



## **UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

### **Maestría y Doctorado en Ciencias Bioquímicas**

Detección, caracterización y clasificación de enzimas 3-deoxi-D-arabino-heptulosoato 7-fosfato (DAHP) sintasas provenientes de una librería metagenómica de suelo

### **T E S I S**

QUE PARA OPTAR POR EL GRADO DE:

**MAESTRO EN CIENCIAS BIOQUÍMICAS**

**PRESENTA:**

Q.B.P. Uriel Gutiérrez Gómez

**TUTOR PRINCIPAL:**

Dr. José Adelfo Escalante Lozada (Instituto de Biotecnología – UNAM)

**COMITÉ TUTOR:**

Dra. María Esperanza Martínez Romero (Centro de Ciencias Genómicas – UNAM)

Dra. Rosa María Gutiérrez Ríos (Instituto de Biotecnología – UNAM)

Cuernavaca, Morelos. 2014



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

El presente trabajo fue realizado bajo la dirección del Dr. José Adelfo Escalante Lozada, en el Departamento de Ingeniería Celular y Biocatálisis del Instituto de Biotecnología de la Universidad Nacional Autónoma de México.

El sustentante Uriel Gutiérrez Gómez fue apoyado económicamente durante sus estudios de maestría con la beca del Consejo Nacional de Ciencia y Tecnología (CONACyT). Asimismo, el proyecto contó con el financiamiento del proyecto CONACyT-Ciencia Básica Proyecto 106428.

*A mis papás, María Dolores Gómez Olvera y Leoncio G. Gutiérrez López, porque ustedes fueron, son y serán siempre mi mayor orgullo y fortaleza.*

*A ti Viridiana, porque eres mi motivación y apoyo para seguir hacia adelante cosechando éxitos, pero sobre todo, sobreponerse a las adversidades.*

## AGRADECIMIENTOS

A los integrantes de mi jurado de titulación.

Al Dr. Adelfo Escalante por permitirme el ingreso al laboratorio y desarrollar mi proyecto de maestría.

A mi comité tutor, las Dras. Esperanza Martínez y Rosa María Gutiérrez, porque sus observaciones y sugerencias contribuyeron de manera importante al desarrollo del proyecto.

A todos mis compañeros y compañeras del laboratorio 4 Bolívar-Gosset, cuyos comentarios, sugerencias y observaciones, dentro y fuera del laboratorio, aportaron mucho al proyecto y contribuyeron a mi formación como persona.

Al personal técnico y administrativo del laboratorio, especialmente a Mercedes Enzaldo (Mechita) y Aurelia González (Aure), porque su trabajo constituye una piedra angular para el laboratorio.

A la Unidad de Síntesis y Secuenciación del Instituto de Biotecnología, así como a las Unidades Universitarias de Secuenciación Masiva de DNA y de Apoyo Bioinformático.

A la Universidad Nacional Autónoma de México y al Instituto de Biotecnología, porque es un honor pertenecer a dos instituciones pilares en el desarrollo científico, tecnológico y humanístico de México.

## RESUMEN

La biosíntesis de compuestos aromáticos en microorganismos se lleva a cabo tanto en metabolismo primario como en metabolismo secundario; sin embargo, en ambos se ha caracterizado una enzima clave que canaliza el flujo de carbono desde metabolismo central y lo dirige a la biosíntesis de compuestos aromáticos, ésta es la 3-deoxi-D-arabino-heptulosoato 7-fosfato (DAHP) sintasa. Esta enzima cataliza la condensación aldólica entre D-eritrosa 4-fosfato (E4P) y fosfoenolpiruvato (PEP) para producir DAHP, un carbohidrato de 7 carbonos. Este compuesto participa como precursor inicial de diversas reacciones que derivan en distintas vías metabólicas, e.g. la vía del shikimato, biosíntesis de fenazinas, vía del aminoshikimato. El gen que codifica para la enzima DAHP sintasa es considerado como gen esencial para el crecimiento. En este trabajo realizamos un análisis de diversidad de secuencias de DAHP sintasas con miembros microbianos representativos de todas las familias conocidas para estas enzimas; este análisis se realizó considerando tanto el contexto genómico y la reconstrucción filogenética de las familias de DAHP sintasas.

Asimismo, a partir de un banco metagenómico realizado con DNA de suelo forestal, se realizó la búsqueda de genes que codifican para la enzima DAHP sintasa mediante complementación funcional utilizando como huésped una cepa de *Escherichia coli* carente de actividad de DAHP sintasa obtenida mediante la delección de los genes cromosomales *aroF*, *aroG* y *aroH*. Se obtuvieron 138 clonas, las cuales fueron analizadas tanto a nivel de secuencia como en cinéticas de crecimiento en medio mínimo con glucosa como única fuente de carbono y energía; dos de estas clonas, denominadas UGGL96 y UGGS2 mostraron fases lag de 3 y 5 horas respectivamente, obteniendo una biomasa máxima de 0.95 y 0.75 g<sub>CDW</sub>/L en 12 y 20 horas. En las mismas condiciones de cultivo, para el caso de una cepa silvestre de *E. coli* MG1655 se obtuvo una biomasa máxima de 0.8 g<sub>CDW</sub>/L con una fase lag de 6 horas. La actividad enzimática relativa de DAHP sintasa de estas clonas son 53 U·mg<sup>-1</sup> (UGGL96) y 30 U·mg<sup>-1</sup> (UGGS2), mientras que la actividad relativa en la cepa MG1655 es de 75 U·mg<sup>-1</sup>. De acuerdo al análisis de secuenciación, en estas clonas se obtuvieron dos insertos (10 y 12 kb) que corresponden a la bacteria *Nitrospira defluvii*, dentro de los cuales se encuentra el gen *aroF*, que codifica para la enzima fosfo-2-dehidro-3-deoxiheptonato aldolasa (nomenclatura alternativa de DAHP sintasa), así como otros genes asociados al contexto genómico de *aroF* en *N. defluvii*, mientras que la secuencia de aminoácidos de esta enzima posee mayor homología con aquellas DAHP sintasas de la subfamilia AroA<sub>1α</sub>.

## ÍNDICE DE CONTENIDO

|   |           |
|---|-----------|
| RESUMEN.....  | iv        |
| <b>I. INTRODUCCIÓN.....</b>   | <b>1</b>  |
| I.1. Síntesis de 3-deoxi-D-arabino-heptulosoato 7-fosfato (DAHP).....   | 1         |
| I.2. Biosíntesis de compuestos aromáticos en el metabolismo microbiano primario. DAHP y vía del shikimato.....  | 2         |
| I.3. DAHP sintasas y KDOP sintasas. Biosíntesis de lipopolisacáridos.....   | 3         |
| I.4. DAHP sintasas en el metabolismo microbiano secundario.....   | 4         |
| I.4.1. Biosíntesis de fenazinas.....  | 4         |
| I.4.2. Biosíntesis de ácido 3-amino-5-hidroxibenzoico (AHBA). Vía del aminoshikimato.....   | 5         |
| I.5. Metagenoma y su aplicación biotecnológica.....   | 7         |
| I.5.1. Diversidad microbiana del suelo.....   | 7         |
| I.5.2. Bancos o librerías metagenómicas.....  | 8         |
| <b>II. ANTECEDENTES.....</b>  | <b>10</b> |
| <b>III. JUSTIFICACIÓN.....</b>  | <b>19</b> |
| <b>IV. HIPÓTESIS.....</b>   | <b>19</b> |
| <b>V. OBJETIVOS.....</b>  | <b>20</b> |
| V.1. Objetivo general.....  | 20        |
| V.2. Objetivos particulares.....  | 20        |
| <b>VI. MATERIALES Y MÉTODOS.....</b>  | <b>21</b> |
| VI.1. Análisis del contexto genómico de DAHP sintasas de las subfamilias AroA <sub>Iα</sub> (e.g. <i>E. coli</i> ) y AroA <sub>Iβ</sub> (e.g. <i>B. subtilis</i> )..... | 21        |
| VI.2. Análisis de secuencias de DAHP sintasas (familias AroA <sub>I</sub> y AroA <sub>II</sub> ), aminoDAHP sintasas y KDOP sintasas.....                               | 21        |
| VI.3. Cepas bacterianas, plásmidos, antibióticos y condiciones de cultivo.....  | 22        |
| VI.4. Extracción de DNA cromosómico y plasmídico.....   | 23        |
| VI.5. Eliminación de genes cromosomales <i>aroF</i> , <i>aroG</i> y <i>aroH</i> en <i>E. coli</i> .....   | 23        |
| VI.6. Amplificación por PCR de genes <i>aroF</i> , <i>aroG</i> y <i>aroH</i> .....  | 24        |
| VI.7. Transformación de la cepa de <i>E. coli</i> $\Delta$ <i>aroF</i> $\Delta$ <i>aroH</i> $\Delta$ <i>aroG::kan</i> con el banco metagenómico de suelo.....           | 25        |
| VI.8. Selección de clonas transformantes T2.1 con recuperación de actividad de DAHP sintasa.....  | 25        |
| VI.9. Análisis de clonas UGG.....   | 26        |
| VI.9.1. Análisis de restricción de plásmidos pUGG.....  | 26        |
| VI.9.2. Secuenciación de insertos contenidos en plásmidos pUGG.....   | 27        |
| VI.9.3. Secuenciación masiva de clonas seleccionadas.....   | 28        |
| VI.9.4. Caracterización fisiológica de las clonas UGG.....  | 28        |
| VI.9.4.1. Cinética de crecimiento.....  | 28        |
| VI.9.4.2. Determinación de actividad enzimática de DAHP sintasa.....  | 28        |

|  |    |
|--|----|
| <b>VII. RESULTADOS</b> .....   | 29 |
| VII.1. Análisis del contexto genómico de DAHP sintasas de las subfamilias AroA <sub>Iα</sub> (e.g. <i>E. coli</i> ) y AroA <sub>Iβ</sub> (e.g. <i>B. subtilis</i> ).....                                   | 29 |
| VII.2. Reconstrucción filogenética y análisis de secuencias de familias representativas de DAHP sintasas (AroA <sub>Iα</sub> , AroA <sub>Iβ</sub> , AroA <sub>II</sub> , KdsA, PhzDAHPS y aminoDAHPS)..... | 33 |
| VII.3. Obtención de cepa de <i>E. coli</i> ΔaroF ΔaroG ΔaroH carente de actividad de DAHP sintasa .....  | 40 |
| VII.4. Transformación de cepa de <i>E. coli</i> ΔaroF ΔaroH ΔaroG::kan con banco metagenómico de suelo.....  | 44 |
| VII.5. Análisis de clonas transformantes T2.1 con recuperación de actividad de DAHP sintasa .....  | 45 |
| VII.5.1 Análisis de restricción de plásmidos pUGG .....  | 45 |
| VII.5.2. Secuenciación de insertos de clonas UGG.....  | 48 |
| VII.5.3. Cinéticas de crecimiento de las clonas UGG.....   | 50 |
| VII.5.4. Determinación de actividad enzimática específica de DAHP sintasa .....  | 52 |
| VII.5.5. Secuenciación masiva de las clonas UGGL96 y UGGS2.....  | 55 |
| <b>VIII. CONCLUSIONES</b> .....  | 71 |
| <b>IX. PERSPECTIVAS</b> .....  | 72 |
| <b>X. REFERENCIAS</b> .....  | 73 |
| <b>XI. APÉNDICE</b> .....  | 77 |



## ÍNDICE DE FIGURAS

|   |    |
|---|----|
| <b>Figura 1.</b> Mecanismo de reacción propuesto de la enzima 3-deoxi-D-arabino-heptulosonato 7-fosfato (DAHP) sintasa.....   | 1  |
| <b>Figura 2.</b> Vía condensada de biosíntesis de compuestos aromáticos en bacterias, hongos, plantas y algunos protozoarios .....  | 2  |
| <b>Figura 3.</b> Reacciones catalizadas por la KDOP y DAHP sintasa .....  | 3  |
| <b>Figura 4.</b> Vía de biosíntesis de fenazinas .....  | 4  |
| <b>Figura 5.</b> Compilación de operones bacterianos involucrados en biosíntesis de fenazinas .....   | 5  |
| <b>Figura 6.</b> Vías del shikimato y aminoshikimato .....  | 6  |
| <b>Figura 7.</b> Arreglo del operón <i>rif</i> en <i>Amycolatopsis mediterranei</i> .....   | 6  |
| <b>Figura 8.</b> Reacciones demostradas en presencia de extractos crudos de <i>Amycolatopsis mediterranei</i> o expresión heteróloga.....   | 7  |
| <b>Figura 9.</b> Esquema general de construcción y análisis de una librería metagenómica .....  | 9  |
| <b>Figura 10.</b> Alineamiento de las secuencias de aminoácidos de las DAHP sintasas de <i>Solanum tuberosum</i> (papa) y <i>Arabidopsis thaliana</i> (DHS1 y DHS2).....  | 10 |
| <b>Figura 11.</b> Comparación de las secuencias de aminoácidos de las DAHP sintasas de <i>Corynebacterium glutamicum</i> , <i>Escherichia coli</i> y <i>Salmonella typhimurium</i> .....  | 11 |
| <b>Figura 12.</b> Alineamiento múltiple de las secuencias de DAHP sintasas de <i>Streptomyces rimosus</i> , <i>Neurospora crassa</i> y <i>Streptomyces lividans</i> con algunas DAHP sintasas de plantas superiores.....          | 12 |
| <b>Figura 13.</b> Dendrograma (PILEUP) mostrando las relaciones de la familia I de DAHP sintasas.....   | 14 |
| <b>Figura 14.</b> Árbol filogenético de DAHP sintasas de la familia AroA <sub>II</sub> .....  | 15 |
| <b>Figura 15.</b> Esquema de la secuencia de DAHP sintasa de <i>Xanthomonas campestris</i> .....  | 16 |
| <b>Figura 16.</b> Árbol filogenético de homología de DAHP sintasas de las familias AroA <sub>Iα</sub> y AroA <sub>Iβ</sub> con las KDOP sintasas (KdsA).....  | 17 |
| <b>Figura 17.</b> Alineamiento de secuencias de DAHP y KDOP sintasas microbianas.....   | 18 |
| <b>Figura 18.</b> Regiones de alineamiento de oligonucleótidos en sitios adyacentes a los genes <i>aroF</i> , <i>aroG</i> y <i>aroH</i> .....   | 24 |
| <b>Figura 19.</b> Esquema de vector pJET1.2.....  | 27 |
| <b>Figura 20.</b> Análisis de contexto genómico de 79 secuencias de DAHP sintasas de la subfamilia AroA <sub>Iβ</sub> , ortólogas a <i>Bacillus subtilis</i> , dentro de un árbol filogenético de genes 16S rRNA .....            | 31 |
| <b>Figura 21.</b> Análisis de contexto genómico de 58 secuencias de DAHP sintasas de la subfamilia AroA <sub>Iα</sub> , ortólogas a <i>Escherichia coli</i> , dentro de un árbol filogenético de genes 16S rRNA.....              | 32 |
| <b>Figura 22.</b> Esquemas de DAHP sintasas representativas de cada familia conocida .....  | 34 |
| <b>Figura 23.</b> Alineamiento de secuencias de DAHP sintasas de algunas especies vegetales representativas de la familia AroA <sub>II</sub> .....  | 38 |
| <b>Figura 24.</b> Esquema general del patrón de alineamiento y distribución de la estructura primaria de enzimas DAHP sintasas .....  | 38 |
| <b>Figura 25.</b> Árbol filogenético de DAHP sintasas de organismos representativos de todos los grupos conocidos (AroA <sub>Iα</sub> , AroA <sub>Iβ</sub> , AroA <sub>II</sub> , PhzDAHP sintasas, y aminoDAHP sintasas).....    | 39 |
| <b>Figura 26.</b> Distribución de los genes <i>aroF</i> (amarillo), <i>aroG</i> (verde) y <i>aroH</i> (azul) en el cromosoma de <i>Escherichia coli</i> .....   | 41 |
| <b>Figura 27.</b> Productos de PCR en amplificación de gen <i>aroF</i> .....  | 42 |
| <b>Figura 28.</b> Productos de PCR en amplificación de gen <i>aroH</i> .....  | 42 |
| <b>Figura 29.</b> Productos de PCR en amplificación de gen <i>aroH</i> .....  | 43 |
| <b>Figura 30.</b> Productos de PCR en amplificación de gen <i>aroG</i> .....  | 44 |
| <b>Figura 31.</b> Perfiles de restricción de los plásmidos provenientes de las clonas UGGL(1-14) con las enzimas <i>Pst</i> I y <i>Pvu</i> II, en un gel de agarosa al 2.5%, con un marcador de tamaño molecular de 1 kb. ....    | 46 |
| <b>Figura 32.</b> Perfiles de restricción de los plásmidos provenientes de las clonas UGGL (1-14) con las enzimas <i>Bsp</i> M I y <i>Pvu</i> II, en un gel de agarosa al 2.5%, con un marcador de tamaño molecular de 1 kb. .... | 47 |

|   |    |
|---|----|
| <b>Figura 33.</b> Perfiles de restricción de los plásmidos provenientes de las 19 clonas UGG con perfiles únicos en un gel de agarosa al 2.5 % con las enzimas <i>BspM I</i> y <i>Pvu II</i> y un marcador de tamaño molecular de 1 kb. ....  | 48 |
| <b>Figura 34.</b> Perfil de restricción de los plásmidos pUGGL96 y pUGGS2 realizados con la enzima <i>Ade I</i> en un gel de agarosa al 1.5 % utilizando un marcador de tamaño molecular de 1 kb. ....  | 50 |
| <b>Figura 35.</b> Cinéticas de crecimiento en medio mínimo M9S de las clonas UGGL96 y UGGS2 que recuperaron la actividad de DAHP sintasa, así como las cepas control MG1655, T2.1, T2.1 pJLBaroG <sup>trktA</sup> y T2.1 pJET1.2-B.....   | 51 |
| <b>Figura 36.</b> Linealidad de actividad enzimática de DAHP sintasa de las cepas MG1655, T2.1 pJLBaroG <sup>trktA</sup> y las clonas UGGS2 y UGGL96.....   | 53 |
| <b>Figura 37.</b> Actividad enzimática de DAHP sintasa de las cepas MG1655, T2.1 pJLBaroG <sup>trktA</sup> y las clonas UGGS2 y UGGL96.....   | 54 |
| <b>Figura 38.</b> Árbol filogenético de secuencias de DAHP sintasas y los <i>contigs</i> obtenidos para los insertos de las clonas UGGL96 y UGGS2 correspondientes a <i>Nitrospira defluvii</i> .....   | 66 |
| <b>Figura 39.</b> Árbol filogenético de secuencias de 16S rRNA de los miembros bacterianos pertenecientes a la subfamilia de DAHP sintasas AroA <sub>1α</sub> , señalando la ubicación del organismo <i>Nitrospira defluvii</i> .....   | 68 |
| <b>Figura 40. A.</b> Esquema de motivos conservados en la subfamilia AroA <sub>1α</sub> de DAHP sintasas. <b>B.</b> Alineamiento de secuencias de <i>contigs</i> C75, C95, C153 y C189 con la secuencia de aminoácidos de la DAHP sintasa (AroF) de <i>Nitrospira defluvii</i> (352 aa) ..... | 69 |
| <b>Figura 41.</b> Esquema de contexto genómico de gen <i>aroF</i> en el cromosoma de <i>Nitrospira defluvii</i> .....   | 70 |

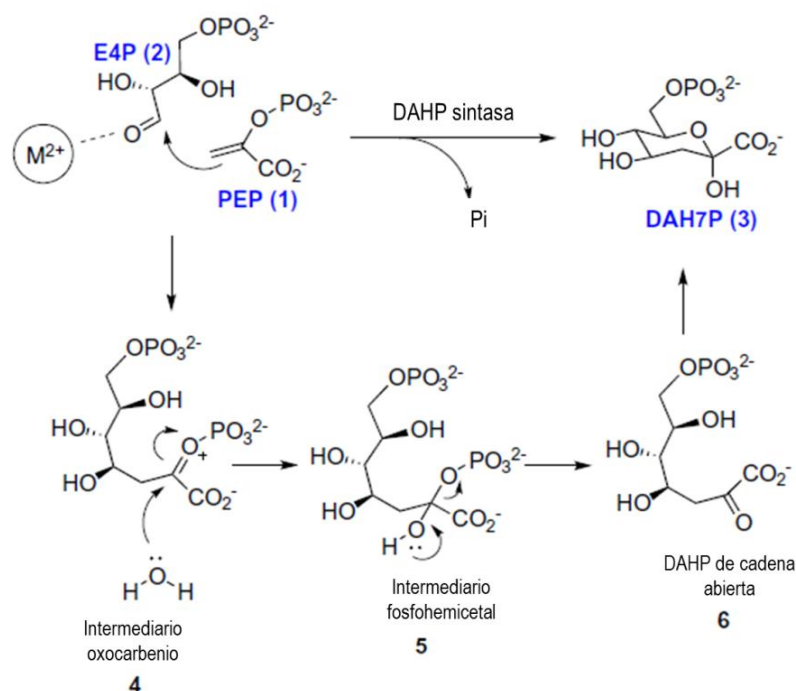
## ÍNDICE DE TABLAS

|  |    |
|--|----|
| <b>Tabla 1.</b> Porcentajes de similitud e identidad entre las secuencias de DAHP sintasas de <i>Escherichia coli</i> , <i>Saccharomyces cerevisiae</i> , <i>Arabidopsis thaliana</i> y <i>Solanum tuberosum</i> (papa).....                   | 11 |
| <b>Tabla 2.</b> Cepas y plásmidos utilizados en este trabajo.....  | 22 |
| <b>Tabla 3.</b> Secuencias de oligonucleótidos para amplificar los genes <i>aroF</i> , <i>aroG</i> y <i>aroH</i> .....   | 24 |
| <b>Tabla 4.</b> Mezcla de reacción para amplificación por PCR (Volumen final: 50 μL) .....   | 25 |
| <b>Tabla 5.</b> Condiciones de reacción para amplificación de genes <i>aroF</i> , <i>aroG</i> y <i>aroH</i> por PCR.....   | 25 |
| <b>Tabla 6.</b> Volúmenes de reacción para doble restricción enzimática (Volumen final: 20 μL) .....   | 26 |
| <b>Tabla 7.</b> Ejemplo de algunas enzimas relacionadas en el contexto genómico de las DAHP sintasas de <i>Escherichia coli</i> (subfamilia (AroA <sub>1α</sub> )) y <i>Bacillus subtilis</i> (subfamilia AroA <sub>1β</sub> ).....            | 33 |
| <b>Tabla 8.</b> Longitud y ubicación de los genes <i>aroF</i> , <i>aroG</i> y <i>aroH</i> , en el cromosoma de <i>Escherichia coli</i> .....   | 40 |
| <b>Tabla 9.</b> Enzimas de restricción con sitios de corte común en los genes <i>aroA</i> ( <i>Bacillus subtilis</i> ) y <i>aroF</i> ( <i>Escherichia coli</i> ).....  | 46 |
| <b>Tabla 10.</b> Tamaño de inserto de secuencias de las clonas UGG.....  | 48 |
| <b>Tabla 11.</b> Resultados de BLASTX realizados a los <i>contigs</i> y <i>scaffolds</i> obtenidos en los insertos de la clona UGGL96.....   | 55 |
| <b>Tabla 12.</b> Resultados de BLASTX realizados a los <i>contigs</i> y <i>scaffolds</i> obtenidos en los insertos de la clona UGGS2.....  | 62 |
| <b>Tabla 13.</b> Parámetros de BLASTX obtenidos para la enzima fosfo-2-dehidro-3-deoxiheptonato aldolasa de <i>Nitrospira defluvii</i> de los <i>contigs</i> C75, C95, C135 y C189 presentes en los insertos de las clonas UGGL96 y UGGS2..... | 64 |

## I. INTRODUCCIÓN

### I.1. Síntesis de 3-deoxi-D-arabino-heptulosoato 7-fosfato (DAH7P)

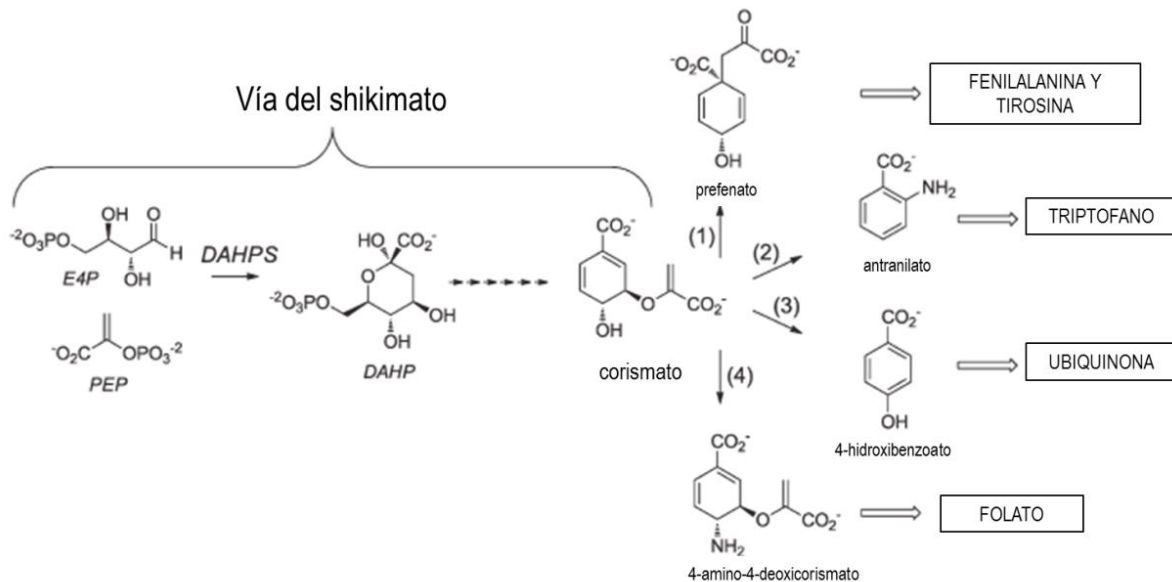
La enzima 3-deoxi-D-arabino-heptulosoato 7-fosfato (DAH7P) sintasa cataliza la condensación aldólica entre fosfoenolpiruvato (PEP) y D-eritrosa 4-fosfato (E4P) para generar 3-deoxi-D-arabino-heptulosoato 7-fosfato, y fosfato inorgánico (Figura 1). Esta reacción es estereoespecífica con respecto a ambos sustratos, la cara *si* del PEP interacciona con la cara *re* del grupo carbonilo de la E4P. Se propone que la reacción inicial del PEP sobre la E4P es promovido mediante una activación tipo ácido de Lewis del grupo aldehído de la E4P por un catión metálico divalente, seguido por la reacción con PEP para formar un intermediario oxocarbenio con geometría trigonal plana en el C2 del PEP. La interacción de un par de electrones del átomo de oxígeno de una molécula de agua con el ión oxocarbenio electrofílico produce un intermediario fosfohemiacetal tetraédrico y la eliminación de un grupo fosfato lleva al DAH7P de cadena abierta [Walker *et al.*, 2011].



**Figura 1.** Mecanismo de reacción propuesto de la enzima 3-deoxi-D-arabino-heptulosoato 7-fosfato (DAH7P) sintasa. E4P: eritrosa 4-fosfato; PEP: fosfoenolpiruvato; Pi: fosfato inorgánico; DAH7P: 3-deoxi-D-arabino-heptulosoato 7-fosfato;  $M^{2+}$ : catión metálico divalente [modificado de Walker *et al.*, 2011].

## I.2. Biosíntesis de compuestos aromáticos en el metabolismo microbiano primario. DAHP y vía del shikimato

En microorganismos (bacterias y hongos), plantas, y en algunos parásitos del phylum Apicomplexa, la biosíntesis de diversos compuestos aromáticos esenciales para el crecimiento se lleva a cabo mediante la vía del shikimato. Esta vía metabólica consiste en una serie de siete reacciones bioquímicas que inicia con la formación de DAHP y cuyo producto final es el corismato (Figura 2). A partir de éste, se derivan múltiples vías metabólicas para la biosíntesis de diversos compuestos tales como aminoácidos, pigmentos, transportadores de electrones, vitaminas y algunos metabolitos secundarios. La presencia de estas vías metabólicas y sus ramificaciones son características propias de algunos microorganismos, por lo que los productos de estas vías son considerados también como metabolitos propios que contribuyen a la individualidad de los organismos [Gosset *et al.*, 2001; Light *et al.*, 2012].

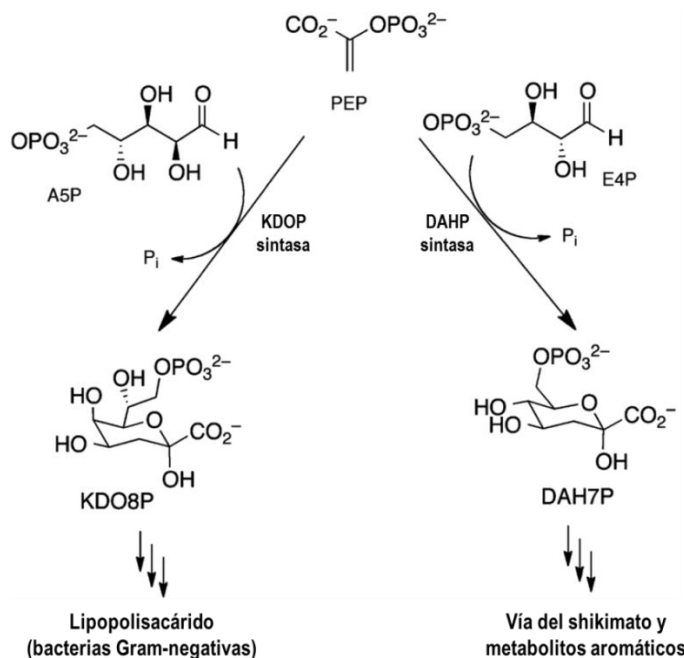


**Figura 2.** Vía condensada de biosíntesis de compuestos aromáticos en bacterias, hongos, plantas y algunos protozoarios. E4P: eritrosa 4-fosfato; PEP: fosfoenolpiruvato; DAHP: 3-deoxi-D-arabino-heptulosa 7-fosfato; DAHPS: DAHP sintasa [modificado de Light *et al.*, 2012].

### I.3. DAHP sintasas y KDOP sintasas. Biosíntesis de lipopolisacáridos.

En bacterias Gram-negativas, un componente esencial para la biosíntesis de lipopolisacárido (LPS) de membrana externa es el 3-deoxi-D-mano-octuloso (KDO). El precursor directo de este compuesto es el 3-deoxi-D-mano-octuloso 8-fosfato (KDOP), el cual es sintetizado por la enzima KDOP sintasa. Esta enzima ha sido caracterizada y estudiada en diversos organismos como *Escherichia coli*, *Neisseria meningitidis*, *Aquifex aeolicus*, *Salmonella typhimurium*, *Aquifex pyrophilus*, *Helicobacter pylori*, *Arabidopsis thaliana* y *Neisseria gonorrhoeae* [Allison *et al.*, 2013].

La enzima KDOP sintasa está relacionada con la enzima DAHP sintasa a nivel evolutivo y estructural así como en sus respectivos mecanismos de reacción. Las KDOP sintasas catalizan la condensación aldólica entre D-arabinosa 5-fosfato (A5P) y fosfoenolpiruvato (PEP) para producir KDOP, mientras que las DAHP sintasas catalizan la condensación entre D-eritrosa 4-fosfato (E4P) y PEP (Figura 3). Algunas KDOP sintasas requieren la presencia de un catión metálico divalente para la catálisis, mientras que otras no [Allison *et al.*, 2013]. A nivel de secuencia de aminoácidos, se ha determinado que las KDOP sintasas están relacionadas con una subfamilia de DAHP sintasas microbianas aproximadamente con un 30% de identidad entre ambas [Subramaniam *et al.*, 1998].

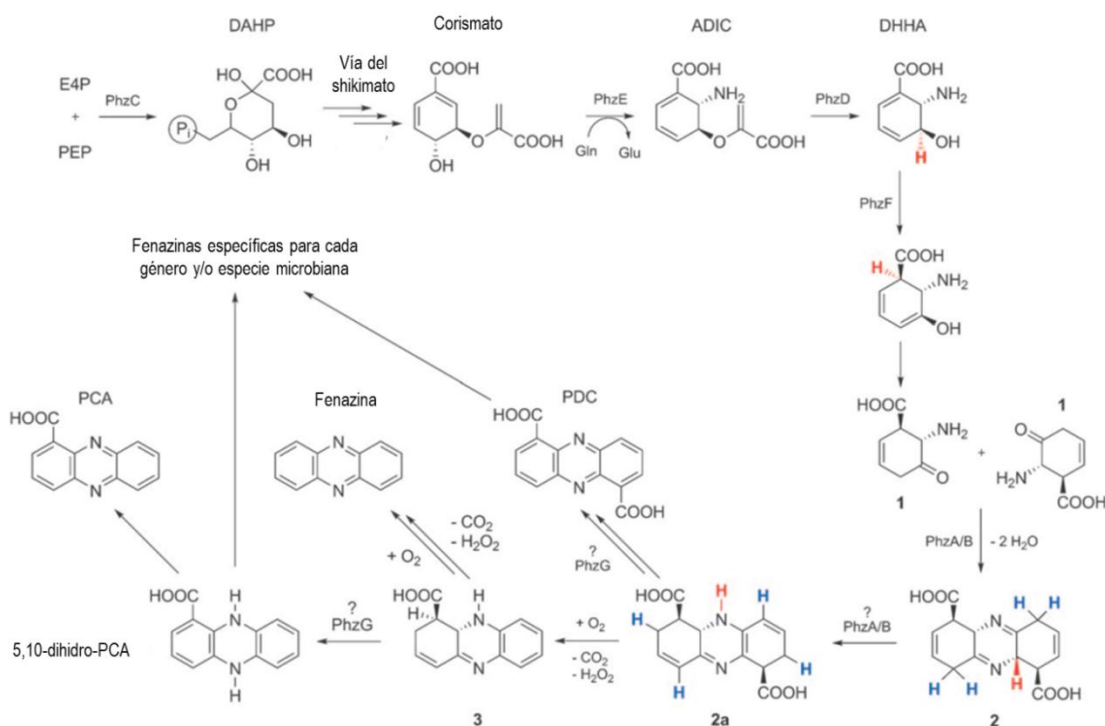


**Figura 3.** Reacciones catalizadas por la KDOP y DAHP sintasa. PEP: fosfoenolpiruvato; A5P: D-arabinosa 5-fosfato; E4P: D-eritrosa 4-fosfato; KDO8P: KDOP sintasa; DAH7P: DAHP sintasa [modificado de Allison *et al.*, 2013].

