Universitat
de les Illes Balears

TESIS DOCTORAL
2018

CATABOLISMO Y TOLERANCIA A HIDROCARBUROS EN EL LINAJE ROSEOBACTER. APROXIMACIÓN MULTIÓMICA

Catalina María Alejandro Marín

Universitat de les Illes Balears

TESIS DOCTORAL
2018

Programa de Doctorado en Microbiologia Ambiental y Biomédica

CATABOLISMO Y TOLERANCIA A HIDROCARBUROS EN EL LINAJE ROSEOBACTER. APROXIMACIÓN MULTIÓMICA

Catalina María Alejandro Marín

Directora: Balbina Nogales Fernández
Director: Rafael Bosch Zaragoza
Tutora: Balbina Nogales Fernández

Doctora por la Universitat de les Illes Balears

LISTADO DE PUBLICACIONES DERIVADAS DE LA TESIS

Alejandro-Marín CM, Bosch R, Nogales B. 2014. Comparative genomics of the protocatechuate branch of the β-ketoadipate pathway in the Roseobacter lineage. Marine Genomics; 17: 25–33.

Bosch R, Brunet-Galmés I, Mas-Lladó M, Alejandro-Marín CM, Jaen-Luchoro D, Prince C, Busquets A, Gomila M, Mulet M, Peña A, Bennasar A, García-Valdés E, Nogales B, Lalucat J. 2014. Degradació aeròbica d'hidrocarburs aromàtics per *Pseudomonas* i Roseobacters. Bolletí Societat Història Natural Balears; 55: 9–23.

“Sens dubte és bo que el món només conegui l’obra bella i no els seus orígens ni les condicions en què ha aparegut, perquè el coneixement de les fonts d’on ha brollat la inspiració potser determinaria el desconcert i l’astorament del lector i l’exquisitat de l’obra restaria sense efecte”

Thomas Mann

AGRADECIMIENTOS

Esta tesis representa la culminación de un proyecto que empezó hace ocho años con mi llegada al laboratorio como alumna colaboradora. A lo largo de estos años son muchas las personas e instituciones que han participado en este trabajo y a las que debo expresar mi gratitud.

En primer lugar, quisiera agradecer esta tesis a todas aquellas personas que han sido importantes en mi vida. A lo largo de la vida conocemos a miles de personas, pero solo unas cuantas nos marcan, nos cambian para bien o para mal. Por ello quisiera agradecer su aportación a toda aquella gente cuya presencia en mi vida fue determinante, sigan o no sigan presentes en ella, puesto que todos ellos han contribuido a convertirme en lo que soy a día de hoy y a mis logros.

Entre las personas que más han aportado al desarrollo y finalización de esta tesis me gustaría destacar a mis jefes, sin los cuales esta tesis no habría sido posible. A Balbina por darme la oportunidad de ser su alumna colaboradora sin conocerme, por confiar en mí para realizar esta tesis doctoral, por los conocimientos que me ha enseñado y por todo el tiempo que ha invertido en ser mi mentora durante estos años. Gracias. También me gustaría agradecer a Rafa sus ideas y su espíritu positivo, que siempre ayuda a ver los problemas que se nos plantean de una forma más sencilla.

También me gustaría destacar el apoyo que he recibido de mis compañeros de laboratorio puesto que mi día a día no habría sido lo mismo sin ellos. A Bel, Maria y Dani, mis compañeros de generación de becarios, puesto que nadie entiende mejor el día a día de un tesista que alguien que está pasando por lo mismo. Por vuestra ayuda y por todas las experiencias vividas, desde contarnos las penas y frustraciones cuando las cosas no salían bien, hasta la euforia cuando por fin salían. A Magda y a Cris, por vuestra predisposición a ayudarnos siempre, ya fuera ayudándonos con nuestros problemas logísticos, escuchando nuestras crisis o dándonos vuestros buenos consejos. A David y a Claudia Prince, por ser la alegría del laboratorio y convertir las meriendas en grandes eventos. A Arantxa y Marga, por lo aprendido con vuestras charlas y consejos. Además también me gustaría agradecer la ayuda que me prestaron otros compañeros del departamento con los que coincidí durante estos años como Sebas, Joana, Lady, Claudia S, Sofía, Inma, Marga M, Toni B y Mohammad y al resto de profesores del departamento.

