidad de *Cupriavidus necator* U2 con los genes seleccionados de

Ralstonia sp. U2 (a excepción de nagI) y una alta identidad de *Pseudomonas sp.* MC1 y *Pseudomonas umsongensis* GO16 con los genes y el operón sal de *Pseudomonas putida* G7.

Abstract (in English, 250 words or less):

The increased presence of xenobiotics in the environment has increased interest in studying the degradation of these compounds, including bioremediation. Within these xenobiotics, the group of polycyclic aromatic hydrocarbons should be highlighted, where naphthalene stands out for being the simplest compound and, therefore, the most studied. The degradation of naphthalene by bacteria is carried out in two phases, although in this case, the study will focus on the lower pathway that can be carried out by two pathways: the catechol pathway and the gentisate pathway. In this work we propose a comparative genomics study of selected genes of the catechol pathway (nahG, nahH and nahI), using *Pseudomonas putida* G7 as a model, and of the gentisate pathway (nagG, nagH, nagI and nagAa) from *Ralstonia* sp. U2. As a result, we obtain a phylogenetic tree for each gene mentioned and, in addition, the prediction and comparison of some selected operons. It has been observed that in the databases for the catechol route, the majority of strains belong to the *Gammaproteobacteria* class while for the gentisic acid route the *Betaproteobacteria* class stands out. Furthermore, a high identity of *Cupriavidus necator* U2 with the selected genes of *Ralstonia* sp. U2 has been observed (except for nagI) and a high identity of *Pseudomonas* sp. MC1 and *Pseudomonas umsongensis* GO16 with the genes and the sal operon of *Pseudomonas putida* G7.

Índice

1. Introducción……..….….2
   1.1. Contexto y justificación del Trabajo…….....2
   1.2. Objetivos del trabajo……..5
   1.3. Enfoque y método seguido……6
   1.4. Planificación del Trabajo……7
   1.5. Breve sumario de productos obtenidos……...….10
   1.6. Breve descripción de los otros capítulos de la memoria……..…10
2. Metodología……..11
3. Resultados…….,…15
4. Conclusiones……29
   4.1. Conclusiones……29
   4.2 Futuras líneas de estudio…….….…30
   4.3 Seguimiento de la planificación……..….30
5. Glosario…….….31
6. Bibliografía……31

Lista de figuras

Figura 1. Descripción del metabolismo del naftaleno……..…3
Figura 2. Genes localizados en el operón *sal* del plásmido NAH7 de *Pseudomonas putida* G7…….……..4
Figura 3. Genes involucrados en la vía del gentisato en *Ralstonia* sp. U2……5
Figura 4. Organización del mes de marzo y abril……..…9
Figura 5. Organización del mes de mayo e inicios de junio……...9
Figura 6. Reacción catalizada por la enzima salicilato 5-hidroxilasa (S5H)……11
Figura 7. Similitud entre la vía superior de la degradación del naftaleno y la ruta del gentisato……..12
Figura 8. Árbol filogénico basado en la enzima nagG……16
Figura 9. Árbol filogénico basado en la enzima nagH……18
Figura 10. Árbol filogenético basado en la enzima nagAa……19
Figura 11. Árbol filogenético basado en la enzima nagI……21
Figura 12. Comparación de operones putativos de *Cupriavidus necator* U2.…22
Figura 13. Árbol filogenético basado en la enzima nahG……...24
Figura 14. Árbol filogenético basado en la enzima nahH……..25
Figura 15. Árbol filogenético basado en la enzima nahI……26
Figura 16. Predicciones de operones de la ruta del catecol……28

Lista de tablas

Tabla 1. Cepas incluidas en la predicción de operones…….27

1 Introducción

1.1 Contexto y justificación del Trabajo

El aumento de las actividades antropogénicas, como la industrialización, la combustión de combustibles fósiles y el uso de fumigantes, implica una mayor liberación de compuestos contaminantes al medio ambiente que pueden perdurar muchos años sin ser degradados y conllevan un peligro para la salud de los seres vivos[1]. Dentro de estos compuestos debemos destacar el grupo de los hidrocarburos aromáticos policíclicos (HAPs), donde se encuentra el naftaleno, que ha sido ampliamente estudiado debido a su simplicidad molecular[2]. El naftaleno se encuentra ubicuo en el ambiente a causa de su alta volatilidad y se utiliza para la fabricación de pesticidas y plásticos[3]. La exposición humana aguda a este producto se relaciona con anemia hemolítica y, en grandes dosis, puede llegar a causar edema cerebral y fallo renal[4,5]. Adicionalmente, las últimas dos décadas se ha debatido sobre el poder carcinogénico del naftaleno y la Agencia de Protección Ambiental de los Estados Unidos (EPA) lo ha clasificado como Grupo C, posible carcinógeno humano[1,6].

La eliminación de los xenobióticos del medio ambiente por métodos tradicionales, como la degradación química o la volatilización, no suelen eliminar el problema al completo y pueden causar más daño ambiental. Por ello, la biorremediación se suele considerar un método más eficaz para eliminar los compuestos contaminantes[7,8]. Este proceso se lleva a cabo por microorganismos que son capaces de utilizar compuestos orgánicos contaminantes como fuente de carbono y energía, obteniendo productos inocuos como CO_2 y H_2O[9]. En este ámbito, se ha utilizado el naftaleno como modelo de estudio para analizar el metabolismo y la degradación de los HAPs, debido a que su estructura molecular simple, formada únicamente por dos anillos de benceno, facilita su estudio.

El metabolismo del naftaleno se puede dividir en dos fases, la vía superior y la vía inferior (**Figura 1**). La vía superior se caracteriza por la conversión del naftaleno al salicilato, por medio de varias enzimas. A continuación, este salicilato entrará en la vía inferior del metabolismo, que se divide en dos rutas, la ruta del catecol y la ruta del ácido gentísico. Además, la ruptura del catecol se puede realizar en la posición *meta*, obteniendo acetaldehído y piruvato como producto final, o en la posición *orto*, obteniendo succinato y acetil-CoA[10].

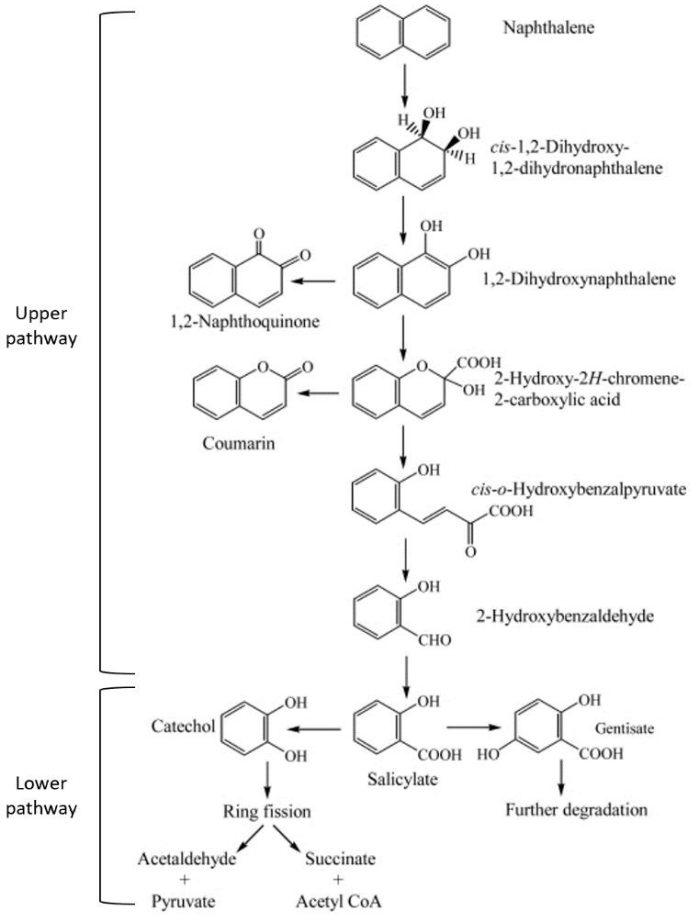

Figura 1. Descripción del metabolismo del naftaleno, indicando la ruta superior y la ruta inferior[71].

La mayoría de las bacterias degradadoras de naftaleno que se conocen hasta el momento pertenecen al grupo de las Gram-negativas, siendo el género *Pseudomonas* el más estudiado y caracterizado hasta el momento[11–15]. Además, en estas bacterias los genes catabólicos suelen encontrarse en operones localizados en plásmidos móviles, que pueden transferirse a otras cepas y favorecer la evolución y adaptación al medio[15,16].

Las cepas más estudiadas dentro del género *Pseudomonas* son *Pseudomonas putida* G7 (plásmido NAH7) y *Pseudomonas putida* NCIB 9816-4 (plásmido pDGT1)[17,18]. En ambas cepas, los genes involucrados en la degradación del naftaleno se encuentran localizados en dos operones distintos: el operón *nah* (o *nah1*) y el operón *sal* (o *nah2*). El operón *nah* contiene los genes involucrados en la fase superior de la degradación del naftaleno (nahAa a nahD), en la que no se entrará en detalle en este trabajo. Por otro lado, el operón *sal*, incluye los genes que permiten la oxidación del salicilato a catecol, y la meta-escisión de este para obtener intermediarios del ciclo del ácido tricarboxilico[19,20]. Dentro de este operón encontramos una serie de genes (nahG a nahO), pero el más

estudiado y destacado es el gen nahG ya que codifica la enzima salicilato hidroxilasa que permite el paso de salicilato a catecol (**Figura 2**)[21].

A)

	Gen	Enzima	
operón *sal* (NAH 7)	nahG	*salicylate hydroxylase*	→ Salicilato a catecol
	nahT	*chloroplast-type ferredoxin*	Catecol a piruvato + acetaldehído
	nahH	*catechol 2,3-dioxygenase*	
	nahI	*2-hydroxymuconic semialdehyde dehydrogenase*	
	nahN	*2-hydroxymuconic semialdehyde hydrolase*	
	nahL	*2-exopent-4-enoate hydratase*	
	nahM	*acetaldehyde dehydrogenase*	
	nahO	*4-hydroxy-2-oxovalerate*	
	nahK	*4-oxolocrotonate decarboxylase*	
	nahJ	*4-oxalocrotonate tautomerase*	

B)

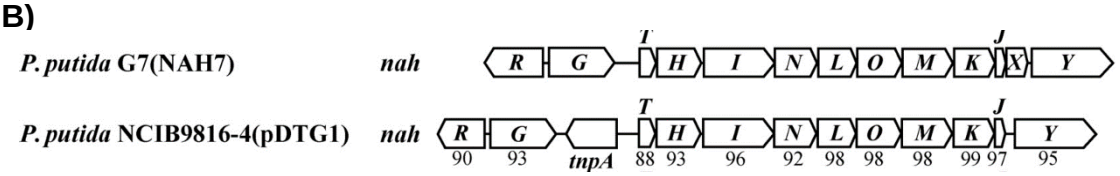

Figura 2. Genes localizados en el operón *sal* del plásmido NAH7 de *Pseudomonas putida* G7. (A) Tabla con el nombre de los genes y las enzimas involucradas en la degradación del salicilato por la vía del catecol[72]. (B) Representación gráfica del operón *sal* de *P. putida* G7 y *P. putida* NCIB 9816-4[73].

Por otro lado, y como se ha mencionado anteriormente, también existe la posibilidad de degradar el salicilato por la vía del ácido gentísico. En este ámbito se incluyen varias bacterias Gram-positivas, como las cepas del género *Rhodococcus*, aunque *Ralstonia* sp. U2 (originalmente clasificada como *Pseudomonas* sp.) destaca por ser la primera bacteria caracterizada que lleva a cabo la vía del gentisato para degradar el salicilato (**Figura 1**)[14,22]. En este caso, destacan los genes nagG y nagH, que codifican, junto a nagAa y nagAb, el complejo multienzimático salicilato 5-hidroxilasa que permite el paso de salicilato a gentisato (**Figura 3**)[22].

A)

Vía gentisato en *Ralstonia* sp. U2	Gen	Función
	nagG	*Salicylate-5-hydroxylase (large oxygenase component)*
	nagH	*Salicylate-5-hydroxylase (small oxygenase component)*
	nagAa	*Ferredoxin reductase*
	nagAb	*Ferredoxin*
	nagI	*Gentisate* 1,2-*dioxygenase*
	nagL	*Maleylpyruvate isomerase*
	nagK	*Fumarylpyruvate hydrolase*

B)

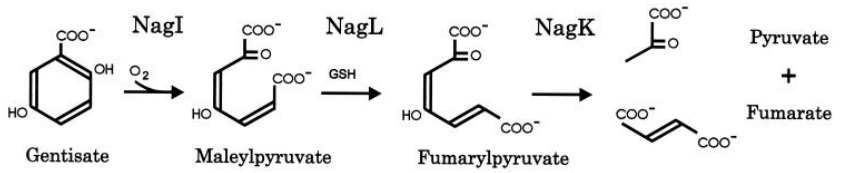

Figura 3. Genes involucrados en la vía del gentisato en *Ralstonia* sp. U2. (A) Tabla resumiendo los genes implicados y su función[72]. (B) Representación gráfica de la degradación del gentisato hasta piruvato y fumarato, indicando las enzimas que catalizan cada paso[22].

Las vías metabólicas de las bacterias y otros procariotas han evolucionado y siguen evolucionando mediante modificaciones y reordenamientos de su ADN. Estas modificaciones pueden ir desde un cambio de pequeño de pocos nucleótidos a variaciones más extensas producidas por procesos como deleciones, duplicaciones e inserciones[23,24]. Al obtener más información sobre la organización sobre los genes involucrados en un cierto metabolismo, podemos tener la oportunidad de comprender mejor la evolución de los genes, operones y clústeres catabólicos. Por ello, con este trabajo se pretende ampliar el conocimiento sobre la homología y las modificaciones de los genes relacionados con el metabolismo del salicilato, diferenciando la ruta del catecol y la ruta del ácido gentísico. De esta manera, se podría obtener más información sobre la relación evolutiva de este conjunto de genes.

1.2 Objetivos del Trabajo

Objetivos generales:

1. Analizar el metabolismo del salicilato (ruta inferior de la degradación del naftaleno) y las diferencias entre la ruta del catecol y la ruta del ácido gentísico.
2. Realizar un estudio de genómica comparativa con los genes relacionados con la ruta del catecol y la ruta del ácido gentísico.

Objetivos específicos:

1. Detallar las enzimas que se encuentran involucradas en el metabolismo del salicilato. En primer lugar, se tendrá en cuenta la ruta del ácido gentísico y posteriormente la ruta del catecol.
2. Obtener las secuencias de las enzimas destacadas anteriormente (a partir de organismos modelo).
3. Obtener las secuencias homólogas para cada enzima mediante herramientas bioinformáticas como BLAST.
4. Realizar una comparación mediante alineamientos múltiples de las secuencias homólogas de cada enzima, empleando el software MEGA.
5. Estudiar la relación filogenética para cada enzima.

1.3 Enfoque y método seguido

En primer lugar, se realiza un trabajo con un enfoque teórico que se basa en una búsqueda bibliográfica para obtener información detallada sobre las dos rutas del metabolismo del salicilato, anotando las enzimas que se encuentran involucradas en esta degradación. Además, también se buscarán las cepas bacterianas más estudiadas para utilizarlas como modelo para obtener las secuencias de los genes y sus enzimas. A partir de aquí, se pasará a un trabajo más experimental, en el que se emplearán herramientas bioinformáticas con el objetivo de buscar secuencias homólogas de las enzimas y realizar alineamientos múltiples y arboles filogenéticos.

Se iniciará el trabajo con la ruta del ácido gentísico ya que se ha estudiado en menor medida que la ruta del catecol y resulta interesante profundizar más en este ámbito. Tras finalizar la ruta del ácido gentísico se procederá a analizar la ruta del catecol.

Las herramientas bioinformáticas que se utilizarán para este trabajo se han seleccionado por su popularidad en estudios de genómica comparativa, su disponibilidad gratuita *online* y por ser herramientas que se han introducido en mayor o menor medida durante el transcurso del máster.

Para realizar la búsqueda de las secuencias homólogas se empleará la herramienta informática PSI-BLAST (*Position-Specific Iterative* BLAST). BLAST nos permite obtener una lista de secuencias similares a nuestra secuencia de interés, mediante una base de datos y una serie de alineamientos locales[25]. PSI-BLAST es un método especifico de la herramienta BLAST, que nos permite realizar una serie de iteraciones para detectar relaciones más distantes entre proteínas[26].

Para realizar el alineamiento múltiple de las secuencias homólogas y crear los árboles filogenéticos se empleará MEGA, un *software* centrado en el análisis filogenético y en el análisis estadístico, que es fácil de usar gracias a su interfaz gráfica de usuario[27]. MEGA contiene una herramienta para realizar múltiples alineamientos de secuencias de ADN o de proteínas, utilizando algoritmos de *ClustalW* o *Muscle*.