## I.4. DAHP sintasas en el metabolismo microbiano secundario

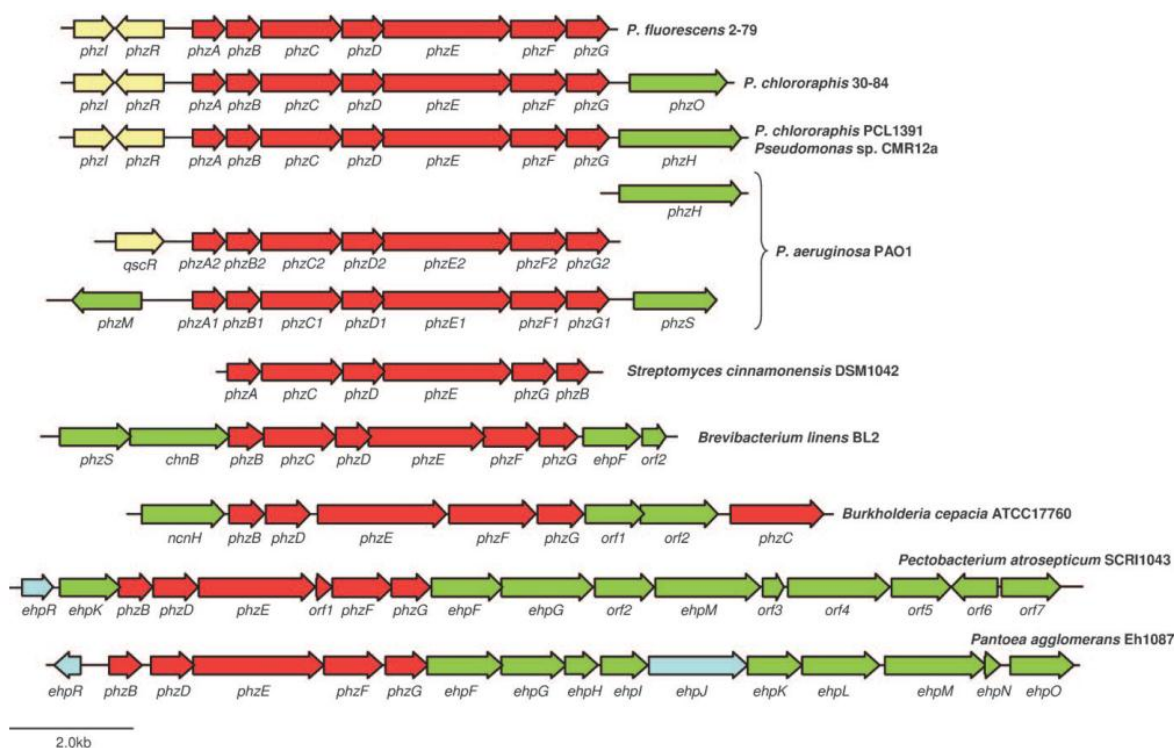
### I.4.1. Biosíntesis de fenazinas

En algunos microorganismos, existe una vía metabólica relacionada con la vía del shikimato, la cual está involucrada en la producción de compuestos heterocíclicos denominados fenazinas (Figura 4); estos compuestos son metabolitos secundarios y se producen principalmente en organismos de los Phyla Actinobacteria y Proteobacteria (clados  $\beta$ -Proteobacteria y  $\gamma$ -Proteobacteria). La distribución de los genes involucrados en la producción de fenazinas ha sido descrita ampliamente en algunos miembros de los géneros *Streptomyces*, *Burkholderia*, *Pseudomonas*, *Pectobacterium* y *Brevibacterium*. Los compuestos derivados de fenazinas poseen distintas propiedades que les confieren diferentes funciones tales como actividad antimicrobiana (bacteriana y/o fúngica), citotoxicidad en células eucariontes y transporte de electrones [Mentel *et al.*, 2009].



**Figura 4.** Vía de biosíntesis de fenazinas. E4P: eritrosa 4-fosfato; PEP: fosfoenol piruvato; DAHP: 3-deoxi-D-arabino-heptuloseonato 7-fosfato; ADIC: 2-amino-2-desoxicorismato; DHHA: *trans*-2,3-dihidro-3-hidroxi-antranilato; PCA: ácido fenazín-1-carboxílico; PDC: ácido fenazín-1,6-dicarboxílico [modificado de Mentel *et al.*, 2009].

Los genes involucrados en la producción de fenazinas se agrupan en un operón denominado *phz* que contiene todos los genes encargados para la síntesis de fenazinas (Figura 5). Dentro del operón *phz*, en algunos organismos, se encuentra el gen *phzC* que codifica para una DAHP sintasa; ésta tiene gran similitud con las DAHP sintasas de plantas superiores a nivel de secuencia, y asimismo ésta carece de control alostérico. La expresión de los genes que codifican para las enzimas involucradas en la biosíntesis de fenazinas se encuentra regulada por sistemas de respuesta dependientes de moléculas de señalización o autoinductores (*i.e.* sistema sensor de quórum) [Mentel *et al.*, 2009; Mavrodi *et al.*, 2010].

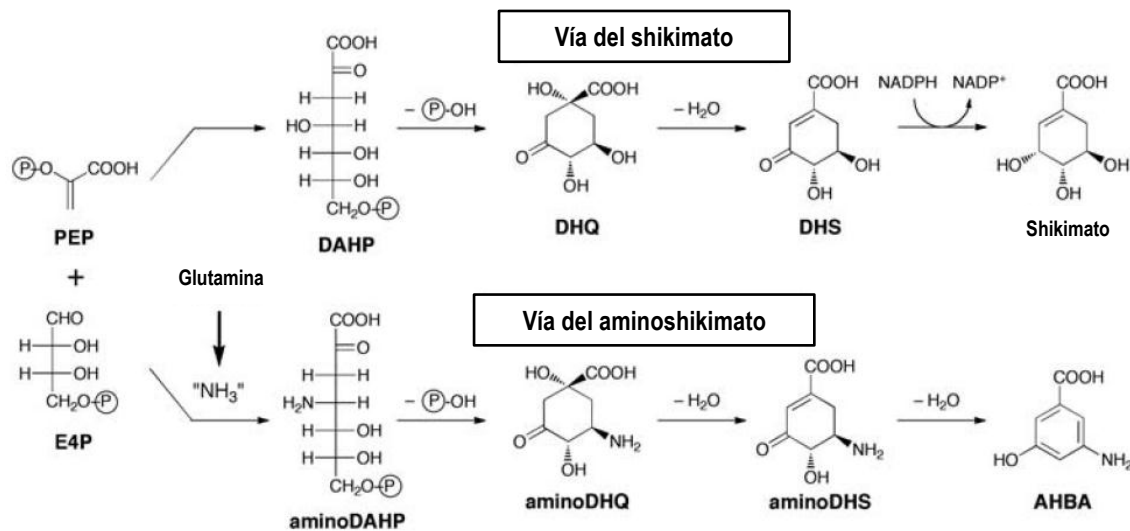


**Figura 5.** Compilación de operones bacterianos involucrados en biosíntesis de fenazinas. Los colores indican las funciones (ROJO: biosíntesis del núcleo de fenazinas; VERDE: modificación de fenazinas; CYAN: transporte y resistencia). Notar que la DAHP sintasa no está conservada en todos los operones *phz* y que la duplicación de los genes *phzA/phzB* está restringida a los organismos del género *Pseudomonas* [tomado de Mentel *et al.*, 2009].

#### 1.4.2. Biosíntesis de ácido 3-amino-5-hidroxibenzoico (AHBA). Vía del aminoshikimato

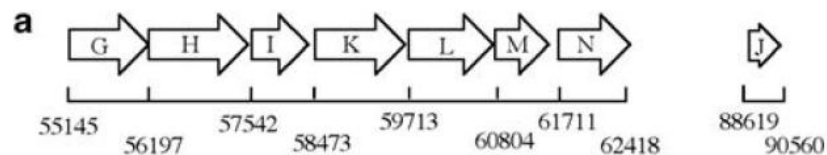
Existe una vía metabólica que se relaciona con la vía del shikimato y se encuentra presente en microorganismos como *Amycolatopsis mediterranei*, *Actinosynnema pretiosum*, y *Streptomyces collinus*. Esta vía metabólica se denomina vía del aminoshikimato, cuyo producto final es el ácido 3-amino-5-hidroxibenzoico (AHBA). Este compuesto es un intermediario clave del cual derivan muchos

antibióticos como la ansamicina, mitomicina, rifamicina, etc. Existe gran similitud entre los compuestos intermediarios de las vías del shikimato y aminoshikimato, en ésta última, el primer compuesto que es producido es el aminoDAHP, el cual es sintetizado por la enzima aminoDAHP sintasa mediante la condensación de E4P y PEP, además de una transferencia de un grupo amino producto de una reacción de transaminación con el aminoácido glutamina (Figura 6). El primer compuesto de la vía del aminoshikimato es el 4-amino-3,4-dideoxi-D-arabino-heptulosonato 7-fosfato, conocido como aminoDAHP [Guo & Frost, 2002; Floss *et al.*, 2011].



**Figura 6.** Vías del shikimato y aminoshikimato. PEP: fosfoenolpiruvato; E4P: eritrosa 4-fosfato; DAHP: 3-deoxi-D-arabino-heptulosonato 7-fosfato; DHQ: dehidroquinato; DHS: dehidroshikimato; AHBA: ácido 3-amino-5-hidroxibenzoico [modificado de Floss *et al.*, 2011].

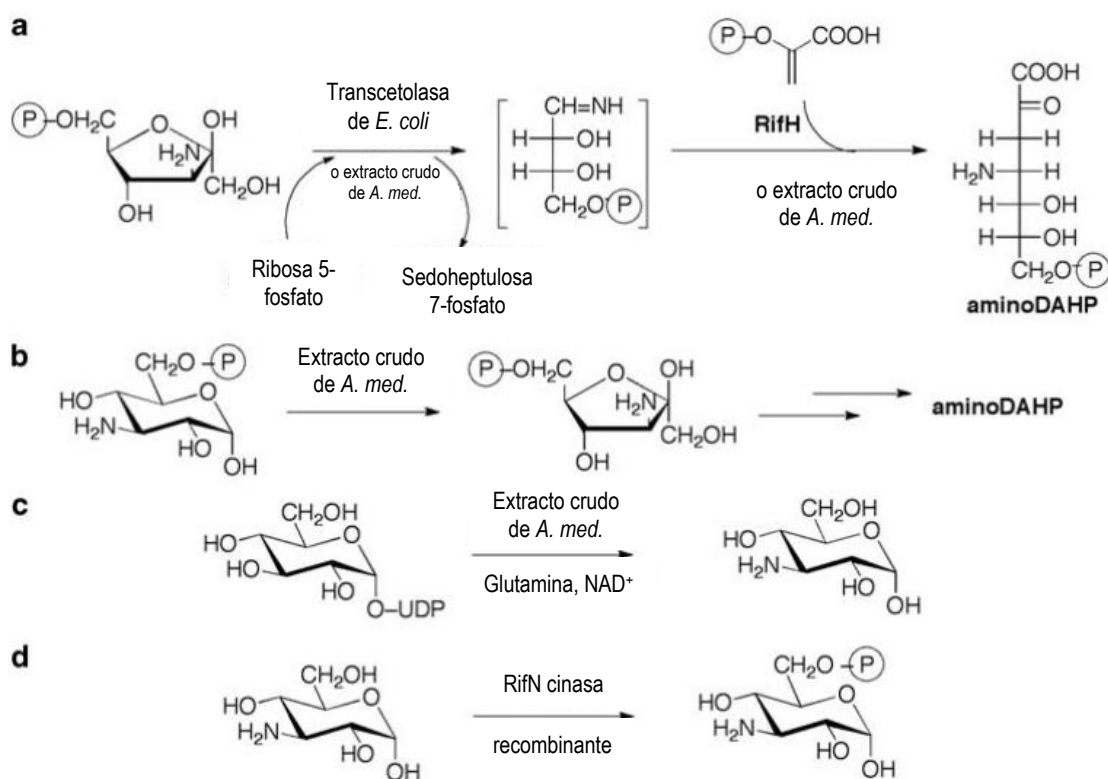
En *A. mediterranei*, los genes que participan en la biosíntesis del AHBA se encuentran agrupados en un operón denominado *rif* (Figura 7), dentro del cual la aminoDAHP sintasa se encuentra codificada en el gen *rifH*, la cual cataliza la reacción de condensación entre fosfoenolpiruvato (PEP) y 1-deoxi-1-imino-D-eritrosa 4-fosfato (iminoE4P) [Floss *et al.*, 2011].



**Figura 7.** Arreglo del operón *rif* en *Amycolatopsis mediterranei* [tomado de Floss *et al.*, 2011].



La formación del iminoE4P inicia a partir de UDP-glucosa, la cual mediante una reacción de transaminación, con glutamina como donador de grupo amino, es convertida en kanosamina. Posteriormente, mediante fosforilación enzimática, se obtiene kanosamina 6-fosfato. Este compuesto es transformado en su isómero para obtener 3-amino-3-deoxi-D-fructosa 6-fosfato, el cual mediante una transcetolasa (en el caso de *E. coli*) o en presencia de extracto crudo de *A. mediterranei*, es transformado en iminoE4P [Floss *et al.*, 2011].



**Figura 8.** Reacciones demostradas en presencia de extractos crudos de *Amycolatopsis mediterranei* o expresión heteróloga. (a) Formación de aminoDAHP a partir de 3-amino-3-deoxifruktosa 6-fosfato; (b) formación de aminoDAHP a partir de kanosamina 6-fosfato; (c) formación de kanosamina a partir de UDP-glucosa; (d) formación de kanosamina 6-fosfato a partir de kanosamina con proteína recombinante RifN [modificado de Floss *et al.*, 2011].

## I.5. Metagenoma y su aplicación biotecnológica

### I.5.1. Diversidad microbiana del suelo

El suelo es uno de los ambientes microbianos más complejos debido al tamaño y diversidad de comunidades microbianas que pueden existir en él. Para el estudio de esta diversidad, se han empleado diversas estrategias, siendo el crecimiento directo en medios de cultivo una de las

primeras en utilizarse; sin embargo, considerando los estudios que se han realizado a partir de cultivo directo, sólo del 0.1 al 1.0% del total de la microbiota se considera cultivable. Debido a esto ha sido crucial el desarrollo de estrategias de estudio independientes al cultivo que permitan un análisis más detallado de la diversidad microbiana, siendo una de éstas el estudio a nivel genómico [Daniel, 2005]. Una de las herramientas recientemente utilizadas en los análisis de diversidad microbiana del suelo es el estudio del metagenoma, definido éste como el conjunto del material genético de la microbiota de un ambiente determinado que, en teoría, representa el genoma de todos los microorganismos presentes, en este caso, en una muestra de suelo [Handelsman *et al.*, 1998].

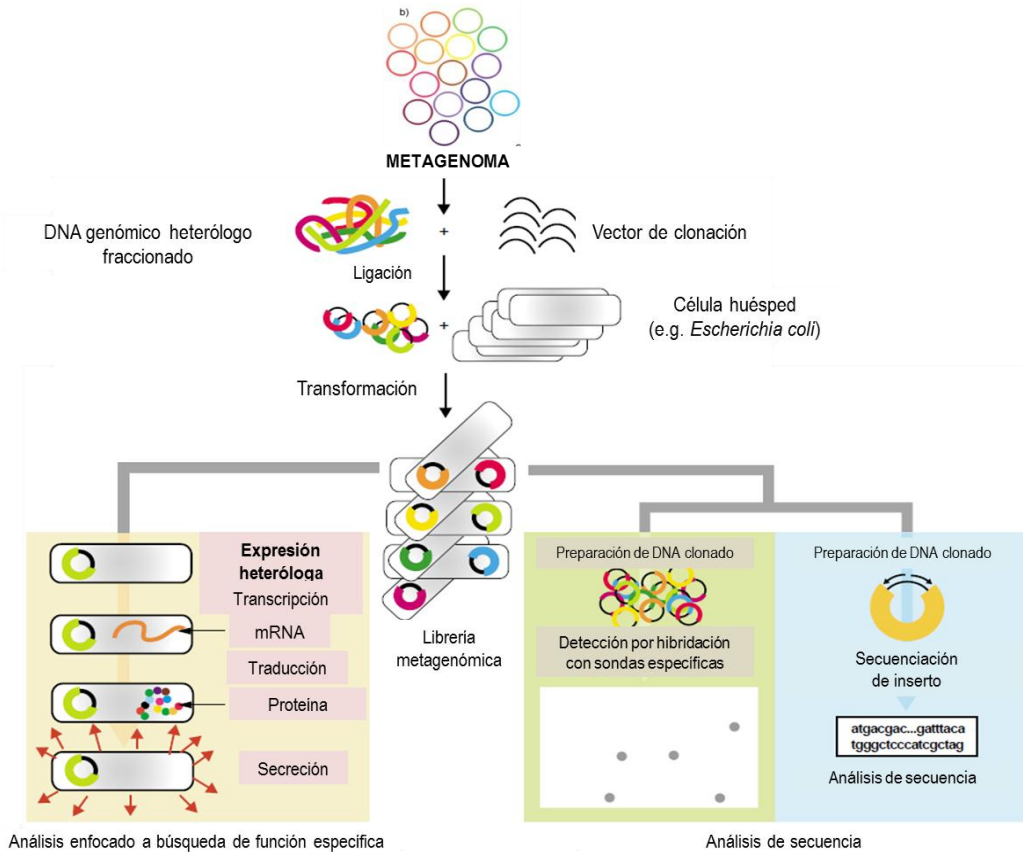
### **I.5.2. Bancos o librerías metagenómicas**

El estudio del metagenoma consiste en la construcción de librerías o bancos metagenómicos, que se inicia a partir de la fragmentación del DNA metagenómico mediante enzimas de restricción o acción mecánica, posterior a la inserción de los fragmentos de DNA en un vector apropiado, y finalmente la transformación del vector recombinante en un hospedero adecuado. En el caso del metagenoma de suelo, las librerías de genes clonados en plásmidos, cósmidos, fósidos o BACs como vectores han permitido detectar nuevos genes que codifican para enzimas y antibióticos con potencial impacto industrial. Algunos de estos genes han mostrado escasa homología con genes previamente reportados, derivando en el interés para utilizar genes clonados a partir de comunidades del suelo que codifican para lipasas, proteasas, oxidorreductasas, amilasas, antibióticos o enzimas que confieren resistencia a antibióticos y proteínas membranales [Daniel, 2005].

Mientras que la amplificación de genes por reacción en cadena de la polimerasa (PCR) requiere de un conocimiento previo de la secuencia del gen blanco para el diseño de iniciadores, la extracción directa y clonación del DNA metagenómico puede, en teoría, permitir el acceso a genes de cualquier secuencia o función. La clonación directa de genes ofrece la oportunidad de detectar operones que codifican para vías metabólicas que pueden dirigir la síntesis de moléculas químicamente complejas. También es posible obtener la información de la secuencia de los genes que flanquean al gen de particular interés, proporcionando así un panorama sobre el contexto

genómico del gen o de la relación filogenética del organismo del cual derivaron estos genes [Schloss & Handelsman, 2003].

El estudio de un banco metagenómico puede abordarse desde dos enfoques: la búsqueda de una función específica o su análisis a nivel de secuencias (Figura 9). El análisis dirigido a la búsqueda de una función específica es iniciado por la identificación de clonas que expresan un fenotipo deseado, seguido de una caracterización de las clonas activas por análisis bioquímico y de secuencia. La estrategia dirigida al análisis de secuencia radica en el uso de secuencias conservadas de DNA utilizadas como sondas de hibridación o iniciadores de PCR para la detección de secuencias de interés dentro de una librería metagenómica [Schloss & Handelsman, 2003].



**Figura 9.** Esquema general de construcción y análisis de una librería metagenómica [modificado de Schloss & Handelsman, 2003].

Dentro de la búsqueda de función o actividad específica en un banco metagenómico existe un estudio denominado complementación heteróloga; éste consiste en la exploración de genes

heterólogos para lograr la complementación genética y funcional en una célula huésped, resultando en la expresión de un gen que es esencial para el crecimiento bajo condiciones específicas de selección. Esta técnica es considerada de alta especificidad y selectividad [Ekkers *et al.*, 2012].

## II. ANTECEDENTES

Los primeros estudios sobre homología de la enzima DAHP sintasa fueron realizados a nivel de secuencia de aminoácidos realizando estudios de comparación y análisis de similitud entre las secuencias de DAHP sintasas microbianas y DAHP sintasas de plantas superiores. En 1991, Keith *et al.* realizaron un estudio de complementación funcional en *Saccharomyces cerevisiae* con una DAHP sintasas de *Arabidopsis thaliana*; en el análisis de similitud entre las secuencias de aminoácidos de las DAHP sintasas de *Solanum tuberosum* (papa), *A. thaliana*, *S. cerevisiae* y *E. coli* se observó que la similitud es comparativamente baja entre las DAHP sintasas de origen microbiano y las de plantas superiores (Figura 10 y Tabla 1).

```

DHS1      MALSNASS----LSTRSIYGGDLSHRPSNRQSSFT-----FHPAVNTKPKSVNL--VTAVHAAEPARNAVSVKESVASSSSGALK--WTPESWKLKALQLPDYPNAN
DHS2      -----MVTLNASSPLTTKS.LPYRHAPRRPISFSP.F...STD.KKSTQ.ASA.-----K.SL...S.....D.K
POTATO    ...ST.TTNSL.PN..LVQNQPLLPSPLKNAFFSNNSTKTVR.VQPISAVHS.DSN--KIPIVSDK.SKSPF.AATATTAPAPAVT.TE.AVD...S.....E....E

DHS1      ELESVLKTIIEAFPPIVFAGEARNLEERLADAAGVGFLLQGGDCAESFKEFNATNIRDTRVLLQMSIVLTFGGQVPIKVGMRMAGQFAKPRSDAFEKDFVKLPSYKGDN
DHS2      DVD..Q.LSS.....K..DK.GQ..M.Q..M.....N.....GV..M...L.....L.P.....R...
POTATO    ..R.....DE.....S.....GE..M.R.....N.....I...GA..M...M.....S.....R...

DHS1      INGDTFDEKSRIPDPNRMIRAYTQSAATLNLLRAFATGGYAAIQRVTOQWNLDFVEQEQADRYQELANRVDEALGFMSACGLGTDHPLMTTDFYTSHECLLLPYEQSLTR
DHS2      ...A.....H..V.....V.....M..S.....TQH...G...R.....G.A..TSA.....E.W.....A...
POTATO    V...A..V...T...Q.L...C.....M..IN.....TDH...G...R...S.....T.A..TM...K..E.W.....

DHS1      LDSTSGLYDCSAHMVWCGERTRLDGAHVEFLRGIANPLGIKVSNKMDPFELVKLVEILNPNNKGRITVIVRMGAENMRVKLPHLIRAVRRSGQIVTWVCDPMHGNTIK
DHS2      E.....L.V.....D..V.S...I...Q.....N.....GA.....S.....M
POTATO    R.....FL.V.....D...SA...I...Q..A...I.T.....A.....S.....

DHS1      STCGLKTRAFDSILAEVRAFLDVHEQEGSHAGGIHLEMTGQNVTECIGGSRVTYDDLSSRYHTHCDPRLNASQSLELAFIVAERLKRRTGSQRVS
DHS2      APG....S..A.R..L...F...D...FP..V.....V...I..N.....L..GNLPSIGV
POTATO    AP....P...R...F...D...P..V.....F.....SLSLNV

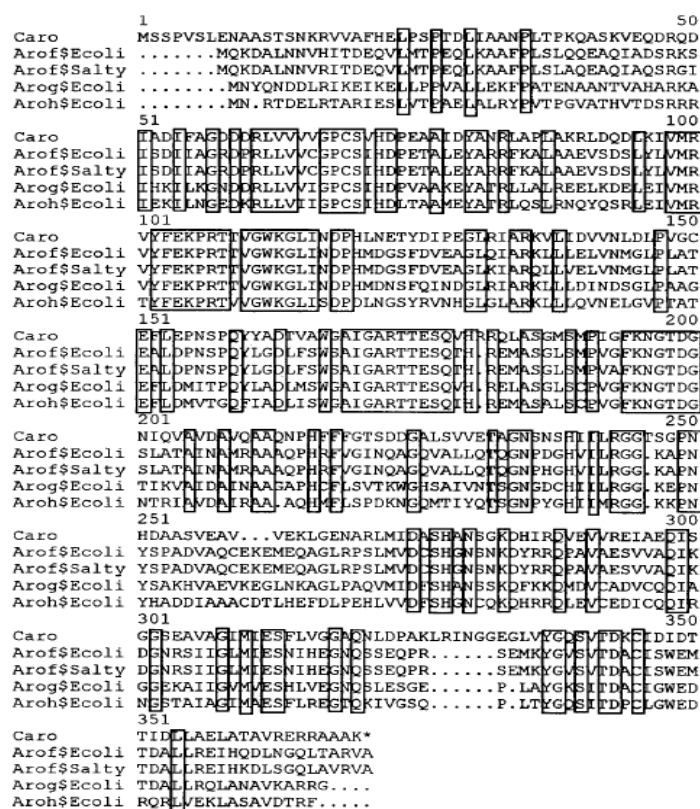
```

**Figura 10.** Alineamiento de las secuencias de aminoácidos de las DAHP sintasas de *Solanum tuberosum* (papa) y *Arabidopsis thaliana* (DHS1 y DHS2). Los puntos representan identidad con la secuencia DHS1 [tomado de Keith *et al.*, 1991].

**Tabla 1.** Porcentajes de similitud e identidad entre las secuencias de DAHP sintasas de *Escherichia coli*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana* y *Solanum tuberosum* (papa). Los números en paréntesis representan el porcentaje de identidad; los números sin paréntesis representan el porcentaje de similitud [tomado de Keith *et al.*, 1991].

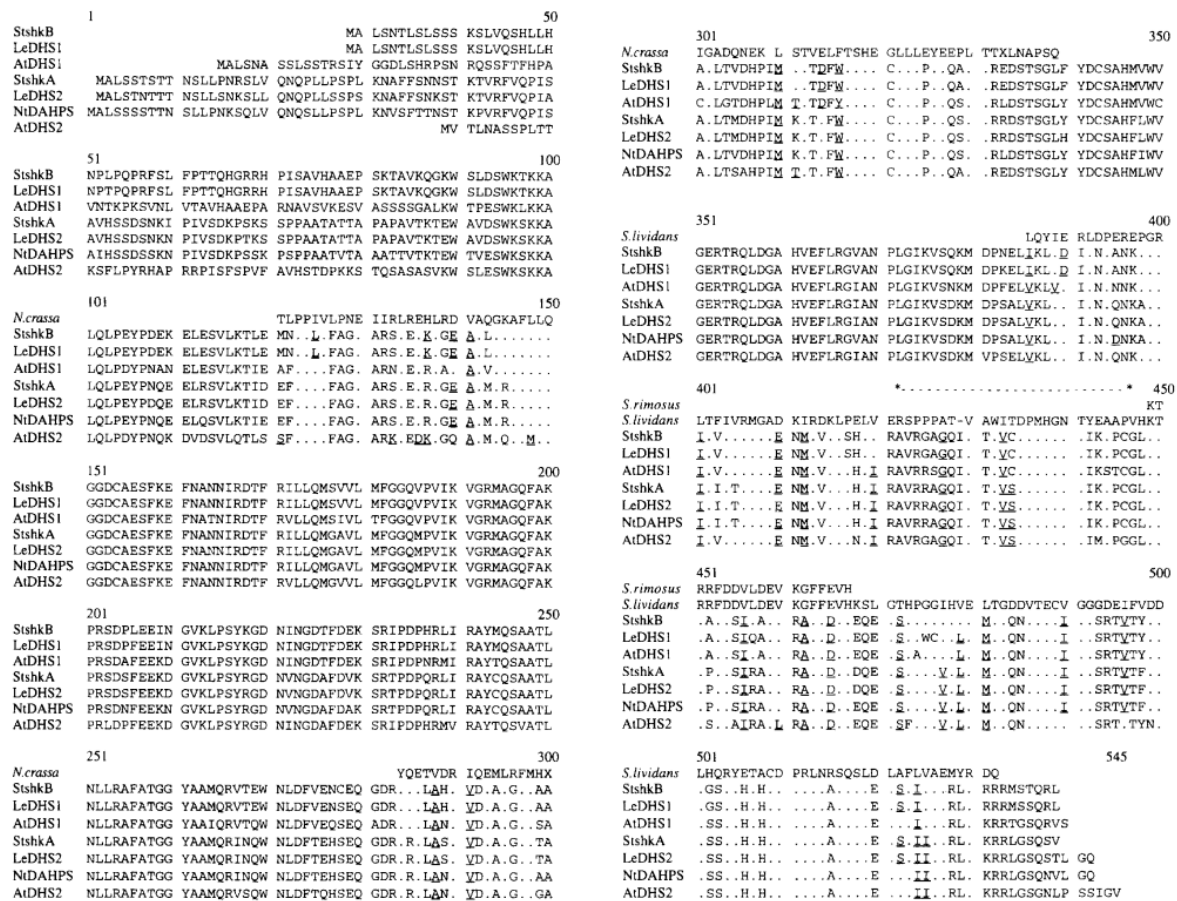
|       | <i>Arabidopsis</i><br>DHS2 | Potato<br>λBD21 | <i>E. coli</i> |         |         | Yeast<br>ARO3 |
|-------|----------------------------|-----------------|----------------|---------|---------|---------------|
|       |                            |                 | AroF           | AroG    | AroH    |               |
| DHS1  | 85 (76)                    | 82 (73)         | 35 (13)        | 35 (13) | 30 (11) | 36 (16)       |
| DHS2  |                            | 86 (79)         | 35 (12)        | 39 (12) | 35 (13) | 35 (15)       |
| λBD21 |                            |                 | 37 (14)        | 32 (11) | 30 (13) | 37 (16)       |
| AroF  |                            |                 |                | 71 (53) | 67 (46) | 71 (49)       |
| AroG  |                            |                 |                |         | 72 (56) | 72 (55)       |
| AroH  |                            |                 |                |         |         | 63 (43)       |

En 1993, Chen *et al.* realizaron un estudio de complementación funcional y expresión heteróloga de la DAHP sintasa de una cepa de *Corynebacterium glutamicum* en *E. coli*, y una comparación entre las secuencias de aminoácidos de DAHP sintasas de *C. glutamicum*, *E. coli* y *Salmonella typhimurium*, mostrando una identidad promedio del 50% entre éstas (Figura 11).



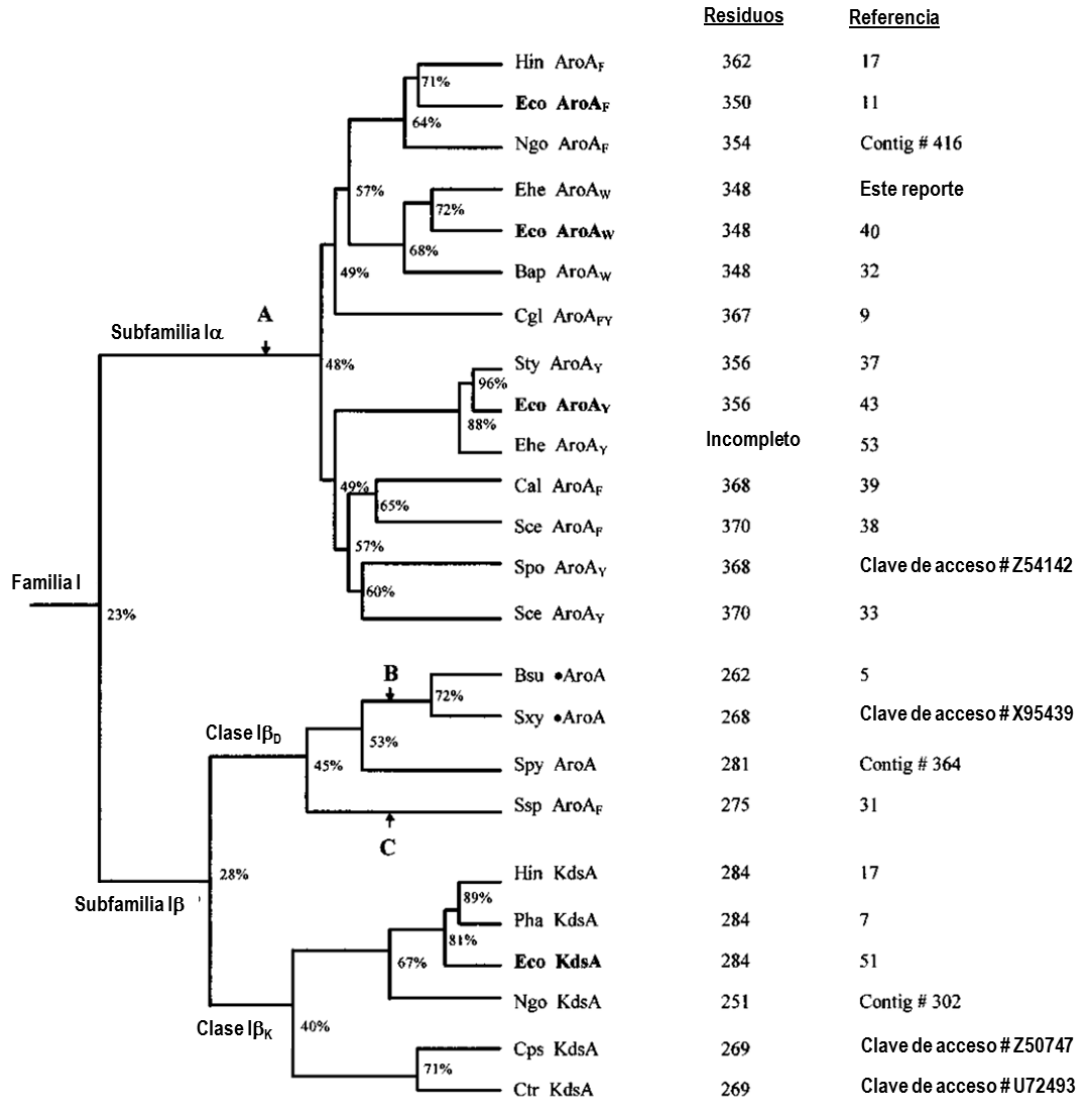
**Figura 11.** Comparación de las secuencias de aminoácidos de las DAHP sintasas de *Corynebacterium glutamicum*, *Escherichia coli* y *Salmonella typhimurium*. Aquellos aminoácidos idénticos se encuentran rodeados de rectángulos. Caro, DAHP sintasa de *C. glutamicum*; Arof\$Ecoli, DAHP sintasa de *E. coli* sensible a inhibición por tirosina; Arof\$Salty, DAHP sintasa de *S. typhimurium* sensible a inhibición por tirosina; Arog\$Ecoli, DAHP sintasa de *E. coli* sensible a inhibición por fenilalanina; Aroh\$Ecoli, DAHP sintasa de *E. coli* sensible a inhibición por triptófano [tomado de Chen *et al.*, 1993].

Posteriormente, y en base a los estudios previos, Walker *et al.* propusieron una clasificación para las DAHP sintasas, agrupándolas como DAHP sintasas tipo I (origen microbiano), con una subunidad monofuncional  $M_r$  de aproximadamente 39 000, y DAHP sintasas tipo II (plantas superiores) con una subunidad monofuncional  $M_r$  de aproximadamente 54 000. Sin embargo, en su estudio se obtuvo que las secuencias de aminoácidos de las DAHP sintasas de *Streptomyces coelicolor*, *Streptomyces rimosus* y *Neurospora crassa* eran más similares a las DAHP sintasas de plantas superiores que a las microbianas (Figura 12), planteando así la necesidad de elaborar esquemas de clasificación más complejos basados en la relación filogenética de las secuencias de DAHP sintasas.