Debo agradecer a Joseph la oportunidad que me dio para realizar una estancia en su laboratorio de la universidad de Warwick, por el tiempo que me dedicó y por permitirme aprender nuevas técnicas que han sido de gran utilidad para la realización de esta tesis. A Mar, por su inestimable ayuda para adaptarme a la universidad de Warwick y por todo el tiempo que invirtió enseñándome. Y a mis compañeros de la universidad de Warwick Despoina, Gabriel, Miriam, Fabrizio, Alevtina, Vinko, Amandeep y Robyn por integrarme en su grupo desde el primer día y por su ayuda, mi estancia en tierras inglesas no habría sido tan acogedora sin ellos.

También me gustaría agradecer la ayuda que me han prestado nuestros tres oficiales de laboratorio Ángel, Xisco y Guillem y nuestro conserge Macià, puesto que con vuestro trabajo, buena disponibilidad y charlas habéis facilitado y animado mi trabajo.

No puedo concluir mis agradecimientos sin incluir a mis padres y a mi hermana, por estar allí, por confiar en mí y escucharme en mis crisis, a pesar de la sobredimensión de estas. Gracias.

La investigación realizada en esta tesis se ha financiado con los proyectos CTM2011-24886 y CTM2015-70180-R del Ministerio de Economia, Industria y Competitividad (cofinanciados con Fondos Feder) y una ayuda a grupos competitivos de la Conselleria d'Educació, Cultura i Universitats (también con cofinanciación FEDER).

También me gustaría agradecer la beca FPI (BES-2012-054941) y la beca para estancias breves FPI (EEBB-I-16-10940) al Ministerio de Economia, Industria y Competitividad que ha aportado los fondos necesarios para que pudiera dedicarme a la realización de esta tesis.

ÍNDICE

LISTADO DE FIGURAS I
LISTADO DE TABLAS V
LISTADO DE ABREVIATURAS VII
RESUMEN XI
INTRODUCCIÓN 1
1. El linaje Roseobacter 3
2. Degradación aeróbica de compuestos aromáticos 7
3. Evidencias de la degradación de compuestos aromáticos por bacterias del linaje Roseobacter 11
4. Potencial para la degradación de hidrocarburos en el ambiente 14
5. Mecanismos de tolerancia a compuestos aromáticos 16
HIPÓTESIS Y OBJETIVOS 21
MATERIAL Y MÉTODOS 25
1. Predicción de genes para la degradación y tolerancia de compuestos aromáticos en genomas del linaje Roseobacter 27
2. Cultivos de aislados del linaje Roseobacter 30
2.1. Determinación de la capacidad de crecimiento con compuestos aromáticos como fuente de carbono y energía 30
2.2. Cultivos para el análisis proteómico preliminar de tolerancia a naftaleno y tolueno 31
2.3. Cultivos para el análisis de tolerancia a tolueno y naftaleno por proteómica masiva (*shot-gun*) 31
2.4. Cultivos del mutante de *R. pomeroyi* DSS-3^T defectivo para el gen codificante de la proteína SPO2567 32
3. Procesamiento de muestras para análisis proteómico 32
3.1. Fraccionamiento y electroforesis SDS-PAGE en los experimentos preliminares 32
3.2. Procesamiento de muestras para proteómica masiva 33
4. Análisis de los datos de proteómica masiva 35
5. Generación de un mutante de *R. pomeroyi* DSS-3^T defectivo para el gen codificante de la proteína SPO2567 36
6. Análisis *in silico* de las características de la proteína SPO2567 39
7. Generación de metagenomas de muestras marinas 40
7.1. Toma de muestras 40
7.2. Experimentos de microcosmos 40
7.2.1. Preparación de los microcosmos 40
7.2.2. Análisis de polimorfismos de longitud de fragmentos de restricción terminales (T-RFLP) 41