1.4 Planificación del Trabajo

1.4.1 Tareas

Como se ha mencionado anteriormente, se iniciará el estudio con la ruta del ácido gentísico y posteriormente se continuará con la ruta del catecol. Las tareas a realizar son similares en ambos casos.

<u>Ruta ácido gentísico</u>

1. Búsqueda bibliográfica de los genes y enzimas involucrados en la ruta del ácido gentísico.
2. Establecer un organismo modelo y obtener las secuencias de las enzimas.
3. Obtener las secuencias homólogas de las enzimas de la ruta del ácido gentísico mediante PSI-BLAST y crear la base de datos.
4. Realizar el alineamiento múltiple empleando el *software* MEGA.
5. Obtener los árboles filogenéticos las enzimas de la ruta del ácido gentísico.

<u>Ruta del catecol</u>

1. Búsqueda bibliográfica de los genes y enzimas involucrados en la ruta del catecol
2. Establecer un organismo modelo y obtener las secuencias de las enzimas
3. Obtener las secuencias homólogas de las enzimas de la ruta del catecol mediante PSI-BLAST y crear la base de datos.
4. Realizar el alineamiento múltiple empleando el *software* MEGA
5. Obtener los árboles filogenéticos para las enzimas de la ruta del catecol.

1.4.2 Calendario

PEC 1: Plan de trabajo	Fecha finalización
Búsqueda bibliográfica sobre el metabolismo salicilato	16/03/2021
Definir objetivos del trabajo	16/03/2021
Buscar herramientas informáticas a utilizar	16/03/2021
PEC 2: Fase 1 del desarrollo del trabajo (gentisato)	Fecha finalización
Buscar enzimas de la ruta del ácido gentísico y organismo modelo	18/03/2021
Realizar la búsqueda de secuencias homólogas para las enzimas	31/03/2021
Alineamientos múltiples	07/04/2021
Creación de árboles filogénicos	19/04/2021
PEC 3: Fase 2 del desarrollo del trabajo (catecol)	Fecha finalización
Buscar enzimas de la ruta del catecol y organismo modelo	21/04/2021
Realizar la búsqueda de secuencias homólogas para las enzimas	04/05/2021
Alineamientos múltiples	10/05/2021
Creación de árboles filogénicos	17/05/2021
PEC 4: Cierre de la memoria	Fecha finalización
Corrección de posibles errores o falta de información	25/05/2021
Finalizar memoria escrita	08/06/2021
PEC 5a: Elaboración de la presentación	Fecha finalización
Elaboración de las diapositivas a presentar	11/06/2021
Practicar presentación	13/06/2021
PEC 5b: Defensa pública	Fecha finalización
	23/06/2021

Tabla 1. Calendario de las actividades a realizar

En la **Figura 4** y la **Figura 5** se muestra un diagrama de Gantt para favorecer la visualización de los meses de trabajo. En rojo se marcan los hitos que corresponden a las entregas de las diferentes PECs.

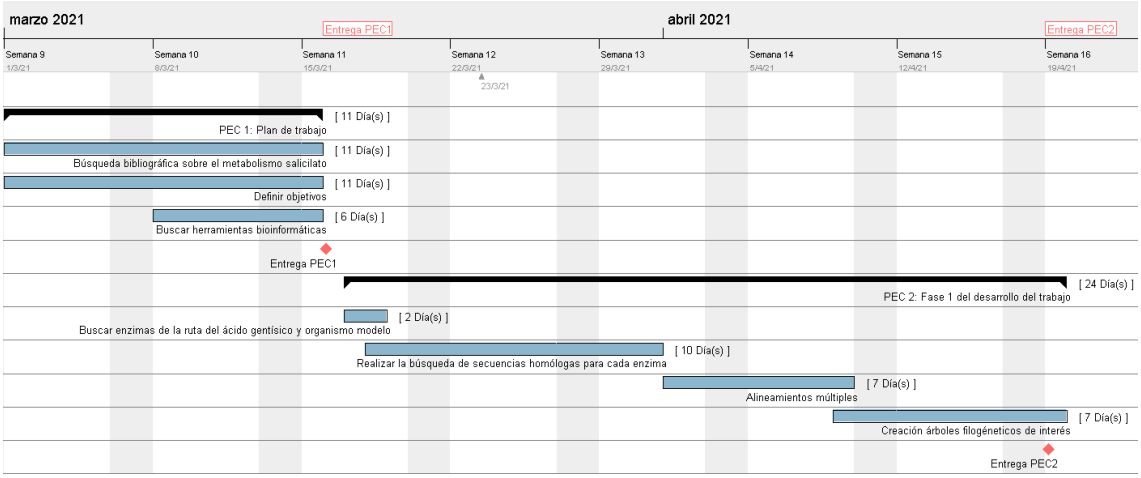

Figura 4. Organización del mes de marzo y abril

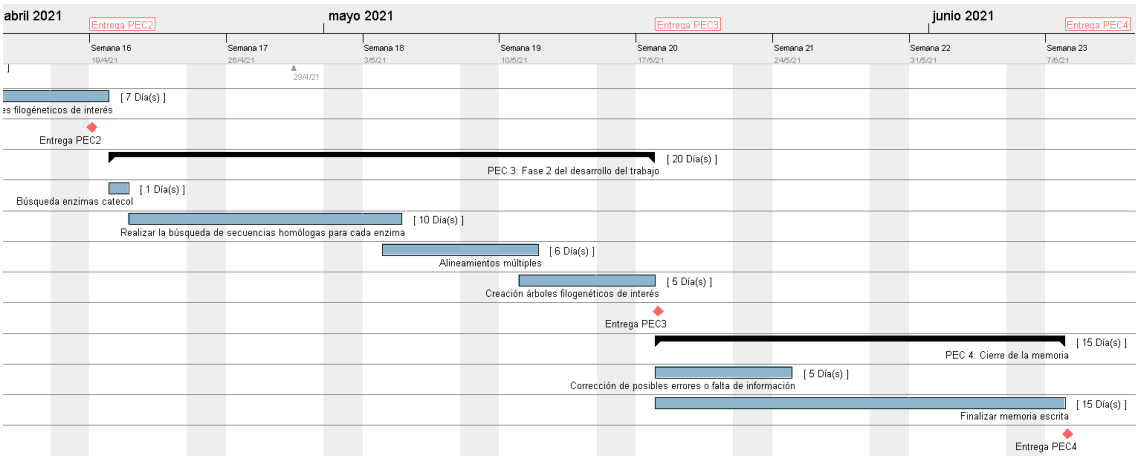

Figura 5. Organización del mes de mayo e inicios de junio.

1.4.3 Análisis de riesgos

Existen varias complicaciones que se podrían presentar a lo largo de la realización del trabajo. El principal problema podría ser el tiempo limitado para realizar las comparaciones de las secuencias descritas para cada enzima. En ese caso, se trataría de realizar un corte justificado o reducir el número de enzimas analizadas.

Por otro lado, también se podrían presentar complicaciones como la mala elección de microorganismos modelo o la falta de práctica con las herramientas a utilizar. Además, podrían surgir nuevas dudas o intereses durante el trabajo que involucren el uso de otras herramientas bioinformáticas.

1.5 Breve sumario de contribuciones y productos obtenidos

Mediante este trabajo se obtendrá:

1. Un plan de trabajo en el que se especifique los objetivos iniciales del trabajo y la planificación temporal de las tareas a realizar. Además, también se tendrán en cuenta los posibles riesgos y complicaciones que puedan surgir durante el progreso del trabajo.
2. Una memoria donde se detallará el trabajo realizado y los resultados obtenidos.
3. Se realizará una presentación que será explicada oralmente (de manera virtual) para explicar de manera resumida el desarrollo del trabajo y los resultados obtenidos

1.6 Breve descripción de los otros capítulos de la memoria

El Trabajo Fin de Máster se encontrará dividido en una serie de apartados, en los que se incluyen:

1. Introducción: sección inicial donde se contextualiza el tema a tratar durante el trabajo.
2. Materiales y métodos: descripción del proceso llevado a cabo y las herramientas bioinformáticas utilizadas.
3. Resultados: presentación de los productos obtenidos mediante las herramientas bioinformáticas mencionadas y discusión.
4. Conclusiones: ideas extraídas tras finalizar el trabajo.
5. Bibliografía: Contiene todas las fuentes utilizadas durante el trabajo, incluyendo artículos científicos, libros y páginas web.
6. Anexo: Se incluye toda la información extensa que rompería la presentación ordenada del trabajo y los materiales complementarios

2 Metodología

Búsqueda bibliográfica ácido gentísico: cepa modelo y selección de enzimas

En primer lugar, el trabajo presenta un enfoque teórico en el que se lleva a cabo una búsqueda bibliográfica con la finalidad de seleccionar una cepa modelo para estudiar la ruta del ácido gentísico y las enzimas involucradas. Se ha seleccionado la cepa *Ralstonia* sp. U2 como modelo al ser una de las más estudiadas en este ámbito, junto a *Polaromonas naphthalenivorans* CJ2[22,23,28].

Anteriormente se han mencionado las enzimas descritas en *Ralstonia* sp. U2 que actúan en la vía del gentisato (**Figura 3**). Del total de enzimas se han seleccionado 4 para el estudio de genómica comparativa: nagG, nagH, nagAa y nagI.

- nagG, nagH y nagAa: forman parte de la enzima multicomponente salicilato 5 - hidroxilasa (EC 1.14.13.172)[29], que cataliza el primer paso de la ruta del ácido gentísico, la hidroxilación del salicilato a gentisato (**Figura 6)**[30]**.**

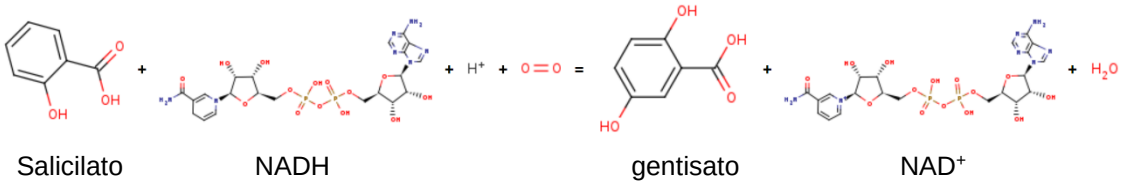

Figura 6. Reacción catalizada por la enzima salicilato 5-hidroxilasa (S5H)[30].

La hidroxilación del salicilato a gentisato se realiza gracias a nagGH, que actúa como monooxigenasa, asociada a una cadena de transporte de electrones que consiste en una ferredoxina (nagAb) y una ferredoxina reductasa (nagAa)[28].

Para el presente trabajo, además de nagG y nagH, se ha decidido estudiar nagAa al haberse establecido que existe relación con la enzima multicomponente naftaleno 1,2-dioxigenasa (NDO), involucrada en el primer paso de la ruta superior de la degradación del naftaleno (**Figura 7**). Existen evidencias de que la salicilato 5-hidroxilasa y la naftaleno 1,2-dioxigenasa comparten la cadena de electrones, por lo que tanto nagAb como nagAa pertenecerían a la ruta del ácido gentísico y a la ruta superior de la degradación del naftaleno[28].

- **nagI:** gentisato 1,2-dioxigenasa que cataliza la reacción la oxidación del gentisato a maleylpiruvato (EC 1.13.11.4)[31].

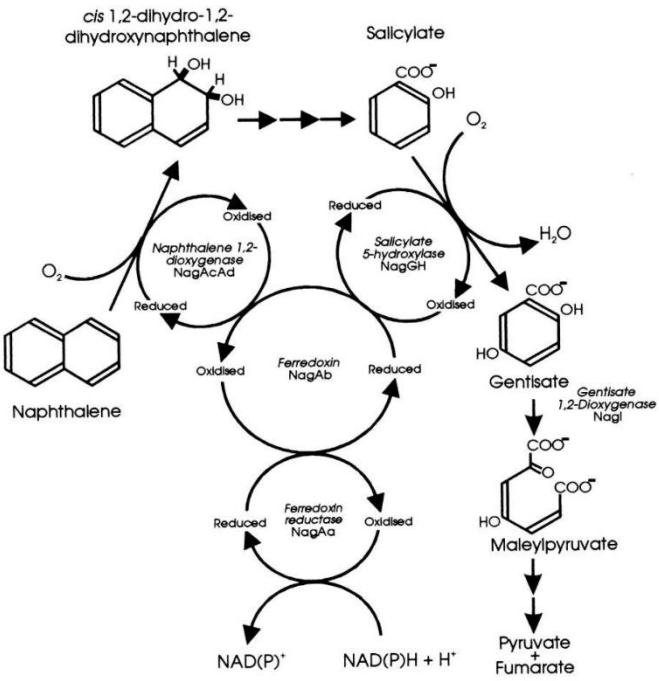

Figura 7. Conexión entre la reacción de la salicilato 5-hidroxilasa y la naftaleno 1,2-dioxigenasa, al compartir a nagAa y nagAb[28].

Búsqueda bibliográfica ruta catecol: cepa modelo y selección de enzimas

Para el caso de la ruta del catecol, se escogió la cepa *Pseudomonas putida* G7, específicamente el plásmido NAH7, como modelo para obtener las secuencias potencialmente homólogas. Se ha seleccionado esta cepa debido a que ha servido de modelo para muchos estudios publicados actualmente y el operón se encuentra bien caracterizado[32–34]. Al haber seleccionado a *P. putida* G7 como modelo, el estudio se centrará en la ruptura del catecol en la posición *meta*.

Anteriormente se han mencionado las enzimas descritas en *Pseudomonas putida* G7 que actúan en la vía del catecol (**Figura 3**). Del total de enzimas se han seleccionado 3 para el estudio de genómica comparativa: nahG, nahH y nahI.

- nahG (salicilato hidroxilasa): se ha seleccionado ya que realiza la primera reacción de la degradación del ácido salicílico, el paso a catecol mediante una hidroxilación (EC:1.14.13.1)[30].
- nahH (catecol 2,3-dioxigenasa): cataliza la oxidación del catecol a 2-hidroximuconato semialdehído, el siguiente paso tras la actividad de nahG (EC:1.13.11.2)[31].

- nahI (2-hidroximucónico semialdehído deshidrogenasa): convierte el 2-hidroximucónico semialdehído a 2-hidroxipenta-2,3-dienoato.

Búsqueda de secuencias homólogas y creación de la base de datos.

Mediante PSI-BLAST se realizó la búsqueda de las secuencias homólogas para las 4 enzimas seleccionadas de *Ralstonia* sp. U2, utilizando los números de acceso AAD12607.1 (nagG), AAD12608.1 (nagH), AAD12610.1 (nagAa) y AAD12619.1 (nagI)[35]. Por otro lado, en relación a la ruta del catecol, se realizó la búsqueda de las secuencias homólogas para las 3 enzimas de *Pseudomonas putida* G7 empleando los números de acceso BAE92165.1 (nahG), BAE92167.1 (nahH) y BAE92168.1 (nahI)[33].

Para todas las enzimas mencionadas se realizaron 4 iteraciones en PSI-BLAST y se realizó una filtración indicando un mínimo del 60% de identidad y un mínimo de 97% de *Query Cover*. En cuanto al valor E, se indicó un límite de 10^{-60} para todas las enzimas a excepción de nagG, ya que en la primera búsqueda se obtuvo gran cantidad de secuencias no relacionadas con la ruta de degradación del naftaleno, por lo que se decidió disminuir el valor E a 10^{-200}.

Tras obtener las bases de datos, también se realiza una búsqueda de las secuencias FASTA empleando la base de datos de *National Center for Biotechnology Information* (NCBI)[36].

Alineamientos múltiples y creación de árboles filogenéticos

Las secuencias FASTA obtenidas anteriormente se someten al análisis mediante el *software* MEGA X. Se realiza un alineamiento múltiple para cada enzima, utilizando el método MUSCLE, ya que se ha determinado que proporciona mejor precisión y velocidad que el método CLUSTALW que se ha utilizado de manera tradicional[37].

Tras obtener los alineamientos múltiples para las enzimas mencionadas, se procede a realizar los árboles filogenéticos con el mismo *software* MEGA X[38]. Para ello existen diferentes métodos, incluyendo *Maximum Likelihood Tree*, *Neighbor-Joining Tree* y *Minimum-Evolution Tree*. En este caso, se trabajará con el método *Maximum Likelihood Tree* para todos los árboles filogenéticos, ya que es el método filogenético más comúnmente utilizado en trabajos de investigación[39,40]. Una de las desventajas del *Maximum Likelihood tree* es que puede llegar a ser lento y computacionalmente costoso, pero al disponer de bases de datos de aproximadamente 40 secuencias, esto no supondría un gran problema.