**Figura 12.** Alineamiento múltiple de las secuencias de DAHP sintasas de *Streptomyces rimosus*, *Neurospora crassa* y *Streptomyces lividans* con algunas DAHP sintasas de plantas superiores. AtDHS1 y AtDHS2, *Arabidopsis thaliana*; StshkA y StshkB, papa; NiDAHPS, tabaco; LeDHS1 y LeDHS2, jitomate. Los puntos representan identidad respecto a las secuencias de *Neurospora crassa* o *Streptomyces lividans* y las sustituciones conservativas de aminoácidos se encuentran subrayadas [tomado de Walker *et al.*, 1996].

A partir de un estudio sobre la especificidad de sustrato de la enzima 3-deoxi-D-mano-octulosonato 8-fosfato (KDOP) sintasa de *Neisseria gonorrhoeae*, Subramaniam *et al.* (1998) realizaron una nueva clasificación de la familia de DAHP sintasas tipo I (DAHP sintasas microbianas). En este análisis se divide a esta familia en dos subfamilias, ilustradas en el dendrograma de la Figura 13. La subfamilia I $\alpha$  comprende únicamente a DAHP sintasas, mientras que la subfamilia I $\beta$  agrupa tanto a DAHP sintasas (denominadas I $\beta_D$ ) como KDOP sintasas (denominadas I $\beta_K$ ). Las proteínas de la subfamilia I $\beta$  tienen un peso molecular de 28.5 a 30.8 kDa, las cuales son de menor tamaño que las de la subfamilia I $\alpha$ . Los niveles de identidad promedio entre algunos miembros de las subfamilias I $\alpha$  y I $\beta_D$ , o entre miembros de las subfamilias I $\alpha$  y I $\beta_K$  son suficientemente bajos que la homología no es apreciable. Sin embargo, en su alineamiento múltiple de proteínas se muestra que las tres subfamilias comprenden una familia de proteínas homólogas que comparten 17 residuos invariables de aminoácidos. Aunque los miembros de la clase I $\beta_D$  son funcionalmente equivalentes a los miembros de la subfamilia I $\alpha$  en el sentido de que catalizan la reacción de DAHP sintasa, en general exhiben mayor similitud con los miembros de la clase I $\beta_K$ . La clase I $\beta_D$  de DAHP sintasas posee 17 residuos que son conservados con la subfamilia I $\alpha$ , pero no con la clase I $\beta_K$  de KDOP sintasas. La clase I $\beta_D$  de DAHP sintasas posee 19 residuos adicionales que son conservados en la clase I $\beta_K$  de KDOP sintasas, pero no en la subfamilia I $\alpha$  de DAHP sintasas. Este estudio proporciona la primera clasificación de DAHP sintasas en la que incluye las secuencias de KDOP sintasas.



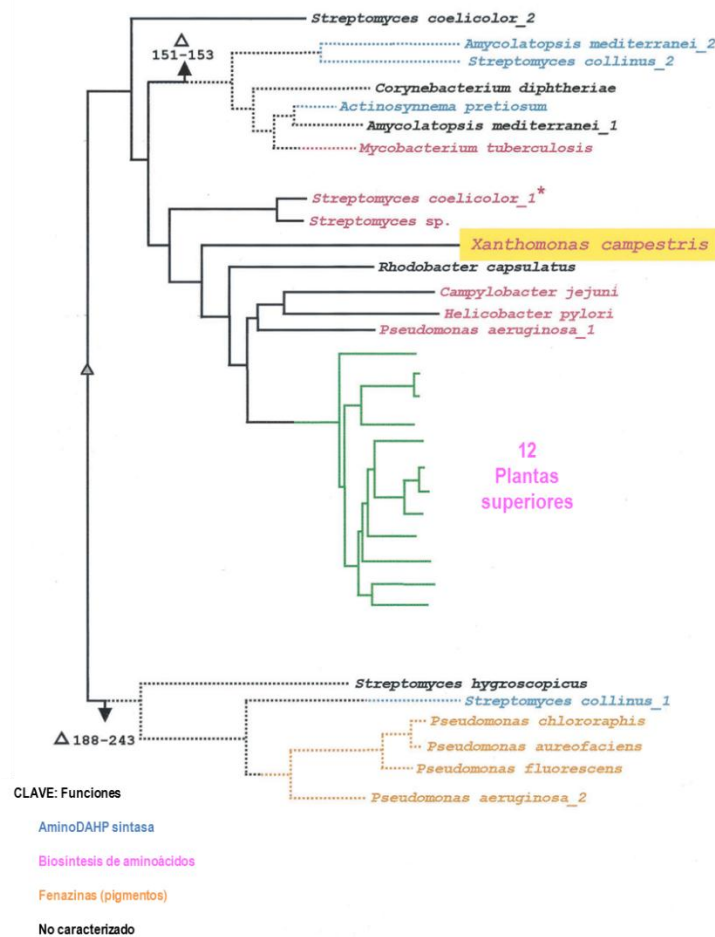
**Figura 13.** Dendrograma (PILEUP) mostrando las relaciones de la familia I de DAHP sintasas. Los porcentajes promedio de identidad de las secuencias deducidas de aminoácidos están indicados en los nodos. Abreviaturas: Hin, *Haemophilus influenzae*; Eco, *Escherichia coli*; Ngo, *Neisseria gonorrhoeae*; Ehe, *Erwinia herbicola*; Bap, *Buchnera aphidicola*; Cgl, *Corynebacterium glutamicum*; Sty, *Salmonella typhimurium*; Cal, *Candida albicans*; Sce, *Saccharomyces cerevisiae*; Spo, *Schizosaccharomyces pombe*; Bsu, *Bacillus subtilis*; Sxy, *Staphylococcus xylosus*; Spy, *Streptococcus pyogenes*; Ssp, *Synechocystis* sp.; Pha, *Pasteurella haemolytica*; Cps, *Chlamydia psittaci*; y Ctr, *Chlamydia trachomatis* [adaptado de Subramaniam *et al.*, 1998].

Posteriormente Gosset *et al.* (2001) realizaron una nueva clasificación de DAHP sintasas, reorganizándolas en dos familias: AroA<sub>I</sub> y AroA<sub>II</sub>. La familia AroA<sub>I</sub> se divide en dos subfamilias: AroA<sub>Iα</sub> (representado por bacterias Gram-negativas con parálogos diferencialmente regulados, como en el caso de *E. coli*) y AroA<sub>Iβ</sub>, como las 3-deoxi-D-mano-octuloso-8-fosfato (KDOP) sintasas o DAHP sintasas de bacterias Gram-positivas como *Bacillus subtilis*. La familia AroA<sub>II</sub> está constituida



básicamente por DAHP sintasas de plantas superiores. Como en el caso de *Streptomyces* spp. y *N. crassa*, se relacionó la DAHP sintasa de *Xanthomonas campestris* con las de la familia AroA<sub>II</sub>, estableciendo una relación filogenética y evolutiva más cercana a las DAHP sintasas de plantas superiores.

El estudio de Gosset *et al.* (2001) sobre la DAHP sintasa de *X. campestris* destaca también la presencia de algunos motivos relevantes en esta proteína (Figura 15). Particularmente se distingue la presencia del motivo 140-KPRS-143 en la DAHP sintasa de *X. campestris*, el cual constituye un motivo conservado en las DAHP sintasas de la familia AroA<sub>II</sub>; además que al compararlo con la familia AroA<sub>I</sub> (e.g. *E. coli*) se localiza también una homología en el motivo 97-KPRT-100. Asimismo, otro motivo de interés es el 374-DxxHxN-379 de *X. campestris* con el 265-DxxHxN-270 de *E. coli* (familia AroA<sub>I</sub>).



**Figura 14.** Árbol filogenético de DAHP sintasas de la familia AroA<sub>II</sub>. Un alineamiento múltiple fue obtenido mediante el programa Multalin, y fue analizado por el método *Neighbor-Joining* utilizando el programa Phylip [tomado de Gosset *et al.*, 2001].

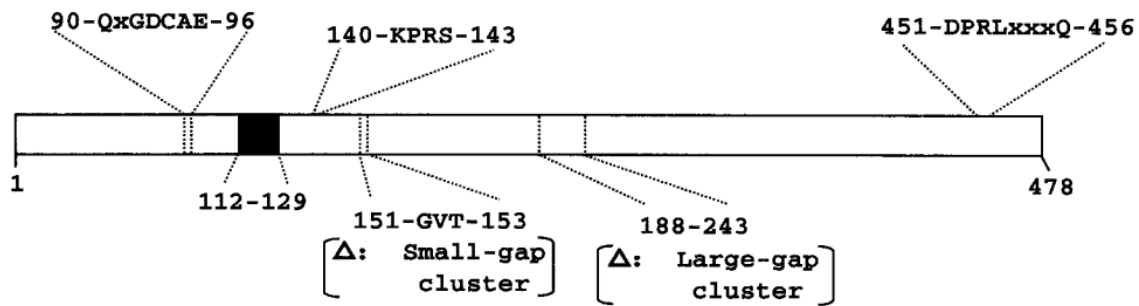
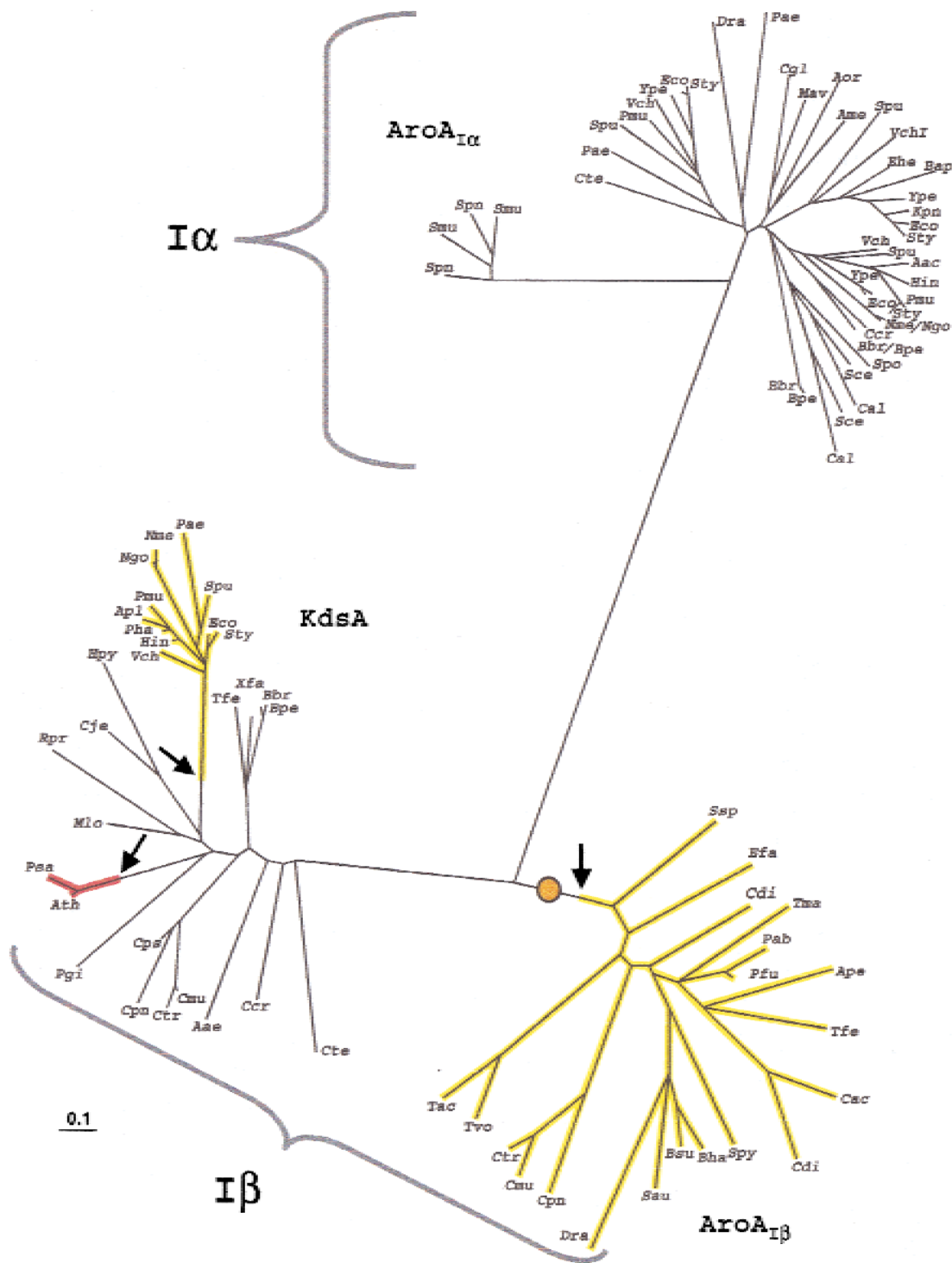
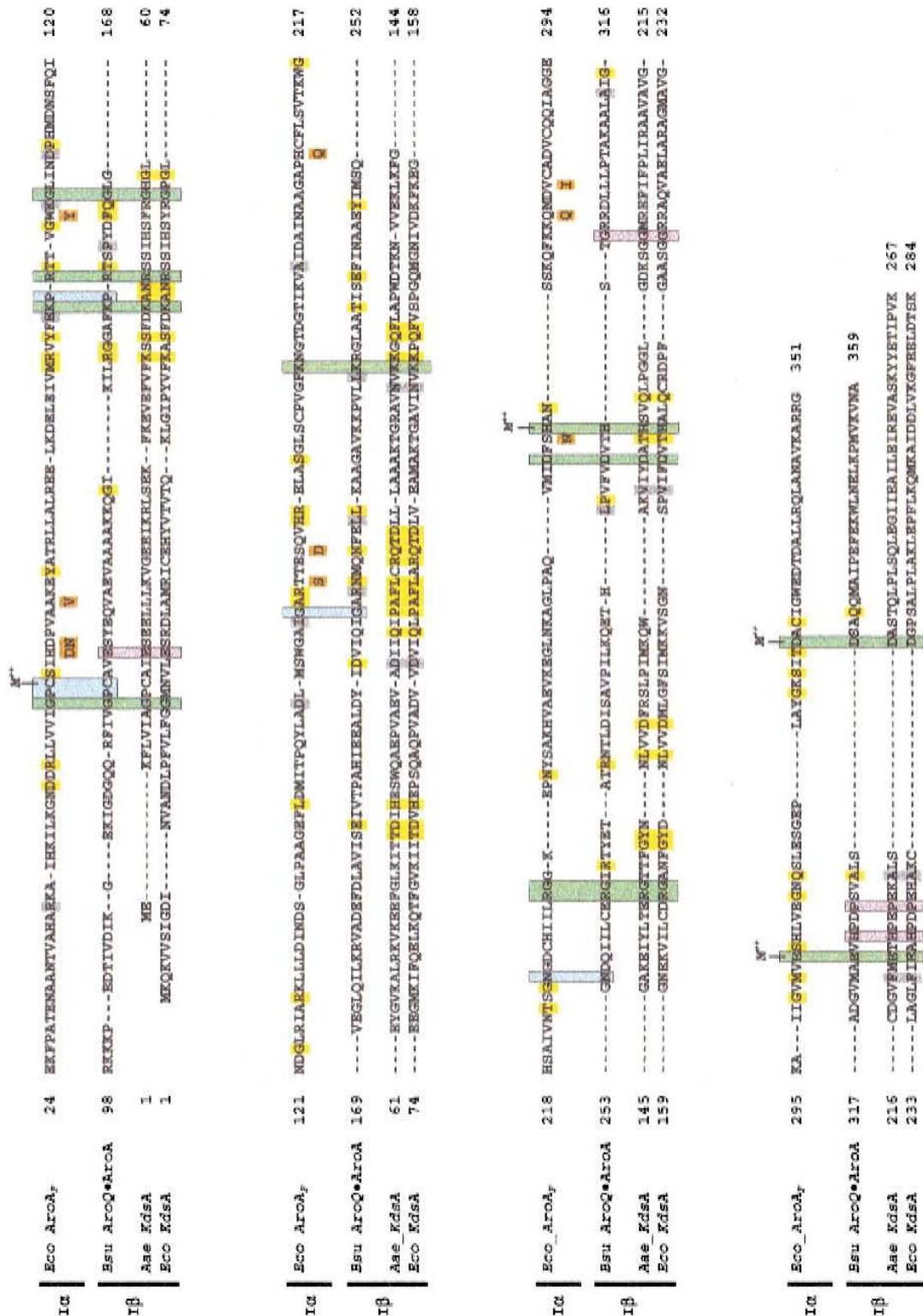


Figura 15. Esquema de la secuencia de DAHP sintasa de *Xanthomonas campestris* [tomado de Gosset *et al.*, 2001].

La clasificación propuesta por Gosset *et al.* (2001) ha servido como modelo para posteriores estudios sobre DAHP sintasas; sin embargo, habiendo realizado una búsqueda en la literatura, no se ha modificado este esquema de clasificación de DAHP sintasas. Posteriormente Jensen *et al.* (2002) realizaron una clasificación del grupo AroA<sub>Iβ</sub> en el que se demuestra la divergencia filogenética entre las DAHP sintasas de este grupo y aquellas enzimas que catalizan la biosíntesis de KDOP; en este análisis filogenético se presenta la cercanía del grupo de DAHP sintasas AroA<sub>Iβ</sub> hacia el grupo de KDOP sintasas (denominado KdsA), así como una clara divergencia con el grupo de DAHP sintasas AroA<sub>Iα</sub>. En este estudio también se realiza un alineamiento de algunas secuencias de DAHP y KDOP sintasas microbianas representativas de las familias caracterizadas hasta el momento, resaltando los residuos de aminoácidos conservados entre estas secuencias (Figuras 16 y 17).



**Figura 16.** Árbol filogenético de homología de DAHP sintasas de las familias AroA<sub>Iα</sub> y AroA<sub>Iβ</sub> con las KDOP sintasas (KdsA). El linaje que lleva a las secuencias de dos KDOP sintasas de plantas superiores se resalta en color rojo. La raíz propuesta se representa con un círculo amarillo, y los eventos evolutivos que llevaron a la independencia de metales para la actividad enzimática son señalados con flechas [tomado de Jensen *et al.*, 2002].



**Figura 17.** Alineamiento de secuencias de DAHP y KDOP sintasas microbianas. Eco AroA<sub>f</sub> (DAHP sintasa de *Escherichia coli* representativa de la subfamilia AroA<sub>1α</sub>), Bsu AroQ•AroA (DAHP sintasa de *Bacillus subtilis* representativa de la subfamilia AroA<sub>1β</sub>), Aae KdsA (KDOP sintasa de *Aquifex aeolicus*) y Eco KdsA (KDOP sintasa de *Escherichia coli*) [tomado de Jensen *et al.*, 2002].

Actualmente existen pocos reportes sobre estudios dedicados a la búsqueda metagenómica dirigida hacia genes involucrados en la biosíntesis de compuestos aromáticos (metabolismo microbiano primario). Algunos de los estudios más recientes con respecto al estudio de librerías metagenómicas y a la vía de biosíntesis de compuestos aromáticos se han efectuado sobre genes de la vía biosintética de triptófano [Li *et al.*, 2005], búsqueda de una enzima 5-enoilpiruvilshikimato 3-fosfato sintasa (EPSPS) resistente a glifosato [Jin *et al.*, 2007] y de la enzima prefenato deshidrogenasa [Jiang *et al.*, 2013]. Sin embargo, la búsqueda de genes que codifican para la enzima DAHP sintasa a partir de un banco metagenómico no ha sido abordada en estudios recientes.

### III. JUSTIFICACIÓN

Debido a la importancia del compuesto 3-deoxi-D-*arabino*-heptulosoato 7-fosfato (DAHP) como precursor en la producción de diversos compuestos aromáticos esenciales para el metabolismo microbiano, siendo algunos de éstos de interés biotecnológico, y ante los estudios de diversos grupos de investigación sobre la caracterización y clasificación de DAHP sintasas, este estudio pretende realizar una búsqueda de los genes que codifican para estas enzimas en un banco metagenómico de suelo; asimismo, de manera paralela, se realizará un análisis de diversidad con el propósito de ampliar el conocimiento sobre la evolución, distribución y clasificación de las diferentes familias de DAHP sintasas conocidas.

### IV. HIPÓTESIS

La presencia de una gran diversidad genómica y microbiana (cultivable o no cultivable) en el suelo permitirá identificar DAHP sintasas de origen microbiano mediante complementación funcional y expresión heteróloga en *E. coli*, a partir de un banco metagenómico de suelo.

## V. OBJETIVOS

### V.1. Objetivo general

Detectar y caracterizar genes que codifican para DAHP sintasas microbianas presentes en un banco metagenómico de suelo.

### V.2. Objetivos particulares

- Construcción de un marco de referencia filogenético sobre la diversidad de familias de DAHP sintasas (AroA<sub>α</sub>, AroA<sub>β</sub>, KdsA, PhzDAHPS y aminoDAHPS):
  - Análisis del contexto genómico de secuencias de DAHP sintasas
  - Reconstrucción y distribución filogenética de familias de DAHP sintasas
- Obtención de una cepa de *E. coli*  $\Delta$ aroF  $\Delta$ aroG  $\Delta$ aroH carente de actividad de DAHP sintasa.
- Detección de DAHP sintasas a partir de DNA metagenómico de suelo por complementación funcional de una cepa de *E. coli*  $\Delta$ aroF  $\Delta$ aroG  $\Delta$ aroH.
- Caracterización y análisis de clonas detectadas:
  - Análisis de plásmidos en función de perfiles de restricción enzimática
  - Secuenciación y búsqueda de identidad de insertos
  - Determinación de cinética de crecimiento de clonas
  - Determinación de actividad enzimática específica de DAHP sintasa
- Análisis filogenético de secuencias de insertos dentro del contexto de secuencias representativas de todos los grupos de DAHP sintasas

## VI. MATERIALES Y MÉTODOS

### VI.1. Análisis del contexto genómico de DAHP sintasas de las subfamilias AroA<sub>Iα</sub> (e.g. *E. coli*) y AroA<sub>Iβ</sub> (e.g. *B. subtilis*)

En bacterias, los genes adyacentes localizados a un gen específico (contexto genómico) en muchos casos están relacionados entre sí a nivel de bioquímica y fisiología celular. La ubicación de estos genes en operones, así como el cambio en la configuración de éstos a través del tiempo, puede proporcionar información sobre eventos de divergencia evolutiva, o bien conducir a la identificación de genes ortólogos en otros organismos; asimismo, la conservación o cambio en la función biológica es otro dato importante a considerar en el estudio de estos genes.

De acuerdo a los estudios previos de clasificación de las DAHP sintasas, fueron seleccionadas una secuencia de aminoácidos de DAHP sintasa representativa del grupo AroA<sub>Iα</sub> (gen *aroF* de *E. coli*) y otra del grupo AroA<sub>Iβ</sub> (gen *aroA* de *B. subtilis*). Estas secuencias fueron analizadas mediante el programa GeConT I ([http://bioinfo.ibt.unam.mx/gecont\\_1/](http://bioinfo.ibt.unam.mx/gecont_1/), versión 2012) [Ciria *et al.*, 2004], ésta interfase permite visualizar el contexto genómico de un gen o grupo de genes y sus ortólogos en todos los genomas completamente secuenciados. Posteriormente, las DAHP sintasas desplegadas fueron agrupadas en árboles filogenéticos organizados en función de la secuencia del gen 16S rRNA de los organismos que corresponden estas DAHP sintasas.

### VI.2. Análisis de secuencias de DAHP sintasas (familias AroA<sub>I</sub> y AroA<sub>II</sub>), aminoDAHP sintasas y KDOP sintasas

Se realizó un análisis a partir de las secuencias de aminoácidos de las DAHP sintasas y KDOP sintasas obtenidas mediante el análisis del contexto genómico de las DAHP sintasas de *E. coli* y *B. subtilis* con el programa GeConT I. Además, fueron incluidas algunas secuencias de DAHP sintasas representativas de plantas superiores [*A. thaliana*, *S. tuberosum* (papa), *Solanum lycopersicum* (jitomate), *Nicotiana tabacum* (tabaco), *Zea mays* (maíz), y *Petroselinum crispum* (perejil)], así como las aminoDAHP sintasas conocidas y las DAHP sintasas asociadas a la síntesis de fenazinas.

### VI.3. Cepas bacterianas, plásmidos, antibióticos y condiciones de cultivo

Las cepas y plásmidos utilizados en este trabajo se enlistan en la Tabla 2, destacando algunas de sus características genotípicas relevantes. Para los cultivos de las cepas fueron utilizados los medios LB, M9, M9S y ARO. En la recuperación de células posterior al proceso de electroporación, se utilizó el medio SOC. Cuando fue requerido, se adicionaron a los medios los antibióticos cicloheximida (Chx), ampicilina (Ap), carbenicilina (Cb), kanamicina (Km) y tetraciclina (Tc); éstos dos últimos a una concentración de 15 y 30 µg/mL respectivamente, mientras que la carbenicilina y la cicloheximida fue utilizada a una concentración de 100 µg/mL en los experimentos de complementación funcional, y la ampicilina a 200 µg/mL en la escisión de gen de resistencia a kanamicina. La composición de los medios empleados en este trabajo se presenta en el Apéndice.

**Tabla 2.** Cepas y plásmidos utilizados en este trabajo.

| Cepas/Plásmidos              | Características relevantes   | Referencia                              |
|------------------------------|--|---|
| <b>Cepas</b>                 |  |   |
| <i>Escherichia coli</i> K-12 |  |   |
| BW25113                      | <i>rmB3 ΔlacZ4787 hsdR514 Δ(araBAD)567 Δ(rhaBAD)568 rph-1</i>  |   |
| BW25113 $\Delta aroF::kan$   | Gen cromosomal <i>aroF</i> interrumpido por gen de resistencia a kanamicina  | Baba <i>et al.</i> , 2006               |
| BW25113 $\Delta aroG::kan$   | Gen cromosomal <i>aroG</i> interrumpido por gen de resistencia a kanamicina  | Baba <i>et al.</i> , 2006               |
| BW25113 $\Delta aroH::kan$   | Gen cromosomal <i>aroH</i> interrumpido por gen de resistencia a kanamicina  | Baba <i>et al.</i> , 2006               |
| MG1655                       | F <sup>-</sup> $\lambda$ - <i>rph-1</i>  |   |
| T1.1, T1.2, T1.3             | Cepas BW25113 $\Delta aroF \Delta aroH$ producto de la primera transducción generalizada con el bacteriófago P1vir   | Este trabajo                            |
| T2.1, T2.2, T2.3, T2.4       | Cepas BW25113 $\Delta aroF \Delta aroH \Delta aroG::kan$ producto de la segunda transducción generalizada con el bacteriófago P1vir  | Este trabajo                            |
| <b>Plásmidos</b>             |  |   |
| Banco pJET/MS                | Metagenoma de suelo forestal de Corredor Biológico "Chichinautzin" clonado en vector pJET1.2/Blunt [replicón <i>rep</i> de plásmido pMB1, gen <i>bla</i> que codifica para una $\beta$ -lactamasa (Cb <sup>R</sup> o Ap <sup>R</sup> ), transcripción de inserto clonado bajo control de promotor de RNA polimerasa de bacteriófago T7], fragmentos clonados de tamaño molecular aproximado entre 5 y 10 kb, tamaño aproximado de banco: 0.43 Gb de DNA. | Sirena-Sánchez, 2009                    |
| pJLBaroG <sup>fbr</sup> tktA | Derivado de pJLBaroG <sup>fbr</sup> , que contiene el gen <i>tktA</i> con su promotor nativo   | Balderas-Hernández <i>et al.</i> , 2009 |
| pCP20                        | FLP <sup>+</sup> , $\lambda$ cI857 <sup>+</sup> , $\lambda$ pR Rep <sup>ts</sup> , Ap <sup>R</sup> , Cm <sup>R</sup>   | Cherepanov & Wackernagel, 1995          |
| pJET1.2-B                    | Vector pJET1.2 carente de un fragmento de 50 pb dentro del sitio de multiclonación   | Este trabajo                            |



#### **VI.4. Extracción de DNA cromosómico y plasmídico**

La extracción de DNA cromosómico fue realizada mediante el UltraClean® Microbial DNA Isolation Kit de la compañía MO-BIO Laboratories, Inc., siguiendo las instrucciones del fabricante. La extracción de DNA plasmídico fue realizada mediante el procedimiento de lisis alcalina [Birboim *et al.*, 1983]. A partir de un cultivo crecido a saturación, se tomó una alícuota de 1.5 mL para separar las células por centrifugación a 13 000 x *g* por 2 minutos. Después de separar el sobrenadante, se resuspendieron las células en 150  $\mu$ L de solución I fría y adicionando 5  $\mu$ L de RNasa; esta mezcla se incubó por 5 minutos a temperatura ambiente. Posteriormente se adicionó 350  $\mu$ L de la solución II, seguido de una mezcla por inversión y una incubación en hielo por 10 minutos. Después se agregaron 250  $\mu$ L de la solución III fría, se mezcló por inversión y la mezcla fue incubada en hielo por 10 minutos. Para separar el DNA plasmídico del DNA cromosómico y proteínas, se centrifugó a 13 000 x *g* por 15 minutos, transfiriendo el sobrenadante a un tubo nuevo. Para precipitar y concentrar el DNA plasmídico, se adicionó un volumen de 1 mL de isopropanol, posteriormente se mezcló por inversión y se centrifugó a 13 000 x *g* por 15 minutos. Después se realizó un lavado con 1 mL de etanol al 70 %, y finalmente se resuspendió en un volumen de 50  $\mu$ L de agua miliQ. La composición de las soluciones empleadas en este procedimiento se describen en el Apéndice de este trabajo.

#### **VI.5. Eliminación de genes cromosomales *aroF*, *aroG* y *aroH* en *E. coli***

En *E. coli*, los genes *aroF*, *aroG* y *aroH* codifican para la biosíntesis de tres DAHP sintetas, cada una de éstas se encuentra bajo control alostérico por los tres aminoácidos aromáticos: la DAHP sintasa sintetizada a partir del gen *aroF* se encuentra regulada por el aminoácido tirosina, el producto del gen *aroG* es regulado por fenilalanina, y el producto de *aroH* se regula por triptófano. Para obtener una cepa de *E. coli* carente de actividad de DAHP sintasa (DAHPS<sup>-</sup>) es necesario eliminar o inactivar los genes cromosomales *aroF*, *aroG* y *aroH*. La eliminación de estos genes se llevó a cabo mediante transducción generalizada, modificada de la propuesta por Miller en 1992, utilizando el bacteriófago P1*vir* como partícula transductante; fueron seleccionadas las cepas de *E. coli* BW25113  $\Delta$ *aroF*::*kan*,  $\Delta$ *aroG*::*kan*, y  $\Delta$ *aroH*::*kan* como cepas donadoras en la técnica. Las cepas obtenidas fueron evaluadas por PCR para verificar la eliminación de los genes o la inserción

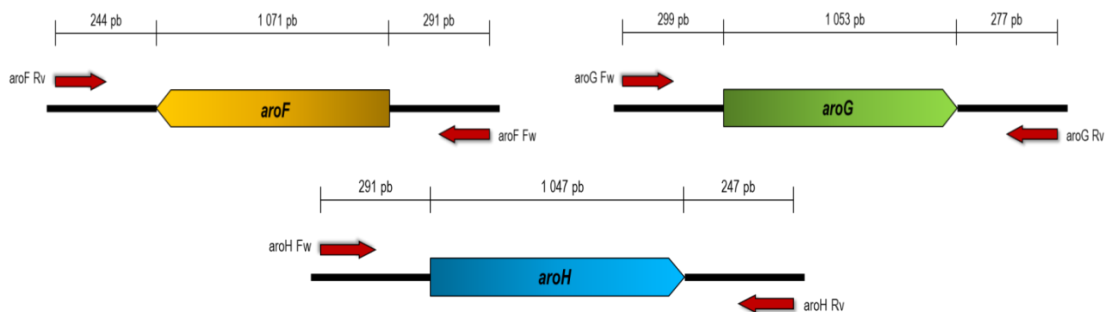
de gen de resistencia a kanamicina. Cuando fue requerido, la eliminación del gen de resistencia a kanamicina fue realizada mediante la inserción de plásmido pCP20 a las cepas generadas mediante la transducción [Cherepanov & Wackernagel, 1995].

## VI.6. Amplificación por PCR de genes *aroF*, *aroG* y *aroH*

Se diseñaron tres pares de oligonucleótidos para amplificar la región codificante de los genes *aroF*, *aroG* y *aroH*. Estos iniciadores amplifican desde una región aproximada de 300 pb antes del codón de inicio ATG hasta aproximadamente 300 pb después del codón de término para cada gen. Por esta característica, los iniciadores también fueron utilizados para detectar la eliminación del gen ( $\Delta aroF$ ,  $\Delta aroG$  o  $\Delta aroH$ ) o la inserción del gen de resistencia a kanamicina ( $\Delta aroF::kan$ ,  $\Delta aroG::kan$  o  $\Delta aroH::kan$ ). La reacción de PCR fue preparada siguiendo el protocolo señalado para la *Taq* DNA polimerasa (recombinante) de la compañía Fermentas (Tabla 3), trabajando con las condiciones de reacción de la Tabla 4. Los productos de amplificación fueron verificados en un gel de agarosa al 1%.