7.2.3. Recuento de procariotas totales e hibridación *in situ* con sondas fluorescentes (FISH) 41
7.2.4. Análisis de limitación de nitrógeno en las aguas de los microcosmos 41
7.3. Extracción de ácidos nucleicos totales para generar metagenomas 42
8. Análisis de los metagenomas 43
CAPÍTULO 1: POTENCIAL PARA EL CATABOLISMO DE COMPUESTOS AROMÁTICOS EN MIEMBROS DEL LINAJE ROSEOBACTER 45
1. Genes para la degradación de compuestos aromáticos en genomas del linaje Roseobacter 47
2. Vías de degradación de protocatecuato y catecol 50
2.1. Rotura *orto* (Vía del beta-cetoadipato) 50
2.1.1. Rama del protocatecuato de la vía del beta-cetoadipato 50
2.1.2. Rama del catecol de la vía del beta-cetoadipato 60
2.2. Rotura *meta* 63
2.2.1. Rotura *meta* de protocatecuato 63
2.2.2. Rotura *meta* de catecol 64
2.3. Rutas periféricas de degradación hacia protocatecuato 65
2.3.1. Degradación de ferulato y vanillato 65
2.3.2. Degradación de ftalato, isoftalato y tereftalato 68
3. Vía de degradación de homogentisato 70
4. Vía de degradación de fenilacetil-CoA 75
5. Vía de degradación de homoprotocatecuato 81
6. Vía de degradación del gentisato 85
7. Vía de degradación de benzoil-CoA 90
8. Resumen del potencial degradador de compuestos aromáticos presente en genomas del linaje Roseobacter 92
9. Crecimiento de miembros del linaje Roseobacter con hidrocarburos como fuente de carbono y energía 96
CAPÍTULO 2: TOLERANCIA A HIDROCARBUROS EN MIEMBROS DEL LINAJE ROSEOBACTER 99
1. Aproximación proteómica a la respuesta a naftaleno y tolueno en tres miembros del linaje Roseobacter 101
1.1. Cambios proteicos globales asociados a la respuesta a naftaleno y tolueno 104
2. Análisis detallado de las proteínas observadas en la respuesta a naftaleno 110
2.1. Proteínas de membrana 110
2.2. Otras proteínas 116
3. Aproximación al estudio de la función de la proteína SPO2567 de *Ruegeria pomeroyi* DSS-3^T 121
3.1. Análisis bioinformático 121
3.2. Generación de un mutante de *R. pomeroyi* DSS-3^T defectivo en SPO2567 123
CAPÍTULO 3: METAGENÓMICA DE COMUNIDADES MICROBIANAS MARINAS 127

1. Experimentos de microcosmos de laboratorio 129
2. Análisis de los metagenomas 132
2.1. Características de los metagenomas 132
2.2. Caracterización taxonómica de los metagenomas 134
2.3. Caracterización funcional de los metagenomas 141
2.4. Caracterización del potencial catabólico de compuestos aromáticos en los metagenomas 147
CONCLUSIONES 163
BIBLIOGRAFÍA 169
ANEXOS 191
ANEXO 1: FIGURAS SUPLEMENTARIAS 193
ANEXO 2: TABLAS SUPLEMENTARIAS 197

LISTADO DE FIGURAS

Figura I1. Grupos filogenómicos descritos en el linaje Roseobacter.

Figura I2. Vías de degradación de compuestos monoaromáticos detectadas en genomas del linaje Roseobacter.

Figura I3. Mecanismos implicados en la tolerancia a tolueno en *Pseudomonas putida* DOT-T1E.

Figura M1. Esquema del fragmento de cromosoma de *R. pomeroyi* DSS-3^T que contiene el gen codificante para la proteína SPO2567.

Figura 1.1. Rutas centrales de degradación de protocatecuato y catecol: vía del beta-cetoadipato (rotura *orto*) y rutas de degradación por rotura *meta*.

Figura 1.2. Detección de genes codificantes para las proteínas involucradas en la rama del protocatecuato de la vía del beta-cetoadipato en genomas del linaje Roseobacter.

Figura 1.3. Organización de los genes de la ruta del protocatecuato de la vía del beta-cetoadipato en los genomas del linaje Roseobacter.

Figura 1.4. Organización de los genes *pca* en regiones separadas del *core pcaCHG* en los genomas del linaje Roseobacter.