Por otro lado, existen varios enfoques para estimar la confianza de los árboles filogenéticos, pero en este caso se selecciona el método *Bootstrap* con 500 replicaciones[41,42]. Lo óptimo en este caso sería realizar 1.000 replicaciones para obtener una mayor precisión, pero debido al tiempo limitado y a las capacidades del ordenador, se seleccionan 500 replicaciones.

Predicción de operones y comparación

Tras analizar los árboles filogenéticos resultantes, se llevó a cabo una predicción y comparación de operones que no estaba prevista desde un primer momento. Por un lado, relacionado con la ruta del ácido gentísico, se llevó a cabo una predicción de operones de *Cupriavidus necator* U2 y posteriormente se realizó una comparación con el operón nag de *Ralstonia* sp. U2, localizado en el plásmido pWWU2. Por otro lado, para las cepas de la ruta del catecol, se realizó una predicción de operones de las cepas más cercanas a *Pseudomonas putida* G7, incluyendo las siguientes cepas:

- *Pseudomonas putida* NCIB 9816-4 (plásmido pDTG1)
- *Pseudomonas putida* MC1 (plásmido pYIC),
- *Pseudomonas umsongensis* GO16 (plásmido pENK22),
- *Pseudomonas putida* ND6 (plásmido pND6-1).
- *Pseudomonas fluorescens* PC20 (plásmido IncP-9)

Los operones son una característica muy estudiada de los genomas de organismos procariotas, al permitir la regulación y la transcripción conjunta de genes relacionados funcionalmente. El análisis y la comparación entre operones puede aportar información adicional a los árboles filogenéticos y ayudar a la detección de posibles transferencias genéticas horizontales[43].

En primer lugar, se realizó una búsqueda de las secuencias FASTA del genoma completo de las cepas de interés y se empleó la herramienta FGENESB de *Softberry* para realizar una predicción de los operones. FGENESB es el programa más preciso para realizar la predicción de genes y operones *ab initio*, utilizando un algoritmo de predicción basado en modelos de cadenas de Markov de regiones codificantes y zonas de terminación[44,45]. Para la predicción de los operones de *Cupriavidus necator* U2 se seleccionó la categoría “BACTERIAL generic” como organismo más cercano, mientras que para las predicciones de las cepas de *Pseudomonas* se seleccionó *Pseudomonas putida* KT2440 como organismo más cercano. Al introducir las secuencias FASTA del genoma completo se obtiene una lista de los genes donde se indica si pertenecen a una unidad transcripcional (Tu) o a un operón (Op). Tras obtener esta lista se realiza la búsqueda del operón de interés, gracias a las coordenadas de los genes involucrados.

Al haber obtenido la predicción de operones para las cepas de interés, se visualizaron con la herramienta SnapGene, que contiene una interfaz intuitiva para la visualización de secuencias de ADN y aminoácidos[46]. En este caso concreto, se utilizó SnapGene para extraer las secuencias GenBank de los operones de interés. Finalmente, se empleó la herramienta *Easyfig* 2.2.5 para incluir las secuencias GenBank de las predicciones de operones y realizar las comparaciones de operones. *Easyfig* 2.2.5 es una aplicación de Python que

permite obtener figuras de comparación de múltiples regiones genómicas[47]. Al introducir los archivos de anotación Genbank, *Easyfig* nos permite realizar un BLAST entre las secuencias y posteriormente obtener la figura comparativa.

3 Resultados

Secuencias homólogas

Tras finalizar la búsqueda y filtración de las secuencias homólogas para la ruta del ácido gentísico, se obtienen un total de 45 secuencias para nagG, 40 para nagH, 38 para nagAa y 44 para nagI (**Anexo 1**).

Para el caso de nagI, se añadieron secuencias adicionales de la enzima gentisato 1,2-dioxigenasa del género de *Rhodococcus* a la base de datos, que no se obtuvieron como resultado del PSI-BLAST. Se incluyeron estas secuencias adicionales debido a que el género *Rhodococcus* está siendo estudiado por su capacidad de degradación del naftaleno por la vía del gentisato y por la habilidad de adaptarse a ambientes contaminados[3,48].

En el caso de las enzimas de la ruta del catecol, se obtienen un total de 38 secuencias homólogas para nahG, 40 para nahH y 40 para nahI (**Anexo 2**).

Arboles filogenéticos: ácido gentísico

En la **Figura 8** se puede observar el árbol filogenético resultado de comparar las secuencias proteicas del componente grande de la salicilato 5-hidroxilasa (nagG). Lo primero que destaca al analizar el árbol es que todas las bacterias incluidas pertenecen a la clase *Betaproteobacteria*, por lo que no se ha encontrado una alta homología con bacterias de clase *Alphaproteobacteria* y *Gammaproteobacteria*.

Si entramos en más detalle, destacan las bacterias del orden *Burkholderiales*, a excepción del género *Aromatoleum* y *Azoarcus* que pertenecen al orden *Rhodocyclales*. Dentro del orden *Burkholderiales* destacan dos familias: la familia *Burkholderiaceae y la familia Comamonadaceae.* Por un lado, de la familia *Burkholderiaceae* podemos encontrar los géneros *Burkholderia*, *Ralstonia* y *Pandorea* en el árbol filogenético, mientras que de la familia *Comamonadaceae* destacan los géneros *Comamonas*, *Variovorax* y *Curvibacter*.

Las cepas bacterianas para las que existen estudios publicados que confirman la capacidad de degradación del naftaleno (puntos negros) se encuentran agrupadas mayoritariamente en la zona inferior del árbol, aunque el valor *Bootstrap* de algunos nodos no supera el 50%. Por otro lado, también se incluyen otras cepas que son conocidas por degradar compuestos diferentes al naftaleno, como es el caso de *Burkholderia* sp. C3, conocida por degradar otro hidrocarburo aromático policíclico denominado fenantreno[49].

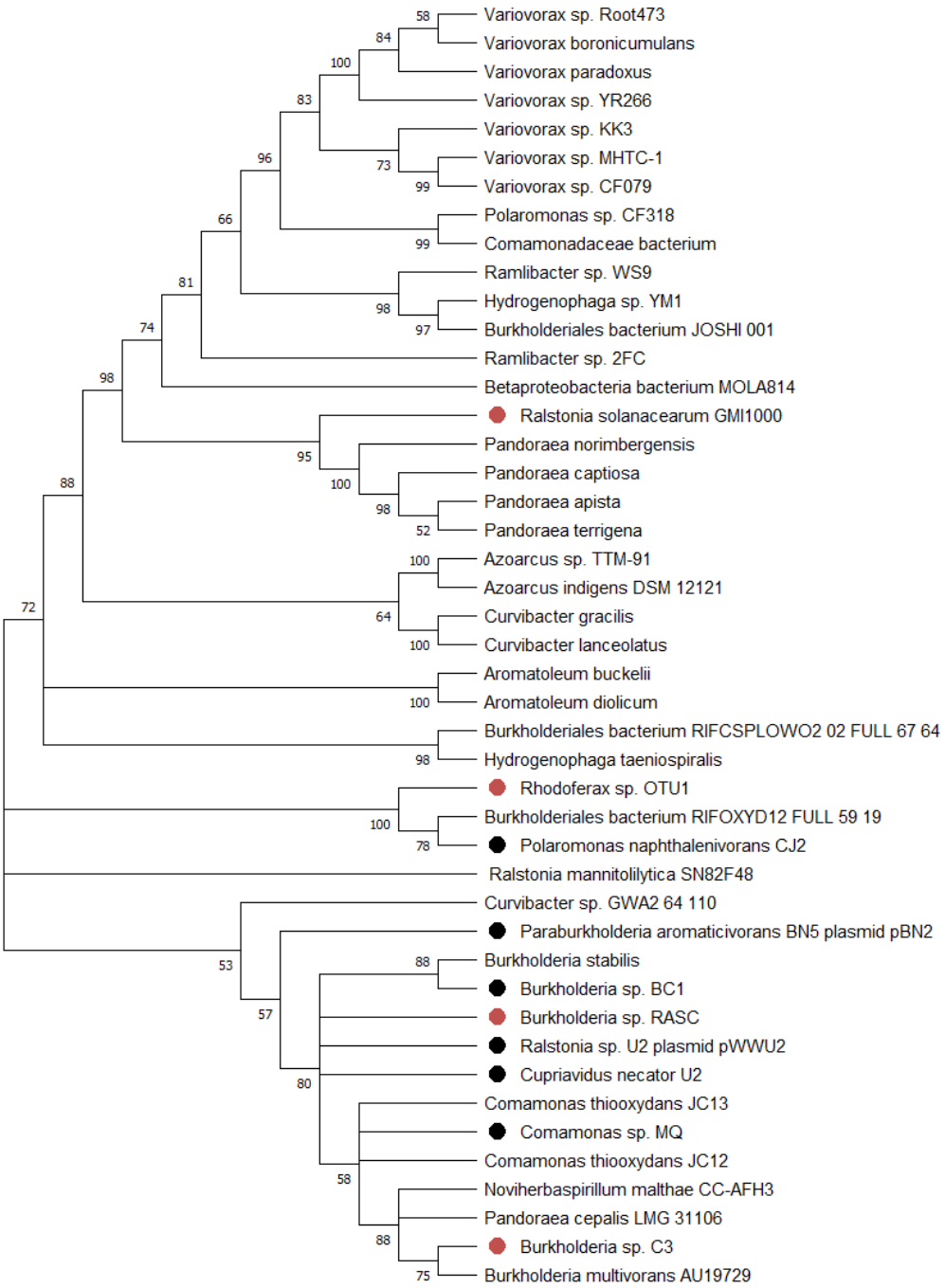

Figura 8. Árbol filogénico basado en la enzima nagG inferido mediante *Maximum Likelihood method* [50]. El porcentaje de réplicas de árboles en los que los taxones asociados se agruparon mediante la técnica *Boostrap* (500 réplicas) se muestra junto a las ramas[41]. Los árboles iniciales para la búsqueda heurística se obtuvieron automáticamente aplicando los algoritmos *Neighbour-Join* y *BioNJ* a una matriz de distancias por pares estimadas utilizando el modelo JTT, y luego seleccionando la topología con un valor de probabilidad de registro superior. Este análisis involucró 45 secuencias de aminoácidos. Hubo un total de 453 posiciones en el conjunto de datos final. El análisis evolutivo se llevó a cabo con MEGAX[38]. Mediante un punto negro se marcan aquellas cepas bacterianas para las cuales existen estudios publicados que confirman la capacidad de degradar naftaleno. Mediante un punto rojo se muestran las cepas bacterianas para las cuales existe evidencia de la degradación de otros compuestos diferentes al naftaleno.

La vía de degradación del fenantreno muestra cierta similitud con la ruta de degradación del naftaleno, ya que las vías superiores conllevan a la síntesis de *1-Hydroxy-2-naphthanoic acid* (1H2NA) que se puede degradar por la vía del naftaleno, incluyendo la ruta del catecol y del ácido gentísico[51]. Por otro lado, *Burkholderia* sp. RASC es conocida por su capacidad de degradación del 2,4-diclorofenol (2,4-DCP) y *Rhodoferax* sp. OTU1 por degradar el sulfulano[52,53].

Cabe mencionar que existe una agrupación clara para las secuencias del género *Variovorax*, tal y como podemos observar en el extremo superior del árbol, siendo así los organismos más lejanos respecto al organismo molde *Ralstonia* sp. U2. Las cepas de *Variovorax* se pueden encontrar con frecuencia en entornos contaminados y se ha descrito que muchas de ellas codifican genes para dioxigenasas hidrolizantes de anillos aromáticos[54,55]. Actualmente se ha identificado la capacidad de *Variovorax* de degradar benceno[56], e existen indicios de la degradación de naftaleno[57,58], pero hasta el momento no se han publicado estudios que caractericen las rutas de degradación de cepas específicas de este género.

Por otro lado, en la **Figura 9** se observa el árbol filogenético para la oxigenasa pequeña (nagH) del complejo salicilato 5-hidroxilasa. Destacar que las cepas del árbol resultante son muy similares a las del árbol de nagG, a excepción del género *Limnohabitans* y las especies adicionales del género *Paraburkholderiales*, que no se habían incluido en el árbol anterior. En este caso observamos más valores *Bootstrap* inferiores al 50%, sobre todo en la zona inferior donde se encuentra *Ralstonia* sp. U2.

En el caso del árbol filogenético para la ferredoxina reductasa, nagAa (**Figura 10**), podemos observar cómo se incluyen cepas del género *Pseudomonas*, que no habían aparecido hasta el momento. Como se ha comentado anteriormente, existe evidencia de que el complejo naftaleno 1,2-dioxigenasa (NDO) y el complejo salicilato 5-hidroxilasa (S5H) comparten cadena de electrones, formada por NagAa y NagAb[28] (**Figura 7**). Por esta razón, al realizar el PSI-BLAST se han encontrado secuencias homólogas relacionadas con el complejo naftaleno 1,2-dioxigenasa, como las secuencias del género *Pseudomonas.* Este caso sería un ejemplo de eficiencia genética, ya que tanto la ruta superior como la ruta inferior comparten dos proteínas transportadores de electrones[28].

Tras haber comparado los 3 árboles mencionados hasta ahora, relacionados con la enzima salicilato 5-hidroxilasa, se observa que, de manera general, las especies más cercanas a *Ralstonia* sp. U2 suelen ser *Burkholderia* sp. BC1, *Comamonas* sp. MQ, *Cupriavidus necator U2, Burkholderia RASC, y Cupriavidus necator U2.* Sin embargo, estos resultados muestran unos valores *Bootstrap* que no superan el 50%, por lo no aportan una elevada confianza.

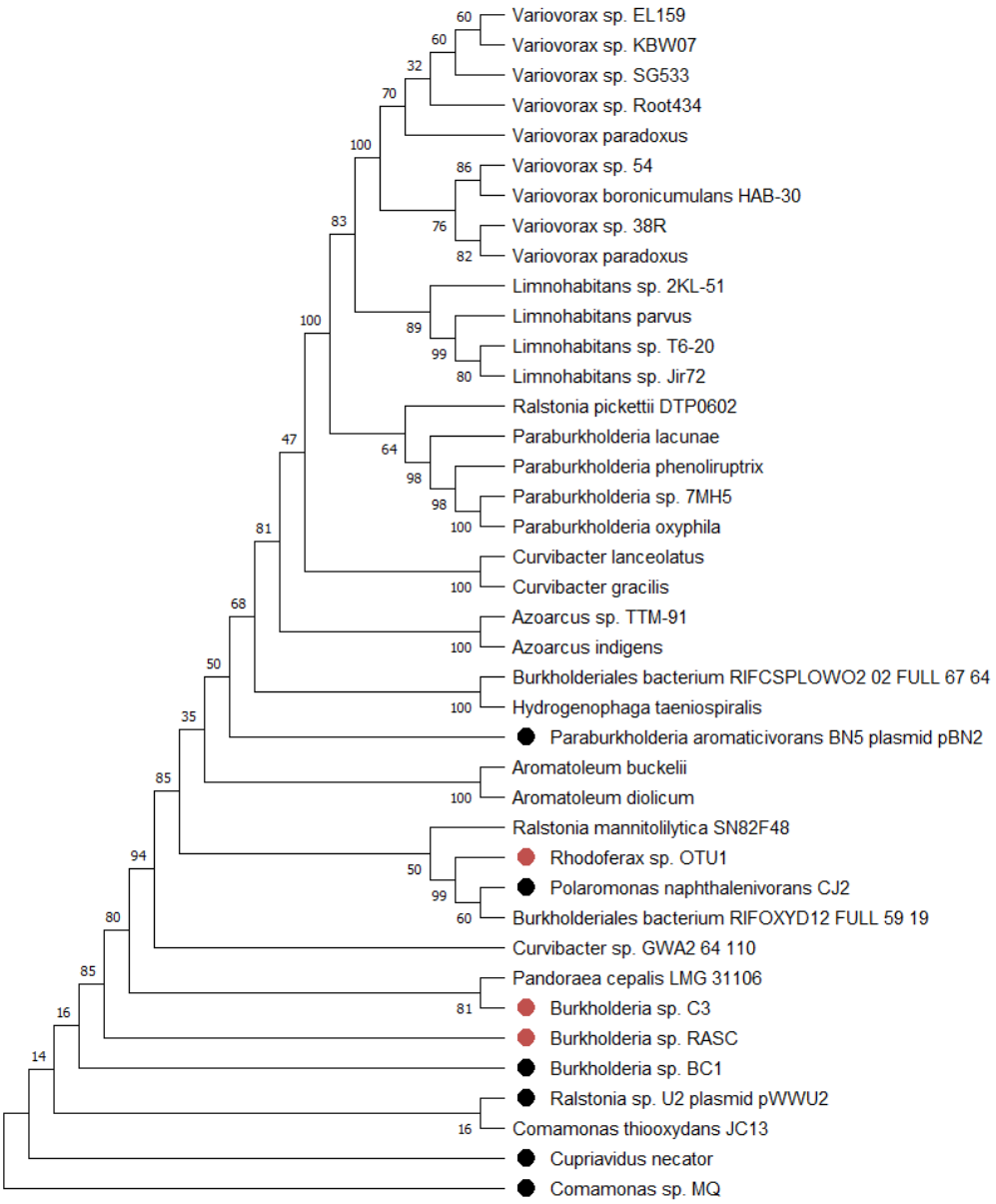

Figura 9. Árbol filogénico basado en la enzima nagH, inferido mediante *Maximum Likelihood method* [50]. El porcentaje de réplicas de árboles en los que los taxones asociados se agruparon mediante la técnica *Bootstrap* (500 réplicas) se muestra junto a las ramas[41]. Los árboles iniciales para la búsqueda heurística se obtuvieron automáticamente aplicando los algoritmos *Neighbour-Join* y *BioNJ* a una matriz de distancias por pares estimadas utilizando el modelo JTT, y luego seleccionando la topología con un valor de probabilidad de registro superior. Este análisis involucró 40 secuencias de aminoácidos. Hubo un total de 171 posiciones en el conjunto de datos final. El análisis evolutivo se llevó a cabo con MEGA X[38]. Los puntos negros y rojos presentan el mismo significado que en la figura 8.