**Tabla 3.** Secuencias de oligonucleótidos para amplificar los genes *aroF*, *aroG* y *aroH*.

| Oligonucleótido | Secuencia                      |
|-----------------|--------------------------------|
| aroF Fw         | 5' – AGGTTGGCGCGTGAGTATAA – 3' |
| aroF Rv         | 5' – CACCCGACGCAAACATCCT – 3'  |
| aroG Fw         | 5' – GGTCGTTATGTCGGATAACC – 3' |
| aroG Rv         | 5' – GTCCCGCTGGTGTATGAGTT – 3' |
| aroH Fw         | 5' – AGGATGGAAGTCGCGGAAGT – 3' |
| aroH Rv         | 5' – TTGCCATCCGGACCTAACAG – 3' |



**Figura 18.** Regiones de alineamiento de oligonucleótidos en sitios adyacentes a los genes *aroF*, *aroG* y *aroH*.

**Tabla 4.** Mezcla de reacción para amplificación por PCR (Volumen final: 50  $\mu$ L)

|                           |                        |
|---------------------------|------------------------|
| Buffer <i>Taq</i> 10X     | 5 $\mu$ L              |
| Mezcla dNTP (10 mM)       | 1 $\mu$ L              |
| Oligonucleótido Fw        | 0.5 $\mu$ M            |
| Oligonucleótido Rv        | 0.5 $\mu$ M            |
| MgCl <sub>2</sub> 25 mM   | 2 mM                   |
| DNA templado              | 1 $\mu$ g              |
| <i>Taq</i> DNA polimerasa | 1.25 U                 |
| Agua                      | Completar a 50 $\mu$ L |

**Tabla 5.** Condiciones de reacción para amplificación de genes *aroF*, *aroG* y *aroH* por PCR.

| Temperatura | Tiempo    | Ciclos |
|-------------|-----------|--------|
| 95 °C       | 5 minutos | 1      |
| 95 °C       | 1 minuto  |        |
| 55 °C       | 1 minuto  | 30     |
| 72 °C       | 1 minuto  |        |
| 72 °C       | 5 minutos | 1      |

### VI.7. Transformación de la cepa de *E. coli* $\Delta$ *aroF* $\Delta$ *aroH* $\Delta$ *aroG::kan* con el banco metagenómico de suelo

Se utilizó la cepa T2.1 con genotipo  $\Delta$ *aroF*  $\Delta$ *aroH*  $\Delta$ *aroG::kan* para buscar la complementación funcional con el banco metagenómico. La transformación se llevó a cabo mediante electroporación y el volumen de recuperación final fue de 500  $\mu$ L. Previo a la siembra en medio mínimo, este volumen de recuperación se sometió a lavados con medio mínimo M9 para la eliminación completa de medio SOC. Se utilizaron los medios ARO y M9 (sólido y líquido), así como en medio M9S (sólido y líquido) para realizar el cultivo y selección de células T2.1 transformantes que contengan una secuencia que codifique para una DAHP sintasa que permita el crecimiento por complementación funcional. Como pruebas control, se realizó la transformación de la cepa T2.1 con los plásmidos pJET1.2-B y pJLBaroG<sup>fbr</sup>ktA.

### VI.8. Selección de clonas transformantes T2.1 con recuperación de actividad de DAHP sintasa

Para este modelo de búsqueda, la evidencia de complementación funcional fue la capacidad de crecimiento bacteriano en los cultivos. En caso del medio mínimo en matraz, la selección de

clonas transformantes se realizó mediante una siembra posterior en agar M9S. De acuerdo al desarrollo observado en el cultivo líquido, y previo a la siembra en agar, se realizaron las diluciones apropiadas para obtener colonias aisladas para la selección y evitar la saturación de colonias sobre la superficie del agar. Las clonas obtenidas mediante esta selección se denominarán UGGL. Asimismo, se realizó una selección de transformantes en siembra directa en agar M9S, denominando a estas clonas UGGS.

## VI.9. Análisis de clonas UGG

### VI.9.1. Análisis de restricción de plásmidos pUGG

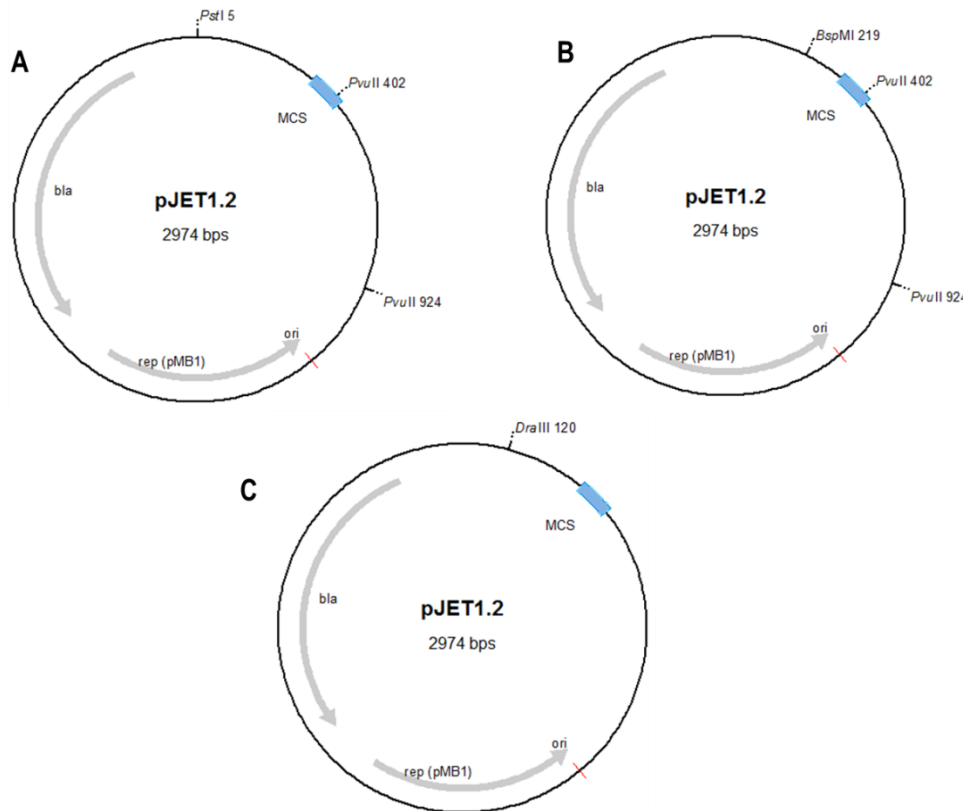
Para las clonas UGG obtenidas, se realizó la extracción de plásmido para cada una. Posteriormente se efectuó un análisis de restricción enzimática con dos combinaciones de enzimas (doble digestión): *Pst* I – *Pvu* II, y *Bsp*M I – *Pvu* II (ambas de Thermo-Scientific®). Los perfiles de restricción con estas enzimas constituyeron el primer criterio de selección para la detección de clonas de inserto único. De manera general, todas las enzimas fueron incubadas a 37 °C por dos horas. Posterior al tiempo de incubación, las muestras fueron analizadas en un gel de agarosa al 2.5%. Los volúmenes de reacción utilizados en estas reacciones se señalan en la Tabla 6.

**Tabla 6.** Volúmenes de reacción para doble restricción enzimática (Volumen final: 20  $\mu$ L).

|                                  |                                      |            |
|----------------------------------|--------------------------------------|------------|
| Agua                             |                                      | 12 $\mu$ L |
| Buffer 10X                       |                                      | 2 $\mu$ L  |
| Enzima (10 U/ $\mu$ L)           | 2 $\mu$ L (1 $\mu$ L de cada enzima) |            |
| DNA (0.5 – 1.0 $\mu$ g/ $\mu$ L) |                                      | 4 $\mu$ L  |

Particularmente para las clonas UGGL96 y UGGS2, se realizó posteriormente una restricción enzimática con la enzima *Ade* I (*Dra* III) de la marca Thermo-Scientific®, con la cual se realizó la determinación aproximada del tamaño de inserto en estas dos clonas. Esta restricción fue incubada a 37 °C por una hora, manteniendo los volúmenes de reacción de la Tabla 6, analizando los perfiles de restricción en un gel de agarosa al 1.5%

Las enzimas de restricción utilizadas tienen distintos sitios de corte en el plásmido pJET1.2 (Figura 19), de tal forma que al obtener los perfiles de restricción estos sitios fueron considerados para determinar el tamaño probable de los insertos que contienen.



**Figura 19.** Esquema de vector pJET1.2 señalando algunas de sus características importantes como el gen *bla* que confiere resistencia a ampicilina/carbenicilina, el replicón *rep* de plásmido pMB1, el sitio de multiclonación (MCS) y los sitios de corte para las combinaciones de enzimas empleadas en los experimentos de restricción enzimática. **A.** Enzimas *Pst* I y *Pvu* II; **B.** Enzimas *Bsp*M I y *Pvu* II; **C.** Enzima *Ade* I (*Dra* III).

### VI.9.2. Secuenciación de insertos contenidos en plásmidos pUGG

Los insertos de las clonas UGGL (1-14, 96) y UGGS (1a, 2, 3a, 4) fueron seleccionados para ser secuenciados en función del patrón único de restricción mostrado. Para esto, el DNA plasmídico de cada clona fue purificado utilizando el kit de purificación GeneJET™ PCR Purification Kit de la marca comercial Fermentas®, siguiendo las instrucciones del proveedor, empleando los iniciadores 23-mer y 24-mer con sitios de unión en el vector pJET1.2 en las regiones 310-332 (23-mer) y 428-405 (24-mer). Una vez purificados los templados, se determinaron sus concentraciones para posteriormente preparar la mezcla de reacción para secuenciación de acuerdo a las especificaciones

de la Unidad de Secuenciación del Instituto de Biotecnología – UNAM: 33 – 50 ng de DNA plasmídico/ $\mu\text{L}$  (templado) y 1  $\mu\text{L}$  de iniciadores a una concentración de 10 pmol/mL, esto en un volumen final de 16  $\mu\text{L}$ . La muestra fue secuenciada en un equipo Perkin Elmer/Applied Biosystems Modelo 3730 por el método Taq FS Dye Terminator Cycle Sequencing Fluorescence-Based Sequencing.

### **VI.9.3. Secuenciación masiva de clonas seleccionadas**

En aquellas clonas únicas cuyo inserto sea tamaño mayor a 1 000 pb, se secuenciaron empleando una técnica de secuenciación masiva de los plásmidos/insertos seleccionados. Este procedimiento fue realizado en el CINVESTAV Unidad Irapuato mediante la preparación de bibliotecas siguiendo el protocolo de Nextera® XT DNA de la compañía Illumina; asimismo, la secuenciación de estas bibliotecas fue realizada en el sistema MiSeq de la misma compañía mediante secuenciación por síntesis. Las secuencias obtenidas fueron procesadas en la Unidad de Apoyo Bioinformático del Instituto de Biotecnología para generar *contigs* y *scaffolds* para cada plásmido secuenciado.

### **VI.9.4 Caracterización fisiológica de las clonas UGG**

#### **VI.9.4.1. Cinética de crecimiento**

Fueron realizadas las caracterizaciones del crecimiento en medio mínimo M9S de cada una de las 19 clonas con inserto único. Estas se realizaron en matraces bafleados de 250 mL con 50 mL de medio, incubando en condiciones aerobias a 37 °C en agitación constante de 300 rpm, adicionando glucosa a una concentración de 2 g/L como única fuente de carbono y energía. Todas las cinéticas fueron iniciadas a una densidad celular de 0.1 D.O.<sub>600</sub>.

#### **VI.9.4.2. Determinación de actividad enzimática de DAHP sintasa**

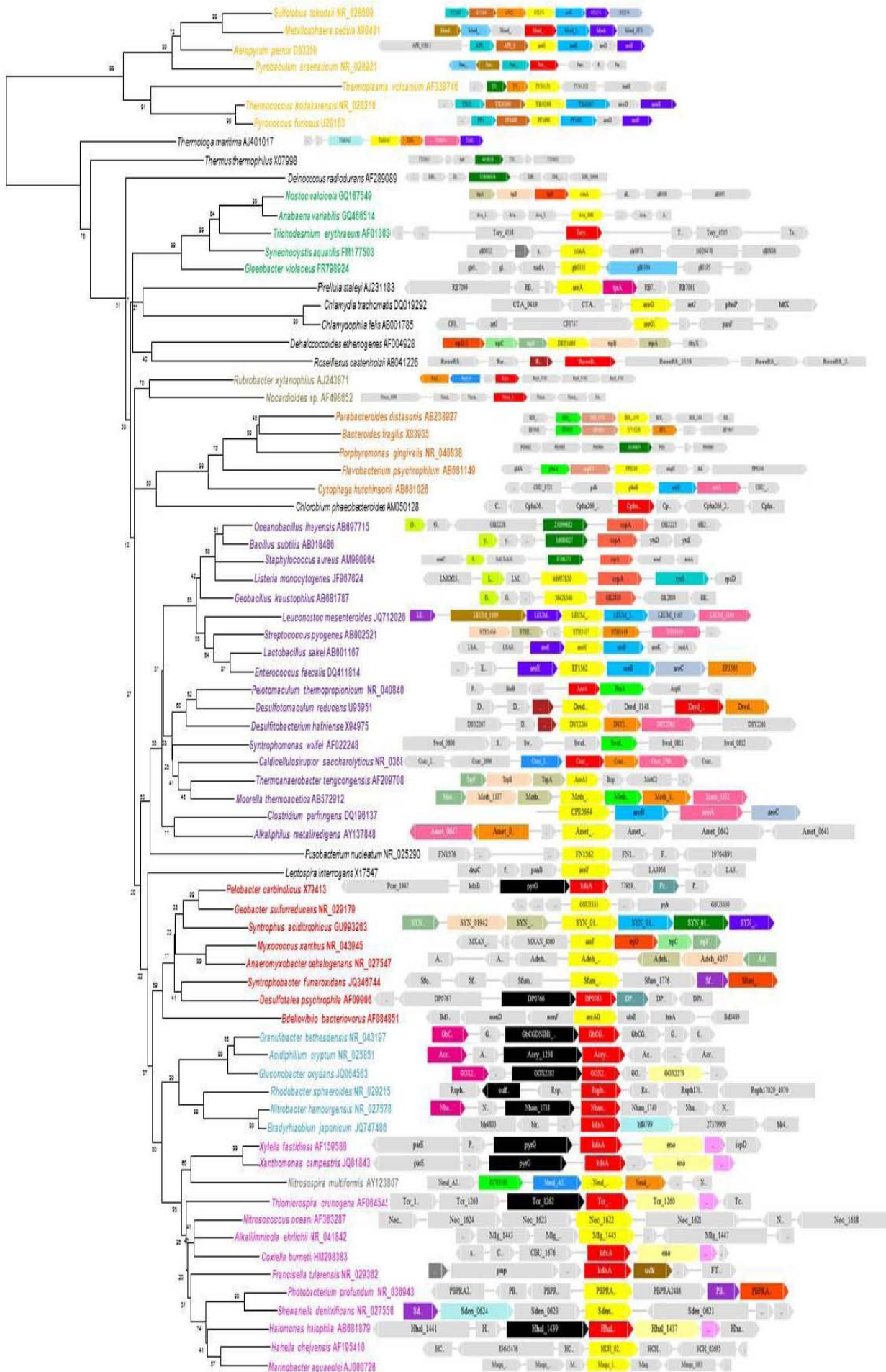
Una vez establecida la cinética de crecimiento característica para cada clona, se realizó la determinación de actividad enzimática de DAHP sintasa a partir del extracto celular de cada clona. Las células fueron cosechadas en la etapa tardía del crecimiento exponencial, posteriormente fueron

lavadas y resuspendidas en solución EPPS 50 mM pH 8.5. Las células fueron lisadas mediante sonicación con un equipo MSE Soniprep 150, sometiéndolas a tres pulsos de 30 segundos cada uno, dejando reposar en hielo al menos 30 segundos entre cada pulso. Los restos celulares fueron separados mediante centrifugación a 12 000 x g; este extracto fue utilizado para la determinación de actividad enzimática (ver procedimiento en apéndice). La cuantificación de DAHP fue realizada por el método de ácido tiobarbitúrico [Waravdekar & Saslaw, 1959] mientras que la cuantificación de proteína fue realizada por el método de Bradford [Bradford, 1976]. Es importante señalar que la actividad específica de DAHP sintasa obtenida en estos experimentos se expresa en función de la cantidad de proteína total (mg) obtenida de los extractos crudos.

## VII. RESULTADOS

### VII.1. Análisis del contexto genómico de DAHP sintasas de las subfamilias AroA<sub>α</sub> (e.g. *E. coli*) y AroA<sub>β</sub> (e.g. *B. subtilis*)

El programa GeConT I es una interfase diseñada para visualizar el contexto genómico de un gen o de un grupo de genes con sus ortólogos en todos los genomas completamente secuenciados [Ciria *et al.*, 2004]. De acuerdo a las clasificaciones propuestas para las DAHP sintasas, se eligió una secuencia representativa de la subfamilia AroA<sub>α</sub> (gen *aroF* de *E. coli*) y de la subfamilia AroA<sub>β</sub> (gen *aroA* de *B. subtilis*) para efectuar un estudio preliminar de búsqueda de secuencias de DAHP sintasas en genomas totalmente secuenciados y la identificación de ortólogos en otros organismos. Con la finalidad de realizar un estudio preliminar de las DAHP sintasas con secuencia y función conocidas, fueron elegidas las DAHP sintasas de *B. subtilis* y *E. coli*. Para cada una de éstas dos, el análisis del programa GeConT I mostró un número determinado de DAHP sintasas ortólogas; en algunos casos, el estudio desplegó más de una especie para un género específico. En estos casos, se buscaron posibles diferencias entre los contextos genómicos de las distintas especies de un mismo género, pero al ser idénticos entre éstas, seleccionamos solo una especie representativa de cada género. En base a esta consideración, la DAHP sintasa de *B. subtilis* desplegó 79 ortólogos, mientras que la DAHP sintasa de *E. coli* mostró 58.

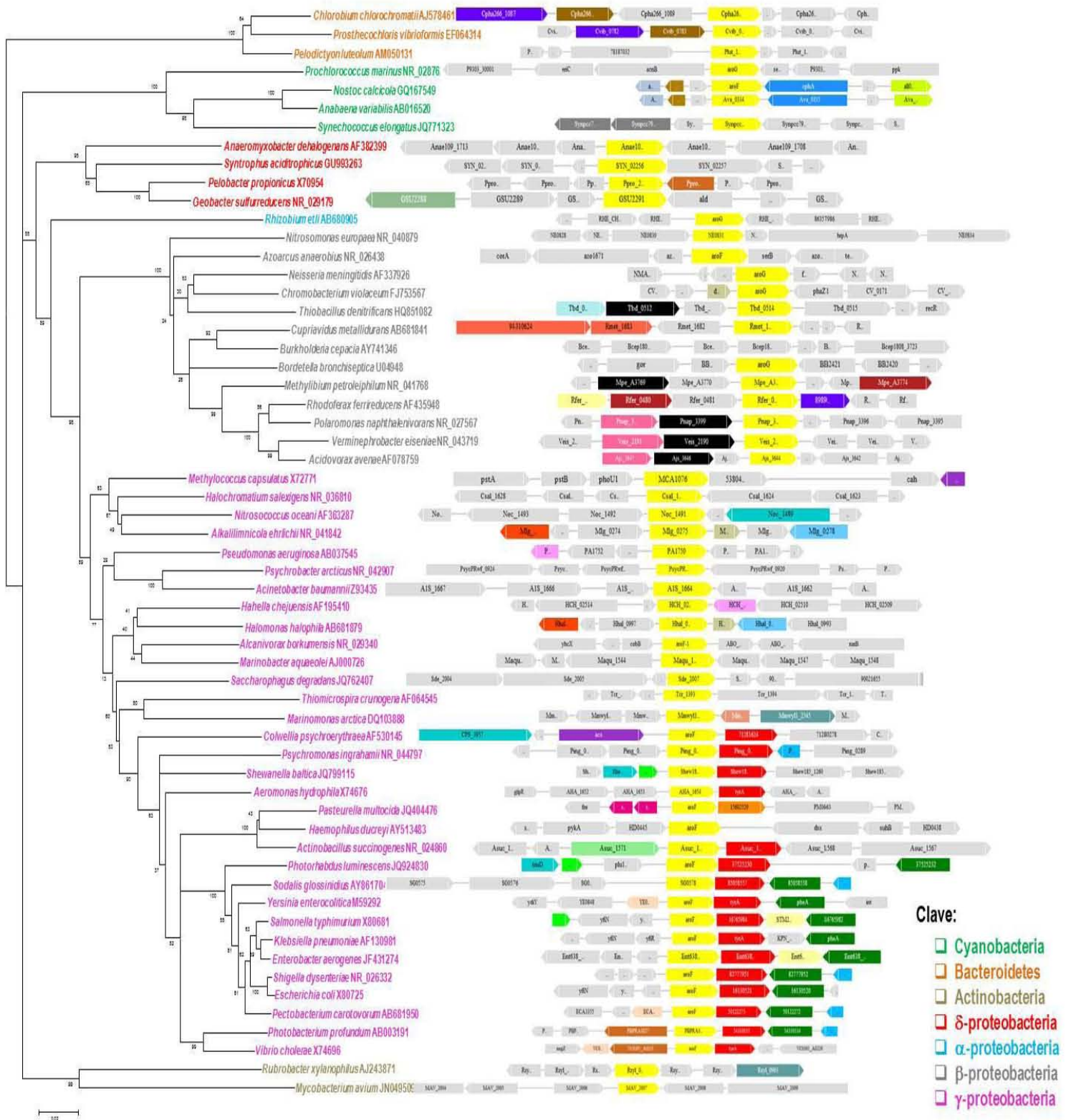


Clave:

- Archaea
- Cyanobacteria
- Bacteroidetes
- Actinobacteria
- Firmicutes
- δ-proteobacteria
- α-proteobacteria
- β-proteobacteria
- γ-proteobacteria



**Figura 20.** Análisis de contexto genómico de 79 secuencias de DAHP sintasas de la subfamilia AroA<sub>1β</sub>, ortólogas a *Bacillus subtilis*, dentro de un árbol filogenético de genes 16S rRNA. Los esquemas de contexto genómico se muestran a la derecha del árbol filogenético desplegando 3 genes a cada extremo del gen de DAHP sintasa. De acuerdo al programa GeConT I, cada color desplegado en el esquema de contexto genómico corresponde a una función enzimática específica, las DAHP sintasas se representan con color amarillo y las KDOP sintasas con color rojo. La historia evolutiva fue inferida utilizando el método *Neighbor-Joining*. Se muestra el árbol óptimo con una suma de longitud de brazo de 5.18755651. El porcentaje de árboles réplica en el que los taxa fueron agrupados en la prueba bootstrap (1 000 réplicas) se señala a un costado de las ramas. Las distancias evolutivas fueron calculadas utilizando el método de Kimura 2 y están expresadas en unidades de número de sustituciones de base por sitio. El análisis incluyó 79 secuencias nucleotídicas. Todas las posiciones que contenían espacios o datos perdidos fueron eliminadas. En el grupo final de datos hubieron un total de 977 posiciones. El análisis filogenético fue realizado con el programa MEGA5.



**Figura 21.** Análisis de contexto genómico de 58 secuencias de DAHP sintasas de la subfamilia AroA<sub>α</sub>, ortólogas a *Escherichia coli*, dentro de un árbol filogenético de genes 16S rRNA. Los esquemas de contexto genómico se muestran a la derecha del árbol filogenético desplegando 3 genes a cada extremo del gen de DAHP sintasa. De acuerdo al programa GeConT I, cada color desplegado en el esquema de contexto genómico corresponde a una función enzimática específica, las DAHP sintasas se representan con color amarillo. La historia evolutiva fue inferida utilizando el método *Neighbor-Joining*. Se muestra el árbol óptimo con una suma de longitud de brazo de 3.07872887. El porcentaje de árboles réplica en el que los taxa fueron agrupados en la prueba bootstrap (1 000 réplicas) se señala a un costado de las ramas. Las distancias evolutivas fueron calculadas utilizando el método de Kimura 2 y están expresadas en unidades de número de sustituciones de base por sitio. El análisis incluyó 59 secuencias nucleotídicas. En el grupo final de datos hubo un total de 1 117 posiciones. Todas las posiciones que contenían espacios o datos perdidos fueron eliminadas. El análisis filogenético fue realizado con el programa MEGA5

En base a la distribución filogenética del gen 16S rRNA de los organismos con DAHP sintasa ortóloga a la de *B. subtilis*, se observa una gran diversidad de grupos microbianos comprendiendo desde el reino Archaea hasta la clase  $\gamma$ -Proteobacteria. Por otra parte, las DAHP sintasas de la subfamilia AroA<sub>α</sub> se encuentran principalmente dentro del phylum Proteobacteria. Esto puede ser un indicio que reafirme las hipótesis de otros grupos de investigación que proponen que la subfamilia de DAHP sintasas AroA<sub>β</sub> es la más ampliamente distribuida en la naturaleza, y la más ancestral comparada con la subfamilia AroA<sub>α</sub>.

También se puede apreciar que los ortólogos no sólo comprenden a DAHP sintasas, sino que también se observa una gran similitud con las KDOP sintasas, con las cuales hay una gran similitud a nivel de secuencia de aminoácidos con la subfamilia AroA<sub>β</sub>, y que en esencia catalizan la misma reacción de condensación aldólica que las DAHP sintasas, aunque la KDOP sintasas condensan PEP y pentosas. De acuerdo al árbol filogenético, la distribución de las KDOP sintasas se ubica principalmente en el phylum Proteobacteria.

Los elementos genéticos adyacentes a un gen de interés también pueden estar relacionados con su función biológica dentro de la célula. De acuerdo al GeConT I, el contexto genómico de algunos ortólogos a las DAHP sintasas de *B. subtilis* y *E. coli* se asocia con las rutas biosintéticas de aminoácidos aromáticos (fenilalanina, triptofano y tirosina) (Tabla 7); cabe destacar que aunque el contexto genómico de un gran número de ortólogos no se asocia con rutas biosintéticas de compuestos aromáticos, la distribución de los genes adyacentes aún es similar en grupos microbianos de distancia evolutiva cercana.

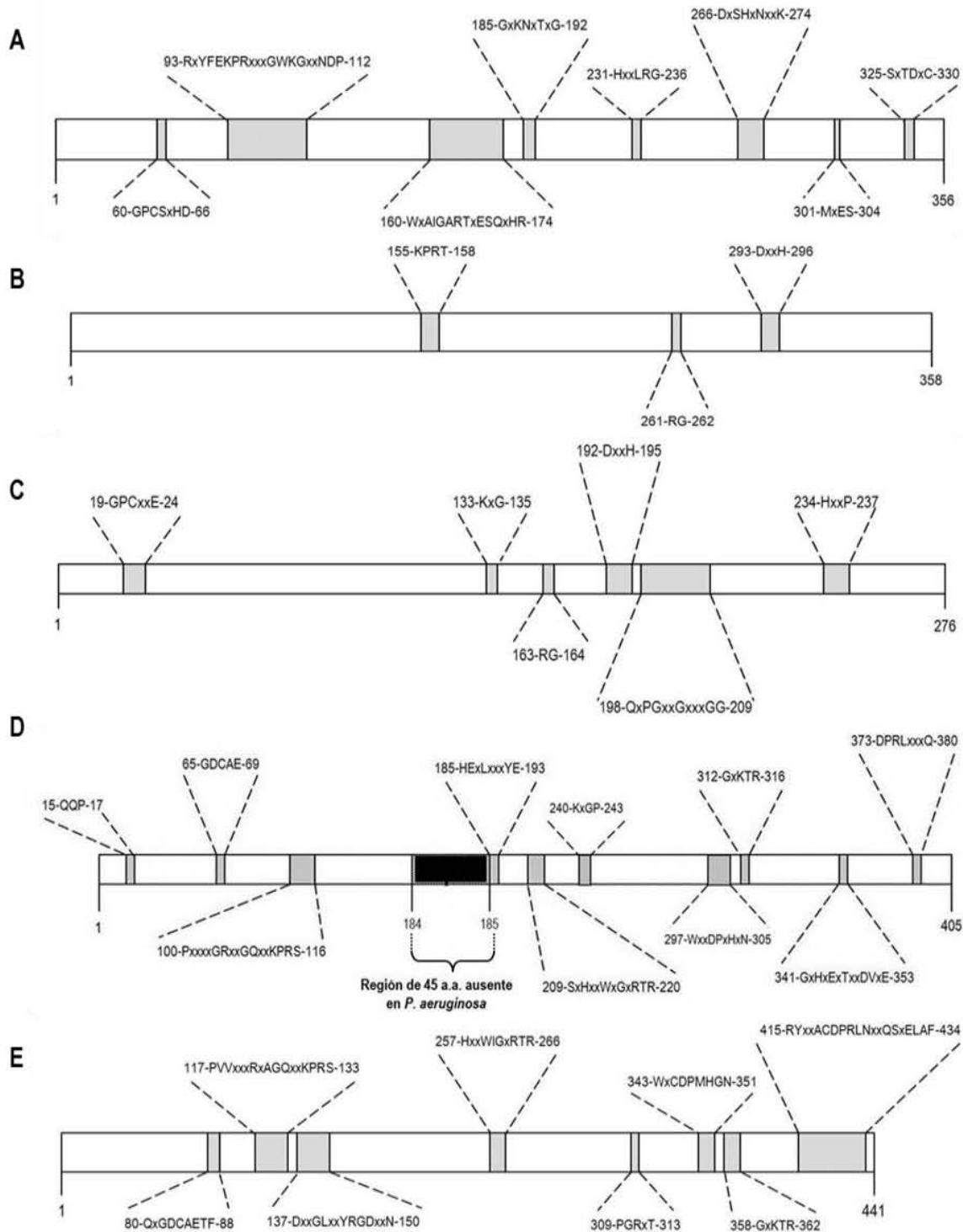
**Tabla 7.** Ejemplo de algunas enzimas relacionadas en el contexto genómico de las DAHP sintasas de *Escherichia coli* (subfamilia AroA<sub>Iα</sub>) y *Bacillus subtilis* (subfamilia AroA<sub>Iβ</sub>).

| Subfamilia AroA <sub>Iα</sub>         | Subfamilia AroA <sub>Iβ</sub>                 |
|---------------------------------------|---|
| Sulfito reductasa                     | Transcetolasa                                 |
| Transglicosilasa lítica               | 3-dehidroquinato sintasa                      |
| ATPasa                                | 3-dehidroquinato dehidratasa                  |
| Peptidilprolil isomerasa              | Shikimato dehidrogenasa                       |
| Cianoficina sintetasa                 | Corismato mutasa                              |
| Cianoficinasa                         | Corismato sintasa                             |
| Aconitasa                             | Prefenato dehidrogenasa                       |
| Canal de ion cloruro                  | Tirosil-tRNA sintetasa                        |
| Nitrorreductasa                       | Fosforribosiltransferasa                      |
| Polifosfato cinasa                    | Transportadores de aminoácidos                |
| FAD dehidrogenasa                     | Antranilato/ <i>p</i> -aminobenzoato sintasas |
| Poli(3-hidroxiobutirato) depolimerasa | Indol-3-glicerol fosfato sintasa              |
| Dihidrofolato reductasa               | Enzima activadora de piruvato-formato liasa   |
| Proteína supresora DnaK               | Piruvato cinasa                               |
| Fosfoserina fosfatasa                 | Triptofano sintasa                            |
| Histidinol fosfatasa                  | ATP pirofosfatasa (biosíntesis de tiamina)    |
| Serinproteasa periplásmica            | Aspartato/tirosina aminotransferasa           |
| Prefenato dehidrogenasa               | Prefenato dehidratasa                         |
| Acil-CoA sintetetasas                 | 5-enolpiruvilshikimato 3-fosfato sintasa      |
| Receptores de membrana externa (Fe)   | ATPasa/GTPasa                                 |
| Uroporfirinógeno-III metilasa         | Antranilato fosforribosiltransferasa          |
| tRNA (guanina-N1)-metiltransferasa    | Fosforribosilantranilato isomerasa            |
| Prefenato dehidratasa                 | Quinolinato sintasa                           |
| Gluconolactonasa                      | CTP sintasa (UTO-amonio liasa)                |
| Regulador de asimilación de fosfato   | Permeasas                                     |
| Proteínas tipo tiorredoxina           | Triosafofosfato isomerasa                     |
| Proteína tipo ferritina               | Enolasa                                       |
| Corismato mutasa                      | Uridina cinasa                                |
| β-N-acetilhexosaminidasa              |   |

## VII.2. Reconstrucción filogenética y análisis de secuencias de familias representativas de DAHP sintasas (AroA<sub>Iα</sub>, AroA<sub>Iβ</sub>, AroA<sub>II</sub>, KdsA, PhzDAHPS y aminoDAHPS)

Del estudio efectuado con el programa GeConT se obtuvieron 58 secuencias de DAHP sintasas pertenecientes a la familia AroA<sub>Iα</sub> y 79 de la familia AroA<sub>Iβ</sub>. Para realizar la reconstrucción filogenética, se obtuvieron las secuencias de DAHP sintasas de plantas superiores (familia AroA<sub>II</sub>), así como aquellas relacionadas con biosíntesis de fenazinas (PhzDAHPS) y la vía de aminoshikimato (aminoDAHPS). En primer lugar, se generó un alineamiento de las secuencias para distinguir el patrón de alineamiento de las principales familias de DAHP sintasas, así como la

distinción de posibles motivos o regiones conservadas para todos los grupos. Estos alineamientos fueron obtenidos mediante el programa ClustalW utilizando la matriz Gonnet para proteínas en el programa MEGA 5.2.



**Figura 22.** Esquemas de DAHP sintasas representativas de cada familia conocida. Los motivos conservados señalados aquí son correspondientes a cada una de estas familias. Los residuos de aminoácidos variables se representan con la

letra "X". A, subfamilia AroA<sub>Iα</sub>, representada por la DAHP sintasa (AroF) de *E. coli*; B, subfamilia AroA<sub>Iβ</sub>, representada por la DAHP sintasa de *B. subtilis*; C, familia KdsA representada por la KDOP sintasa de *Xanthomonas campestris*; D, PhzDAHPS representada por la DAHP sintasa (PhzC) de *Pseudomonas aeruginosa*; E, aminoDAHPS representada por la aminoDAHPS sintasa (RifH) de *Amycolatopsis mediterranei*.

En las figuras anteriores, se representan las familias conocidas de DAHP sintasas. Para distinguir los motivos conservados característicos en cada una de éstas, se eligió un miembro microbiano representativo para cada familia o subfamilia con la finalidad de establecer con mayor claridad la ubicación aproximada de los motivos y la longitud promedio de aminoácidos que componen a estas DAHP sintasas.

En el caso de la subfamilia AroA<sub>Iα</sub>, ésta se encuentra representada por la DAHP sintasa (AroF) de *E. coli* y el alineamiento fue realizado con 58 secuencias de aminoácidos. Las enzimas de esta subfamilia se componen por 356 aa en promedio. Se encontraron 8 motivos conservados en estas DAHP sintasas, distribuidos en prácticamente toda la longitud de la secuencia. Las DAHP sintasas de esta subfamilia, y particularmente la DAHP sintasas de *E. coli*, han sido las más estudiadas, e incluso se ha logrado obtener la estructura terciaria de la proteína [Shumilin *et al.*, 1999]. Se conoce que los aminoácidos que participan en la interacción con PEP son principalmente lisina (Lys), arginina (Arg) e histidina (His); en el caso de la DAHP sintasa (AroG) de *E. coli*, estos aminoácidos son Arg92, Lys97, Arg165, Lys186, Arg234 e His268. Esta subfamilia de DAHP sintasas se caracteriza también por el requerimiento de cationes divalentes para la biocatálisis enzimática; en el caso de la DAHP sintasa (AroG) de *E. coli* los aminoácidos que participan en la interacción con metales son Cys61, His268, Glu302, Asp327. En el alineamiento realizado, tanto los aminoácidos encargados en la interacción con PEP como con metales divalentes se encuentran incluidos en los motivos conservados en toda la subfamilia, visualizando entonces la conservación de estos aminoácidos en todas las DAHP sintasas de los miembros microbianos incluidos en este estudio.