Figura 1.5. Organización de los genes de la rotura *orto* del catecol en los genomas del linaje Roseobacter.

Figura 1.6. Organización de los genes de la rotura *meta* del protocatecuato en los genomas del linaje Roseobacter.

Figura 1.7. Rutas de degradación periféricas que convergen en protocatecuato.

Figura 1.8. Organización de los genes de la vía del ferulato en los genomas del linaje Roseobacter.

Figura 1.9. Organización de los genes de las vías del ftalato, isoftalato y tereftalato en los genomas del linaje Roseobacter.

Figura 1.10. Vía de degradación de homogentisato y de compuestos precursores.

Figura 1.11. Detección de genes codificantes para las proteínas involucradas en la vía del homogentisato en genomas del linaje Roseobacter.

Figura 1.12. Organización de los genes de la vía del homogentisato en genomas del linaje Roseobacter.

Figura 1.13. Vía de degradación de fenilacetil-CoA y de compuestos cuya degradación procede por esta vía.

Figura 1.14. Detección de genes codificantes para las proteínas involucradas en la vía del fenilacetil-CoA en genomas del linaje Roseobacter.

Figura 1.15. Organización de los genes de la vía del fenilacetil-CoA en los genomas del linaje Roseobacter.

Figura 1.16. Organización de los genes para la degradación de fenilacetil-CoA separadas del *core paaABCDE* en los genomas del linaje Roseobacter.

Figura 1.17. Vía de degradación de homoprotocatecuato.

Figura 1.18. Detección de genes codificantes para las proteínas involucradas en la vía del homoprotocatecuato en genomas del linaje Roseobacter.

Figura 1.19. Organización de los genes de la vía del homoprotecuato en los genomas del linaje Roseobacter.

Figura 1.20. Vías de degradación del gentisato, naftaleno, salicilato y 3-hidroxibenzoato.

Figura 1.21. Organización de los genes de la vía del gentisato en los genomas del linaje Roseobacter.

Figura 1.22. Vía de degradación de benzoil-CoA.

Figura 1.23. Organización de los genes de la vía del benzoil-CoA en los genomas del linaje Roseobacter.

Figura 1.24. Mapa catabólico global para la degradación de compuestos aromáticos en el linaje Roseobacter.

Figura 1.25. Compuestos que potencialmente pueden ser degradados por miembros del linaje Roseobacter en base al estudio de predicción genómica.

Figura 2.1. Densidad óptica de los cultivos en los tratamientos con hidrocarburos.

Figura 2.2. Comparación global de los proteomas analizados en este estudio.

Figura 2.3. Cambios producidos tras la exposición a naftaleno en diferentes categorías de proteínas en los microorganismos estudiados.

Figura 2.4. Estructura de la región que contiene los genes codificantes para las proteínas implicadas en la degradación de compuestos aromáticos en *C. aestuarii* 357.

Figura 2.5. Crecimiento en cultivo líquido de la cepa silvestre *R. pomeroyi* DSS-3^{T} y el mutante SPO2567^{-}.

Figura 2.6. Cambios proteicos asociados a la exposición a naftaleno en la cepa silvestre y el mutante SPO2567^{-} de *R. pomeroyi* DSS-3^{T}.

Figura 3.1. Seguimiento de las comunidades microbianas en los microcosmos de abril de 2013 y M1 de abril de 2014.

Figura 3.2. Comparación de la composición de las comunidades microbianas en los microcosmos mediante T-RFLP.

Figura 3.3. Abundancia relativa de las secuencias asignadas a familias del dominio *Bacteria* en los metagenomas analizados.

Figura 3.4. Asignación funcional de las lecturas de los metagenomas a las familias de enzimas para degradación de compuestos aromáticos de la base de datos AromaDeg.

Figura 3.5. Comparación global de los metagenomas en base a la asignación funcional con la base de datos AromaDeg.

Figura 3.6. Grupos de enzimas que discriminan los metagenomas ambientales.

Figura 3.7. Grupos de enzimas que discriminan los metagenomas de los microcosmos.