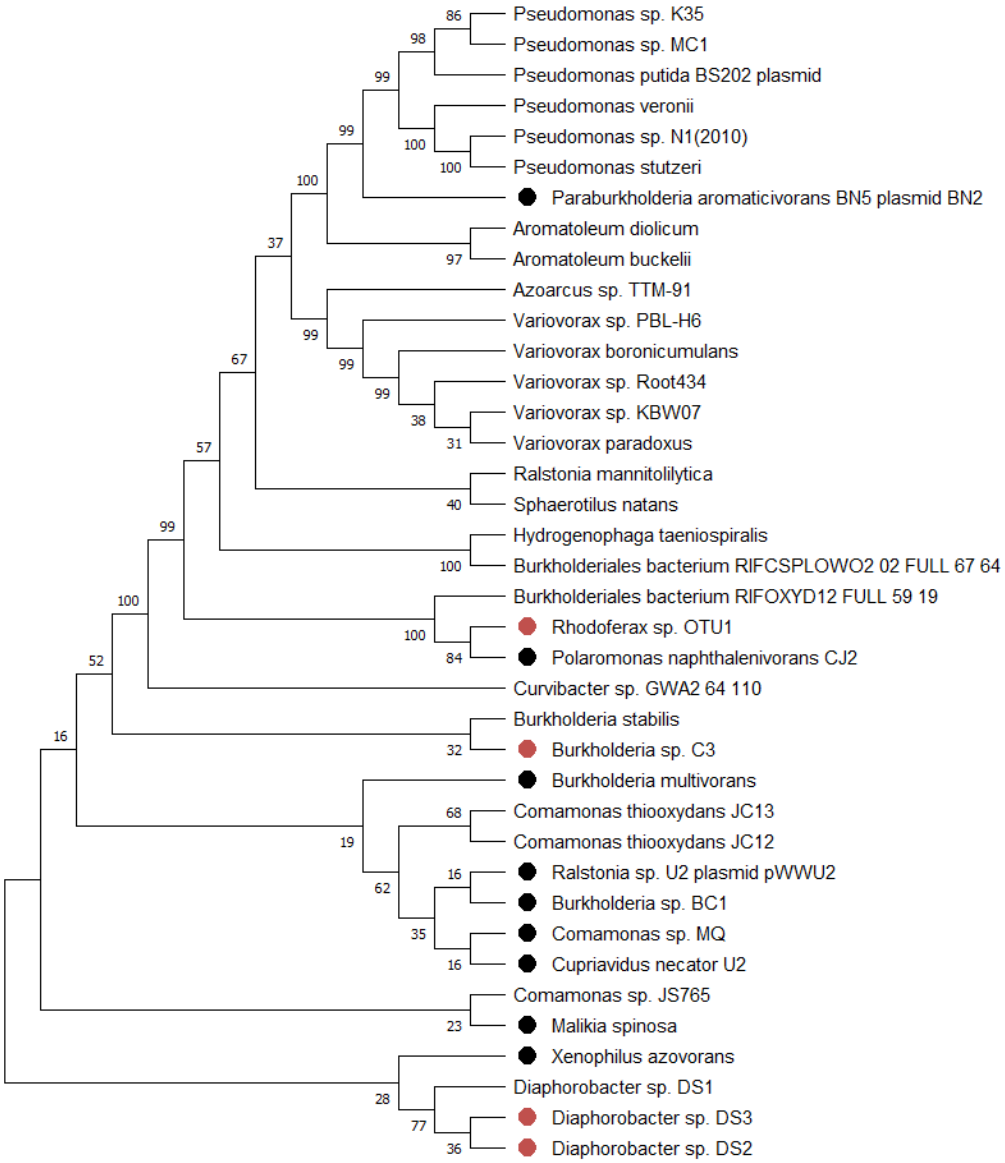

Figura 10. Árbol filogénico basado en la enzima nagAa inferido mediante *Maximum Likelihood method* [50]. El porcentaje de réplicas de árboles en los que los taxones asociados se agruparon mediante la técnica *Bootstrap* (500 réplicas) se muestra junto a las ramas[41]. Los árboles iniciales para la búsqueda heurística se obtuvieron automáticamente aplicando los algoritmos Neighbour-Join y BioNJ a una matriz de distancias por pares estimadas utilizando el modelo JTT, y luego seleccionando la topología con un valor de probabilidad de registro superior. Este análisis involucró 38 secuencias de aminoácidos. Hubo un total de 335 posiciones en el conjunto de datos final. El análisis evolutivo se llevó a cabo con MEGAX[38]. Los puntos negros y rojos presentan el mismo significado que en la figura 8.

Finalmente se presenta el árbol filogenético correspondiente a las secuencias homólogas de la enzima gentisato 1,2-dioxigenasa o nagI (**Figura 11**). Como se ha mencionado anteriormente, se han incluido secuencias adicionales del género *Rhodococcus*, que se encuentran agrupadas en el extremo superior, siendo el género más distante a *Ralstonia* sp. U2. De manera general, se observan resultados similares a los árboles filogenéticos anteriores, incluyendo únicamente bacterias de la clase *Betaproteobacteria (*sin tener en cuenta al género *Rhodococcus)* y destacando las familias *Burkholderiaceae y Comamonadaceae*. Sin embargo, destaca el cambio en la cepa *Cupriavidus necator* U2, que en este caso se encuentra más distante de *Ralstonia* sp. U2 y más cercana a las especies del género *Rhodococcus*.

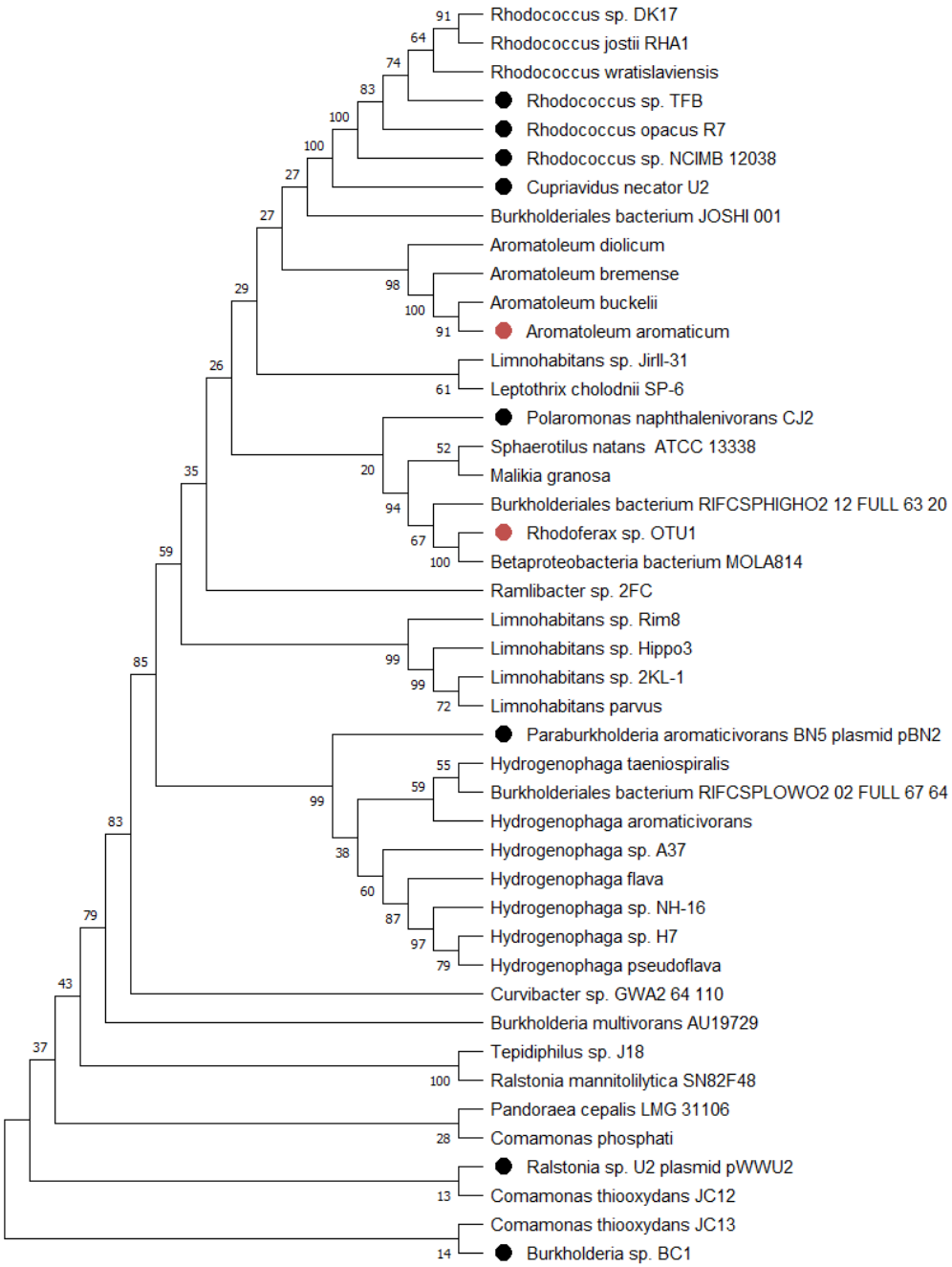

Figura 11. Árbol filogénico basado en la enzima nagI inferido mediante *Maximum Likelihood method* [50]. El porcentaje de réplicas de árboles en los que los taxones asociados se agruparon mediante la técnica *Bootstrap* (500 réplicas) se muestra junto a las ramas[41]. Los árboles iniciales para la búsqueda heurística se obtuvieron automáticamente aplicando los algoritmos Neighbour-Join y BioNJ a una matriz de distancias por pares estimadas utilizando el modelo JTT, y luego seleccionando la topología con un valor de probabilidad de registro superior. Este análisis involucró 44 secuencias de aminoácidos. Hubo un total de 390 posiciones en el conjunto de datos final. El análisis evolutivo se llevó a cabo con MEGAX[38]. Los puntos negros y rojos presentan el mismo significado que en la figura 8.

Predicción de operones y comparación con operón nag

A continuación, se muestran los resultados de la predicción de operones para la cepa *Cupriavidus necator* U2 y la comparación con el operón nag de *Ralstonia* sp. U2, que se encuentra bien caracterizado (**Figura 12**).

Para el caso de *Cupriavidus necator* U2, la herramienta FGENESB predice 2 operones relacionados con la degradación del naftaleno (**Anexo 3**). El primer operón está formado por 11 genes, donde se incluyen los componentes de la salicilato 5-hidroxilasa (nagG, nagH, nagAa y nagAb), mientras que el segundo operón está formado por 5 genes e incluye a la gentisato 1,2-dioxigenasa (nagI). Además, se puede observar como el operón 1 presenta una alta identidad (cercana al 100%), mientras que los genes del operón 2 presentan un gris más claro, indicando que presentan una identidad más baja que ronda el 65%. Estos resultados coinciden con lo observado en los árboles filogenéticos, ya que *Cupriavidus* necator U2 se encontraba muy cercana a *Ralstonia* sp. U2 en los árboles de nagG, nagH y nagAa. Sin embargo, se observaba más distancia evolutiva en el árbol de nagI, donde se acercaba más al género *Rhodococcus*.

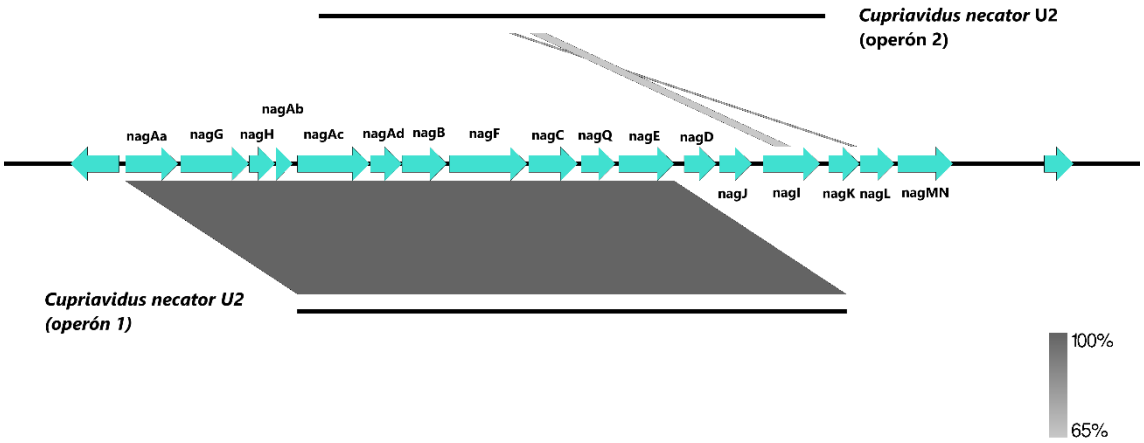

Figura 12. Comparación de dos operones putativos de *Cupriavidus necator* U2 y el operón nag de *Ralstonia* sp. U2 (plásmido pWWU2).

Árboles filogenéticos: ruta catecol

El árbol filogenético resultado de comparar las secuencias homólogas obtenidas para nahG (**Figura 13**) muestra una gran variedad de cepas del género *Pseudomonas.* Estos resultados concuerdan con lo esperado, ya que una gran parte de los microorganismos degradadores de naftaleno aislados y estudiados hasta el momento pertenecen al género *Pseudomonas*[59]. Adicionalmente, ya se ha mencionado anteriormente que los genes del catabolismo del naftaleno se encuentran a menudo localizados en plásmidos conjugativos, que pueden estar

relacionados con la transferencia horizontal de genes y con la evolución de las poblaciones bacterianas[12,59]. En este caso destaca una agrupación de genes codificados en plásmidos en la zona central del árbol filogenético, donde se incluyen plásmidos conocidos y muy estudiados como NAH7 y pDTG1[9,32,60].

Por otro lado, en el extremo superior se incluyen cepas pertenecientes al género *Acinetobacter* y *Halomonas*, los cuales también se han relacionado con la degradación de compuestos aromáticos y la biorremedación[61,62]. Específicamente, el género *Halomonas* comprende un conjunto de bacterias aerobias halófilas y halotolerantes, que habitualmente se localizan en ambientes salinos[57]. Por esta razón, ha surgido interés en estudiar este género en el ámbito de la biorremediación del agua, ya que podría presentar un mayor potencial para degradar los hidrocarburos en ambientes marinos que las cepas convencionales[63–66]. Sin embargo, todavía falta más información sobre los genes involucrados en esta degradación y los posibles operones.

En el segundo árbol se establecen las relaciones evolutivas de las secuencias homólogas para nahH de *Pseudomonas putida* G7 (**Figura 14**). Al igual que en árbol mostrado anteriormente, destacan las cepas del género *Pseudomonas*, muchas de las cuales coinciden con las cepas observadas para nahG. Además, se observan especies del género *Marinobacter* que no habían aparecido anteriormente. Este género se caracteriza por haberse detectado ampliamente en ambientes salinos contaminados con petróleo, lo que ha aumentado el interés por estudiar su posible papel biorremediador, pese a que todavía falta información sobre los genes degradadores de HAP que contiene[21,67]. En este ámbito, hasta el momento solo se ha informado de una cepa, (*Marinobacter* sp. NCE312), que contiene una naftaleno 1,2-dioxigenasa similar a nahAa de *Pseudomonas putida G7*[68]*.* Adicionalmente, a diferencia del árbol para nahG, entre la agrupación de las cepas del género *Pseudomonas* se incluye una cepa del género *Achromobacter*. Esta cepa específica, *Achromobacter* sp. BP3, ha sido estudiada por su capacidad de degradar el bifenilo, otro hidrocarburo policíclico aromático que está formado por dos anillos aromáticos. Por tanto, podemos observar una cierta relación evolutiva entre la enzima catecol 2,3-dioxigenasa que actúa en la ruta de degradación del naftaleno y la catecol 2,3-dioxigenasa que actúa en la ruta de degradación del bifenilo.