Para el análisis de subfamilia AroA<sub>Iβ</sub> se realizó un alineamiento para DAHP sintasas y otro para KDOP sintasas; para el primero se utilizó la DAHPS de *B. subtilis* como referencia, y para el segundo se utilizó la KDOP sintasa de *X. campestris*. En general, las DAHP sintasas de la subfamilia AroA<sub>Iβ</sub> están constituidas por 358 aa, número muy similar a las de la subfamilia AroA<sub>Iα</sub>. Además de *E. coli*, otra estructura terciaria de DAHP sintasa caracterizada es la de *Thermotoga maritima*

[Shumilin et al., 2004]; aunque ésta última se clasifica dentro de la subfamilia AroA<sub>β</sub>, se ha observado que comparte los mismos aminoácidos que la DAHP sintasa de *E. coli* en la interacción con PEP. Observando el alineamiento de las secuencias de la subfamilia AroA<sub>β</sub> se encontraron tres motivos conservados en los 56 miembros microbianos analizados, éstos son 155-KPRT-158, 261-RG-262 y 293-DxxH-296. Considerando que los aminoácidos que participan en la interacción con PEP coinciden tanto en *E. coli* como en *T. maritima*, se puede inferir que algunos de estos aminoácidos se encuentran en los motivos 155-KPRT-158 y 293-DxxH-296. Aunque no se encuentra señalado en el esquema de alineamiento, es importante considerar que el aminoácido Asp333 se encuentra en todas las secuencias de la subfamilia AroA<sub>β</sub>; este aminoácido se encuentra cercano al Asp327 reportado para la subfamilia AroA<sub>α</sub> que participa en la interacción con metales divalentes. De manera general, se ha postulado que las DAHP sintasas de la subfamilia AroA<sub>β</sub> no requieren metales como cofactor en la reacción de biosíntesis de DAHP; sin embargo, un estudio de caracterización de la DAHP sintasa de *T. maritima* sugiere que esta enzima sí requiere un metal como cofactor. A reserva de estudios más detallados sobre la caracterización enzimática de otras DAHP sintasas de la subfamilia AroA<sub>β</sub>, la presencia del aminoácido Asp333 podría tener alguna participación en la interacción con metales divalentes, tal y como sucede con el Asp327 en las DAHP sintasas de la subfamilia AroA<sub>α</sub>. Las KDOP sintasas catalizan la condensación de PEP con A5P, y aunque a nivel de reacción bioquímica comparten alguna similitud con las DAHP sintasas, existe también una estrecha relación a nivel de secuencia de aminoácidos principalmente con la subfamilia AroA<sub>β</sub>. Esta relación se aprecia también en la presencia de motivos conservados, ya que los motivos 163-RG-164 y 192-DxxH-195 de la familia KdsA son idénticos a los identificados en la subfamilia AroA<sub>β</sub>. La presencia del motivo DxxH tanto en DAHP como en KDOP sintasas es sugerente de una participación importante de éste en la interacción con el sustrato común de ambas enzimas: el PEP. De manera alterna, el motivo 19-GPCxxE-24 encontrado en la familia KdsA es muy similar al motivo 60-GPCxHD-66 de la subfamilia AroA<sub>α</sub>; en este motivo se encuentra el aminoácido Cys61, que en la DAHP sintasa de *E. coli* es responsable de la interacción con metales divalentes. La presencia de este motivo en ambas familias podría justificar la existencia de KDOP sintasas que requieren metales como cofactor para la biocatálisis. Las 23 secuencias de KDOP sintasas analizadas están constituidas por aproximadamente 276 aa; sin embargo, tomando como referencia la KDOP sintasa de *X. campestris*, 10 secuencias de KDOP sintasas carecen de una región de 10 aa en su

secuencia; este fragmento se ubica entre la región Gln198 y Gly208 de la KDOP sintasas de *X. campestris*.

Para el caso de las DAHP sintasas que participan en la biosíntesis de fenazinas (PhzDAHPS) se obtuvieron 17 secuencias conocidas, y se utilizó la secuencia PhzC de *P. aeruginosa* como referencia de análisis en el esquema de alineamiento. Comparado con las DAHP sintasas de la familia AroA<sub>I</sub>, las enzimas de la familia PhzDAHPS se encuentran constituidas por un número mayor de aminoácidos; en promedio, las PhzDAHPS contienen 405 aa, aunque algunas contienen un fragmento adicional de 45 aa relativamente conservado a nivel de secuencia, por lo que algunas PhzDAHPS contienen 450 aa. Este fragmento adicional, tomando como referencia a PhzC de *P. aeruginosa*, se encuentra entre los aminoácidos Ser184 e His185. En esta familia se pueden distinguir los motivos 113-KPRS-116, 297-WxxDPxHxN-305 y 373-DPRLxxxQ-380, los cuales son característicos de las DAHP sintasas relacionadas con plantas superiores (familia AroA<sub>II</sub>). Los dos primeros motivos se relacionan tanto con la interacción con PEP y E4P, como con metales divalentes.

La familia de aminoDAHP sintasas (aminoDAHPS) fue analizada tomando como referencia a RifH de *Amycolatopsis mediterranei*, considerando 12 secuencias de aminoácidos para el alineamiento. Estas DAHP sintasas tienen una composición de aminoácidos similar en número a las PhzDAHPS, teniendo en promedio 441 a.a. en su secuencia. De los grupos microbianos analizados, éste grupo de enzimas contiene el mayor número de motivos conservados entre sí; asimismo, las aminoDAHPS comparten muchos motivos con la familia PhzDAHPS, e.g. los motivos GDCAE, KPRS, GxRTR, GxKTR y DPRL. Se realizó también un alineamiento de las secuencias de DAHPS de seis plantas superiores, *A. thaliana*, *Solanum lycopersicum*, *S. tuberosum*, *Nicotiana tabacum*, *Zea mays* y *Petroselinum crispum*, representativas de la familia AroA<sub>II</sub>, para identificar los motivos conservados entre estas enzimas. En el alineamiento se observa que aproximadamente el 85% de la secuencia de DAHPS de cada planta es idéntica a las demás, por lo que la familia AroA<sub>II</sub> se podría considerar como la más conservada dentro del universo de DAHPS conocidas. Cabe destacar que la región variable de las DAHPS se localiza en la primera cuarta parte de las secuencias, por lo que esta región contribuye de manera importante a la individualidad de cada DAHPS. Estas enzimas se

constituyen en promedio por 520 aa, siendo las de mayor tamaño comparadas con las demás familias.

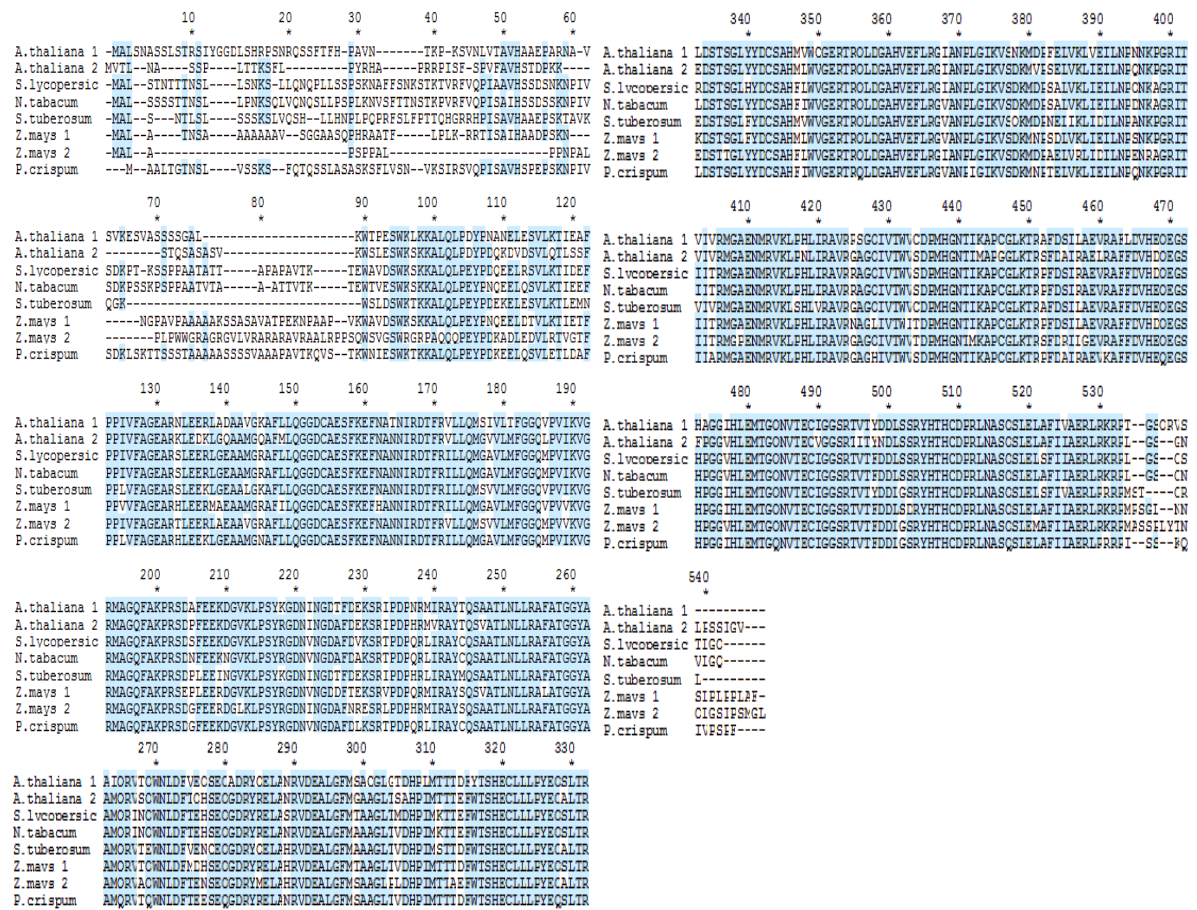


Figura 23. Alineamiento de secuencias de DAHP sintasas de algunas especies vegetales representativas de la familia *AroAII*. Las regiones conservadas se distinguen con sombreado azul.

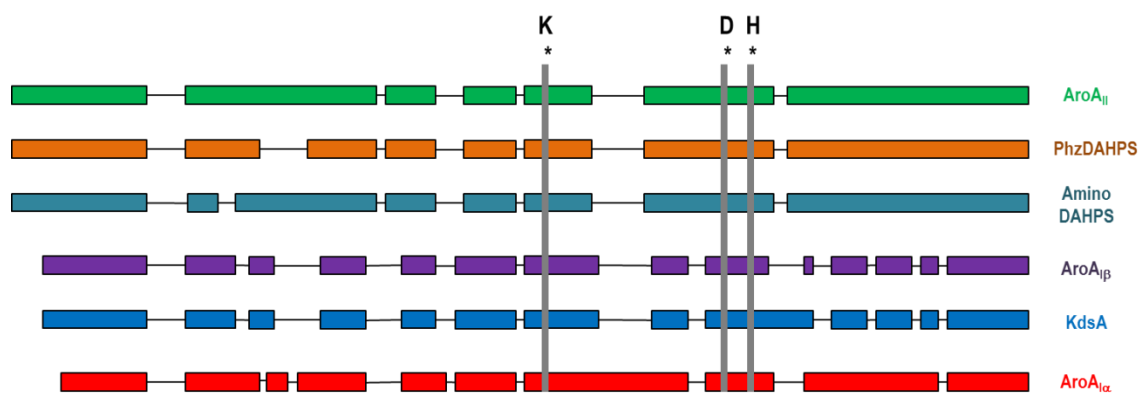
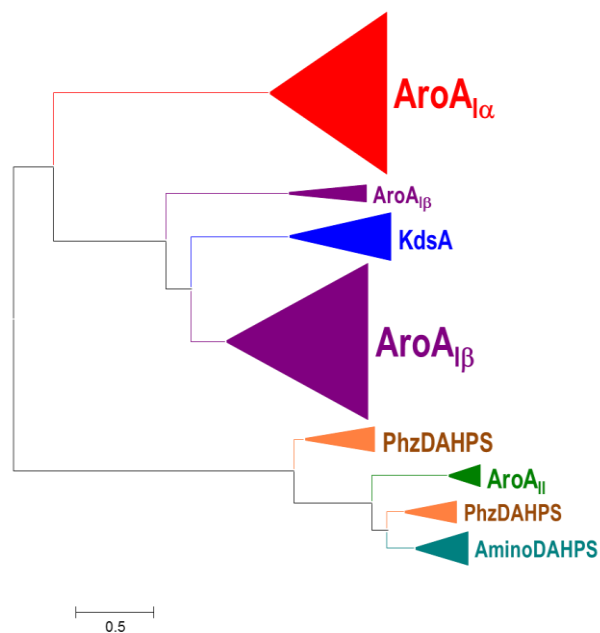


Figura 24. Esquema general del patrón de alineamiento y distribución de la estructura primaria de enzimas DAHP sintasas. Se señalan los 5 grupos principales de clasificación y los tres aminoácidos (lisina, aspartato e histidina) que son conservados en todas las familias.



En base al alineamiento realizado de las secuencias de DAHP sintasas, se determinó que cada grupo o familia posee regiones características o motivos compartidos identificables, pero que solamente existen tres aminoácidos conservados en todas las DAHP sintasas, los cuales son Lys, Asp, e His, que se señalan en el esquema. En la subfamilia AroA<sub>Iα</sub> estos aminoácidos se encuentran en las regiones Lys187, Asp266 e His269; en la subfamilia AroA<sub>Iβ</sub> en las regiones Lys231, Asp293 e His296; en la familia KdsA en las regiones Lys133, Asp192 e His195; en las PhzDAHPS en Lys240, Asp300 e His303; y en las aminoDAHPS, en Lys286, Asp346 e His349. En todas las familias de DAHPS, el motivo DxxH, que incluye tanto al Asp como His, se encuentra conservado. El patrón de alineamiento es considerablemente conservado en las DAHP sintasas de la familia AroA<sub>II</sub>, PhzDAHPS y aminoDAHPS, ubicando la principal modificación en la primera tercera parte del alineamiento. Es posible que esta región de modificación, junto con las diferencias en las secuencias de aminoácidos, contribuyan de manera importante a la diversidad catalítica de estos grupos de DAHP sintasas. De igual manera, se podría proponer esta perspectiva para la subfamilia AroA<sub>Iβ</sub> y KdsA, en donde las regiones de alineamiento son muy similares y cuya principal diferencia, además de algunos sitios de la secuencia de aminoácidos, es la presencia de una región de aminoácidos adicional en la familia KdsA con respecto a la AroA<sub>Iβ</sub>.



**Figura 25.** Árbol filogenético de DAHP sintasas de organismos representativos de todos los grupos conocidos (AroA<sub>Iα</sub>, AroA<sub>Iβ</sub>, AroA<sub>II</sub>, PhzDAHPS sintasas, y aminoDAHPS sintasas). La historia evolutiva fue inferida utilizando el método *Maximum Likelihood* basado en el modelo de matrices de Jones-Taylor-Thornton (JTT). El árbol mostrado es el de mayor

logaritmo de probabilidad (-81992.7997). Los árboles iniciales para la búsqueda heurística fueron obtenidos aplicando el método *Neighbor-Joining* a una matriz estimada de distancia por parejas utilizando un modelo JTT. El análisis incluyó 174 secuencias de aminoácidos. Todas las posiciones ambiguas fueron removidas para cada par de secuencia. En el grupo final de datos hubo un total de 711 posiciones. El análisis filogenético fue realizado con el programa MEGA5.

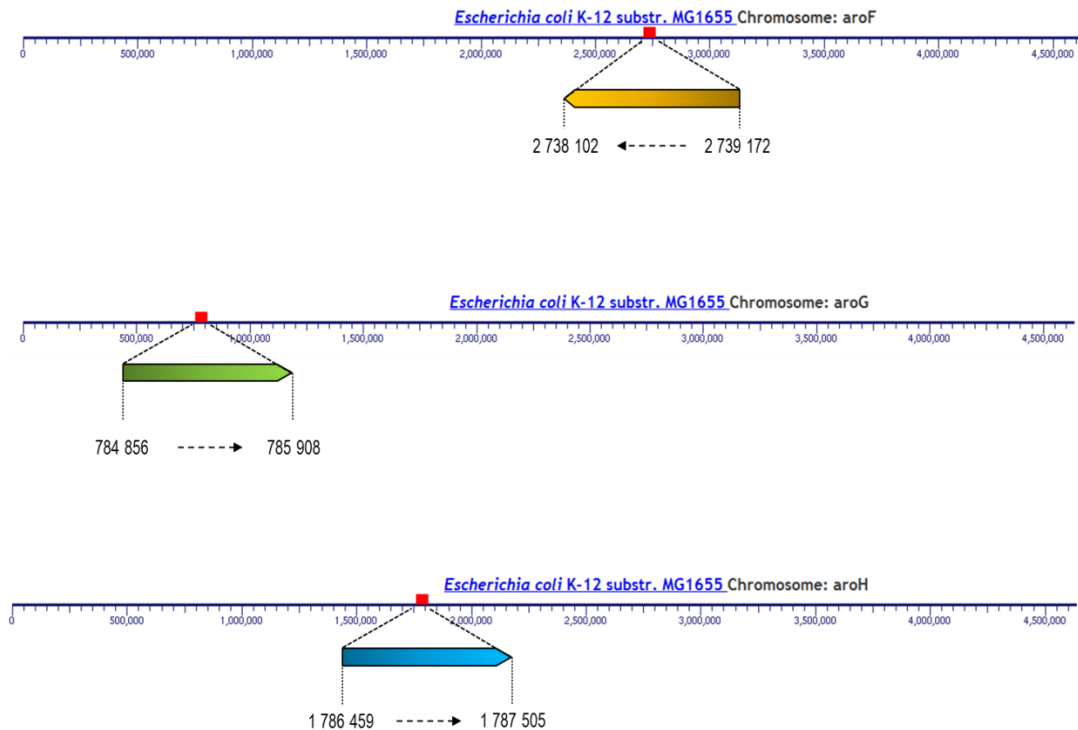
En el árbol filogenético de la Figura 23 se muestra la distribución de los grupos de DAHP sintasas que existen actualmente. Como otros estudios previos han demostrado, existe una clara diferencia a nivel de secuencia entre las DAHP sintasas denominadas AroA<sub>I</sub> (AroA<sub>Iα</sub> y AroA<sub>Iβ</sub>) y AroA<sub>II</sub>, este último grupo representado por las DAHP sintasas de plantas superiores. Asimismo, se ha reportado la similitud entre el grupo AroA<sub>II</sub> con las DAHP sintasas involucradas en la síntesis de fenazinas (PhzDAHPS) y aminoDAHPS sintasas (aminoDAHPS). En este árbol también se ilustra la relación que existe entre las DAHP sintasas del grupo con el grupo de KDOP sintasas (KdsA). La topología de los grupos de DAHP sintasas aquí representados sugiere que la separación temprana de los grupos AroA<sub>II</sub>, PhzDAHPS y aminoDAHPS ocurrió a partir de un evento evolutivo de divergencia inicial que derivó en la separación y evolución independiente con respecto a las demás DAHP sintasas microbianas. Analizando las DAHP sintasas microbianas del metabolismo primario, el grupo AroA<sub>Iα</sub> evolucionó de manera independiente y paralela a los grupos AroA<sub>Iβ</sub> y KdsA.

### VII.3. Obtención de cepa de *E. coli* Δ*aroF* Δ*aroG* Δ*aroH* carente de actividad de DAHP sintasa

Los genes parálogos *aroF*, *aroG* y *aroH*, poseen similar número de bases entre ellos, y además se encuentran ubicados en sitios distantes en el cromosoma de *E. coli*.

**Tabla 8.** Longitud y ubicación de los genes *aroF*, *aroG* y *aroH*, en el cromosoma de *Escherichia coli*.

| Gen         | Longitud (pb) | Ubicación                  |
|-------------|---------------|----------------------------|
| <i>aroF</i> | 1 071         | 2 738 102 ←----- 2 739 172 |
| <i>aroG</i> | 1 053         | 784 856 -----> 785 908     |
| <i>aroH</i> | 1 047         | 1 786 459 -----> 1 787 505 |

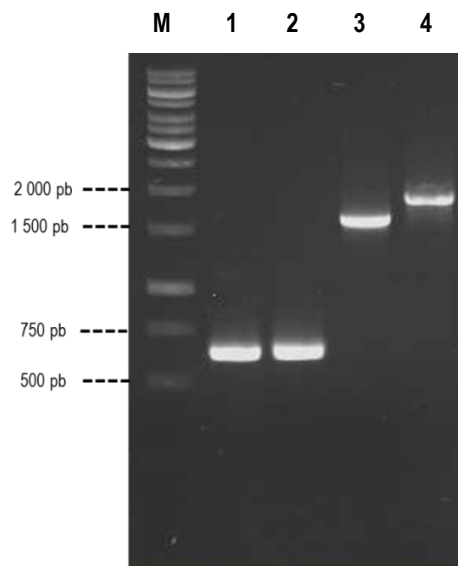


**Figura 26.** Distribución de los genes *aroF* (amarillo), *aroG* (verde) y *aroH* (azul) en el cromosoma de *Escherichia coli*.

Una de las técnicas para efectuar la interrupción de genes es la transducción generalizada, la cual involucra la transferencia de fragmentos aleatorios de DNA cromosómico de una cepa donadora a una cepa receptora mediante un bacteriófago; así, los fragmentos transferidos son recombinados por homología dentro del cromosoma de la cepa receptora. Si en la transducción se utilizan cepas donadoras con los genes de interés interrumpidos por genes de resistencia a algún antibiótico, se generarán cepas receptoras resistentes al antibiótico producto de la interrupción del gen blanco con el gen de resistencia en la cepa receptora. Por lo tanto, fueron seleccionadas tres cepas de *E. coli* K-12 BW25113 ( $\Delta aroF::kan$ ,  $\Delta aroG::kan$  y  $\Delta aroH::kan$ ) [Baba *et al.*, 2006] cada una con un gen de resistencia a kanamicina interrumpiendo los genes blanco y utilizando el bacteriófago P1*vir* como partícula transductante [Miller, 1992].

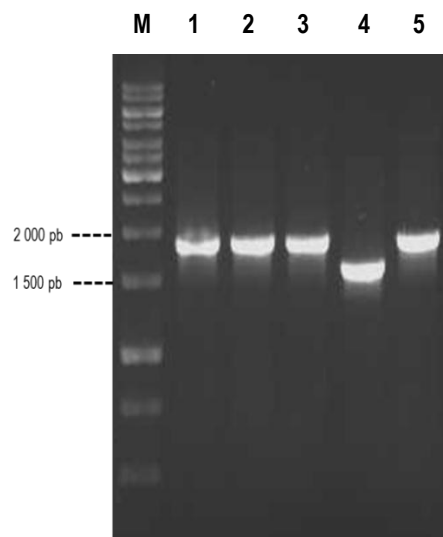
En la primera transducción se utilizó la cepa BW25113  $\Delta aroH::kan$  como donadora y un derivado de la cepa BW25113  $\Delta aroF::kan$  como receptora; este derivado se obtuvo transformando a la cepa receptora con el plásmido pCP20 para escindir el gen de resistencia a kanamicina,

obteniendo dos cepas BW25113  $\Delta aroF$  (Figura 27), seleccionando una de éstas como cepa receptora.



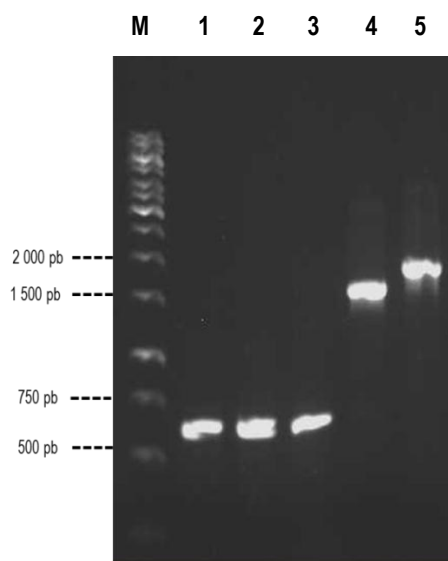
**Figura 27.** Productos de PCR en amplificación de gen *aroF*. M) Marcador de tamaño molecular de 1 kb; 1 & 2) cepas BW25113  $\Delta aroF$ ; 3) cepa MG1655; 4) cepa BW25113  $\Delta aroF::kan$ .

Como resultado del primer experimento de transducción, se obtuvieron tres cepas BW25113  $\Delta aroF \Delta aroH::kan$ , identificando la presencia del gen de resistencia a kanamicina interrumpiendo el gen *aroH* mediante PCR (Figura 28).



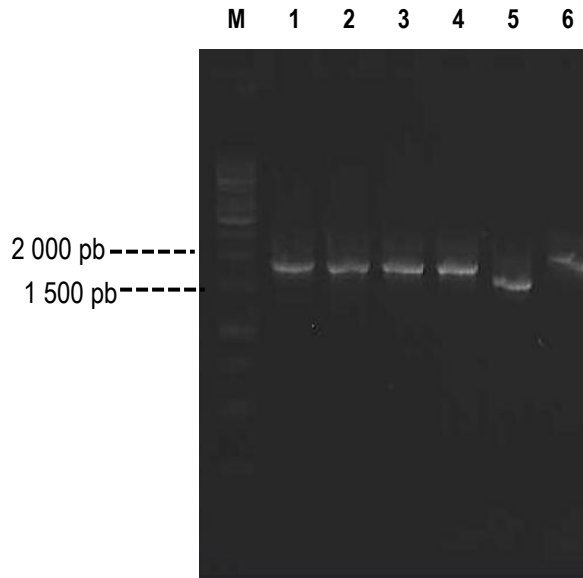
**Figura 28.** Productos de PCR en amplificación de gen *aroH*. M) Marcador de tamaño molecular de 1 kb; 1, 2 & 3) cepas BW25113  $\Delta aroF \Delta aroH::kan$ ; 4) cepa MG1655; 5) BW25113  $\Delta aroH::kan$ .

Para utilizar la cepa BW25113  $\Delta aroF \Delta aroH::kan$  como receptora en un segundo experimento de transducción, ésta fue transformada con el plásmido pCP20 para escindir el gen de resistencia a kanamicina; así se obtuvieron tres cepas BW25113  $\Delta aroF \Delta aroH$ , denominadas cepas T1.1, T1.2 y T1.3 (Figura 29).



**Figura 29.** Productos de PCR en amplificación de gen *aroH*. M) Marcador de tamaño molecular de 1 kb; 1, 2 & 3) cepas BW25113  $\Delta aroF \Delta aroH$ ; 4) cepa MG1655; 5) BW25113  $\Delta aroH::kan$ .

Al verificar la delección del gen *aroH*, se eligió la cepa T1.1 como cepa receptora para un segundo experimento de transducción, usando a la cepa BW25113  $\Delta aroG::kan$  como donadora; de esta manera, se obtuvieron cuatro cepas BW25113  $\Delta aroF \Delta aroH \Delta aroG::kan$ , denominándolas como T2.1, T2.2, T2.3 y T2.4. Al verificar el producto del gen *aroG*, se eligió una de las cepas obtenidas para continuar con el experimento de transformación y complementación funcional, seleccionando a la cepa T2.1.



**Figura 30.** Productos de PCR en amplificación de gen *aroG*. M) Marcador de tamaño molecular de 1 kb; 1, 2, 3 & 4) cepas BW25113  $\Delta$ *aroF*  $\Delta$ *aroH*  $\Delta$ *aroG::kan*; 5) cepa MG1655; 6) BW25113  $\Delta$ *aroG::kan*.

#### VII.4. Transformación de cepa de *E. coli* $\Delta$ *aroF* $\Delta$ *aroH* $\Delta$ *aroG::kan* con banco metagenómico de suelo

Al deletar los genes *aroF*, *aroG* y *aroH*, se elimina completamente la actividad de la DAHP sintasa en *E. coli*; esta actividad es esencial para el crecimiento de *E. coli* en medio mínimo, por lo que la detección de clonas positivas estará basada en el crecimiento microbiano en medio mineral con glucosa como única fuente de carbono y energía. Fue elegido el medio M9 para comenzar con los experimentos de complementación funcional. Previo a la transformación con el banco metagenómico, se eligió un control positivo para verificar el comportamiento del crecimiento de la cepa T2.1 en el medio M9, para esto se eligieron los plásmidos pJLBaroG<sup>fbr</sup> y pJLBaroG<sup>fbr</sup>*tkTA*; ambos contienen un gen que codifica para una DAHP sintasa de *E. coli* resistente a inhibición por retroalimentación. La cepa transformada tanto con uno o con otro plásmido mostraron crecimiento microbiano en agar M9. Debido a que no se encontró diferencia significativa en el comportamiento de la cepa, fue utilizado solamente el plásmido pJLBaroG<sup>fbr</sup>*tkTA* como control positivo. Posterior a esto, se realizaron los experimentos de selección de clonas transformadas con el banco metagenómico.

Sin embargo, después de algunos experimentos de transformación, no se observó crecimiento microbiano en el medio M9. Analizando su composición química, el medio M9 carece de algunos cofactores esenciales que podrían dificultar la selección de clonas positivas; por lo tanto, se realizaron los experimentos de complementación en un medio mineral cuya composición es más enriquecida que el medio M9, eligiendo así el medio mineral ARO. Este medio posee ácido *p*-aminobenzoico (PABA), ácido *p*-hidroxibenzoico (PHBA) y ácido 2,3-dihidroxibenzoico (2,3-DHBA), los cuales son cofactores derivados de la vía del shikimato. Con la finalidad de disminuir los tiempos de incubación y acelerar la selección de clonas, se realizaron los experimentos de transformación en matraces con medio ARO en vez de utilizar agar.

Después de varios experimentos de transformación con el banco metagenómico, no se detectaron clonas positivas que complementaran la función de DAHP sintasa en el medio ARO; inclusive, tampoco hubo crecimiento de la cepa T2.1 transformada con el plásmido pJLBaroG<sup>fbr</sup>tktA, la cual es nuestro control positivo de complementación. Por lo tanto, realizamos una adaptación del medio M9, manteniendo esta base mineral y adicionando los precursores PABA, PHBA y 2,3-DHBA del medio ARO; este medio se denominó M9S. Este medio fue utilizado tanto en los experimentos de complementación con el plásmido pJLBaroG<sup>fbr</sup>tktA así como con el banco metagenómico, tanto en agar como en matraz, obteniendo resultados satisfactorios en ambos experimentos.

#### **VII.5. Análisis de clonas transformantes T2.1 con recuperación de actividad de DAHP sintasa**

A partir de un cultivo de células T2.1 transformadas con el banco metagenómico en 50 mL de medio M9S, se realizó una resiembra y selección de clonas únicas en agar M9S Km<sup>15</sup>Cb<sup>100</sup>Chx<sup>100</sup>; estas clonas se denominaron UGGL. Asimismo, se identificaron algunas clonas en la siembra directa en agar M9S, denominando a estas clonas UGGS. En total fueron seleccionadas 133 clonas UGGL, mientras que en la siembra directa en agar, se obtuvieron 5 clonas UGGS.

##### **VII.5.1 Análisis de restricción de plásmidos pUGG**

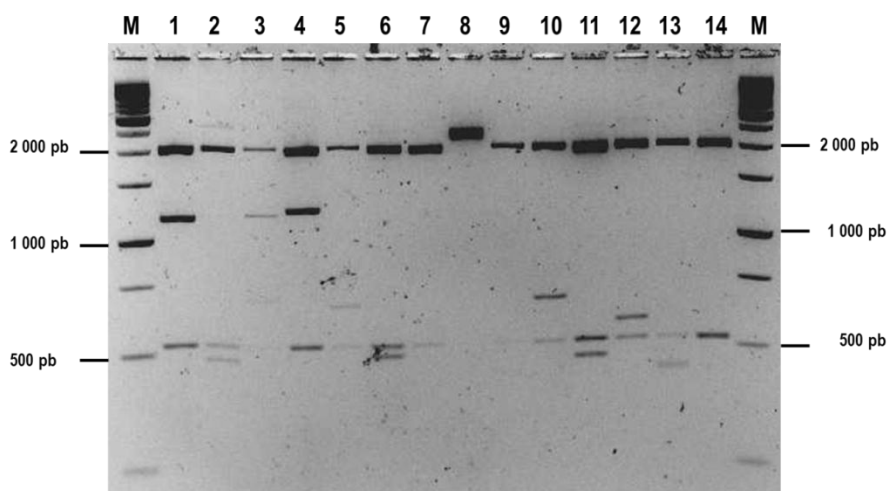
Se realizó el análisis de clonas mediante perfiles de restricción, esto con la finalidad de establecer patrones únicos de restricción que indiquen la presencia de insertos únicos dentro de los plásmidos. A partir de cada perfil de restricción único se seleccionó una clona para secuenciar el

inserto correspondiente. Para realizar la selección de las enzimas de restricción a emplear, se seleccionaron y obtuvieron las secuencias nucleotídicas de los genes que codifican para DAHP sintasas en *B. subtilis* (gen *aroA*), representativa de la subfamilia AroA<sub>β</sub>, y en *E. coli* (gen *aroF*), representativa de la subfamilia AroA<sub>α</sub>, las cuales son representativas de las dos principales subfamilias de DAHP sintasas microbianas. Con estas secuencias, se realizó una búsqueda de sitios de corte para enzimas de restricción disponibles que actúen en ambas secuencias (Tabla 9).

**Tabla 9.** Enzimas de restricción con sitios de corte común en los genes *aroA* (*Bacillus subtilis*) y *aroF* (*Escherichia coli*).

| Enzima de restricción         | Sitio de corte en gen <i>aroA</i> (pb) | Sitio de corte en gen <i>aroF</i> (pb) |
|-------------------------------|--|--|
| <i>Mun</i> I ( <i>Mfe</i> I)  | 314                                    | 485                                    |
| <i>MspA</i> I                 | 416                                    | 1 048                                  |
| <i>Pvu</i> II                 | 416                                    | 1 048                                  |
| <i>Bsp1286</i> I              | 474                                    | 491                                    |
| <i>BstX</i> I                 | 489                                    | 336                                    |
| <i>BssH</i> II                | 625                                    | 603                                    |
| <i>Pst</i> I                  | 413                                    | 370                                    |
| <i>BspM</i> I ( <i>Bve</i> I) | 396                                    | 651                                    |

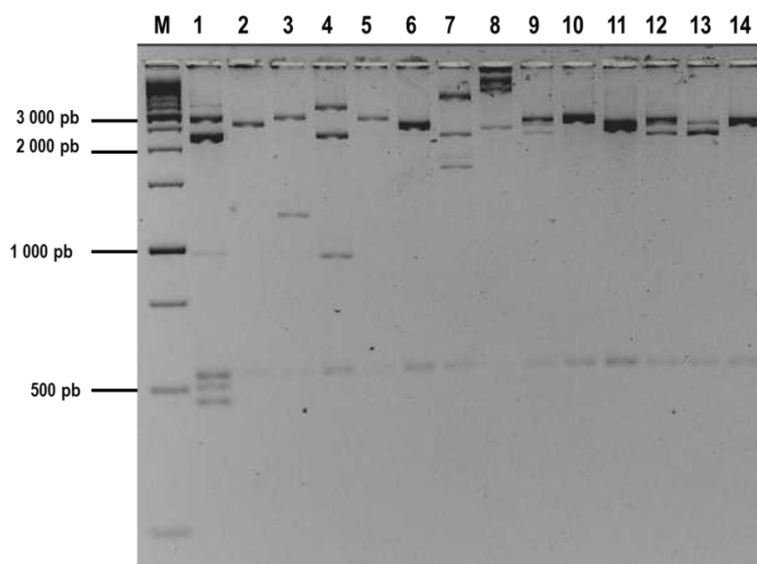
De acuerdo a las posibles enzimas para obtener los perfiles de restricción, fueron seleccionadas las enzimas *Pst* I y *Pvu* II (Figura 29). Asimismo, se eligieron 14 clonas para evaluar la heterogeneidad de los perfiles de restricción que pueden obtenerse con las enzimas *Pst* I y *Pvu* II.