Figura 3.8. Tipo de compuestos sobre los que actuarían los enzimas correspondientes a las secuencias génicas que discriminan metagenomas control de muestras contaminadas con hidrocarburos.

LISTADO DE TABLAS

Tabla M1. Categorías funcionales de los genes involucrados en la degradación de los compuestos aromáticos estudiados.

Tabla M2. Organismos modelo para la detección en genomas del linaje Roseobacter de los genes implicados en las rutas centrales y periféricas del catabolismo de compuestos aromáticos.

Tabla M3. Organismos modelo para la detección en genomas de Roseobacter de genes para bombas de tipo RND implicadas en tolerancia a disolventes orgánicos.

Tabla M4. Cebadores utilizados para la construcción del mutante SPO2567^{-} de *R. pomeroyi* DSS-3^{T}.

Tabla 1.1. Detección de los genes codificantes para las enzimas principales de las vías centrales de degradación de compuestos aromáticos en genomas del linaje Roseobacter.

Tabla 1.2. Detección de vías centrales completas para la degradación de compuestos aromáticos en genomas del linaje Roseobacter.

Tabla 1.3. Crecimiento en medio mineral con diferentes compuestos aromáticos como fuente de carbono y energía.

Tabla 2.1. Proteínas detectadas en los experimentos con naftaleno y tolueno en los tres microorganismos estudiados.

Tabla 2.2. Clasificación en categorías funcionales de las proteínas significativas con función asignada (KEGG) del proteoma total tras exposición a naftaleno.

Tabla 2.3. Clasificación de las proteínas de membrana del proteoma total tras exposición a naftaleno.

Tabla 3.1. Datos de los metagenomas generados en este estudio.

Tabla 3.2. Detalle de las lecturas analizadas con MG-RAST de los metagenomas generados en este estudio.

Tabla 3.3. Asignación de las secuencias de los metagenomas a los diferentes dominios.

Tabla 3.4. Abundancia relativa de las lecturas asignadas a bacterias del linaje Roseobacter en los metagenomas analizados en este estudio.

Tabla 3.5. Abundancia relativa de las lecturas asignadas a diferentes categorías funcionales.

Tabla 3.6. Asignación de lecturas de los metagenomas según la base de datos AromaDeg.

Tabla 3.7. Abundancia relativa de lecturas asignadas a las superfamilias de enzimas para degradación de compuestos aromáticos de la base de datos AromaDeg en los metagenomas.

Tabla 3.8. Sustratos de los grupos de enzimas importantes para diferenciar los metagenomas de microcosmos.

Tabla 3.9. Asignación taxonómica de las lecturas asignadas a grupos de enzimas de AromaDeg que diferencian los metagenomas ambientales.

LISTADO DE ABREVIATURAS

4-ABS: 4-aminobencenosulfonato

3,4-DHSA: 3,4-dihidroxi-9,10-secoandrost-1,3,5(10)-trieno-9,17-diona

AAnP: Fototrofía anoxigénica aeróbica

ABC: Sistema de transporte de casete de unión a ATP (*ATP-binding cassette transporter system*)

ANOVA: Análisis de la varianza

ARNr: Ácido ribonucleico ribosómico

ATP: Adenosín trifosfato

BLAST: Basic Local Alignment Search Tool

CFAs: Ácidos grasos de ciclopropano

CoA: Coenzima A

COG: Agrupación de grupos ortólogos

CTI: Isomerasa *cis-trans*

CTP: Proteína de transporte de tricarboxilato (*tricarboxylate transport protein*)

DAPI: 4′,6-diamidino-2′-fenilindol dihidrocloruro

DMSP: Dimetil sulfoniopropionato

DMT: Sistema de transporte de la superfamila metabolitos de drogas (*drug–metabolite superfamily transporter system*)

DO_{600}: Densidad óptica a una longitud de onda de 600 nm

DTT: Ditiotreitol

DWH: Deepwater Horizon

EDTA: Ácido etilendiaminotetraacético

EXDO: Extradiol dioxigenasas de la superfamilia quelación de oxígeno adyacente (*extradiol dioxygenases of the vicinal oxygen chelate superfamily*)