Finalmente, en el último árbol se muestran las relaciones de las secuencias homólogas obtenidas para nahI de *Pseudomonas putida* G7 (**Figura 15**). Podemos visualizar resultados muy similares a los anteriores dos árboles, ya que se repiten muchas de las cepas del género *Pseudomonas* y vuelve a aparecer el género *Marinobacter* y *Halomonas*.

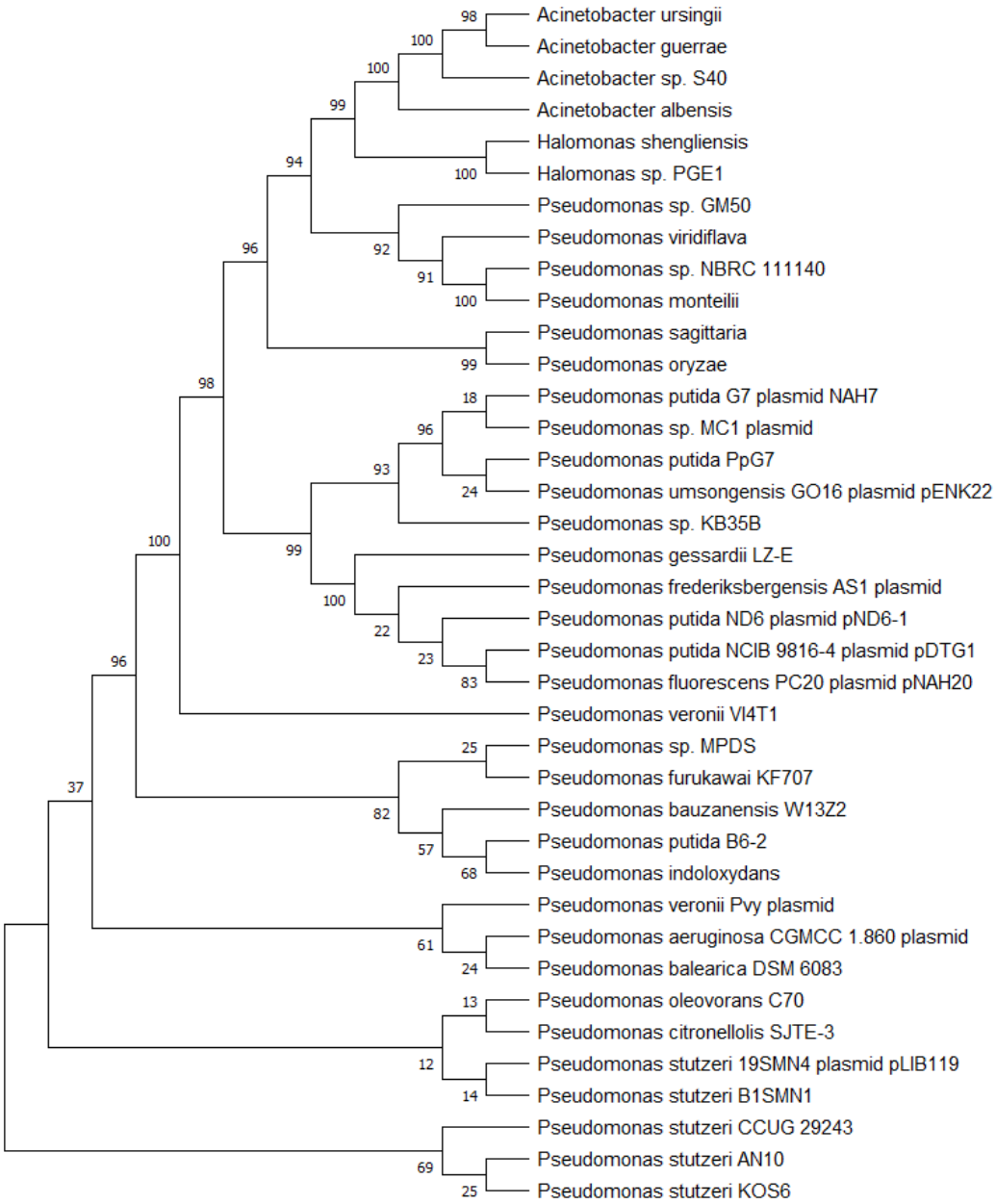

Figura 13. Árbol filogénico basado en la enzima nahG inferido mediante *Maximum Likelihood method*[50]. El porcentaje de réplicas de árboles en los que los taxones asociados se agruparon mediante la técnica *Boostrap* (500 réplicas) se muestra junto a las ramas[41]. Los árboles iniciales para la búsqueda heurística se obtuvieron automáticamente aplicando los algoritmos *Neighbour-Join* y *BioNJ* a una matriz de distancias por pares estimadas utilizando el modelo JTT, y luego seleccionando la topología con un valor de probabilidad de registro superior. Este análisis involucró 38 secuencias de aminoácidos. Hubo un total de 443 posiciones en el conjunto de datos final. El análisis evolutivo se llevó a cabo con MEGAX[38]

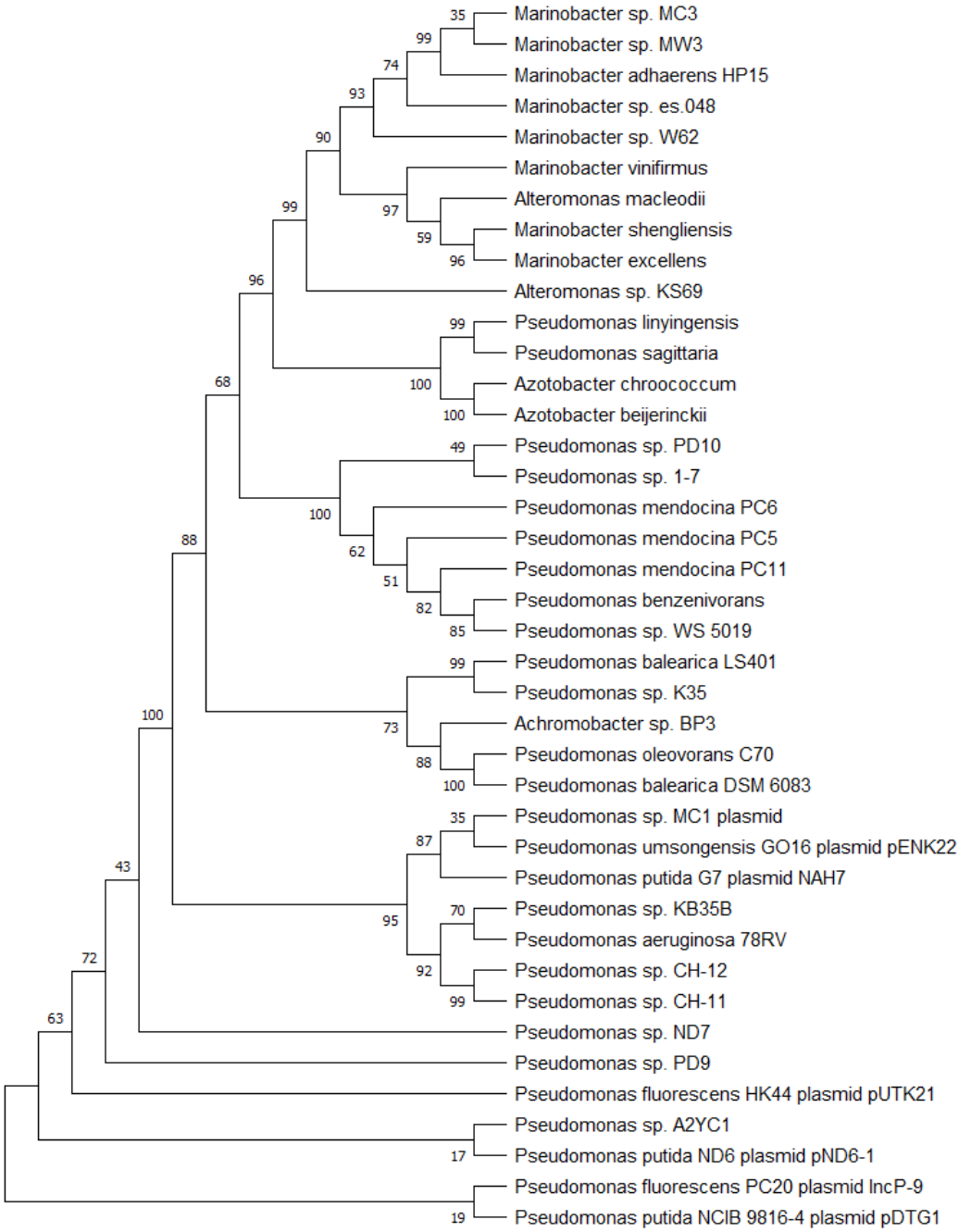

Figura 14. Árbol filogénico basado en la enzima nahH inferido mediante *Maximum Likelihood method*[50]. El porcentaje de réplicas de árboles en los que los taxones asociados se agruparon mediante la técnica *Boostrap* (500 réplicas) se muestra junto a las ramas[41]. Los árboles iniciales para la búsqueda heurística se obtuvieron automáticamente aplicando los algoritmos *Neighbour-Join* y *BioNJ* a una matriz de distancias por pares estimadas utilizando el modelo JTT, y luego seleccionando la topología con un valor de probabilidad de registro superior. Este análisis involucró 40 secuencias de aminoácidos. Hubo un total de 307 posiciones en el conjunto de datos final. El análisis evolutivo se llevó a cabo con MEGAX[38].

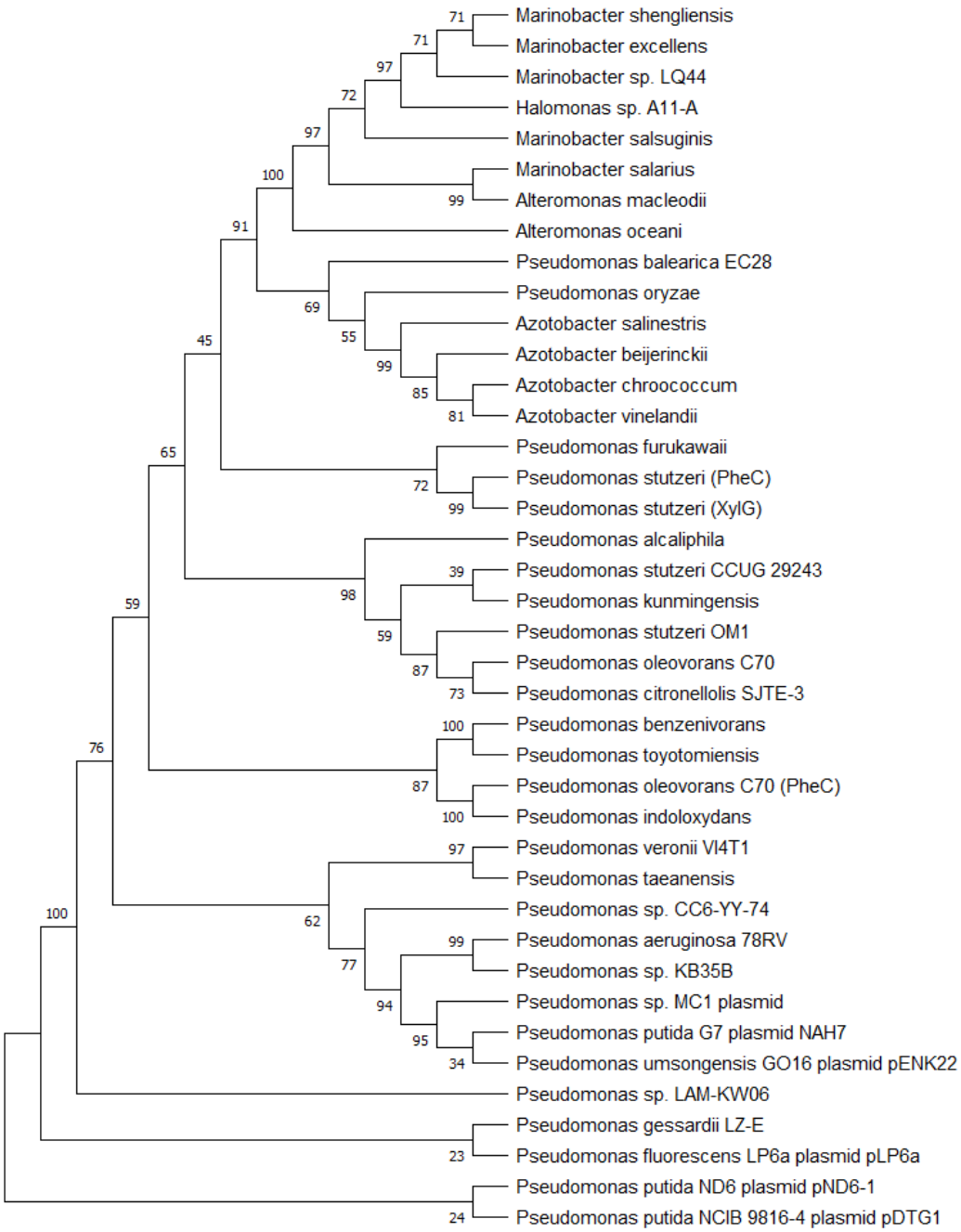

Figura 15. Árbol filogénico basado en la enzima nahI inferido mediante *Maximum Likelihood method*[50]. El porcentaje de réplicas de árboles en los que los taxones asociados se agruparon mediante la técnica *Boostrap* (500 réplicas) se muestra junto a las ramas[41]. Los árboles iniciales para la búsqueda heurística se obtuvieron automáticamente aplicando los algoritmos *Neighbour-Join* y *BioNJ* a una matriz de distancias por pares estimadas utilizando el modelo JTT, y luego seleccionando la topología con un valor de probabilidad de registro superior. Este análisis involucró 40 secuencias de aminoácidos. Hubo un total de 486 posiciones en el conjunto de datos final. El análisis evolutivo se llevó a cabo con MEGAX[38].

De manera general, en los 3 árboles filogenéticos, podemos observar algunas características en común que llaman la atención. Por un lado, prácticamente todas las bacterias incluidas en ellos pertenecen a la clase *Gammaproteobacteria*, a excepción de la cepa *Achromobacter* sp. B3 que pertenece a la clase *Betaproteobacteria*. Aquí observamos una diferencia clara con los árboles filogenéticos de la ruta del ácido gentísico, ya que en ese caso la mayoría de las cepas pertenecían a la clase *Betaproteobacteria*.

Adicionalmente, en los 3 árboles se observa que las cepas evolutivamente más cercanas a *Pseudomonas putida* G7 son *Pseudomonas* sp. MC1 y *Pseudomonas umsongensis* GO16, de las cuales existen estudios publicados que confirman la degradación del naftaleno. Estas cepas muestran un 100% de identidad con las secuencias estudiadas de *Pseudomonas putida* G7 (plásmido NAH7).

Por un lado, *Pseudomonas* sp. MC1 es una bacteria adaptada al frío con capacidad de degradar el naftaleno, característica que ha llamado la atención ya que facilitaría la biodegradación de contaminantes en ambientes marinos con bajas temperaturas[61]. Además, el plásmido pYIC que contiene los genes catabólicos es muy similar al plásmido NAH7 de *Pseudomonas putida* G7[60,69]. Por otro lado, *Pseudomonas umsongensis* GO16 ha sido muy estudiada en el ámbito de la degradación del ácido tereftálico derivado del PET (tereftalato de polietileno). Adicionalmente, se ha confirmado que esta cepa presenta la capacidad de degradar PAHs como el naftaleno (por la vía *meta* del catecol) gracias a los genes catabólicos del plásmido pENK22[70].

Predicción y comparación de operones

Como resultado de la predicción de genes mediante FGENESB se obtienen 6 operones que contienen entre 8 y 10 genes, que se pueden observar en la **Tabla 1**. Un detalle a tener en cuenta es que el resultado de la predicción del operón de *Pseudomonas putida* G7 no coincide con la bibliografía publicada, ya que la predicción considera al gen nahG como unidad transcripcional (Tu), fuera del operón. Esta variación podría ser debida a que al seleccionar a *Pseudomonas putida* KT2440 como organismo más cercano, no se tiene en cuenta que el operón se encuentra localizado en un plásmido. Lo mismo se ha observado con los otros organismos estudiados en esta predicción de operones, ya que tampoco incluían al gen nahG en el operón.