**Figura 31.** Perfiles de restricción de los plásmidos provenientes de las clonas UGGL(1-14) con las enzimas *Pst* I y *Pvu* II, en un gel de agarosa al 2.5%, con un marcador de tamaño molecular de 1 kb.



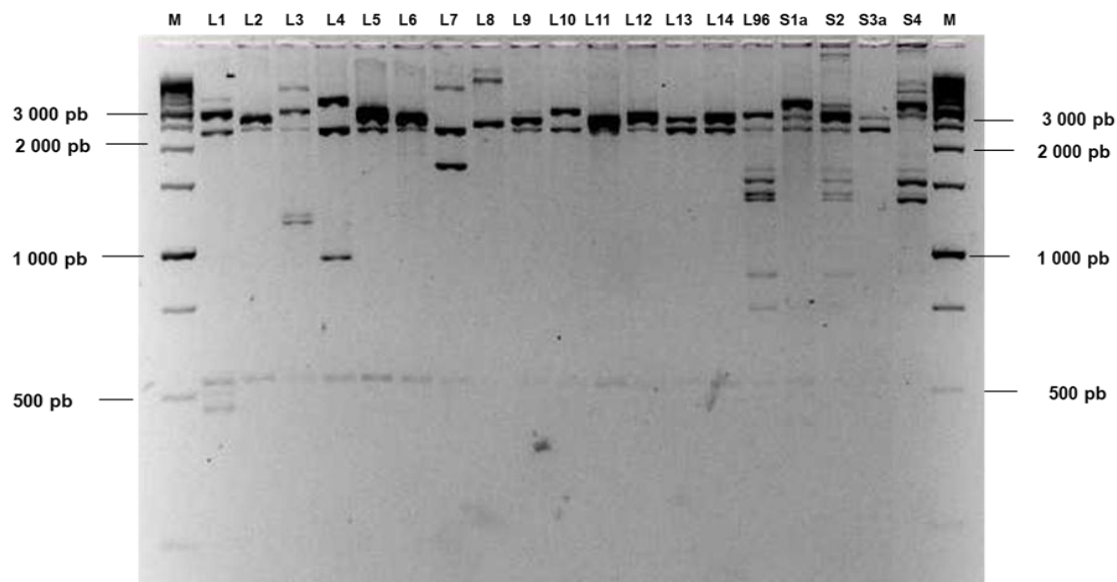
Los patrones de restricción obtenidos con las enzimas *Pst* I y *Pvu* II mostraron que las 14 clonas tienen patrones de restricción diferentes entre sí y, en consecuencia, tienen insertos únicos en cada plásmido que restituyen la actividad de DAHP sintasa y el crecimiento de las clonas. Sin embargo, se realizó un nuevo análisis de restricción con las enzimas *Pvu* II y *Bsp*M I con la finalidad de buscar patrones de restricción más heterogéneos para cada una de estas 14 clonas, y así compararlo con el patrón anterior.



**Figura 32.** Perfiles de restricción de los plásmidos provenientes de las clonas UGGL (1-14) con las enzimas *Bsp*M I y *Pvu* II, en un gel de agarosa al 2.5%, con un marcador de tamaño molecular de 1 kb.

Como se puede observar en la Figura 32, de la restricción con las enzimas *Bsp*M I y *Pvu* II, se confirma que estas 14 clonas tienen un inserto distinto en sus plásmidos pUGGL. Sin embargo se obtienen patrones de restricción más heterogéneos para cada clona con esta combinación de enzimas, por lo que en experimentos posteriores de análisis de restricción, se utilizarán las enzimas *Bsp*M I y *Pvu* II.

Una vez establecido el uso de las enzimas señaladas en el párrafo anterior, se procedió a realizar un análisis de las demás clonas restantes para determinar, como en el grupo anterior, las posibles clonas con insertos únicos en los plásmidos. Después del análisis del patrón de restricción se obtuvieron 19 perfiles únicos cuyo DNA fue secuenciado.



**Figura 33.** Perfiles de restricción de los plásmidos provenientes de las 19 clonas UGG con perfiles únicos en un gel de agarosa al 2.5 % con las enzimas *BspM I* y *Pvu II* y un marcador de tamaño molecular de 1 kb.

### VII.5.2. Secuenciación de insertos de clonas UGG

La Tabla 10 muestra el tamaño de los insertos secuenciados provenientes de las clonas UGGL(1-14, 96) y UGGS (1a, 2, 3a, 4).

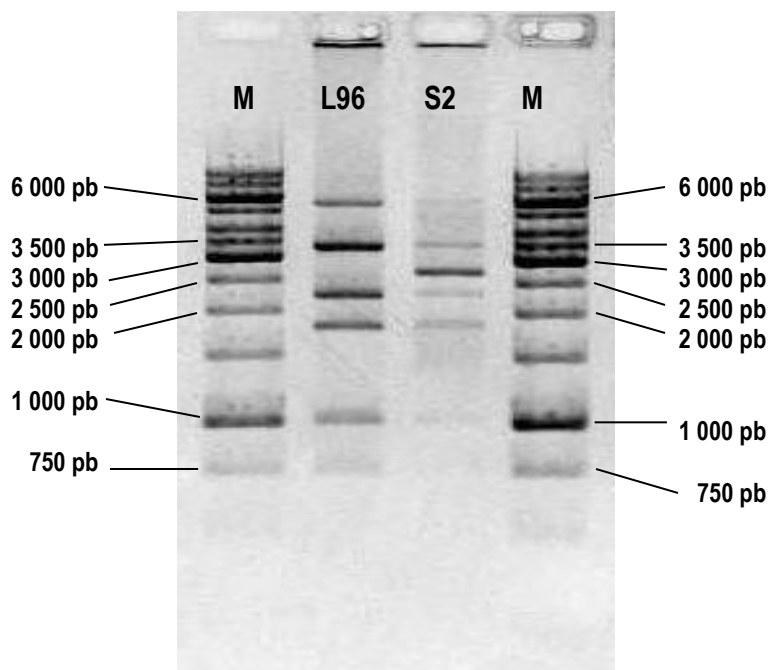
**Tabla 10.** Tamaño de inserto de secuencias de las clonas UGG.

| Secuencia | Tamaño de inserto (pb) | Secuencia | Tamaño de inserto (pb) |
|-----------|------------------------|-----------|------------------------|
| UGGL1     | 786                    | UGGL11    | 85                     |
| UGGL2     | 90                     | UGGL12    | 195                    |
| UGGL3     | ≈ 1 512                | UGGL13    | 490                    |
| UGGL4     | 867                    | UGGL14    | 127                    |
| UGGL5     | 272                    | UGGL96    | > 1 565 *              |
| UGGL6     | 104                    | UGGS1a    | 1 142                  |
| UGGL7     | 1 626                  | UGGS2     | > 1 246 *              |
| UGGL8     | > 770                  | UGGS3a    | 298                    |
| UGGL9     | 224                    | UGGS4     | > 748                  |
| UGGL10    | 292                    |           |                        |

\* El tamaño de la secuencia de las clonas UGGL96 y UGGS2 corresponde a un evento de secuenciación empleando los iniciadores 23-mer y 24-mer y no corresponde al tamaño de la secuencia completa.

De acuerdo a los resultados de la sección de los perfiles de restricción únicos y el tamaño de las secuencias obtenidas, se puede concluir que el criterio de selección por complementación de actividad en la triple mutante de DAHP sintasa en medio mínimo derivó en la obtención de 19 clonas con secuencias únicas pero algunas con un tamaño muy pequeño (entre 85 y 90 pb). Estas clonas (e.g. UGGL2) es claro que no corresponden al tamaño esperado de un gen que codifica para una DAHP sintasa, aproximadamente de 1 000 pb. Para los propósitos de nuestro proyecto, fueron de nuestro particular interés las clonas que tuvieron un inserto mayor o igual a 1 000 pb, así como aquellas clonas que presentaron un inserto con una secuencia única. No obstante que no se obtuvo la secuencia completa del inserto de la clona UGGS4, y a pesar de que las clonas UGGL96 y UGGS4 mostraron un perfil de restricción aparentemente distinto, las secuencias obtenidas del inserto de ambas clonas son idénticas; por lo tanto, fueron seleccionadas la clonas UGGL96 y UGGS2 para determinar sus secuencias completas.

Para elegir la estrategia para secuenciar, se realizó una estimación del tamaño de inserto mediante perfil de restricción. Para esto fue seleccionada la enzima *Ade I*, la cual tiene solo un sitio de corte en el vector pJET1.2, considerando también la posibilidad que dentro de los insertos existan sitios de corte para esta enzima. De acuerdo al perfil de restricción obtenido (Figura 32), y considerando que el tamaño del vector pJET1.2 es de 2 974 pb, los insertos de las clonas UGGL96 y UGGS2 tienen un tamaño de inserto aproximado de 10 y 12 kb respectivamente. Se realizaron varios intentos de continuar con la secuenciación de los insertos por medio de *primer-walking* sin obtener resultados, por lo que se decidió secuenciar completamente estos dos plásmidos por secuenciación masiva.

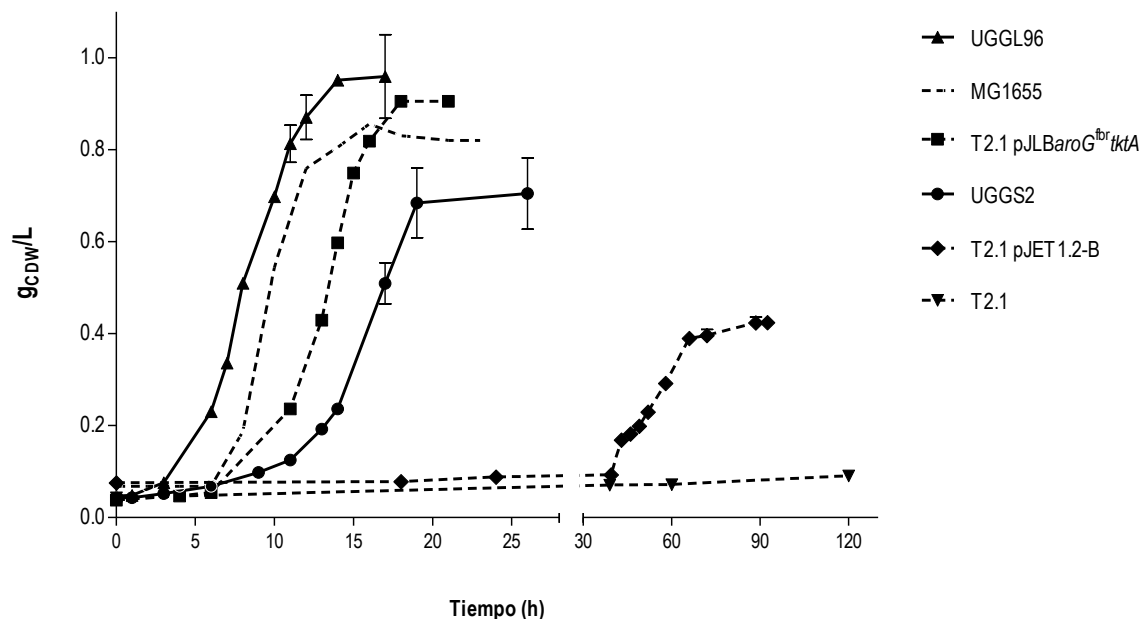


**Figura 34.** Perfil de restricción de los plásmidos pUGGL96 y pUGGS2 realizados con la enzima *Ade I* en un gel de agarosa al 1.5 % utilizando un marcador de tamaño molecular de 1 kb.

### VII.5.3. Cinéticas de crecimiento de las clonas UGG

El medio M9S con glucosa como única fuente de carbono y energía fue utilizado para analizar el crecimiento de las clonas UGGL96 y UGGS2. El control positivo considerado fue la cepa T2.1 transformada con el plásmido pJLBaroG<sup>fbr</sup>tktA, el cual contiene el gen *aroG<sup>fbr</sup>*, que codifica para una DAHPS de *E. coli* resistente a retroinhibición por fenilalanina. Asimismo, también fue contemplada la cepa T2.1 transformada con el vector pJET1.2 como un control negativo. Sin embargo, debido al diseño del vector pJET1.2, no fue posible obtener un vector íntegro circularizado sin inserto; esto debido a que la presencia de un inserto dentro del sitio de multiclonación también interrumpe el marco de lectura del gen letal *eco47IR*, permitiendo la selección positiva de clonas con inserto. De tal manera que, al circularizar el vector, se reintegró el marco de lectura del gen letal *eco47IR*, impidiendo la selección de clonas con vector pJET1.2 circularizado. Para solucionar este problema, fue escindido un fragmento de 50 pb dentro del gen *eco47IR* con la enzima *Bgl* II, interrumpiendo así el marco de lectura de este gen y permitiendo la recircularización del plásmido. Este plásmido carente de estos 50 pb se denominó pJET1.2-B. La cepa T2.1 transformada con este vector se tomó como un control adicional a los experimentos de transformación. Es importante señalar que al final de cada cinética de crecimiento se realizó la amplificación por PCR de los genes

interrumpidos *aroF*, *aroG* y *aroH* para descartar la posibilidad de que se presentara algún cambio en éstos durante la cinética de crecimiento.



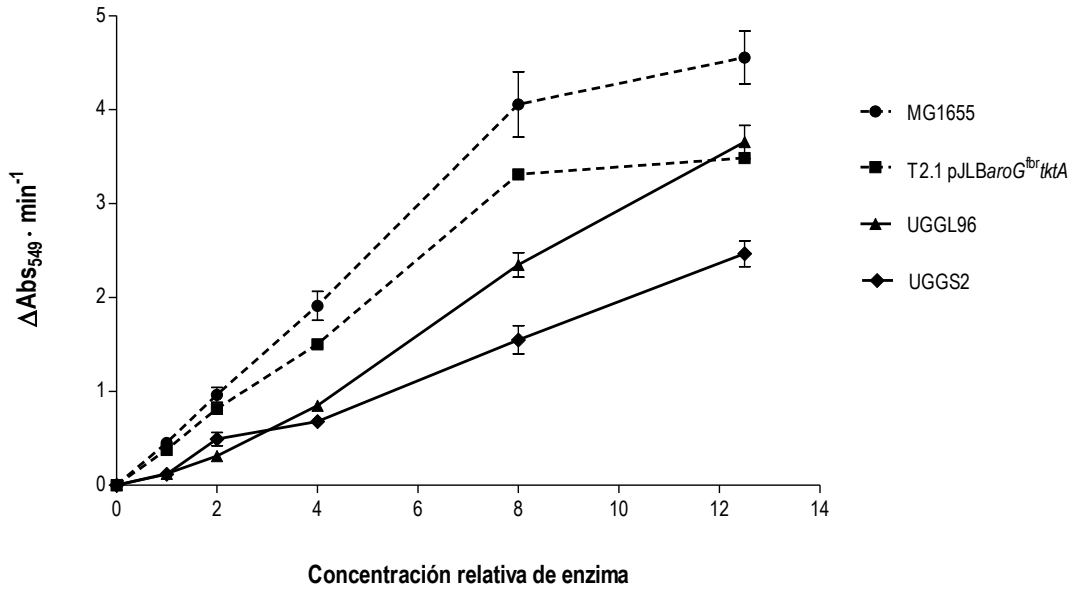
**Figura 35.** Cinéticas de crecimiento en medio mínimo M9S de las clonas UGGL96 y UGGS2 que recuperaron la actividad de DAHP sintasa, así como las cepas control MG1655, T2.1, T2.1 pJLBaroG<sup>brtkA</sup> y T2.1 pJET1.2-B. Todos los cultivos se realizaron por duplicado en matraces bafleados de 250 mL con 50 mL de medio mínimo M9S con 2 g/L de glucosa como única fuente de carbono, incubados en agitación constante de 300 rpm a 37 °C. Los cultivos fueron iniciados a una densidad óptica de 0.1 D.O.<sub>600</sub>.

En todas las cepas utilizadas en las cinéticas de crecimiento no se observaron cambios en los amplicones de los genes interrumpidos *aroF*, *aroG* y *aroH*, por lo que no hubo modificaciones cromosomales en estos genes a lo largo del experimento. En el cultivo de la cepa T2.1 no se presentó crecimiento, a pesar de que el cultivo permaneció en incubación por 120 horas continuas. Esto permite descartar cualquier evento de supresión de las deleciones de los genes *aroF*, *aroG* y *aroH*. En el caso de la cepa transformada con el plásmido pJET1.2-B, se observó crecimiento después de las 30 h de cultivo alcanzando un máximo de biomasa de 0.4 g<sub>CDW</sub>/L, correspondiente a menos de la mitad de la biomasa obtenida para la cepa MG1655. Por el comportamiento de esta cepa podemos inferir que el vector pJET1.2-B tiene efecto sobre el crecimiento de la cepa T2.1; sin embargo, por el momento no contamos con evidencia que nos permita atribuir y explicar el efecto exacto que ejerce este plásmido en el fondo genético de la cepa T2.1 para recuperar la actividad de DAHP sintasa.

El crecimiento microbiano en la cepa T2.1 pJET1.2-B permitió establecer un criterio de selección de clonas para la determinación de actividad enzimática; tomando como referencia la duración de la fase lag de crecimiento en esta cepa, se eligieron solamente las clonas que presentaran una fase lag menor a 30 h de duración. En este caso, las únicas dos clonas que cumplieron este criterio fueron la UGGL96 y UGGS2. El crecimiento observado en estas clonas es interesante, particularmente el de la clona UGGL96, ya que tiene una fase lag relativamente corta, inclusive menor a la cepa MG1655 y T2.1 pJLBaroG<sup>fbtktA</sup> en las condiciones de cultivo establecidas; este crecimiento característico de estas cepas se puede atribuir al contenido de los insertos en los plásmidos.

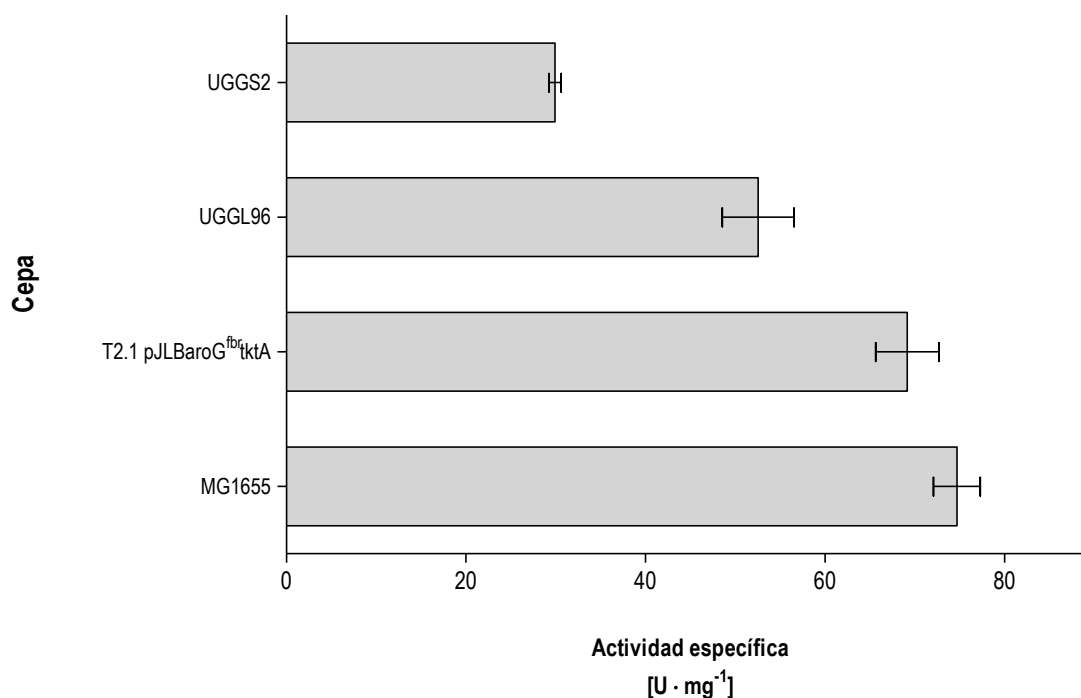
#### **VII.5.4. Determinación de actividad enzimática específica de DAHP sintasa**

La determinación de actividad enzimática específica de DAHP sintasa fue realizada para las clonas UGGL96 y UGGS2 a partir del extracto celular crudo obtenido de un cultivo celular en medio mínimo M9S; las células fueron cosechadas en la etapa exponencial tardía de su crecimiento. Previo a la cuantificación de actividad, se determinó el intervalo de linealidad para la actividad enzimática, realizando determinaciones a distintas concentraciones relativas de enzima; para realizar los cálculos de cuantificación de proteína se eligió el punto donde se registrara el mayor valor de absorbancia y, que a su vez, conservara la relación lineal entre la cantidad de extracto (enzima) y absorbancia (Figura 36).



**Figura 36.** Linealidad de actividad enzimática de DAHP sintasa de las cepas MG1655, T2.1 pJLBaroG<sup>fbr</sup>tktA y las clonas UGGS2 y UGGL96.

Para las clonas probadas, se obtuvo un intervalo lineal de actividad hasta la concentración relativa final de 12.5; mientras que para las cepas de referencia MG1655 y T2.1 pJLBaroG<sup>fbr</sup>tktA, la linealidad se conservó solamente hasta la concentración relativa de 8. En cuanto a estas dos últimas cepas, es importante mencionar la similitud en el comportamiento de linealidad; recordando también que el gen *aroG<sup>fbr</sup>* es una variante del gen *aroG* silvestre de *E. coli* cuya modificación principal es la resistencia a inhibición alostérica por fenilalanina, y además que este producto AroG posee más del 80% de actividad enzimática de DAHP sintasa en *E. coli*. Al establecer los puntos de máxima actividad enzimática, dentro del intervalo de linealidad, se procedió a realizar el cálculo de actividad enzimática específica (Figura 37).



**Figura 37.** Actividad enzimática de DAHP sintasa de las cepas MG1655, T2.1 pJLBaroG<sup>fbr</sup>tkkA y las clonas UGGS2 y UGGL96. La equivalencia de la unidad enzimática de DAHP sintasa es: 1 U = 1 nmol<sub>DAHP</sub> · min<sup>-1</sup>.

De acuerdo a los resultados de la figura 34, se observa que las cepas MG1655 y T2.1 pJLBaroG<sup>fbr</sup>tkkA tienen los valores más elevados de actividad enzimática específica (entre 70 – 75 U · mg<sup>-1</sup>), mientras que la clona UGGL96 registró un valor de actividad aproximado de 60 U, y las clonas UGGS2 y UGGS4 se detectaron valores cercanos a las 35 U · mg<sup>-1</sup>. Sin conocer la secuencia de los insertos de las clonas estudiadas, es posible que la diferencia en el crecimiento bacteriano sea atribuible a algún gen que se encuentra dentro de los insertos, el cual le confiere una ventaja para el crecimiento. Otro factor que pudiera estar involucrado es el origen de las secuencias de los insertos, ya que la expresión de genes que corresponden a organismos filogenéticamente más cercanos a la bacteria huésped ( $\gamma$ -Proteobacteria en el caso de *E. coli*) podría corresponder a una transcripción y traducción más eficiente que se reflejaría en la disminución de la fase lag de crecimiento



## VII.5.5. Secuenciación masiva de las clonas UGGL96 y UGGS2

Las clonas UGGL96 y UGGS2 fueron seleccionadas para su secuenciación completa por medio de una técnica de secuenciación masiva. En el proceso de secuenciación, se obtuvieron también secuencias tanto de *E. coli* como del vector pJET1.2; éstas fueron depuradas y excluidas para considerar solamente las secuencias del inserto. Mediante un análisis BLASTX se determinaron la identidad de los *contigs* y *scaffolds* obtenidos en la secuenciación. En las Tablas 11 y 12 se muestran los mejores *hits* de BLASTX para *contigs* y *scaffolds* obtenidos correspondientes a los insertos de UGGL96 (Tabla 11) y UGGS2 (Tabla 12).

**Tabla 11.** Resultados de BLASTX realizados a los *contigs* y *scaffolds* obtenidos en los insertos de la clona UGGL96.

| Contig/Scaffold | Inicio | Final | Descripción                                | Organismo                         | Referencia NCBI |
|-----------------|--------|-------|--|-----------------------------------|-----------------|
| C288            | 151    | 2     | Proteína hipotética CRE_01315              | <i>Caenorhabditis remanei</i>     | XP_003094876.1  |
| C317            | 180    | 1     | Factor 2 de liberación de cadena peptídica | <i>Salmonella enterica</i>        | ESH65619.1      |
| C539            | 381    | 1     | Glicosiltransferasa                        | <i>Thiocapsa marina</i>           | WP_007191375.1  |
| C541            | 1      | 381   | Proteína hipotética ACD_39C00256G0001      | Bacteria no cultivable            | EKD84083.1      |
| C543            | 103    | 2     | Proteína hipotética NIDE2968               | <i>Nitrospira defluvii</i>        | YP_003798589.1  |
| C545            | 381    | 1     | Fosfo-2-dehidro-3-deoxiheptonato aldolasa  | <i>Nitrospira defluvii</i>        | YP_003798590.1  |
| C547            | 3      | 356   | Proteína hipotética N47_F15770             | <i>Desulfobacterium</i> sp.       | CBX29882.1      |
| C549            | 379    | 47    | Glicosiltransferasa                        | <i>Solibacter usitatus</i>        | YP_822169.1     |
| C551            | 366    | 1     | Glicosiltransferasa                        | <i>Solibacter usitatus</i>        | YP_822169.1     |
| C553            | 2      | 379   | Glicosiltransferasa                        | <i>Thiocapsa marina</i>           | WP_007191375.1  |
| C555            | 380    | 132   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa  | <i>Nitrospira defluvii</i>        | YP_003798590.1  |
| C557            | 330    | 142   | Proteína hipotética NIDE2968               | <i>Nitrospira defluvii</i>        | YP_003798589.1  |
| C559            | 324    | 67    | Permeasa/transportador ABC                 | <i>Mycobacterium tusciae</i>      | WP_006243305.1  |
| C561            | 3      | 380   | Proteína hipotética N47_F15770             | <i>Desulfobacterium</i> sp.       | CBX29882.1      |
| C563            | 381    | 1     | Fosfo-2-dehidro-3-deoxiheptonato aldolasa  | <i>Nitrospira defluvii</i>        | YP_003798590.1  |
| C565            | 174    | 1     | Fosfo-2-dehidro-3-deoxiheptonato aldolasa  | <i>Nitrospira defluvii</i>        | YP_003798590.1  |
| C567            | 1      | 342   | Receptor TonB-dependiente                  | <i>Nitrospira defluvii</i>        | YP_003798588.1  |
| C569            | 2      | 382   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa  | <i>Nitrospira defluvii</i>        | YP_003798590.1  |
| C571            | 379    | 2     | Fosfo-2-dehidro-3-deoxiheptonato aldolasa  | <i>Nitrospira defluvii</i>        | YP_003798590.1  |
| C573            | 263    | 9     | Receptor TonB-dependiente                  | <i>Nitrospira defluvii</i>        | YP_003798588.1  |
| C575            | 5      | 379   | Glicosiltransferasa                        | <i>Desulfomicrobium baculatum</i> | YP_003158688.1  |
| C577            | 380    | 111   | Receptor TonB-dependiente                  | <i>Nitrospira defluvii</i>        | YP_003798588.1  |
| C579            | 381    | 1     | Proteína hipotética N47_F15770             | <i>Desulfobacterium</i> sp.       | CBX29882.1      |
| C581            | 3      | 380   | Proteína hipotética N47_F15770             | <i>Desulfobacterium</i> sp.       | CBX29882.1      |
| C583            | 1      | 381   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa  | <i>Nitrospira defluvii</i>        | YP_003798590.1  |
| C585            | 2      | 382   | Glicosiltransferasa                        | <i>Desulfomicrobium baculatum</i> | YP_003158688.1  |
| C587            | 3      | 380   | Celobiosa fosforilasa                      | <i>Thiocystis violascens</i>      | YP_006414616.1  |
| C589            | 221    | 3     | Proteína hipotética NIDE2968               | <i>Nitrospira defluvii</i>        | YP_003798589.1  |

|      |     |     |   |                                  |                |
|------|-----|-----|---|----------------------------------|----------------|
| C591 | 1   | 381 | Proteína hipotética                       | <i>Lamprocystis purpurea</i>     | WP_020506171.1 |
| C593 | 2   | 319 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C595 | 35  | 382 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C597 | 274 | 2   | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C599 | 131 | 3   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C601 | 2   | 202 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C603 | 16  | 204 | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>     | CCA73309.1     |
| C605 | 215 | 382 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C607 | 381 | 16  | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>     | YP_006414616.1 |
| C611 | 3   | 380 | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C613 | 381 | 1   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C615 | 365 | 3   | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>     | YP_006414616.1 |
| C617 | 288 | 383 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C619 | 94  | 381 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C621 | 383 | 3   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C623 | 383 | 3   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C625 | 3   | 383 | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C627 | 382 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C631 | 1   | 382 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C633 | 282 | 383 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C635 | 183 | 383 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C637 | 3   | 383 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C639 | 9   | 383 | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C641 | 329 | 72  | Permeasa/transportador ABC                | <i>Mycobacterium tusciae</i>     | WP_006243305.1 |
| C647 | 116 | 3   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C651 | 381 | 1   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C653 | 381 | 211 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C655 | 379 | 158 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C657 | 383 | 3   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C659 | 1   | 165 | $\beta$ -1-2-glucano sintasa cíclica      | <i>Collimonas fungivorans</i>    | YP_004752371.1 |
| C663 | 3   | 383 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C665 | 306 | 1   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C669 | 132 | 383 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C671 | 3   | 383 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C673 | 3   | 188 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C675 | 383 | 3   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C677 | 2   | 382 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C679 | 1   | 210 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C681 | 1   | 381 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C683 | 383 | 3   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C685 | 383 | 141 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C687 | 382 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C691 | 4   | 375 | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C693 | 383 | 3   | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C695 | 5   | 184 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C697 | 144 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C699 | 61  | 381 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C701 | 381 | 250 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C703 | 3   | 383 | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |

|      |     |     |   |                                   |                |
|------|-----|-----|---|-----------------------------------|----------------|
| C705 | 3   | 383 | Glicosiltransferasa                       | <i>Azoarcus</i> sp.               | YP_007551292.1 |
| C707 | 164 | 382 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>        | YP_003798588.1 |
| C709 | 82  | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C711 | 182 | 379 | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>      | YP_006414616.1 |
| C713 | 160 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C715 | 88  | 381 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C717 | 382 | 212 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C719 | 1   | 192 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C721 | 1   | 384 | Glicosiltransferasa                       | <i>Thiocapsa marina</i>           | WP_007191375.1 |
| C723 | 296 | 3   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>        | YP_003798588.1 |
| C725 | 1   | 384 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C727 | 382 | 110 | $\beta$ -1-2-glucano sintasa cíclica      | <i>Collimonas fungivorans</i>     | YP_004752371.1 |
| C729 | 3   | 380 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i>  | WP_020878070.1 |
| C731 | 383 | 210 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>        | YP_003798589.1 |
| C733 | 19  | 384 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>        | YP_003798588.1 |
| C735 | 6   | 380 | Glicosiltransferasa                       | <i>Desulfomicrobium baculatum</i> | YP_003158688.1 |
| C737 | 202 | 384 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>        | YP_003798589.1 |
| C739 | 2   | 253 | Proteína hipotética PHACADRAFT_199064     | <i>Phanerochaete carnosa</i>      | EKM51552.1     |
| C741 | 384 | 1   | Glicosiltransferasa                       | <i>Azoarcus</i> sp.               | YP_007551292.1 |
| C743 | 382 | 2   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.       | CBX29882.1     |
| C745 | 2   | 382 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C747 | 384 | 1   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i>  | WP_020878070.1 |
| C751 | 313 | 2   | Glicosiltransferasa                       | Bacteria no cultivable            | EKD34483.1     |
| C753 | 384 | 1   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.       | CBX29882.1     |
| C755 | 382 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C757 | 383 | 9   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.       | CBX29882.1     |
| C761 | 1   | 231 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>        | YP_003798588.1 |
| C763 | 8   | 283 | Permeasa/transportador ABC                | <i>Mycobacterium tusciae</i>      | WP_006243305.1 |
| C765 | 211 | 381 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>        | YP_003798589.1 |
| C769 | 384 | 4   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.       | CBX29882.1     |
| C771 | 382 | 5   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i>  | WP_020878070.1 |
| C773 | 2   | 382 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i>  | WP_020878070.1 |
| C775 | 384 | 115 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C779 | 60  | 383 | Glicosiltransferasa                       | Bacteria no cultivable            | EKD34483.1     |
| C783 | 382 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C785 | 384 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C787 | 5   | 382 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>        | YP_003798588.1 |
| C789 | 1   | 120 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>        | YP_003798589.1 |
| C791 | 3   | 383 | Proteína hipotética ACD_39C00256G0001     | Bacteria no cultivable            | EKD84083.1     |
| C795 | 382 | 236 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>        | YP_003798589.1 |
| C797 | 2   | 328 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.       | CBX29882.1     |
| C799 | 383 | 3   | Glicosiltransferasa                       | <i>Thiocapsa marina</i>           | WP_007191375.1 |
| C801 | 382 | 11  | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.       | CBX29882.1     |
| C803 | 273 | 1   | Glicosiltransferasa                       | Bacteria no cultivable            | EKD34483.1     |
| C805 | 382 | 2   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i>  | WP_020878070.1 |
| C807 | 2   | 382 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i>  | WP_020878070.1 |
| C809 | 382 | 32  | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C813 | 1   | 384 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.       | CBX29882.1     |
| C817 | 321 | 1   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>        | YP_003798588.1 |