Organismo	Plásmido	Secuencia completa	Operón	Coordenadas	Comentario
Pseudomonas putida G7	NAH7	AB237655	9 genes	46628 - 54197	NahG fuera del operón (Tu)
Pseudomonas umsongensis GO16	pENK22	CP044408	10 genes	63.460 - 72.732	NahG fuera del operón (Tu)
Pseudomonas sp. MC1	pYIC1	JN248563	9 genes	46.273 - 53.842	NahG fuera del operón (Tu)
Pseudomonas fluorescens PC20	IncP-9	AAW81680	8 genes	41.478 - 48.614	NahG fuera del operón (Tu)
Pseudomonas putida NCIB 9816-4	pDTG1	AF491307.2	10 genes	41.478 - 49.126	NahG en operón diferente
Pseudomonas putida ND6	pND6-1	AY208917	8 genes	81.283 - 88.419	NahG fuera del operón (Tu)

Tabla 1. Características de las cepas incluidas en la predicción de operones.

En la **Figura 16** se muestra el resultado de la comparación de los operones de las cepas más cercanas a *Pseudomonas putida* G7. Estos operones incluyen las enzimas involucradas en la ruta inferior del naftaleno, la ruta del catecol (exceptuando a nagG). Se observa claramente como existe una alta similitud e identidad entre *Pseudomonas putida* G7, *Pseudomonas putida* MC1 y *Pseudomonas umsongensis* GO16, tal y como se ha comentado en el apartado anterior. En cambio, para las otras cepas se observan zonas con un color gris más claro (menor porcentaje de identidad) y más cambios y variaciones en el orden de los genes incluidos en los operones.

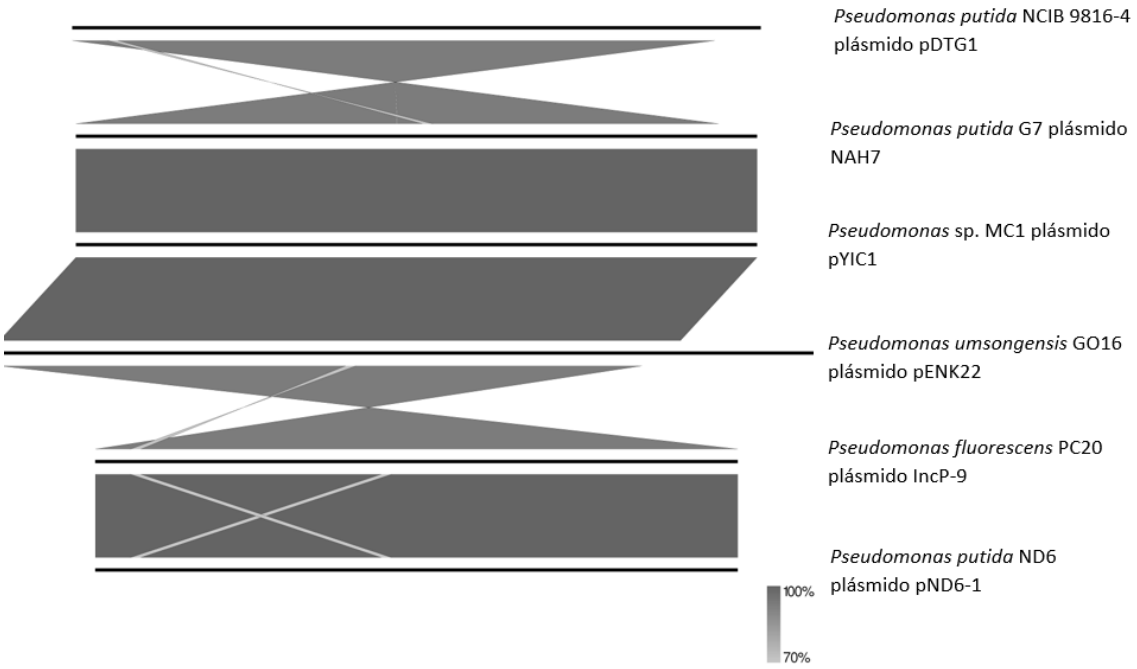

Figura 16. Comparación de las predicciones de operones de la ruta del catecol de cepas de *Pseudomonas*

4 Conclusiones

4.1 Conclusiones

Mediante el Trabajo Fin de Máster realizado se ha obtenido una descripción de la ruta del catecol y del gentisato, incluyendo los genes que participan en estas vías y las cepas más estudiadas. A partir de ahí, se ha realizado un estudio de genómica comparativa, que ha permitido la obtención de árboles filogenéticos para estudiar la relación evolutiva de varias enzimas.

Por un lado, las bases de datos obtenidas para la ruta del ácido gentísico (nagG, nagH, nagI y nagAa) muestran una abundancia de cepas pertenecientes a la clase *Betaproteobacteria*, destacando la familia *Burkholderiaceae* y *Comamonadaceae*. Además, en el árbol filogenético de la enzima nagAa se han incluido cepas del género *Pseudomonas*, que concuerda con la evidencia de que el complejo naftaleno 1,2-dioxigenasa (NDO) y el complejo salicilato 5-hidroxilasa (S5H) comparten cadena de electrones, formada por NagAa y NagAb[28]. Adicionalmente, se ha observado que las enzimas nagG, nagH y nagAa de *Cupriavidus necator* U2 presentan una alta homología con las mismas enzimas de *Ralstonia* sp. U2, pero en cambio, nagI muestra más distancia evolutiva y un menor porcentaje de identidad. Estos resultados coinciden con la predicción y comparación de operones, donde se ha observado que *Cupriavidus necator* U2 contiene los genes de la salicilato 5-hidroxilasa en un operón que ronda el 100% de identidad con el operón sal de *Pseudomonas putida* G7, y el gen nagI en otro operón con un 65% de identidad.

Por otro lado, para la ruta del catecol, las secuencias potencialmente homólogas incluidas para nahG, nahH y nahI se caracterizan por pertenecer a las cepas de la clase *Gammaproteobacteria*, destacando el género *Pseudomonas*. Adicionalmente, en este caso se observa mayor número de secuencias incluidas en plásmidos, lo que podría indicar la transferencia horizontal de este tipo de genes. Tanto en los árboles filogenéticos como en la predicción y comparación de operones se ha observado que *Pseudomonas* sp. MC1 (plásmido pYIC1) y *Pseudomonas umsongensis* GO16 (plásmido pENK22) presentan una alta identidad y cercanía evolutiva con los genes seleccionados de *Pseudomonas putida* G7 (plásmido NAH7). Finalmente, destacar que las predicciones de operones de las cepas seleccionadas de *Pseudomonas* no incluían el gen nahG, lo que no coincide con la bibliografía publicada[22,32]. Este defecto podría ser debido a que al seleccionar a *Pseudomonas putida* KT2440 como organismo más cercano en FGENESB, no se ha tenido en cuenta que los genes estudiados se encuentran localizados en plásmidos.

Resumiendo, se han obtenido las siguientes conclusiones del trabajo:

1. Las bases de datos obtenidas para nagG, nagH y nagI incluyen principalmente a bacterias de clase *Betaproteobacteria*, mientras que las de nahG, nahH y nahI incluyen a bacterias de la clase *Gammaproteobacteria*.
2. La base de datos de nagAa incluye secuencias de *Pseudomonas* relacionadas con el complejo naftaleno 1,2-dioxigenasa, debido a que comparte cadena de electrones con la salicilato 5-hidroxilasa[28].
3. Según las predicciones, *Cupriavidus necator* U2 contiene a la salicilato 5-hidroxilasa y a la gentisato 1,2-dioxigenasa en operones distintos, lo que concuerda con la diferencia de distancia evolutiva que se ha detectado en los árboles filogenéticos.
4. Los operones de *Pseudomonas* sp. MC1 y *Pseudomonas umsongensis* GO16 presentan una alta similitud con el operón sal de *Pseudomonas putida* G7, lo que coincide con la cercanía evolutiva observada en los árboles filogenéticos.

4.2 Futuras líneas de estudio

En un futuro sería interesante incluir las enzimas restantes de la ruta del catecol y la ruta del ácido gentísico en el estudio, ya que debido al tiempo limitado no ha sido posible en este trabajo. Además, también sería interesante incluir más secuencias en la base de datos obtenida por PSI-BLAST, pero al tener tiempo limitado se realizó una filtración de la identidad y del *Query Cover*.

Por otro lado, actualmente hay pocos estudios publicados que se centren exclusivamente en la degradación del naftaleno por *Pseudomonas* sp. MC1 y *Pseudomonas umsongensis* GO16[60,70], por lo que sería interesante centrar más atención en estas dos cepas tras observar la alta identidad y homología que presentan con el operón sal de *Pseudomonas putida* G7. Además, *Pseudomonas* sp. MC1 puede resultar interesante al conocer que es una cepa adaptada al frío, por lo que podría facilitar la biorremediación de naftaleno en ambientes marinos y oceánicos de bajas temperaturas[60].

4.3 Seguimiento de la planificación

En general se ha realizado un buen seguimiento de la planificación y no ha habido problemas mayores. Si se hubiera dispuesto de más tiempo para el estudio, lo óptimo habría sido incluir a todas las enzimas de ambas rutas, pero al tener tiempo limitado se escogieron aquellas enzimas más destacables. Además, en un principio se plantearon unos objetivos iniciales y a medida que ha avanzado el trabajo se han incluido nuevos objetivos y nuevas herramientas bioinformáticas, como FGENESB, *SnapGene* y *Easyfig* para realizar la predicción y comparación de operones. Posiblemente, si se hubieran planificado

desde un principio estas tareas, se podría haber organizado mejor el tiempo para profundizar más e incluir un mayor número de cepas para la predicción y la compasión de operones.

5 Glosario

- HAPs: hidrocarburos aromáticos policíclicos, grupo amplio de compuestos producidos principalmente por la combustión de petróleo, carbón y gas. Se caracterizan por estar compuestos únicamente por carbono e hidrógeno, formando varios anillos aromáticos.
- Xenobiótico: compuesto químico que no se procede del metabolismo de los seres vivos.
- FASTA: formato de texto que incluye secuencias de nucleótidos o de aminoácidos representados por códigos de una sola letra.
- NDO: naftaleno 1,2-dioxigenasa, complejo enzimático implicado en la ruta superior del naftaleno.
- S5H: salicilato 5-hidroxilasa, complejo enzimático implicado en el primer paso de la ruta del ácido gentísico.
- Tu: Unidad transcripcional
- Op: Operón

6 Bibliografía

1. Phale, P. S., Shah, B. A. & Malhotra, H. Variability in assembly of degradation operons for naphthalene and its derivative, carbaryl, suggests mobilization through horizontal gene transfer. *Genes (Basel).* **10**, 569 (2019).
2. More MG, U. S. Biological Degradation of Naphthalene: A New Era. *J. Bioremediation Biodegrad.* **04**, (2013).
3. Pathak, A. *et al.* Comparative genomics and metabolic analysis reveals peculiar characteristics of *rhodococcus opacus* strain M213 particularly for naphthalene degradation. *PLoS One* **11**, (2016).
4. Valaes, T., Doxiadis, S. A. & Fessas, P. Acute hemolysis due to naphthalene inhalation. *J. Pediatr.* **63**, 904–915 (1963).
5. Gupta, R., Singhal, P. C., Muthusethupathy, M. A., Malik, A. K. & Chugh, K. S. Cerebral oedema and renal failure following naphthalene poisoning. *J. Assoc. Physicians India* **27**, 347–348 (1979).
6. EPA. Naphthalene - US EPA. https://www.epa.gov/sites/production/files/2016-09/documents/naphthalene.pdf (2000).

7. Hilyard, E. J., Jones-Meehan, J. M., Spargo, B. J. & Hill, R. T. Enrichment, isolation, and phylogenetic identification of polycyclic aromatic hydrocarbon-degrading bacteria from Elizabeth River sediments. *Appl. Environ. Microbiol.* **74**, 1176–1182 (2008).
8. Huang, H. *et al.* The naphthalene catabolic protein NahG plays a key role in hexavalent chromium reduction in *Pseudomonas brassicacearum* LZ-4. *Sci. Rep.* **7**, (2017).
9. Dennis, J. J. & Zylstra, G. J. Complete sequence and genetic organization of pDTG1, the 83 kilobase naphthalene degradation plasmid from *Pseudomonas putida* strain NCIB 9816-4. *J. Mol. Biol.* **341**, 753–68 (2004).
10. Aghapour, A. A., Moussavi, G. & Yaghmaeian, K. Biological degradation of catechol in wastewater using the sequencing continuous-inflow reactor (SCR). *J. Environ. Heal. Sci. Eng.* **11**, (2013).
11. Lalucat, J., Bennasar, A., Bosch, R., García-Valdés, E. & Palleroni, N. J. Biology of *Pseudomonas stutzeri*. *Microbiol. Mol. Biol. Rev.* **70**, 510–547 (2006).
12. Jawad, F. & Al, H. EFFICIENCY OF NAPHTHALENE DEGRADATION BY PSEUDOMONAS BEARING NAH-PLASMIDS IN MODEL SOIL SYSTEM. *J. Vet.Med.Sci.* **10**, (2016).
13. Van Der Meer, J. R. Evolution of novel metabolic pathways for the degradation of chloroaromatic compounds. in *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* (1997). doi:10.1023/A:1000166400935.
14. Fuenmayor, S. L., Wild, M., Boyes, A. L. & Williams, P. A. A gene cluster encoding steps in conversion of naphthalene to gentisate in *Pseudomonas* sp. strain U2. *J. Bacteriol.* **180(9)**, 2522–30 (1998).
15. Izmalkova, T. Y. *et al.* The organization of naphthalene degradation genes in *Pseudomonas putida* strain AK5. *Res. Microbiol.* **164**, 244–253 (2013).
16. Herrick, J. B., Stuart-Keil, K. G., Ghiorse, W. C. & Madsen, E. L. Natural horizontal transfer of a naphthalene dioxygenase gene between bacteria native to a coal tar-contaminated field site. *Appl. Environ. Microbiol.* **63**, 2330–2337 (1997).
17. Serdar, C. M. & Gibson, D. T. Isolation and characterization of altered plasmids in mutant strains of *Pseudomonas putida* NCIB 9816. *Biochem. Biophys. Res. Commun.* **164**, 764–771 (1989).
18. Dunn, N. W. & Gunsalus, I. C. Transmissible plasmid coding early enzymes of naphthalene oxidation in *Pseudomonas putida*. *J. Bacteriol.* **114**, 974–976 (1973).
19. Anokhina, T. O. *et al.* Alternative Naphthalene Metabolic Pathway Includes Formation of ortho-Phthalic Acid and Cinnamic Acid Derivatives in the *Rhodococcus opacus* Strain 3D. *Biochem.* **85**, 355–368 (2020).
20. Yen, K. M., Serdar, C. M. & Gunsalus, I. C. Genetics of naphthalene catabolism in pseudomonads. *Crit. Rev. Microbiol.* **15**, 247–68 (1988).
21. Wang, C., Huang, Y., Zhang, Z., Hao, H. & Wang, H. Absence of the nahG-like gene caused the syntrophic interaction between *Marinobacter* and other microbes in PAH-degrading process. *J. Hazard. Mater.* **384**, 121387 (2020).

22. Zhou, N. Y., Fuenmayor, S. L. & Williams, P. A. nag genes of *Ralstonia* (formerly pseudomonas) sp. Strain U2 encoding enzymes for gentisate catabolism. *J. Bacteriol.* **183**, 700–708 (2001).
23. Jeon, C. O., Park, M., Ro, H. S., Park, W. & Madsen, E. L. The naphthalene catabolic (nag) genes of *Polaromonas naphthalenivorans* CJ2: Evolutionary implications for two gene clusters and novel regulatory control. *Appl. Environ. Microbiol.* **72**, 1086–95 (2006).
24. Osborn, A. M. & Böltner, D. When phage, plasmids, and transposons collide: Genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid* **48**, 202–12 (2002).
25. Ramsay, L. *et al.* A simple sequence repeat-based linkage map of Barley. *Genetics* **156**, 1997–2005 (2000).
26. Bhagwat, M. & Aravind, L. PSI-BLAST tutorial. in *Comparative Genomics: Volumes 1 and 2* Chapter 10 (2007). doi:10.1385/1-59745-514-8:177.
27. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X : Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
28. Zhou, N. Y., Al-Dulayymi, J., Baird, M. S. & Williams, P. A. Salicylate 5-hydroxylase from *Ralstonia* sp. strain U2: A monooxygenase with close relationships to and shared electron transport proteins with naphthalene dioxygenase. *J. Bacteriol.* **184**, 1547–1555 (2002).
29. KEGG ENZYME: 1.18.1.7. https://www.genome.jp/dbget-bin/www_bget?ec:1.18.1.7.
30. BRENDA - Information on EC 1.14.13.172 - salicylate 5-hydroxylase. https://www.brenda-enzymes.org/enzyme.php?ecno=1.14.13.172.
31. KEGG ENZYME: 1.13.11.4. https://www.genome.jp/dbget-bin/www_bget?ec:1.13.11.4.
32. Yen, K. M. & Gunsalus, I. C. Plasmid gene organization: naphthalene/salicylate oxidation. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 874–878 (1982).
33. Pseudomonas putida plasmid NAH7 DNA, complete sequence, strain:G7 - Nucleotide - NCBI. https://www.ncbi.nlm.nih.gov/nuccore/AB237655.1.
34. Lee, K., Park, J. & Ahn, I. Effect of additional carbon source on naphthalene biodegradation by *Pseudomonas putida* G7. **105**, 157–167 (2003).
35. Ralstonia sp. U2 plasmid pWWU2 putative chemotaxis protein gene, parti - Nucleotide - NCBI. https://www.ncbi.nlm.nih.gov/nuccore/AF036940.
36. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/.
37. Edgar, R. C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, (2004).
38. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
39. Jin, G., Nakhleh, L., Snir, S. & Tuller, T. Maximum likelihood of phylogenetic networks. *Bioinformatics* **22**, 2604–2611 (2006).
40. Tateno, Y., Takezaki, N. & Nei, M. Relative efficiencies of the maximum-

likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. *Mol. Biol. Evol.* **11**, 261–277 (1994).

41. Felsenstein, J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution (N. Y).* **39**, 783–791 (1985).
42. Soltis, P. S. & Soltis, D. E. Applying the Bootstrap in Phylogeny Reconstruction. *Stat. Sci.* **18**, 256–267 (2003).
43. Neville, B. A. & O'Toole, P. W. Molecular Biology: Genomics. *Encycl. Food Microbiol. Second Ed.* **2**, 770–779 (2014).
44. MolQuest. http://molquest.com/molquest.phtml?group=index&topic=gfind#bact.
45. SoftBerry - FGENESB HELP. http://www.softberry.com/berry.phtml?topic=fgenesb&group=help&subgroup=gfindb.
46. SnapGene | Software for everyday molecular biology. https://www.snapgene.com/.
47. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: A genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).
48. Larkin, M. J., Kulakov, L. A. & Allen, C. C. R. Biodegradation and Rhodococcus - Masters of catabolic versatility. *Curr. Opin. Biotechnol.* **16**, 282–290 (2005).
49. Seo, J. S., Keum, Y. S., Hu, Y., Lee, S. E. & Li, Q. X. Degradation of phenanthrene by *Burkholderia* sp. C3: Initial 1,2- and 3,4-dioxygenation and meta- and ortho-cleavage of naphthalene-1,2-diol. *Biodegradation* **18**, 123–131 (2007).
50. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices. *Bioinformatics* **8**, 275–282 (1992).
51. Tittabutr, P., Kyu, I. & Qing, C. Phn and Nag-like dioxygenases metabolize polycyclic aromatic hydrocarbons in *Burkholderia* sp . C3. *Biodegradation* **22**, 1119–1133 (2011).
52. Suwa, Y. *et al.* Characterization of a Chromosomally Encoded from *Burkholderia* sp . Strain RASC. *Appl. Environ. Microbiol.* **62**, 2464–2469 (1996).
53. Kasanke, C. P., Collins, R. E. & Leigh, M. B. Identification and Characterization of a Dominant Sulfolane-Degrading *Rhodoferax* sp . via Stable Isotope Probing Combined with Metagenomics. *Sci. Rep.* **9**, 1–9 (2019).
54. Satola, B., Wübbeler, J. H. & Steinbüchel, A. Metabolic characteristics of the species *Variovorax paradoxus. Appl. Microbiol. Biotechnol.* **97**, 541–560 (2013).
55. Martirani-Von Abercron, S. M., Marín, P., Solsona-Ferraz, M., Castañeda-Cataña, M. A. & Marqués, S. Naphthalene biodegradation under oxygen-limiting conditions: community dynamics and the relevance of biofilm-forming capacity. *Microb. Biotechnol.* **10**, 1781–1796 (2017).
56. Posman, K. M., DeRito, C. M. & Madsen, E. L. Benzene degradation by a *Variovorax* species within a coal tar-contaminated groundwater microbial community. *Appl. Environ. Microbiol.* **83**, 1–13 (2017).

57. Padmanabhan, P. *et al.* Respiration of 13C-labeled substrates added to soil in the field and subsequent 16S rRNA gene analysis of 13C-labeled soil DNA. *Appl. Environ. Microbiol.* **69**, 1614–1622 (2003).

58. Leigh, M. B. *et al.* Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). *ISME J.* **1**, 134–148 (2007).

59. Kosheleva, A. M. B. I. A. Diversity of Naphthalene Biodegradation Systems in Soil Bacteria. *Handb. Hydrocarb. Lipid Microbiol.* 1155–1163 (2010) doi:10.1007/978-3-540-77587-4.

60. Ahn, E. *et al.* Salicylate degradation by a cold-adapted *Pseudomonas* sp. *Ann. Microbiol.* **67**, 417–424 (2017).

61. Towner, K. ., Bergogne-Bérézin, E. & Fewson, C. . *The Biology of Acinetobacter*. (1991).

62. Jiang, Y., Qi, H. & Zhang, X. M. Co-biodegradation of anthracene and naphthalene by the bacterium *Acinetobacter johnsonii*. *J. Environ. Sci. Heal. - Part A Toxic/Hazardous Subst. Environ. Eng.* **53**, 448–456 (2018).

63. Fathepure, B. Z., Coolen, M. J. L. & Hole, W. Recent studies in microbial degradation of petroleum hydrocarbons in hypersaline environments. *Front. Microbiol.* **5**, 1–16 (2014).

64. Mnif, S., Chamkha, M. & Sayadi, S. Isolation and characterization of *Halomonas* sp. strain C2SS100, a hydrocarbon-degrading bacterium under hypersaline conditions. *J. Appl. Microbiol.* **107**, 785–794 (2009).

65. Cheffi, M. *et al.* Isolation and characterization of a newly naphthalene-degrading *Halomonas pacifica*, strain Cnaph3: biodegradation and biosurfactant production studies. *3 Biotech* **10**, 1–15 (2020).

66. Kim, K. K., Lee, J. S. & Stevens, D. A. Microbiology and epidemiology of *Halomonas* species. *Future Microbiol.* **8**, 1559–1573 (2013).

67. Christen, R., Fernandez, L. & Acquaviva, M. *Marinobacter hydrocarbonoclasticus* gen. nov., sp. nov., a New , Extremely Halotolerant , Hydrocarbon-Degrading Marine Bacterium. *Growth Lakel.* **42**, 568–576 (1992).

68. Hedlund, B. P., Geiselbrecht, A. D. & Staley, J. T. *Marinobacter* strain NCE312 has a *Pseudomonas*-like naphthalene dioxygenase. *FEMS Microbiol. Lett.* **201**, 47–51 (2001).

69. Kyung, A. *et al.* Biochemical and Biophysical Research Communications Crystal structure of cis -dihydrodiol naphthalene dehydrogenase (NahB) from *Pseudomonas* sp . MC1 : Insights into the early binding process of the substrate. *Biochem. Biophys. Res. Commun.* **491**, 403–408 (2017).

70. Narancic, T. *et al.* Genome analysis of the metabolically versatile *Pseudomonas umsongensis* GO16: the genetic basis for PET monomer upcycling into polyhydroxyalkanoates. *Microb. Biotechnol.* (2021) doi:10.1111/1751-7915.13712.

71. Seo, J. S., Keum, Y. S. & Li, Q. X. Bacterial degradation of aromatic compounds. *Int. J. Environ. Res. Public Health* **6**, 278–309 (2009).

72. Paliwal, V., Raju, S. C., Modak, A., Phale, P. S. & Purohit, H. J. *Pseudomonas putida* CSV86: A candidate genome for genetic

bioaugmentation. *PLoS One* **9**, (2014).

73. Sota, M. *et al.* Genomic and functional analysis of the IncP-9 naphthalene-catabolic plasmid NAH7 and its transposon Tn4655 suggests catabolic gene spread by a tyrosine recombinase. *J. Bacteriol.* **188**, 4057–67 (2006).

Anexos

Anexo 1. Base de datos para la ruta del ácido gentísico

Base de datos para nagG

	Query			
Ralstonia sp. U2	AAD12607.1			
Bacteria	**Accesion number**	**% Identity**	**% Query cover**	**Aminoacids**
Comamonas sp. MQ	AEV91667.1	99.53	100	423
Cupriavidus necator U2	MRN22621.1	100.00	100	423
Burkholderia stabilis	VBB16398.1	99.29	100	423
Burkholderia sp. BC1	APP18107.1	99.05	100	423
Curvibacter sp. GWA2_64_110	OGO97131.1	96.42	99	420
Pandoraea cepalis LMG 31106	VVE52022.1	98.11	100	420
Burkholderia sp. C3	ACT53246.1	97.87	100	420
Ralstonia mannolytica SN82F48	AJW44643.1	93.99	98	420
Noviherbaspirillum malthae CC-AFH3	WP_194726741.1	97.16	100	420
Comamonas thiooxydans JC13	KGG89605.1	99.52	97	414
Burkholderiales bacterium RIFCSPLOWO2_02_FULL_67_64	OGB17028.1	90.87	98	421
Rhodoferax sp. OTU1	WP_114968267.1	88.94	98	420
Paraburkholderia aromaticivorans BN5	ASW04055.1	92.36	99	420
Ramlibacter sp. 2FC	WP_137892926.1	83.45	100	433
Hydrogenophaga taeniospiralis	WP_068171437.1	90.87	98	421
Burkholderiales bacterium RIFOXYD12_FULL_59_19	OGB52656.1	90.14	98	420
Burkholderia sp. RASC	AAB09764.1	95.74	100	423
Polaromonas naphthalenivorans CJ2	AAZ93385.1	87.59	98	420
Azoarcus sp. TTM-91	WP_169152001.1	83.33	97	422
Azoarcus indigens DSM 12121	TDN47505.1	83.09	97	422
Curvibacter gracilis	WP_027474071.1	81.50	100	422
Curvibacter lanceolatus	WP_019572233.1	81.50	100	422
Hydrogenophaga sp. YM1	QRR32449.1	81.84	97	419
Betaproteobacteria bacterium MOLA814	ESS14505.1	81.11	97	420
Ramlibacter sp. WS9	ROZ77674.1	80.72	98	416
Burkholderiales bacterium JOSHI_001	EHR72093.1	79.42	97	417
Variovorax sp. KK3	WP_077003437.1	81.36	97	421
Variovorax sp. MHTC-1	RST55588.1	79.20	100	420
Ralstonia solanacearum	WP_042590331.1	80.84	96	418
Variovorax boronicumulans	WP_062477806.1	78.42	98	419
Variovorax paradoxus	WP_153284198.1	78.42	98	419
Variovorax sp. Root473	WP_056580645.1	78.42	98	419
Burkholderia multivorans AU19729	PRF53892.1	97.87	100	420
Polaromonas sp. CF318	WP_007862450.1	79.33	98	431
Variovorax sp. MHTC-1	WP_126836390.1	79.20	100	420
Variovorax sp. CF079	WP_093103137.1	79.62	99	420
Comamonadaceae bacterium	RJP72472.1	77.54	100	424
Variovorax sp. KK3	WP_077003437.1	80.00	99	421
Variovorax sp. YR266	WP_093173844.1	79.38	98	418
Pandoraea apista	VVG71070.1	75.00	99	428
Pandoraea captiosa	VVE66233.1	76.67	99	431
Pandoraea norimbergensis	WP_058375802.1	76.78	99	419
Pandoraea terrigena	VVD80394.1	75.24	99	428
Aromatoleum buckelii	WP_169198436.1	88.70	98	420
Aromatoleum diolicum	WP_169259306.1	88.70	98	420

Base de datos para nagH

	Query			
nagH de *Ralstonia* sp. U2 plasmid pWWU2	AAD12608.1			
Bacteria	***Accession number***	***% Identity***	***% Query cover***	***Aminoacids***
Paraburkholderia aromaticivorans BN5 plasmid pBN2	ASW04054.1	83.85	100%	161
Comamonas sp. MQ	AEV91668.1	100.00	100%	161
Ralstonia mannitolilytica SN82F48	AJW44644.1	89.44	100	161
Polaromonas naphthalenivorans CJ2	AAZ93386.1	86.34	100	161
Curvibacter sp. GWA2_64_110	OGO97132.1	94.41	100	163
Rhodoferax sp. OTU1	WP_114968266.1	87.58	100	161
Burkholderiales bacterium RIFOXYD12_FULL_59_19	OGB52655.1	86.96	100	161
Pandoraea cepalis LMG 31106	VVE52026.1	96.89	100	161
Burkholderia sp. BC1	APP18108.1	100.00	100	161
Comamonas thiooxydans JC13	KGG89606.1	100.00	100	161
Azoarcus communis SWub3 = DSM 12120	NMG71581.1	83.23	100	161
Burkholderia sp. RASC	AAD15560.1	98.76	100	161
Curvibacter lanceolatus	WP_157443256.1	70.37	99	168
Azoarcus communis LMG 5514	NMG48085.1	82.61	100	161
Sphaerotilus natans	SIS04504.1	85.09	99	165
Burkholderia sp. C3	ACT53247.1	95.65	100	161
Curvibacter gracilis	WP_027474070.1	69.75	99	165
Azoarcus sp. TTM-91	NMG36337.1	78.12	99	165
Azoarcus indigens	TDN47506.1	78.12	99	165
Thiothrix sp.	TXH78595.1	75.16	100	165
Burkholderiales bacterium RIFCSPLOWO2_02_FULL_67_	OGB17029.1	81.99	100	163
Hydrogenophaga taeniospiralis	WP_084236052.1	81.99	100	161
Betaproteobacteria bacterium	NBT56047.1	62.11	100	157
Burkholderia sp. lig30	KDB08163.1	64.60	100	157
Malikia spinosa AB6	MYZ52297.1	86.34	100	161
Variovorax sp. 54	PIF72985.1	60.62	99	160
Variovorax boronicumulans	GER10811.1	60.62	99	160
Variovorax paradoxus	QFZ85644.1	60.62	99	160
Variovorax sp. 38R	WP_192196782.1	60.62	99	160
Variovorax sp. EL159	SCX65133.1	61.25	99	159
Variovorax sp. SG533]	NVM69256.1	60.00	99	161
Variovorax sp. Root434	WP_056596716.1	60.00	99	160
Variovorax sp. KBW07	WP_124459121.1	60.00	99	159
Ralstonia pickettii DTP0602	AGW94020.1	61.88	99	158
Aromatoleum buckelii	WP_169198435.1	81.25	99	162
Aromatoleum diolicum	WP_169259307.1	82.51	99	162
Cupriavidus necator U2	WP_042878798.1	61.88	99	162
Limnohabitans sp. T6-20	WP_108285846.1	63.35	100	158
Limnohabitans sp. Jir72	WP_108413328.1	61.49	100	158
Limnohabitans sp. 2KL-51	WP_108352670.1	61.49	100	162
Limnohabitans parvus	WP_108311322.1	61.49	100	158

Base de datos para nagAa

	Query			
nagAa de Ralstonia sp. U2 plasmid pWWU2	AAD12606.1			
Bacteria	**Accesion number**	**% Identity**	**% Query cover**	**Aminoacids**
Cupriavidus necator U2	MRN22620.1	100,00	100	328
Comamonas sp. MQ	AEV91666.1	100,00	100	328
Burkholderia sp. BC1	APP18106.1	100,00	100	328
Burkholderia multivorans	AIO74745.1	99.70	100	328
Comamonas thiooxydans JC13	KGG89604.1	99.70	100	328
Diaphorobacter sp. DS2	TFI47444.1	99.09	100	328
Comamonas sp. JS765	AAL76199.1	99.09	100	328
Diaphorobacter sp. DS3	AGH09229.1	99.08	99	328
Malikia spinosa	WP_161125190.1	98.78	100	328
Burkholderia stabilis	VBB16405.1	98.78	100	328
Diaphorobacter sp. DS1	AGH09219.1	98.78	100	328
Burkholderia sp. C3	ACT53245.1	96.04	100	328
Curvibacter sp. GWA2_64_110	OGO97130.1	85.06	100	328
Hydrogenophaga taeniospiralis	WP_068171435.1	77.13	100	328
Rhodoferax sp. OTU1	WP_114968268.1	76.52	100	328
Burkholderiales bacterium RIFCSPLOWO2_02_FULL_67_64	OGB17027.1	75.91	100	328
Polaromonas naphthalenivorans CJ2	AAZ93384.1	75.00	100	328
Burkholderiales bacterium RIFOXYD12_FULL_59_19	OGB52657.1	73.48	100	328
Ralstonia mannitolilytica SN82F48	WP_045786202.1	71.34	100	328
Sphaerotilus natans	WP_037482963.1	70.43	100	328
Paraburkholderia aromaticivorans BN5 plasmid BN2	ASW04056.1	69.82	100	328
Pseudomonas stutzeri	ADK11285.1	66.46	100	328
Pseudomonas veronii	OPK03992.1	66.16	100	328
Pseudomonas sp. K35	OCX93220.1	67.99	100	328
Pseudomonas putida BS202 PLASMID	AAB62705.1	66.77	100	328
Pseudomonas sp. MC1	AEV45874.1	66.77	100	328
Pseudomonas sp. N1	ADK26579.1	66.46	100	328
Aromatoleum diolicum	WP_169259305.1	71.87	99	328
Aromatoleum buckelii	WP_169198437.1	73.09	99	328
Variovorax sp. KBW07	WP_124459123.1	62.20	99	329
Variovorax boronicumulans	WP_159276909.1	61.89	99	329
Variovorax paradoxus	CAA2105320.1	61.59	99	329
Variovorax sp. PBL-H6	VTU34633.1	61.52	99	329
Variovorax sp. Root434	WP_056596711.1	61.40	100	329