|      |     |     |   |                                  |                |
|------|-----|-----|---|----------------------------------|----------------|
| C819 | 243 | 1   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C821 | 2   | 364 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C823 | 2   | 385 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C825 | 2   | 127 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C827 | 60  | 383 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C829 | 190 | 384 | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>     | YP_006414616.1 |
| C831 | 378 | 4   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C833 | 193 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C835 | 266 | 75  | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>     | CCA73309.1     |
| C837 | 3   | 221 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C841 | 384 | 1   | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>     | YP_006414616.1 |
| C845 | 1   | 381 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C847 | 383 | 99  | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C849 | 382 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C851 | 385 | 128 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C853 | 3   | 383 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C855 | 384 | 1   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C857 | 385 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C859 | 2   | 154 | $\beta$ -1-2-glucano sintasa cíclica      | <i>Collimonas fungivorans</i>    | YP_004752371.1 |
| C861 | 60  | 386 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C863 | 385 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C865 | 80  | 385 | Permeasa/transportador ABC                | <i>Mycobacterium tusciae</i>     | WP_006243305.1 |
| C867 | 41  | 229 | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>     | CCA73309.1     |
| C869 | 386 | 6   | Glicosiltransferasa                       | <i>Solibacter usitatus</i>       | YP_822169.1    |
| C873 | 386 | 3   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C875 | 379 | 191 | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>     | CCA73309.1     |
| C877 | 1   | 381 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C879 | 27  | 215 | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>     | CCA73309.1     |
| C881 | 386 | 3   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C883 | 3   | 386 | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C885 | 202 | 387 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C887 | 136 | 387 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C889 | 1   | 387 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C891 | 140 | 3   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C893 | 4   | 111 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C895 | 240 | 52  | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>     | CCA73309.1     |
| C897 | 2   | 352 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C899 | 12  | 389 | Proteína hipotética                       | <i>Leeia oryzae</i>              | WP_018149968.1 |
| C903 | 116 | 391 | Glicosiltransferasa                       | Bacteria no cultivable           | EKD34483.1     |
| C905 | 273 | 1   | Glicosiltransferasa                       | Bacteria no cultivable           | EKD34483.1     |
| C907 | 388 | 5   | Glicosiltransferasa                       | <i>Solibacter usitatus</i>       | YP_822169.1    |
| C911 | 1   | 390 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C913 | 118 | 390 | Glicosiltransferasa                       | Bacteria no cultivable           | EKD34483.1     |
| C915 | 2   | 391 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C919 | 392 | 3   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C921 | 156 | 1   | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C923 | 2   | 364 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C925 | 365 | 108 | Permeasa/transportador ABC                | <i>Mycobacterium tusciae</i>     | WP_006243305.1 |
| C929 | 396 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |

|       |     |     |   |                                  |                |
|-------|-----|-----|---|----------------------------------|----------------|
| C933  | 325 | 68  | Permeasa/transportador ABC                | <i>Mycobacterium tusciae</i>     | WP_006243305.1 |
| C937  | 372 | 1   | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>     | YP_006414616.1 |
| C939  | 346 | 2   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C941  | 268 | 110 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C943  | 3   | 398 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C945  | 248 | 397 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C947  | 1   | 399 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C951  | 5   | 400 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C959  | 170 | 400 | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>     | YP_006414616.1 |
| C963  | 401 | 231 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C965  | 2   | 403 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C971  | 395 | 3   | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C973  | 403 | 140 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C975  | 286 | 405 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C979  | 2   | 406 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C981  | 3   | 122 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C983  | 405 | 223 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C985  | 405 | 1   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C989  | 250 | 408 | Glicosiltransferasa                       | Bacteria no cultivable           | EKD34483.1     |
| C993  | 409 | 2   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C995  | 3   | 410 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C997  | 410 | 3   | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C999  | 1   | 411 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1001 | 412 | 2   | Glicosiltransferasa                       | <i>Solibacter usitatus</i>       | YP_822169.1    |
| C1003 | 1   | 414 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1005 | 415 | 2   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1007 | 3   | 416 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1009 | 417 | 1   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1011 | 234 | 416 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1015 | 414 | 1   | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C1017 | 421 | 2   | Glicosiltransferasa                       | <i>Thiorhodococcus drewsii</i>   | WP_007041693.1 |
| C1021 | 420 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1025 | 55  | 423 | Glicosiltransferasa                       | <i>Solibacter usitatus</i>       | YP_822169.1    |
| C1031 | 1   | 423 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1033 | 4   | 405 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1041 | 426 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1043 | 154 | 2   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1045 | 428 | 3   | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C1051 | 341 | 3   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1055 | 433 | 2   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C1059 | 1   | 432 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C1061 | 433 | 269 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C1065 | 66  | 437 | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>     | YP_006414616.1 |
| C1067 | 1   | 432 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1069 | 26  | 283 | Permeasa/transportador ABC                | <i>Mycobacterium tusciae</i>     | WP_006243305.1 |
| C1071 | 438 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1073 | 2   | 439 | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C1075 | 165 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1077 | 442 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |

|       |     |     |   |                                  |                |
|-------|-----|-----|---|----------------------------------|----------------|
| C1079 | 249 | 443 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1081 | 149 | 3   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1083 | 104 | 3   | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C1085 | 115 | 273 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C1087 | 2   | 442 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C1103 | 455 | 213 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1105 | 1   | 456 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1107 | 457 | 2   | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C1109 | 459 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1111 | 2   | 385 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1115 | 2   | 460 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1117 | 1   | 264 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C1121 | 241 | 459 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1127 | 464 | 3   | Glicosiltransferasa                       | <i>Thiorhodococcus drewsii</i>   | WP_007041693.1 |
| C1129 | 465 | 253 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1133 | 248 | 469 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1135 | 10  | 468 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1137 | 471 | 121 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1139 | 3   | 473 | Celobiosa fosforilasa                     | <i>Thioflavicoccus mobilis</i>   | YP_007244847.1 |
| C1143 | 11  | 421 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1149 | 476 | 3   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C1151 | 141 | 4   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1153 | 2   | 475 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1155 | 3   | 476 | Proteína hipotética                       | <i>Leeia oryzae</i>              | WP_018149968.1 |
| C1159 | 415 | 203 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C1163 | 3   | 476 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C1165 | 224 | 484 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1167 | 338 | 81  | Permeasa/transportador ABC                | <i>Mycobacterium tusciae</i>     | WP_006243305.1 |
| C1169 | 47  | 304 | Permeasa/transportador ABC                | <i>Mycobacterium tusciae</i>     | WP_006243305.1 |
| C1171 | 266 | 12  | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1173 | 235 | 2   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1179 | 1   | 483 | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>     | YP_006414616.1 |
| C1181 | 337 | 179 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C1183 | 495 | 4   | Glicosiltransferasa                       | <i>Azoarcus</i> sp.              | YP_007551292.1 |
| C1185 | 1   | 495 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1187 | 3   | 410 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1189 | 1   | 498 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1191 | 456 | 199 | Permeasa/transportador ABC                | <i>Mycobacterium tusciae</i>     | WP_006243305.1 |
| C1195 | 499 | 254 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1197 | 48  | 239 | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>     | CCA73309.1     |
| C1199 | 109 | 267 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C1201 | 497 | 3   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C1203 | 502 | 2   | Glicosiltransferasa                       | <i>Thiocapsa marina</i>          | WP_007191375.1 |
| C1205 | 503 | 3   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1207 | 3   | 227 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1209 | 504 | 274 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1213 | 232 | 2   | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>     | YP_006414616.1 |
| C1215 | 3   | 506 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1217 | 474 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |

|       |     |     |   |                                  |                |
|-------|-----|-----|---|----------------------------------|----------------|
| C1219 | 506 | 123 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1221 | 505 | 2   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1223 | 3   | 149 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1225 | 327 | 124 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C1227 | 164 | 3   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1229 | 337 | 507 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1235 | 511 | 113 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1237 | 252 | 506 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1239 | 2   | 511 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1241 | 9   | 512 | Proteína hipotética ACD_39C00256G0001     | Bacteria no cultivable           | EKD84083.1     |
| C1243 | 511 | 2   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C1245 | 511 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1247 | 3   | 512 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1249 | 434 | 4   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1251 | 512 | 3   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1253 | 1   | 354 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1255 | 515 | 9   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1267 | 2   | 226 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1269 | 189 | 392 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C1271 | 239 | 3   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1275 | 32  | 526 | Glicosiltransferasa                       | <i>Solibacter usitatus</i>       | YP_822169.1    |
| C1277 | 528 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1281 | 374 | 532 | Glicosiltransferasa                       | Bacteria no cultivable           | EKD34483.1     |
| C1283 | 2   | 538 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C1285 | 343 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1287 | 3   | 542 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1289 | 3   | 428 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1291 | 1   | 213 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1293 | 1   | 282 | $\beta$ -1-2-glucano sintasa cíclica      | <i>Collimonas fungivorans</i>    | YP_004752371.1 |
| C1295 | 552 | 1   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C1297 | 558 | 1   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i> | WP_020878070.1 |
| C1301 | 562 | 260 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1303 | 213 | 55  | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C1309 | 358 | 2   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1315 | 282 | 4   | Glicosiltransferasa                       | Bacteria no cultivable           | EKD34483.1     |
| C1323 | 2   | 163 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1325 | 412 | 2   | Glicosiltransferasa                       | <i>Solibacter usitatus</i>       | YP_822169.1    |
| C1329 | 1   | 585 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1333 | 591 | 4   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1335 | 1   | 591 | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.      | CBX29882.1     |
| C1341 | 542 | 99  | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1345 | 179 | 382 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C1347 | 2   | 589 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1349 | 1   | 618 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |
| C1351 | 622 | 395 | $\beta$ -1-2-glucano sintasa cíclica      | <i>Collimonas fungivorans</i>    | YP_004752371.1 |
| C1353 | 2   | 595 | tRNA-ribosiltransferasa                   | <i>Bacteroides</i> sp.           | WP_022162476.1 |
| C1357 | 286 | 128 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>       | YP_003798589.1 |
| C1359 | 628 | 317 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>       | YP_003798588.1 |
| C1361 | 628 | 212 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>       | YP_003798590.1 |

|             |       |       |   |                                    |                |
|-------------|-------|-------|---|------------------------------------|----------------|
| C1363       | 632   | 3     | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.        | CBX29882.1     |
| C1367       | 2     | 166   | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>         | YP_003798589.1 |
| C1369       | 349   | 636   | Glicosiltransferasa                       | Bacteria no cultivable             | EKD34483.1     |
| C1371       | 1     | 639   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.        | CBX29882.1     |
| C1375       | 336   | 1     | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>         | YP_003798590.1 |
| C1377       | 113   | 640   | Glicosiltransferasa                       | <i>Solibacter usitatus</i>         | YP_822169.1    |
| C1379       | 642   | 1     | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.        | CBX29882.1     |
| C1383       | 353   | 511   | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>         | YP_003798589.1 |
| C1385       | 274   | 645   | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>       | YP_006414616.1 |
| C1387       | 390   | 647   | Glicosiltransferasa                       | Bacteria no cultivable             | EKD34483.1     |
| C1391       | 382   | 2     | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>       | YP_006414616.1 |
| C1399       | 2     | 667   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>         | YP_003798590.1 |
| C1401       | 672   | 4     | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.        | CBX29882.1     |
| C1407       | 373   | 2     | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>         | YP_003798590.1 |
| C1409       | 3     | 677   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.        | CBX29882.1     |
| C1417       | 3     | 701   | Proteína hipotética ACD_39C00256G0001     | Bacteria no cultivable             | EKD84083.1     |
| C1423       | 321   | 1     | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>         | YP_003798588.1 |
| C1447       | 762   | 1     | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i>   | WP_020878070.1 |
| C1471       | 317   | 3     | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>         | YP_003798590.1 |
| C1479       | 399   | 821   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>         | YP_003798588.1 |
| C1501       | 324   | 1     | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.        | CBX29882.1     |
| C1513       | 987   | 400   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>         | YP_003798590.1 |
| C1561       | 1     | 831   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>         | YP_003798590.1 |
| C1569       | 1 004 | 192   | Aminoglicósido 3'-fosfotransferasa        | <i>Corynebacterium diphtheriae</i> | NP_478145.1    |
| Scaffold 15 | 775   | 1 446 | Proteína hipotética ACD_39C00256G0001     | Bacteria no cultivable             | EKD84083.1     |
| Scaffold 24 | 4 985 | 6 676 | Chaperona HscC                            | <i>Shigella dysenteriae</i>        | WP_000367922.1 |

**Tabla 12.** Resultados de BLASTX realizados a los *contigs* y *scaffolds* obtenidos en los insertos de la clona UGGS2.

| Contig/Scaffold | Inicio | Final | Descripción                               | Organismo                       | Referencia NCBI |
|-----------------|--------|-------|---|---------------------------------|-----------------|
| C43             | 204    | 1     | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>    | CCA73309.1      |
| C47             | 61     | 255   | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>    | CCA73309.1      |
| C49             | 59     | 253   | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>    | CCA73309.1      |
| C59             | 32     | 244   | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>    | CCA73309.1      |
| C61             | 32     | 244   | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>    | CCA73309.1      |
| C65             | 1      | 276   | $\alpha$ -metilacil-CoA racemasa          | <i>Piriformospora indica</i>    | CCA73309.1      |
| C67             | 1      | 381   | Glicosiltransferasa                       | <i>Thiocapsa marina</i>         | WP_007191375.1  |
| C71             | 209    | 3     | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>      | YP_003798590.1  |
| C73             | 5      | 379   | Proteína hipotética ACD_39C00256G0001     | Bacteria no cultivable          | EKD84083.1      |
| C75             | 379    | 2     | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>      | YP_003798590.1  |
| C79             | 310    | 2     | Glicosiltransferasa                       | Bacteria no cultivable          | EKD34483.1      |
| C81             | 382    | 5     | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.     | CBX29882.1      |
| C83             | 2      | 382   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>      | YP_003798590.1  |
| C85             | 261    | 7     | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>      | YP_003798588.1  |
| C87             | 382    | 2     | Celobiosa fosforilasa                     | <i>Thioflaviccoccus mobilis</i> | YP_007244847.1  |
| C89             | 7      | 384   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>      | YP_003798590.1  |
| C91             | 2      | 382   | Glicosiltransferasa                       | <i>Thiocapsa marina</i>         | WP_007191375.1  |
| C93             | 384    | 73    | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>      | YP_003798590.1  |
| C95             | 1      | 384   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>      | YP_003798590.1  |



|      |     |     |   |                                   |                |
|------|-----|-----|---|-----------------------------------|----------------|
| C97  | 383 | 3   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C99  | 67  | 267 | Permeasa/transportador ABC                | <i>Mycobacterium tusciae</i>      | WP_006243305.1 |
| C103 | 381 | 181 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C105 | 329 | 6   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>        | YP_003798588.1 |
| C109 | 122 | 385 | Proteína hipotética                       | <i>Escherichia coli</i>           | WP_000027047.1 |
| C113 | 162 | 383 | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>      | YP_006414616.1 |
| C115 | 384 | 1   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.       | CBX29882.1     |
| C117 | 386 | 27  | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>        | YP_003798588.1 |
| C119 | 4   | 384 | Glicosiltransferasa                       | <i>Azoarcus</i> sp.               | YP_007551292.1 |
| C121 | 300 | 97  | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>        | YP_003798589.1 |
| C125 | 386 | 105 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C127 | 270 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C131 | 51  | 380 | Glicosiltransferasa                       | Bacteria no cultivable            | EKD34483.1 ]   |
| C133 | 3   | 386 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i>  | WP_020878070.1 |
| C135 | 2   | 154 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C137 | 1   | 384 | Glicosiltransferasa                       | <i>Azoarcus</i> sp.               | YP_007551292.1 |
| C139 | 387 | 259 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C143 | 185 | 385 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C145 | 72  | 275 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>        | YP_003798589.1 |
| C147 | 2   | 388 | Glicosiltransferasa                       | <i>Solibacter usitatus</i>        | YP_822169.1    |
| C149 | 127 | 2   | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>        | YP_003798588.1 |
| C151 | 387 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C153 | 389 | 3   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C155 | 2   | 388 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>        | YP_003798588.1 |
| C159 | 1   | 390 | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i>  | WP_020878070.1 |
| C161 | 389 | 6   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.       | CBX29882.1     |
| C165 | 112 | 396 | Glicosiltransferasa                       | Bacteria no cultivable            | EKD34483.1     |
| C169 | 340 | 2   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.       | CBX29882.1     |
| C173 | 402 | 1   | Proteína de función desconocida DUF2329   | <i>Desulfococcus multivorans</i>  | WP_020878070.1 |
| C181 | 404 | 6   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.       | CBX29882.1     |
| C183 | 288 | 1   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C189 | 2   | 412 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C191 | 2   | 259 | Permeasa/transportador ABC                | <i>Mycobacterium tusciae</i>      | WP_006243305.1 |
| C195 | 1   | 420 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C205 | 139 | 396 | Permeasa/transportador ABC                | <i>Mycobacterium tusciae</i>      | WP_006243305.1 |
| C209 | 25  | 432 | Glicosiltransferasa                       | <i>Desulfomicrobium baculatum</i> | YP_003158688.1 |
| C211 | 3   | 134 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>        | YP_003798589.1 |
| C219 | 448 | 2   | Glicosiltransferasa                       | <i>Thiocapsa marina</i>           | WP_007191375.1 |
| C225 | 455 | 3   | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C231 | 458 | 201 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>        | YP_003798588.1 |
| C235 | 469 | 161 | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>        | YP_003798588.1 |
| C237 | 209 | 3   | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>      | YP_006414616.1 |
| C241 | 478 | 2   | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp.       | CBX29882.1     |
| C247 | 5   | 160 | Proteína hipotética                       | <i>Shigella boydii</i>            | WP_000415560.1 |
| C249 | 187 | 2   | Celobiosa fosforilasa                     | <i>Thiocystis violascens</i>      | YP_006414616.1 |
| C251 | 1   | 378 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |
| C253 | 494 | 369 | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>        | YP_003798589.1 |
| C255 | 1   | 501 | Proteína hipotética                       | <i>Leeia oryzae</i>               | WP_018149968.1 |
| C265 | 3   | 482 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>        | YP_003798590.1 |

|            |      |      |   |                             |                |
|------------|------|------|---|-----------------------------|----------------|
| C267       | 511  | 320  | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>  | YP_003798590.1 |
| C271       | 421  | 2    | Receptor TonB-dependiente                 | <i>Nitrospira defluvii</i>  | YP_003798588.1 |
| C273       | 468  | 91   | Proteína hipotética pEA28_01              | <i>Erwinia amylovora</i>    | NP_758764.1    |
| C275       | 203  | 3    | Proteína hipotética NIDE2968              | <i>Nitrospira defluvii</i>  | YP_003798589.1 |
| C281       | 1    | 201  | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>  | YP_003798590.1 |
| C283       | 512  | 114  | GTPasa CgtA                               | <i>Prevotella</i> sp.       | WP_007412191.1 |
| C291       | 8    | 601  | Proteína YfiH                             | <i>Prevotella</i> sp.       | WP_021948498.1 |
| C297       | 724  | 2    | Proteína hipotética N47_F15770            | <i>Desulfobacterium</i> sp. | CBX29882.1     |
| Scaffold 1 | 2087 | 1707 | Glicosiltransferasa                       | <i>Azoarcus</i> sp.         | YP_007551292.1 |
| Scaffold 3 | 1026 | 1460 | Fosfo-2-dehidro-3-deoxiheptonato aldolasa | <i>Nitrospira defluvii</i>  | YP_003798590.1 |

De acuerdo a los resultados de BLASTX, los insertos tienen secuencias de genes que pertenecen a microorganismos como *Nitrospira defluvii*, *Thiocystis violascens*, *Desulfococcus multivorans*, *Azoarcus* sp., *Desulfobacterium* sp., *Mycobacterium tusciae*, *Desulfomicrobium baculatum*, *Solibacter usitatus*, *Thiocapsa marina* y *Piriformospora indica*; sin embargo, en ambos insertos se observan *contigs* cuyas secuencias poseen identidad con la enzima fosfo-2-dehidro-3-deoxiheptonato aldolasa (nomenclatura alterna de DAHP sintasa) y con el receptor TonB-dependiente, ambos de la bacteria *Nitrospira defluvii*. Para el caso de la DAHP sintasa en las clonas UGGL96 y UGGS2, los *contigs* C75, C95, C135 y C189 fueron seleccionados para realizar un alineamiento y determinar los parámetros de similitud con la secuencia depositada en el GenBank para la DAHP sintasa (gen *aroF* en *N. defluvii*) de este organismo (Tabla 13).

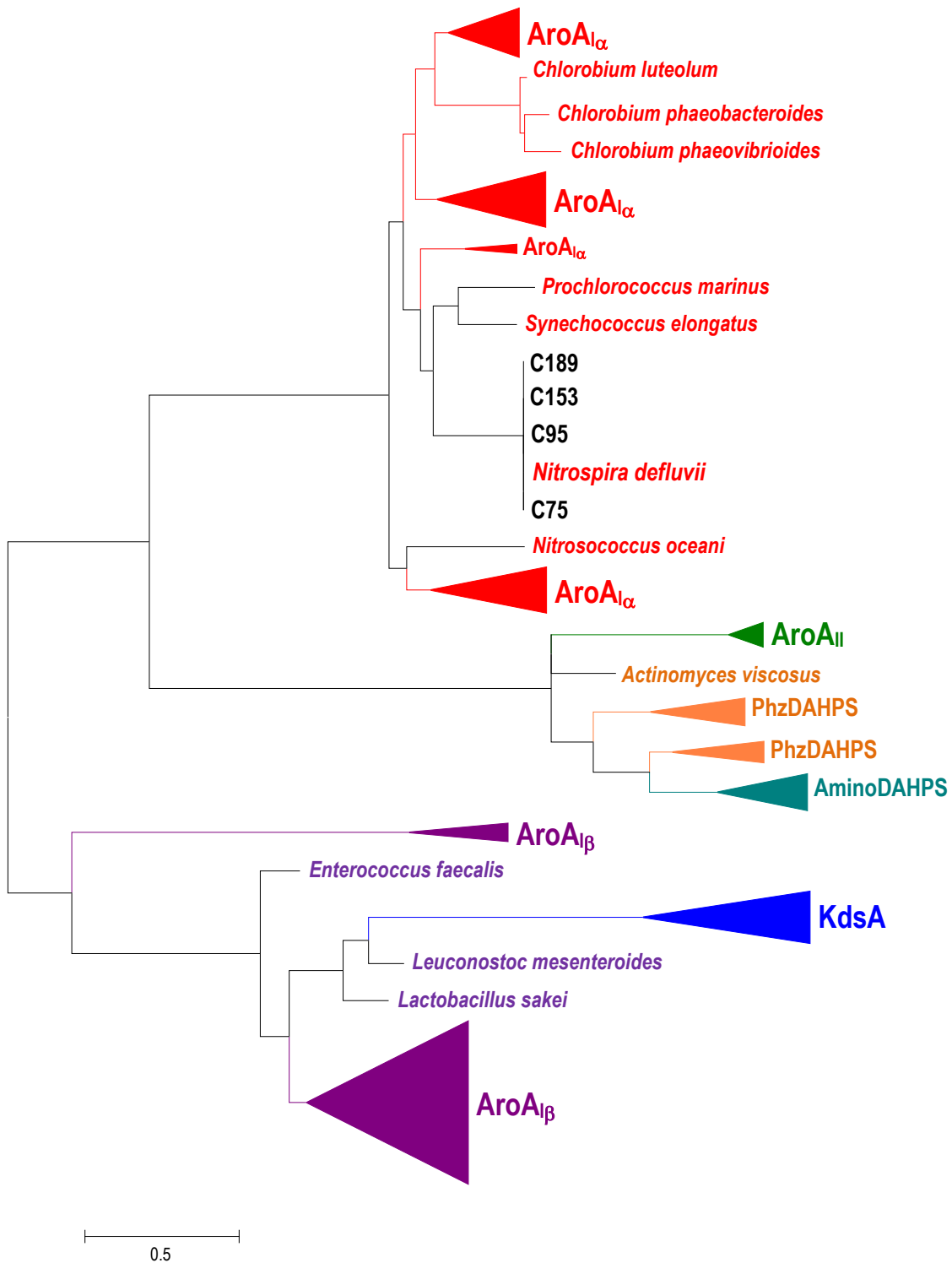
**Tabla 13.** Parámetros de BLASTX obtenidos para la enzima fosfo-2-dehidro-3-deoxiheptonato aldolasa de *Nitrospira defluvii* de los *contigs* C75, C95, C135 y C189 presentes en los insertos de las clonas UGGL96 y UGGS2.

| <b>Contig</b> | <b>Tamaño (pb)</b> | <b>Descripción</b>   | <b>Cobertura (%)</b> | <b>Valor E</b> | <b>Identidad (%)</b> |
|---------------|--------------------|--|----------------------|----------------|----------------------|
| C75           | 381                |  | 99                   | 2e-79          | 96                   |
| C95           | 384                | Fosfo-2-dehidro-3-deoxiheptonato aldolasa ( <i>Nitrospira defluvii</i> ) | 100                  | 4e-82          | 97                   |
| C153          | 389                |  | 99                   | 9e-81          | 95                   |
| C189          | 414                |  | 99                   | 2e-87          | 94                   |

De acuerdo al análisis BLASTX, los *contigs* C75, C95, C153 y C189, cuyo tamaño promedio es de 400 pb, tienen porcentajes de cobertura e identidad muy altos con respecto a la secuencia de DAHP sintasa (gen *aroF*) de *N. defluvii*. Aunque en la secuenciación se identificaron otras enzimas que corresponden a otros microorganismos (i.e. *T. violascens*, *D. multivorans*, *Azoarcus* sp., *Desulfobacterium* sp., *M. tusciae*, *D. baculatum*, *S. usitatus*, *T. marina* y *P. indica*), los valores de cobertura, identidad y valor E indican que el inserto de las clonas UGGL96 y UGGS2 proviene de *N. defluvii*.

Relativamente pocos estudios han sido referidos para las bacterias del género *Nitrospira*; estas bacterias son quimiolitotótrofas que han sido aisladas de suelo, ambientes marinos, cuevas, y recientemente de fuentes termales. Se han encontrado seis especies para este género bacteriano: *N. moscoviensis*, *N. marina*, *N. calida*, *N. bockiana*, *N. japonica* y *N. defluvii*, siendo identificado el primero de éstos en 1986 [Watson *et al.*, 1986; Ehrich *et al.*, 1995; Ushiki *et al.*, 2013]; por su dificultad para ser aislados y enriquecidos en medios de cultivo se consideran como microorganismos fastidiosos. En el caso particular de *N. defluvii*, esta bacteria se encuentra filogenéticamente más relacionada con aquellas bacterias similares al género *Nitrospira* no cultivables [Spieck *et al.*, 2006].

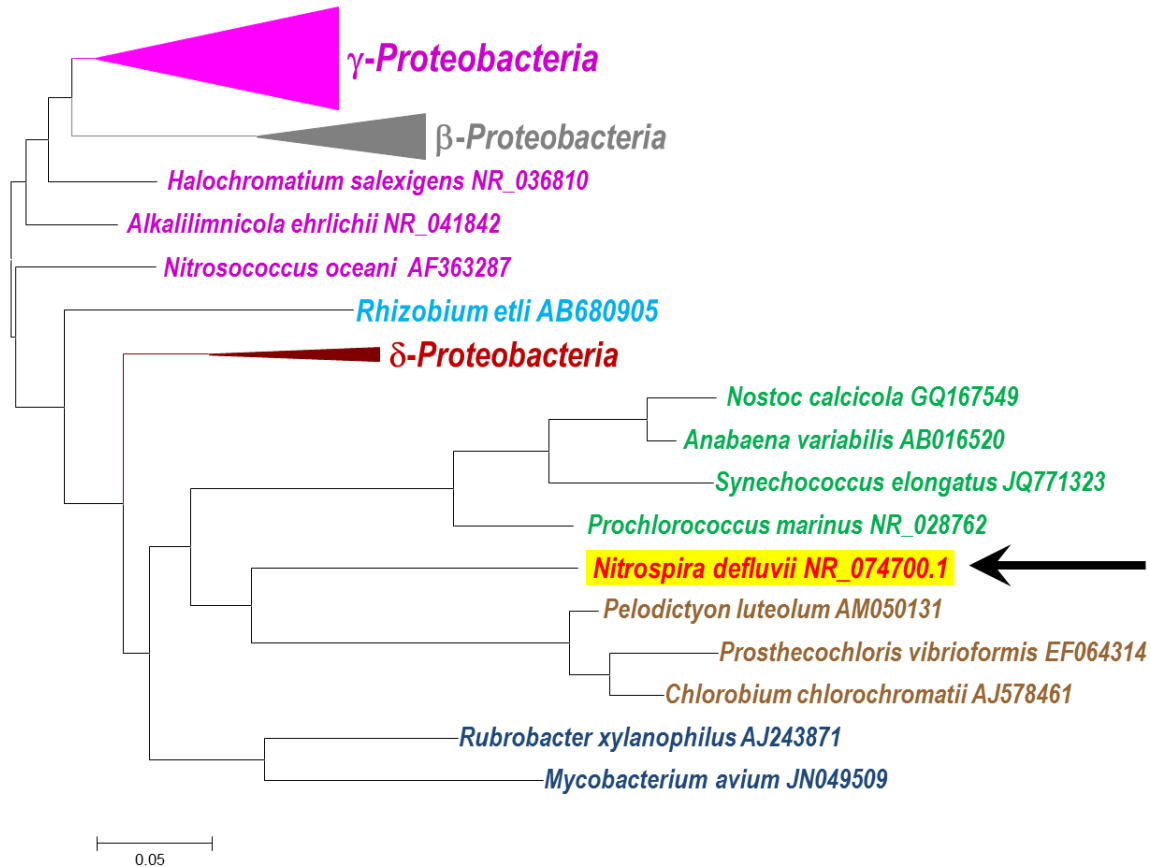
Hasta este momento, el nombre científico de *N. defluvii* está registrado como CANDIDATO en las bases de datos [Spieck *et al.*, 2006], por lo que en un futuro existe la posibilidad de que conserve la misma nomenclatura o que ésta pueda ser modificada; esta bacteria fue caracterizada por primera vez en el año 2006 por el grupo de Spieck *et al.* Respecto a ecología microbiana, las bacterias del género *Nitrospira* participan de manera importante en el metabolismo de nitrógeno, ya que son bacterias nitrato-oxidantes (NOB) las cuales emplean nitratos como donadores de electrones para producir nitritos; esta cualidad metabólica ha sido recientemente aprovechada en plantas de tratamiento de agua, en donde han sido inoculados consorcios de microorganismos NOB, tales como bacterias de los géneros *Nitrobacter*, *Nitrococcus* y *Nitrospira*. Las secuencias de los contigs C75, C95, C153 y C189, así como la secuencia de la DAHP sintasa *N. defluvii* obtenida del GenBank fueron incluidas en un análisis filogenético con las secuencias representativas de familias de DAHP sintasas con la finalidad de clasificarlas al ubicarlas en algunos de los grupos existentes para estas enzimas (Figura 38).



**Figura 38.** Árbol filogenético de secuencias de DAHP sintasas y los *contigs* obtenidos para los insertos de las clonas UGGL96 y UGGS2 correspondientes a *Nitrospira defluvii*. La historia evolutiva fue inferida utilizando el método *Maximum Likelihood* basado en el modelo de matriz Jones-Taylor-Thornton (JTT). El árbol mostrado es el de mayor logaritmo de probabilidad (-31682.7251). Los árboles iniciales para la búsqueda heurística fueron obtenidos automáticamente aplicando los algoritmos *Neighbor-Join* y *BioNJ* a una matriz estimada de distancia por parejas, y luego fue seleccionada la topología con mayor valor de logaritmo de probabilidad. El árbol está dibujado a escala, con la longitud de ramas

medida considerando el número de sustituciones por sitio. El análisis involucró 175 secuencias de aminoácidos. Todas las posiciones que contenían espacios (*gaps*) y datos perdidos fueron eliminadas. Se encontraron un total de 173 posiciones en el grupo de datos finales. El análisis filogenético fue realizado con el programa MEGA5.

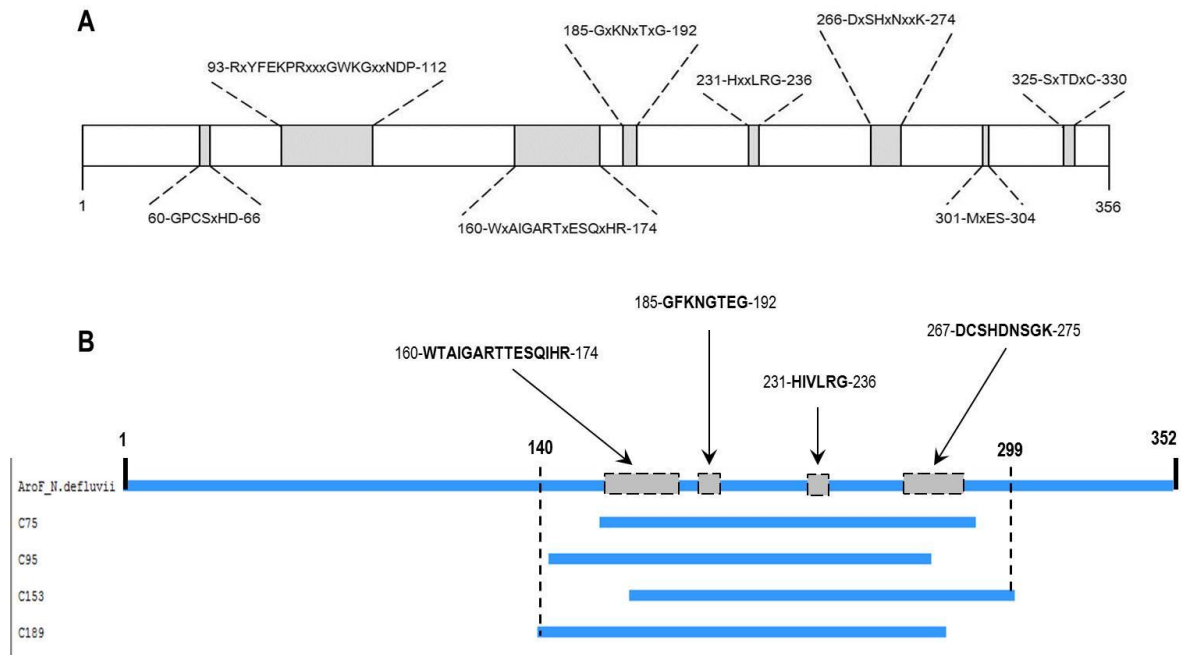
De acuerdo al marco de referencia de DAHP sintasas generado previamente, la secuencia de DAHP sintasa de *N. defluvii* presenta una mayor homología con la subfamilia AroA<sub>1α</sub>. En esta subfamilia principalmente se encuentran las DAHP sintasas de organismo del phylum *Proteobacteria*, sin embargo de acuerdo a la topología del árbol se observa que la secuencia de DAHPS de *N. defluvii* tiene mayor relación filogenética con las DAHPS de cianobacterias como *Synechococcus elongatus* y *Prochlorococcus marinus*. El análisis de contexto genómico realizado con el programa GeConT I para la subfamilia AroA<sub>1α</sub> (Figura 19) no incluyó a la DAHPS de *N. defluvii* dentro de los ortólogos a la DAHPS de *E. coli*; sin embargo, una vez establecido que la DAHPS de *N. defluvii* corresponde a la subfamilia AroA<sub>1α</sub> (Figura 36) y con los resultados del BLASTX obtenidos para los insertos de las clonas UGGL96 y UGGS2, utilizamos la secuencia completa de DAHPS de *N. defluvii* obtenida de GenBank para realizar un nuevo análisis y proponer una relación de contexto genómico entre la DAHP sintasa y las enzimas obtenidas en el análisis BLASTX. Sin embargo, el programa GeConT I no mostró resultados para el contexto genómico de la DAHP sintasa de *N. defluvii*; la búsqueda se repitió en el programa GeConT II y la interfase String (<http://string-db.org/>), obteniendo los mismos resultados. De tal manera que, hasta el momento, no existe información disponible en las bases de datos internacionales sobre el contexto genómico de la DAHP sintasa en *N. defluvii*. Hasta este momento no ha habido ningún estudio o reporte publicado sobre el estudio de la DAHPS de *N. defluvii*, debido quizás a que las bacterias del género *Nitrospira* son relativamente de reciente caracterización y que el principal enfoque de estudio de estos microorganismos ha sido la oxidación de nitritos y su participación en el metabolismo de nitrógeno. Debido a que en el árbol filogenético de los 16S rRNA de los miembros microbianos de la subfamilia AroA<sub>1α</sub> de DAHP sintasas no fue incluido *N. defluvii*, realizamos un nuevo árbol (Figura 39) incluyendo la secuencia de 16S rRNA de *N. defluvii* para determinar la relación evolutiva de esta bacteria con los miembros de la subfamilia AroA<sub>1α</sub>.