Base de datos para nagI

Query				
nagI de *Ralstonia* sp. U2 plasmid pWWU2	AAD12619.1			
Bacteria	***Accession number***	***% Identity***	***% Query cover***	***Aminoacids***
Burkholderia sp. BC1	APP18119.1	100.00	100	355
Comamonas thiooxydans JC13	KGG89617.1	100.00	100	355
Noviherbaspirillum malthae CC-AFH3	WP_194726750.1	96.06	100	355
Burkholderia multivorans AU19729	WP_105798799.1	94.37	100	355
Comamonas phosphati	GGH66302.1	100.00	97	346
Pandoraea cepalis LMG 31106	VVE52159.1	100.00	97	346
Tepidiphilus sp. J18	WP_185957772.1	91.53	99	365
Ralstonia mannitolilytica SN82F48	WP_205624848.1	91.24	99	365
Hydrogenophaga aromaticivorans	WP_177133486.1	82.61	97	345
Polaromonas naphthalenivorans CJ2	AAZ93402.1	79.66	99	351
Hydrogenophaga sp. H7	WP_079367325.1	81.74	97	345
Burkholderiales bacterium RIFCSPLOWO2_02_FULL_67_64	OGB14503.1	83.19	97	345
Hydrogenophaga sp. A37	WP_077334503.1	82.90	97	345
Hydrogenophaga pseudoflava DSM 1084	QBM30445.1	81.16	97	345
Hydrogenophaga taeniospiralis	WP_068175067.1	82.61	97	345
Malikia granosa	WP_105746790.1	76.55	99	351
Burkholderiales bacterium RIFCSPHIGHO2_12_FULL_63_20	OGB01307.1	74.58	99	351
Rhodoferax sp. OTU1	WP_114968258.1	72.88	99	351
Ramlibacter sp. 2FC	WP_137892939.1	82.13	97	349
Curvibacter sp. GWA2_64_110	OGO97139.1	85.63	98	349
Paraburkholderia aromaticivorans	WP_095423752.1	80.58	97	345
Betaproteobacteria bacterium MOLA814	ESS14502.1	70.90	99	355
Burkholderiales bacterium JOSHI_001	EHR71590.1	77.10	97	346
Sphaerotilus natans ATCC 13338	SIS04451.1	72.73	99	356
Limnohabitans sp. Rim8	WP_108358184.1	80.06	97	346
Hydrogenophaga sp. NH-16	WP_127805728.1	81.45	97	345
Hydrogenophaga flava	WP_066263327.1	82.32	97	345
Leptothrix cholodnii SP-6	ACB35924.1	80.46	98	351
Limnohabitans sp. Rim8	WP_108358184.1	80.06	97	346
Aromatoleum buckelii	WP_169198049.1	71.68	97	345
Aromatoleum aromaticum	WP_011236597.1	71.10	97	345
Aromatoleum bremense	WP_169201092.1	71.10	97	345
Limnohabitans parvus	WP_108313683.1	81.40	96	345
Limnohabitans sp. 2KL-1	WP_108330388.1	81.40	96	345
Limnohabitans sp. Hippo3	WP_108426703.1	79.65	96	345
Limnohabitans sp. JirII-31	WP_100143758.1	79.07	96	345
Aromatoleum diolicum	WP_169258611.1	74.42	96	345
Cupriavidus necator U2	MRN22608.1	65.02	96	358
Secuencias adicionales de *Rhodococcus*				
Rhodococcus opacus R7	ABH01038.1			
Rhodococcus sp. TFB	AFE48174.1			
Rhodococcus wratislaviensis	SPZ42482.1			
Rhodococcus sp. NCIMB 12038	WP_087559643.1			
Rhodococcus sp. DK17	WP_016882856.1			
Rhodococcus jostii RHA1	ABG93677.1			

Anexo 2. Base de datos para la ruta del catecol (meta-cleavage)

Base de datos para nahG

Query				
Pseudomonas putida G7 plasmid NAH7	BAE92165.1	100.00	100	434
Name	**Accession Number**	**% Identity**	**Query Cover**	**Aminoacids**
Pseudomonas sp. MC1 plasmid pYIC1	AEV45885.1	100.00	100	434
Pseudomonas umsongensis GO16 plasmid pENK22	QFG27732.1	100.00	100	434
Pseudomonas sp. MPDS	QKJ37192.1	84.56	100	437
Pseudomonas putida NCIB 9816-4 plasmid pDTG-1	AAM09538.1	92.63	100	438
Pseudomonas fluorescens PC20 plasmid pNAH20	AAY21679.2	92.63	100	438
Pseudomonas gessardii LZ-E	ALC77288.1	92.63	100	434
Pseudomonas putida ND6 plasmid pND6-1	AAP44222.1	92.63	100	434
Pseudomonas frederiksbergensis AS1	APV43317.1	92.63	100	434
Pseudomonas oleovorans C70	ANC68233.1	84.33	100	437
Pseudomonas stutzeri 19SMN4	AHY45210.1	84.33	100	437
Pseudomonas citronellolis SJTE-3	ANI13644.1	84.33	100	437
Pseudomonas stutzeri B1SMN1	EPL61045.1	84.33	100	437
Pseudomonas aeruginosa	ACV05012.1	84.33	100	437
Pseudomonas balearica DSM 6083	AJE16216.1	84.33	100	437
Pseudomonas veronii Pvy	QCG68900.1	84.33	100	437
Pseudomonas bauzanensis W13Z2	EZQ19343.1	84.56	100	437
Pseudomonas putida B6-2	ANC83290.1	84.33	100	437
Pseudomonas stutzeri AN10	AAD02146.1	84.10	100	437
Pseudomonas stutzeri CCUG 29243	AFM32555.1	84.10	100	437
Pseudomonas stutzeri KOS6	EWC41239.1	84.10	100	434
Pseudomonas indoloxydans JCM 14246	PTU79838.1	84.10	100	437
Pseudomonas sp. GM50	WP_008013318.1	72.81	97	432
Pseudomonas sp. CFII64	WP_020291297.1	71.46	97	424
Pseudomonas sp. NBRC 111140	WP_060492753.1	71.00	97	427
Pseudomonas sp. KB35B	ABB72206.1	94.47	100	434
Pseudomonas sagittaria	WP_092433702.1	86.98	99	430
Pseudomonas oryzae	WP_090348255.1	86.98	100	430
Pseudomonas veronii	OPK05587.1	86.41	100	434
Pseudomonas furukawaii KF707	AAZ08064.1	83.64	100	437
Pseudomonas viridiflava	WP_122701338.1	69.32	98	426
Pseudomonas monteilii	WP_119369477.1	69.50	97	427
Halomonas shengliensis	WP_089680706.1	57.34	98	439
Halomonas sp. PGE1	WP_169957196.1	56.71	97	439
Acinetobacter sp. S40	WP_200461091.1	49.06	97	423
Acinetobacter ursingii	WP_151832910.1	49.06	97	423
Acinetobacter guerrae	WP_151981890.1	49.06	97	423
Acinetobacter albensis	WP_092721192.1	47.10	98	429

Base de datos para nahH

Query				
Pseudomonas putida G7 plasmid NAH7	BAE92167.1	100.00	100	307
Name	**Accession Number**	**% Identity**	**Query cover**	**Aminoacid**
Pseudomonas sp. MC1 plasmid pYIC1	AEV45887.1	100.00	100	307
Pseudomonas umsongensis GO16 plasmid	QFG27735.1	100.00	100	307
Pseudomonas sp. CH-11	ADM87472.1	97.07	100	307
Pseudomonas sp. KB35B	ABB72208.1	98.70	100	307
Pseudomonas aeruginosa 78RV	TEK52005.1	98.70	100	307
Pseudomonas sp. CH-11	ADM87472.1	97.07	100	307
Pseudomonas fluorescens HK44	EXF91073.1	92.72	100	302
Pseudomonas sp. A2YC1	AAO39661.1	92.51	100	307
Pseudomonas putida NCIB 9816-4 plasmid pDTG1	AAO64305.1	92.52	100	307
Pseudomonas putida ND6 plasmid pND6-1	AAP44220.1	92.51	100	307
Pseudomonas fluorescens PC20 plasmid IncP-9	AAW81680.2	92.51	100	307
Pseudomonas sp. PD9	ABR24794.1	92.51	100	307
Pseudomonas sp. ND7	ABR10832.1	92.18	100	307
Pseudomonas balearica LS401	WP_061340930.1	91.53	100	307
Pseudomonas sp. K35	OCX92113.1	91.53	100	307
Pseudomonas mendocina PC6	AAW81671.1	88.27	100	307
Achromobacter sp. BP3	ACF20629.1	88.27	100	307
Pseudomonas oleovorans C70	ANC68235.1	87.95	100	307
Pseudomonas balearica DSM 6083	AJE16217.1	87.95	100	307
Pseudomonas mendocina PC5	AAW81670.1	87.95	100	307
Pseudomonas mendocina PC11	AAW81676.1	87.95	100	307
Pseudomonas sp. PD10	ABR10837.1	87.62	100	307
Pseudomonas mendocina PC4	AAW81669.1	87.62	100	307
Pseudomonas sp. 1-7	WP_128579868.1	87.62	100	307
Pseudomonas benzenivorans	WP_090439188.1	87.30	100	307
Pseudomonas sp. WS 5019	WP_169970786.1	87.30	100	307
Marinobacter adhaerens HP15	ADP99795.1	81.43	100	307
Marinobacter sp. MC3	MBL3823911.1	81.43	100	307
Marinobacter sp. MW3	MBL3892067.1	81.43	100	307
Marinobacter sp. es.04	WP_088826201.1	81.11	100	307
Marinobacter sp. W62	WP_135954378.1	80.46	100	307
Pseudomonas linyingensis	WP_090313842.1	79.15	100	307
Pseudomonas sagittaria	WP_092435800.1	78.18	100	307
Marinobacter shengliensis	WP_106694359.1	79.15	100	307
Marinobacter vinifirmus	OZC37022.1	78.83	100	307
Marinobacter excellens	WP_061332465.1	78.83	100	307
Azotobacter chroococcum	WP_131298798.1	77.52	100	307
Azotobacter beijerinckii	WP_090731019.1	77.52	100	307
Alteromonas sp. KS69	WP_126873262.1	78.83	100	307
Alteromonas macleodii	WP_156087408.1	78.83	100	307

Base de datos para nahI

Query				
Pseudomonas putida G7 plasmid NAH7	BAE92168.1			
Name	**Accession Number**	**% Identity**	**Query cover**	**Aminoacids**
Pseudomonas gessardii LZ-E	ALC77289.1	100.00	100	486
Pseudomonas fluorescens LP6a plasmid pLP6a	AAM88231.1	100.00	100	486
Pseudomonas putida ND6 plasmid pND6-1	AAP44219.1	100.00	100	486
Pseudomonas putida NCIB 9816-4 plasmid pDTG-1	AAO64304.1	100.00	100	486
Pseudomonas sp. MC1 plasmid pYIC	AEV45888.1	100.00	100	486
Pseudomonas umsongensis plasmid pENK22	QFG27736.1	100.00	100	486
Pseudomonas sp. CC6-YY-74	WP_079204040.1	97.33	100	486
Pseudomonas sp. LAM-KW06	WP_172148005.1	96.71	100	486
Pseudomonas veronii VI4T1	OPK05590.1	93.83	100	486
Pseudomonas taeanensis	WP_025167833.1	95.68	100	486
Pseudomonas aeruginosa 78RV	TEK52004.1	98.35	100	486
Pseudomonas alcaliphila	WP_075746154.1	95.06	100	486
Pseudomonas stutzeri CCUG 29243	AFM32552.1	94.86	100	486
Pseudomonas stutzeri OM1	BAA31265.1	94.86	100	486
Pseudomonas kunmingensis	WP_104098553.1	94.65	100	486
Pseudomonas oleovorans C70	ANC68236.1	93.21	100	486
Pseudomonas citronellolis SJTE-3	ANI13646.1	93.21	100	486
Pseudomonas sp. KB35B	ABB72209.1	97.74	100	486
Pseudomonas furukawaii	WP_096368008.1	94.44	100	486
Pseudomonas benzenivorans	WP_090439184.1	91.56	100	486
Pseudomonas toyotomiensis	WP_059392799.1	91.56	100	486
Pseudomonas veronii VI4T1	OPK05590.1	93.62	100	486
Pseudomonas oleovorans	ANC68222.1	91.56	100	486
Pseudomonas oryzae	WP_090350888.1	90.74	100	486
Pseudomonas balearica	WP_165562441.1	90.12	100	486
Pseudomonas indoloxydans	WP_108233842.1	89.92	100	486
Azotobacter chroococcum	WP_198866623.1	89.09	100	486
Azotobacter salinestris	WP_152388663.1	88.89	100	486
Azotobacter beijerinckii	WP_090897624.1	88.89	100	486
Azotobacter vinelandii	WP_012699539.1	88.27	100	486
Pseudomonas formosensis	WP_090536279.1	87.86	100	486
Marinobacter sp. LQ44	WP_058092969.1	84.98	100	486
Marinobacter shengliensis	WP_106694360.1	84.77	100	486
Marinobacter excellens	WP_061332464.1	84.57	100	486
Marinobacter lutaoensis	WP_176282647.1	84.57	100	486
Marinobacter salsuginis	WP_209399211.1	82.92	100	486
Marinobacter salarius	WP_126813285.1	83.74	100	486
Halomonas sp. A11-A	WP_110070742.1	83.74	100	486
Alteromonas macleodii	WP_156087409.1	83.74	100	486
Alteromonas oceani	WP_123324073.1	82.72	100	486

Anexo 3. Predicción de operones

Predicción de operones de *Cupriavidus necator* U2

```
Prediction of potential genes in microbial genomes
Time:   Tue Jan  1 00:00:00 2005
Seq name: RCOG01000033.1 Cupriavidus necator strain U2 scaffold_32, whole genome shotgun
Length of sequence - 47490 bp
Number of predicted genes - 47
Number of transcription units - 14, operons - 6
    N      Tu/Op   Conserved  S             Start         End    Score
                  pairs(N/Pv)
    1     1 Tu  1     .       -    CDS          2 -       239    144
    2     2 Op  1     .       +    CDS        276 -      1322    853
    3     2 Op  2     .       +    CDS       1352 -      2134    811
    4     2 Op  3     .       +    CDS       2188 -      3345   1150
    5     2 Op  4     .       +    CDS       3351 -      4532   1098
    6     2 Op  5     .       +    CDS       4552 -      5922   1332
    7     2 Op  6     .       +    CDS       5919 -      6758    779
    8     2 Op  7     .       +    CDS       6786 -      7574    874
    9     3 Tu  1     .       -    CDS       7812 -      8831    344
   10     4 Op  1     .       +    CDS       8864 -     10102    797
   11     4 Op  2     .       +    CDS      10110 -     11348    476
   12     4 Op  3     .       +    CDS      11367 -     11687    264
   13     4 Op  4     .       +    CDS      11698 -     12567    618
   14     4 Op  5     .       +    CDS      12564 -     13640    641
   15     4 Op  6     .       +    CDS      13684 -     14862    840
   16     4 Op  7     .       +    CDS      14873 -     15793    778
   17     4 Op  8     .       +    CDS      15796 -     16812    624
   18     4 Op  9     .       +    CDS      16814 -     17563    682
   19     4 Op 10     .       +    CDS      17581 -     18297    440
   20     5 Op  1     .       +    CDS      18458 -     19699   1006
   21     5 Op  2     .       +    CDS      19710 -     20720    779
   22     5 Op  3     .       +    CDS      20720 -     22129    798
   23     5 Op  4     .       +    CDS      22129 -     22773    546
   24     6 Tu  1     .       +    CDS      23056 -     24702   1301
   25     7 Tu  1     .       -    CDS      24669 -     25574    563
```

Operón 1

Operón 2