**Figura 39.** Árbol filogenético de secuencias de 16S rRNA de los miembros bacterianos pertenecientes a la subfamilia de DAHP sintasas AroA<sub>α</sub>, señalando la ubicación del organismo *Nitrospira defluvii*. La historia evolutiva fue inferida utilizando el método *Maximum Likelihood* basado en el modelo de Tamura-Nei. El árbol filogenético mostrado es el de mayor logaritmo de probabilidad (-21403.1702). Los árboles iniciales para la búsqueda heurística fueron obtenidos aplicando el método *Neighbor-Joining* a una matriz de distancias estimada utilizando la aproximación *Maximum Composite Likelihood* (MCL). El árbol está dibujado a escala, con la longitud de ramas medida considerando el número de sustituciones por sitio. Todas las posiciones que contenían espacios (*gaps*) y datos perdidos fueron eliminadas. Se encontraron un total de 1114 posiciones en el grupo de datos finales.

Estudios previos han demostrado que la bacteria *N. defluvii*, y en general el género *Nitrospira*, pertenece a un linaje filogenético distinto que deriva directamente del reino *Bacteria* al phylum *Nitrospirae*. La topografía del árbol filogenético realizado muestra que el género *Nitrospira* constituye un clado independiente que diverge de un ancestro común a los miembros bacterianos de los phyla *Cyanobacteria* y *Bacteroidetes*. De igual forma que en el árbol filogenético de familias de DAHP sintasas, se observa una relación en el 16S rRNA de *N. defluvii* con las cianobacterias *S. elongatus* y *P. marinus*.

Se realizó un alineamiento de estas secuencias para conocer las regiones de homología de los *contigs* C75, C95, C153 y C189 con respecto a la secuencia de aminoácidos de AroF en *N. defluvii*, realizando también una comparación con los motivos conservados presentes en la subfamilia AroA<sub>1α</sub> de DAHP sintasas a la cual pertenece la DAHP sintasa AroF de *N. defluvii* (Figura 40).

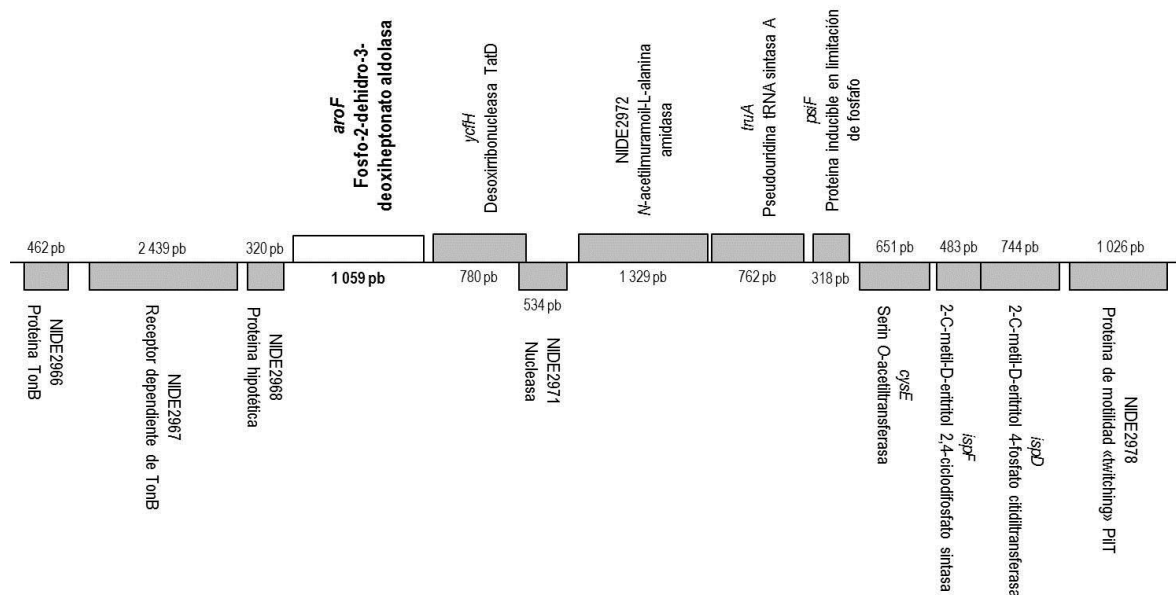


**Figura 40. A.** Esquema de motivos conservados en la subfamilia AroA<sub>1α</sub> de DAHP sintasas, tomando como referencia la DAHP sintasa (AroF) de *Escherichia coli* (356 aa). **B.** Alineamiento de secuencias de *contigs* C75, C95, C153 y C189 con la secuencia de aminoácidos de la DAHP sintasa (AroF) de *Nitrospira defluvii* (352 aa). En la región correspondiente entre los aminoácidos 140 y 299 se indican los motivos que corresponden a los característicos de la subfamilia AroA<sub>1α</sub>.

Los *contigs* C75, C95, C153 y C189 representan cada uno, aproximadamente, una tercera parte de la DAHP sintasa AroF de *N. defluvii*, constituida por 352 aa. La región de homología entre los *contigs* y AroF comprende entre los aminoácidos 140 y 299; dentro de esta región se pueden distinguir cuatro sitios con motivos que corresponden a motivos conservados en toda la subfamilia AroA<sub>1α</sub> de DAHP sintasas. Estos motivos son 160-WTAIGARTTESQIHR-174, 185-GFKNGTEG-192, 231-HIVLRG-236 y 267-DCSHDNSGK-275; la homología de los *contigs* con estos motivos permite afirmar que, considerando la longitud del inserto en las clonas UGGL96 y UGGS2, es muy probable que en el inserto se encuentre el gen *aroF* de *N. defluvii*, así como el contexto genómico asociado a este gen en una extensión máxima de 12 kb. Asimismo, los *contigs* corresponden con motivos que

contienen aminoácidos asociados tanto a interacción con PEP (*i.e.* Arg166, Lys187, Arg235 e His270) y cationes divalentes (*i.e.* His270) característicos de la subfamilia AroA<sub>1α</sub>.

Utilizando los datos obtenidos de la secuenciación de los insertos de las clonas UGGL96 y UGGS2, y complementando con información obtenida de la secuencia del genoma de *N. defluvii*, se realizó una predicción del contexto genómico del gen *aroF* que se encuentra en el inserto de estas dos clonas (Figura 41).



**Figura 41.** Esquema de contexto genómico de gen *aroF* en el cromosoma de *Nitrospira defluvii*. El esquema ilustra una región cromosomal de 12 058 pb, señalando el tamaño de los genes así como la función descrita para cada producto de gen.

De acuerdo a la Figura, el contexto genómico del gen *aroF* se encuentra asociado a distintas funciones y vías metabólicas de *N. defluvii*. Por ejemplo, las proteínas TonB constituyen una clase de transportadores encargados del paso de hierro (Fe) y vitamina B12 a través de la membrana externa de bacterias Gram-negativas por medio de transporte activo; asimismo se ha demostrado que las mutaciones en los transportadores TonB confieren resistencia a infección por bacteriófagos (*e.g.* bacteriófago T1 en *E. coli*) [Postle, 1993]. Asimismo, asociados a *aroF* se encuentran genes que codifican para proteínas citoplásmicas con actividades de DNasa (*e.g.* desoxirribonucleasa TatD) [Wexler *et al.*, 2000], endopeptidasa (*e.g.* N-acetilmuramoi-L-alanin amidasa). También se encuentran genes como *cysE*, el cual codifica para una serin O-acetiltransferasa que participa en la



biosíntesis de L-cisteína a partir de L-serina [Denk & Böck, 1987]; de igual manera se encuentran los genes *ispD* e *ispF*, los cuales codifican para enzimas que participan en la biosíntesis de isoprenoides o terpenos por la vía independiente a mevalonato [Steinbacher *et al.*, 2002].

## VIII. CONCLUSIONES

- Los ortólogos de la secuencia de aminoácidos de la DAHP sintasa de *B. subtilis*, representativa de la subfamilia AroA<sub>β</sub>, se presentan desde grupos microbianos evolutivamente ancestrales como los del reino Archaea hasta algunos miembros de la clase γ-Proteobacteria, implicando una mayor diversidad en grupos microbianos evolutivamente distantes. En cambio, los ortólogos de *E. coli* se agrupan principalmente dentro del phylum *Proteobacteria*; además de que los contextos genómicos de algunas DAHP sintasas de esta subfamilia AroA<sub>α</sub> tienen una gran similitud entre éstos (*e.g.* clase γ-Proteobacteria).
- Particularmente en algunos grupos microbianos, el contexto genómico de las DAHP sintasas se encuentra asociado a genes de vías metabólicas de biosíntesis de aminoácidos aromáticos (*e.g.* reino Archaea; phyla Bacteroidetes, Firmicutes; clase γ-Proteobacteria; géneros *Dehalococcoides*, *Thermotoga*). No obstante que algunos otros ortólogos no se asocian a vías de biosíntesis de aminoácidos aromáticos, su contexto genómico está relativamente conservado entre ellos.
- Los aminoácidos Lys, Asp e His son los únicos conservados en un alineamiento de secuencias representativas de todas las familias de DAHP sintasas. En la subfamilia AroA<sub>α</sub> estos aminoácidos se encuentran en las regiones Lys187, Asp266 e His269; en la subfamilia AroA<sub>β</sub> en las regiones Lys231, Asp293 e His296; en la familia KdsA en las regiones Lys133, Asp192 e His195; en las PhzDAHPS en Lys240, Asp300 e His303; y en las aminoDAHPS, en Lys286, Asp346 e His349. Asimismo, el motivo DxxH, que incluye tanto al aspartato como histidina, se encuentra conservado en todas las familias de DAHP sintasas.

- Producto de los experimentos de transformación se obtuvieron un total de 138 clonas denominadas que recuperaron la actividad de DAHP sintasa; al analizarlas mediante perfil de restricción, se obtuvieron 19 patrones únicos indicativos de insertos únicos.
- Bajo las condiciones de cultivo empleadas, de las 19 clonas únicas obtenidas, las clonas UGGL96 y UGGS2 presentaron el mejor crecimiento en medio mínimo con glucosa, con una fase lag relativamente corta (3 y 5 horas, respectivamente) y obteniendo una biomasa máxima de 0.95 y 0.75 g<sub>CDW</sub>/L en 12 y 20 horas, respectivamente.
- La actividad enzimática específica de las clonas que presentan una menor fase lag en su curva de crecimiento, son 53 U · mg<sup>-1</sup> (UGGL96) y 30 U · mg<sup>-1</sup> (UGGS2); mientras que el valor de actividad enzimática específica obtenido para la cepa control T2.1 pJLBaroG<sup>brtktA</sup> es de 69 U · mg<sup>-1</sup>.
- En las clonas UGGL96 y UGGS2 la complementación funcional fue realizada muy probablemente por la enzima fosfo-2-dehidro-3-deoxiheptonato aldolasa (nomenclatura alternativa a DAHP sintasa), de la bacteria *Nitrospira defluvii*; la secuencia de aminoácidos de esta enzima mostró homología con la subfamilia AroA<sub>α</sub> de DAHP sintasas.

## IX. PERSPECTIVAS

- Establecer la secuencia completa y orden de los genes asociados a la DAHP sintasa de *Nitrospira defluvii* en los insertos de clonas UGGL96 y UGGS2, y compararla con el contexto genómico reportado para el gen *aroF* de esta bacteria.
- Realizar la caracterización enzimática de las DAHP sintasas de la bacteria *N. defluvii* obtenidas de las clonas UGGL96 y UGGS2.
- Subclonar los insertos de DNA encontrados en las clonas UGGL96 y UGGS2 en vectores para una expresión óptima en procesos de fermentación.

- Transferir y probar estos insertos de DNA en cepas productoras de compuestos aromáticos (metabolismo primario o secundario) para experimentos de producción.
- Realizar experimentos de complementación funcional con DAHP sintasas de metabolismo microbiano secundario (aminoDAHP sintasas y de producción de fenazinas) y evaluar su posible función dentro de la vía de metabolismo microbiano primario en *E. coli*.

## X. REFERENCIAS

1. Allison TM, Cochrane F, Jameson GB, Parker EJ (2013) Examining the role of intersubunit contacts in the catalysis by 3-deoxy-D-manno-octulosonate 8-phosphate synthase *Biochemistry* 52:4676-4686.
2. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2:2006-008.
3. Báez-Viveros JL, Osuna J, Hernández-Chávez G, Soberón X, Bolívar F, Gosset G (2004) Metabolic engineering and protein directed evolution increase the yield of L-phenylalanine synthesized from glucose in *Escherichia coli*. *Biotechnol Bioeng* 87:516-524.
4. Balderas-Hernández VE, Sabido-Ramos A, Silva P, Cabrera-Valladares N, Hernández-Chávez G, Báez-Viveros JL, Martínez A, Bolívar F, Gosset G (2009) Metabolic engineering for improving anthranilate synthesis from glucose in *Escherichia coli*. *Microb Cell Fact* 8:19-30.
5. Birck MR, Woodard RW (2001) *Aquifex aeolicus* 3-deoxy-D-manno-2-octulosonic acid 8-phosphate synthase: a new class of KDO 8-P synthase?. *J Mol Evol* 52:205-214.
6. Birnboim HC (1983) A rapid alkaline extraction method for the isolation of plasmid DNA. *Method Enzymol* 100:243-255.
7. Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7:1513-1523.
8. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248-254.

9. Chen CC, Liao CC, Hsu WH (1993) The cloning and nucleotide sequence of a *Corynebacterium glutamicum* 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase gene. *Microbiology Letters* 107:223-230.
10. Cherepanov PP, Wackernagel W (1995) Gene disruption in *Escherichia coli*: Tc<sup>R</sup> and Km<sup>R</sup> cassettes with the option of Flp-catalyzed excision of the antibiotic-resistant determinant. *Gene* 158:9-14.
11. Ciria R, Abreu-Goodger C, Morett E, Merino E (2004) GeConT: gene context analysis. *Bioinformatics* 20: 2307-2308.
12. Daniel R (2005) The metagenomics of soil. *Nat Rev Microbiol* 3:470-478.
13. Denk D, Böck A (1987) L-cysteine biosynthesis in *Escherichia coli*: nucleotide sequence and expression of the serine acetyltransferase (*cysE*) gene from the wild-type and a cysteine-excreting mutant. *J Gen Microbiol* 133:515-525.
14. EcoCyc: Encyclopedia of *Escherichia coli* K-12 Genes and Metabolism (<http://ecocyc.org/>)
15. Ehrich S, Behrens D, Lebedeva E, Ludwig W, Bock E (1995) A new obligately chemolithoautotrophic, nitrite-oxidizing bacterium, *Nitrospira moscoviensis* sp. nov. and its phylogenetic relationship. *Arch Microbiol* 164: 16-23.
16. Ekkers DM, Cretoiu MS, Kielak AM, Van Elsas JD (2012) The great screen anomaly-a new frontier in product discovery through functional metagenomics. *Appl Microbiol Biotechnol* 93:1005-1020.
17. Floss HG, Yu TW, Arakawa K (2011) The biosynthesis of 3-amino-5-hydroxybenzoic acid (AHBA), the precursor of mC<sub>7</sub>N units in ansamycin and mitomycin antibiotics: a review. *J Antibiot* 64:35-44.
18. Gosset G, Bonner CA, Jensen RA (2001) Microbial origin of plant-type 2-keto-3-deoxy-D-arabino-heptulosonate 7-phosphate synthases, exemplified by the chorismate- and tryptophan-regulated enzyme from *Xanthomonas campestris*. *J Bacteriol* 183:4061-4070.
19. Guo J, Frost JW (2002) Biosynthesis of 1-deoxy-1-imino-D-erythrose 4-phosphate: a defining metabolite in the aminoshikimate pathway. *J Am Chem Soc* 124:528-529.
20. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5:245-249.

21. Jensen RA, Xie G, Calhoun DH, Bonner CA (2002) The correct phylogenetic relationship of KdsA (3-deoxy-D-manno-octulosonate 8-phosphate synthase) with one of the two independently evolved classes of AroA (3-deoxy-D-arabino-heptulosonate 7-phosphate synthase). *J Mol Evol* 54:416-423.
22. Jiang C, Yin B, Tang M, Zhao G, He J, Shen P, Wu B (2013) Identification of a metagenome-derived prephenate dehydrogenase gene from an alkaline-polluted soil microorganism. *Anton Leeuw Int J G* 103(6):1209-1219.
23. Jin D, Lu W, Ping S, Zhang W, Chen J, Dun B, Ma R, Zhao Z, Sha J, Li L, Yang Z, Chen M, Lin M (2007) Identification of a new gene encoding EPSPS with high glyphosate resistance from the metagenomic library. *Curr Microbiol* 55:350-355.
24. Jones DT, Taylor WR, Thornton JM (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275-282.
25. Li Y, Wexler M, Richardson DJ, Bond PL, Johnston AWB (2005) Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned *trp* genes. *Environ Microbiol* 7:1927-1936.
26. Light SH, Halavaty AS, Minasov G, Shuvalova L, Anderson WF (2012) Structural analysis of a 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase with an N-terminal chorismate mutase-like regulatory domain. *Protein Sci* 21:887-895.
27. Maixner F, Wagner M, Lückner S, Pelletier E, Schmitz-Esser S, Hace K, Spieck E, Konrat R, Le Paslier D, Daims H (2008) Environmental genomics reveals a functional chlorite dismutase in the nitrite-oxidizing bacterium '*Candidatus Nitrospira defluvii*'. *Environ Microbiol* 10:3043-3056.
28. Mavrodi DV, Peever TL, Mavrodi OV, Parejko JA, Raaijmakers JM, Lemanceau P, Mazurier S, Heide L, Blankenfeldt W, Weller DM, Thomashow LS (2010) Diversity and evolution of phenazine biosynthesis pathway. *Appl Environ Microbiol* 76:866-879.
29. Mentel M, Ahuja EG, Mavrodi DV, Breinbauer R, Thomashow LS, Blankenfeldt W (2009) Of two make one: The biosynthesis of phenazines. *ChemBioChem* 10:2295-2304.
30. Miller JH (1992) A short course in bacterial genetics. A laboratory manual and handbook for *Escherichia coli* and related bacteria. *Cold Spring Harbor Laboratory Press* pp. 268-274.
31. National Center for Biotechnology Information (NCBI) <http://www.ncbi.nlm.nih.gov/>

32. Postle K (1993) TonB protein and energy transduction between membranes. *J Bioenerg Biomembr* 25:591-601.
33. Schloss PD, Handelsman J (2003) Biotechnological prospects from metagenomics. *Curr Opin Biotech* 14:303-310.
34. Shumilin IA, Kretsinger RH, Bauerle RH (1999) Crystal structure of phenylalanine-regulated 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Escherichia coli*. *Structure* 7:865-875.
35. Shumilin IA, Bauerle R, Wu J, Woodard RW, Kretsinger RH (2004) Crystal structure of the reaction complex of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Thermotoga maritima* refines the catalytic mechanism and indicates a new mechanism of allosteric regulation. *J Mol Biol* 341:455-466.
36. Spieck E, Hartwig C, McCormack I, Maixner F, Wagner, Lipski A, Daims H (2006) Selective enrichment and molecular characterization of a previously uncultured *Nitrospira*-like bacterium from activated sludge. *Environ Microbiol* 8: 405-415.
37. Sirena-Sánchez N (2009) Análisis de la diversidad bacteriana no cultivable y de la diversidad metabólica presente en un banco metagenómico de una muestra de suelo. Tesis de Maestría, UNAM. México.
38. Steinbacher S, Kaiser J, Wungsintaweekul J, Hecht S, Eisenreich W, Gerhardt S, Bacher A, Rohdich F (2002) Structure of 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase involved in mevalonate-independent biosynthesis of isoprenoids. *J Mol Biol* 316:79-88.
39. Subramaniam PS, Xie G, Xia T, Jensen RA (1998) Substrate ambiguity of 3-deoxy-D-manno-octulosonate 8-phosphate synthase from *Neisseria gonorrhoeae* in the context of its membership in a protein family containing a subset of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthases. *J Bacteriol* 180:119-127.
40. Tamura K, Nei M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512-526.
41. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28:2731-2739.
42. Ushiki N, Fujitani H, Aoi Y, Tsuneda S (2013) Isolation of *Nitrospira* belonging to sublineage II from a wastewater treatment plant. *Microbes Environ* 28:346-353.

43. Walker GE, Dunbar B, Hunter IS, Nimmo HG, Coggins JR (1996) Evidence for a novel class of microbial 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase in *Streptomyces coelicolor* A3(2), *Streptomyces rimosus* and *Neurospora crassa*. *Microbiology* 142:1973-1982.
44. Walker SR, Jiao W, Parker EJ (2011) Synthesis and evaluation of dual site inhibitors of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase. *Bioorg Med Chem Lett* 21:5092-5097.
45. Waravdekar VS, Saslaw LD (1959) A sensitive colorimetric method for the estimation of 2-deoxy sugars with the use of the malonaldehyde-thiobarbituric acid reaction. *J Biol Chem* 234:1945-1950.
46. Watson SW, Bock E, Valois FW, Waterbury JB, Schlosser U (1986) *Nitrospira marina* gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. *Arch Microbiol* 144:1-7.
47. Wexler M, Sargent K, Jack RL, Stanley NR, Bogsch EG, Robinson C, Berks BC, Palmer T (2000) TatD is a cytoplasmic protein with DNase activity. *J Biol Chem* 275:16717-16722.

## **XI. APÉNDICE**

Este trabajo fue presentado en el XV Congreso de la Sociedad Mexicana de Biotecnología y Bioingeniería y en el 12th International Symposium on Genetics of Industrial Microorganisms, eventos realizados del 23 al 28 de junio de 2013 en Cancún, Quintana Roo, México. El trabajo fue presentado en la modalidad de exposición oral, el día 24 de junio de 2013 a las 16:00 h dentro del área de Biología de Sistemas y Ciencias Ómicas (se anexa la respectiva constancia).



**National Congress of  
Biotechnology and  
Bioengineering**



**GIM2013**  
12th International Symposium on the  
**GENETICS OF INDUSTRIAL  
MICROORGANISMS**

# CERTIFICATE OF PRESENTATION

We hereby certify that

**Uriel Gutiérrez-Gómez, Francisco Bolivar, Adelfo Escalante**

**Delivered the Oral presentation**

Screening, characterization and classification of 3-deoxy-D-arabino-heptulosonate 7-phosphatase (DAMP) synthetase enzymes from a soil metagenomic library

XV National Congress of Biotechnology and Bioengineering and  
12th International Symposium on the Genetics of Industrial Microorganisms,  
June 23rd - 28th, 2013 in Cancun, Quintana Roo, México.

Dr. Gerardo Fuentes  
President of IMBB

Dr. César del Valle Aguilar  
President of the Organizing Committee  
IMBB

Dr. Romina Rodríguez  
President of the Scientific  
Committee IMBB/GIM

Dr. Sergio Sánchez  
President of the Organizing  
Committee GIM

Dr. Stefano Donadio  
President of GIM



## MEDIOS DE CULTIVO

- **Medio LB (1 L)**

| Componente           | Cantidad |
|----------------------|----------|
| Triptona             | 10 g     |
| Extracto de levadura | 5 g      |
| NaCl                 | 5 g      |
| NaOH 1 N             | 1 mL     |

- **Medio M9 (1 L)**

| Componente                       | Cantidad |
|----------------------------------|----------|
| Na <sub>2</sub> HPO <sub>4</sub> | 6 g      |
| KH <sub>2</sub> PO <sub>4</sub>  | 3 g      |
| NH <sub>4</sub> Cl               | 1 g      |
| NaCl                             | 0.5 g    |

ESTERILIZAR, ENFRIAR Y AGREGAR

|                       |        |
|-----------------------|--------|
| CaCl <sub>2</sub> 1 M | 100 µL |
| MgSO <sub>4</sub> 1 M | 1 mL   |
| Glucosa al 20%        | 10 mL  |
| Tiamina (10 mg/mL)    | 100 µL |

- **Medio SOC**

| Componente           | Concentración |
|----------------------|---------------|
| Triptona             | 2 %           |
| Extracto de levadura | 0.5 %         |
| NaCl                 | 10 mM         |
| KCl                  | 2.5 mM        |
| MgCl <sub>2</sub>    | 10 mM         |
| MgSO <sub>4</sub>    | 10 mM         |
| Glucosa              | 20 mM         |

- Medio ARO (1 L)

| Componente                        | Cantidad |
|-----------------------------------|----------|
| Glucosa                           | 1 g      |
| K <sub>2</sub> HPO <sub>4</sub>   | 7 g      |
| KH <sub>2</sub> PO <sub>4</sub>   | 2 g      |
| (NH <sub>4</sub> )SO <sub>4</sub> | 0.5 g    |
| Citrato de amonio férrico         | 0.32 g   |
| NaCl                              | 0.5 g    |
| ESTERILIZAR, ENFRIAR Y AGREGAR    |          |
| Ácido <i>p</i> -aminobenzoico     | 10 µg    |
| Ácido <i>p</i> -hidroxibenzoico   | 10 µg    |
| Ácido 2,3-dihidroxibenzoico       | 10 µg    |
| Tiamina (10 mg/mL)                | 0.5 µg   |
| MgSO <sub>4</sub> 1 M             | 1 mL     |

- Medio M9S (1 L)

| Componente                       | Cantidad |
|----------------------------------|----------|
| Na <sub>2</sub> HPO <sub>4</sub> | 6 g      |
| KH <sub>2</sub> PO <sub>4</sub>  | 3 g      |
| NH <sub>4</sub> Cl               | 1 g      |
| NaCl                             | 0.5 g    |
| ESTERILIZAR, ENFRIAR Y AGREGAR   |          |
| Ácido <i>p</i> -aminobenzoico    | 10 µg    |
| Ácido <i>p</i> -hidroxibenzoico  | 10 µg    |
| Ácido 2,3-dihidroxibenzoico      | 10 µg    |
| Tiamina (10 mg/mL)               | 1 mL     |
| MgSO <sub>4</sub> 1 M            | 1 mL     |
| CaCl <sub>2</sub> 1 M            | 100 µL   |
| Glucosa al 20%                   | 10 mL    |

## SOLUCIONES DEL PROTOCOLO DE EXTRACCIÓN DE DNA PLASMÍDICO POR LISIS ALCALINA (MINIPREP)

### Solución I

8 % Sacarosa  
0.05 M Tris – HCl  
0.5 M EDTA pH 8.0

### Solución II

1 % SDS  
0.2 N NaOH

### Solución III

3 M Acetato de sodio pH 4.8

### Solución TE

10 mM Tris – HCl  
1 mM EDTA pH 8.0

## PROCEDIMIENTO PARA DETERMINACIÓN DE ACTIVIDAD ENZIMÁTICA DE DAHP SINTASA

### A. Obtención de intervalo de linealidad de actividad enzimática

- Preparar una serie de tubos Eppendorf (en hielo) con los siguientes volúmenes de extracto celular, soluciones de EPPS, D-eritrosa 4-fosfato, fosfoenolpiruvato, y cloruro de cobalto. El tubo control para determinar la  $Abs_{549}$  inicial de la muestra, tendrá solamente 200  $\mu$ L del extracto celular.

| EXTRACTO CELULAR<br>( $\mu$ L) | 50 mM<br>EPPS<br>( $\mu$ L) |   |            |   |            |                        |
|--------------------------------|-----------------------------|---|------------|---|------------|------------------------|
| 10                             | 115                         |   | 25 $\mu$ L |   | 25 $\mu$ L | 25 $\mu$ L             |
| 20                             | 105                         | + | E4P 8 mM   | + | PEP 8 mM   | +                      |
| 40                             | 85                          |   |            |   |            | CoCl <sub>2</sub> 4 mM |
| 80                             | 45                          |   |            |   |            |                        |
| 125                            | 0                           |   |            |   |            |                        |

- Incubar esta serie de tubos y el control a 37 °C durante 20 minutos.
- Detener la reacción enzimática adicionando 50  $\mu$ L de solución de ácido tricloroacético al 20 % y comenzar la cuantificación de DAHP (ver más adelante).

## B. Cuantificación de DAHP (ensayo de ácido tiobarbitúrico) en extractos celulares

1. Transferir 200  $\mu\text{L}$  de la muestra a un tubo Eppendorf.
2. Adicionar 50  $\mu\text{L}$  de solución de ácido tricloroacético al 20 %. Mezclar usando vórtex.
3. Centrifugar por 5 minutos para separar el material precipitado.
4. Transferir el sobrenadante (aprox. 200  $\mu\text{L}$ ) a un tubo de ensayo con taparrosca y adicionar 250  $\mu\text{L}$  de solución de peryodato de sodio 0.025 M. Mezclar usando vórtex.
5. Incubar a 42 °C durante 30 minutos.
6. Adicionar 500  $\mu\text{L}$  de solución de arsenito de sodio al 0.2 %. Mezclar usando vórtex y dejar reposar por 2 minutos a temperatura ambiente.
7. Adicionar 2 mL de solución de ácido tiobarbitúrico al 0.3 %. Mezclar usando vórtex.
8. Incubar en un baño de agua hirviendo durante 15 minutos.
9. Leer el sobrenadante rosa, libre de proteína, a una absorbancia de 549 nm.

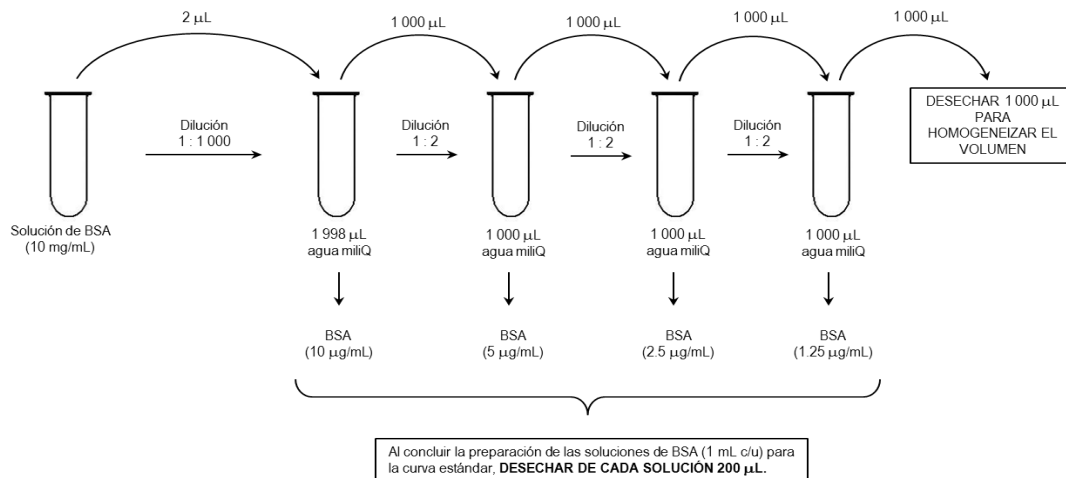
## REACTIVOS PARA CUANTIFICACIÓN DE DAHP

|                              |   |
|------------------------------|---|
| Ácido tricloroacético al 20% | Pesar 2 g y disolver en 10 mL de agua miliQ                                     |
| Peryodato de sodio 0.025 M   | Pesar 0.26 g y aforar a 50 mL con $\text{H}_2\text{SO}_4$ 0.125 N               |
| Arsenito de sodio al 2%      | Pesar 2 g y disolver en 100 mL de HCl 0.5 N                                     |
| Ácido tiobarbitúrico al 0.3% | Pesar 0.75 g y disolver en 250 mL de agua miliQ                                 |
| Ácido sulfúrico 0.125 N      | Medir 0.17 mL de $\text{H}_2\text{SO}_4$ 36.7 N y aforar a 50 mL con agua miliQ |
| Ácido clorhídrico 0.5 N      | Medir 4.1 mL de HCl 12.2 N y aforar a 100 mL con agua miliQ                     |

## C. Cuantificación de proteína total

### • Obtención de curva estándar de albúmina sérica bovina (BSA)

1. Preparar una solución estándar de 10 mg/mL de BSA.
2. Realizar una serie de diluciones para realizar la curva estándar de BSA. Las concentraciones a obtener son 10, 5, 2.5, y 1.25  $\mu\text{g}/\text{mL}$  en un volumen de 1 mL.



3. A cada solución de BSA de la curva estándar, adicionar 200 µL del reactivo de Bradford, agitar e incubar 10 minutos a temperatura ambiente.
4. Leer la absorbancia de cada tubo a 595 nm. Registrar las lecturas y obtener la ecuación de la curva de tendencia lineal para calcular la cantidad de proteína.

- **Determinación de proteína total en extractos celulares**

1. Realizar una dilución de cada extracto a un volumen final de 1 mL.
2. Desechar 200 µL de cada dilución realizada y adicionar un volumen igual (200 µL) de reactivo de Bradford. Agitar e incubar 10 minutos a temperatura ambiente.
3. Leer la absorbancia de las muestras a 595 nm.
4. Para el cálculo de proteína total (mg), tomar en cuenta las siguientes consideraciones:
  - Interpolación en curva estándar, o sustitución en curva de regresión lineal, para obtener la concentración de proteína (µg/mL)
  - Factor de dilución empleado, si es el caso
  - Convertir la concentración de proteína (µg/mL) en cantidad de proteína total (µg):
 
$$(1.25 - 10 \mu\text{g/mL}) \times (0.8 \text{ mL}) = 1 - 8 \mu\text{g}$$
  - Convertir la cantidad de proteína, de microgramos (µg) a miligramos (mg)

### CÁLCULO DE ACTIVIDAD ENZIMÁTICA ESPECÍFICA

$$\text{Actividad específica}_{\text{DAHP sintasa}} = \frac{(\Delta\text{Abs}_{549})(66.8)}{20 \text{ min}} \cdot \frac{1}{\text{mg de proteína}}$$

$$\text{Actividad específica}_{\text{DAHP sintasa}} [=] \frac{\text{nmol}_{\text{DAHP}}}{\text{min} \cdot \text{mg}_{\text{proteína}}}$$