FISH: Hibridación *in situ* con sondas fluorescentes

Gm: Gentamicina

GTA: Agentes de transferencia génica

H: Índice de Shannon

KEGG: Kyoto Encyclopedia of Genes and Genomes

Km: Kanamicina

KO: Ortología KEGG

LB: Medio Luria Bertani

LC-MS/MS: Cromatografía líquida acoplada a espectrometría de masas

MATE: Proteína de expulsion de múltiples antimicrobianos (*multi-antimicrobial extrusion protein*)

MB: Caldo marino

MeSH: Metanotiol

MFP: Proteína de fusión a la membrana

MFS: Sistema de transporte de la supermilia facilitador principal (*major facilitator superfamily transporter system*)

MG-RAST: Metagenomics RAST server

MMM: Medio mineral marino

MMPA: 3-metil-mercaptopropionato

MOD: Materia orgánica disuelta

NCBI: National Center for Biotechnology Information

NOG: Grupos de ortólogos no supervisado

nr: Base de dados de proteínas no redundantes del NCBI

OMG: Gammaproteobacterias marinas oligotróficas

OMP: Factor de la membrana externa

ORF: Marco abierto de lectura

PAST: Paleontological statistics

PBS: Tampón fosfato salino

PCA: Análisis de componentes principales

PCR: Reacción en cadena de la polimerasa

PHYLIP: Phylogeny inference package

Phyre2: Protein homology/analogy recognition engine v2.0

REFSeq: NCBI Reference Sequence Database

RND: Sistema de transporte resistencia-nodulación-división celular (*resistance-nodulation-cell division transporter system*)

ROS: Especies reactivas de oxígeno

SCG: Genómica de única célula (*single-cell genomics*)

SDS: Dodecilsulfato sódico

SDS-PAGE: Electroforesis en gel de poliacrilamida con dodecilsulfato sódico

SMART: Simple Modular Architecture Research Tool

SMR: Proteínas pequeñas de resistencia a múltiples drogas (*small multidrug resistance proteins*)

SOB: Medio super óptimo

SOD: Superóxido dismutasa

STRING: Search Tool for Recurring Instances of Neighbouring Genes

TAE: Tampón Tris-acetato-EDTA

TBE: Tampón Tris-Borato-EDTA

TDA: Ácido tropoditiético

TRAP: Sistema de transporte periplásmico tripartito independiente de ATP (*tripartite ATP-independent periplasmic transporter system*)

T-RF: Fragmento de restricción terminal

T-RFLP: Polimorfismo de longitud de fragmentos de restricción terminales

wt: Cepa silvestre

RESUMEN

Una de las características del linaje Roseobacter es su capacidad para la degradación de compuestos aromáticos. En la presente tesis se ha realizado una aproximación detallada de la capacidad para degradar compuestos aromáticos en el linaje Roseobacter determinando la presencia y organización de 166 genes involucrados en 6 vías de degradación de compuestos centrales anteriorment estudiadas en este grupo (rama del protocatecuato de la vía del beta-cetoadipato, benzoil-CoA, gentisato, fenilacetil-CoA, homoprotocatecuato y homogentisato) así como en las dos ramas de la vía central del catecol, en la rama de la rotura *meta* del anillo aromático del protocatecuato y en todas las vías periféricas que convergen en estas rutas centrales en 61 genomas de Roseobacter. Se determinó que las vías de degradación más detectadas eran la rama del protocatecuato de la vía del beta-cetoadipato y las vías de degradación del homogentisato y fenilacetil-CoA. Además, se observó una gran variabilidad en la organización de los genes para estas rutas catabólicas en los diferentes genomas. También se analizó la capacidad de varios aislados para crecer a expensas de los compuestos aromáticos estudiados como única fuente de carbono y energía y no siempre se cumplieron los resultados esperados en base a la predicción genómica. Por otra parte, no solo se ha estudiado el potencial para el catabolismo de compuestos aromáticos sino también los mecanismos de tolerancia a dos compuestos aromáticos (naftaleno y tolueno) que presentan miembros del linaje Roseobacter. En los cultivos expuestos a naftaleno se detectaron cambios a nivel de proteínas implicadas en el mantenimiento de la estabilidad de la membrana, la respuesta al estrés oxidativo y el transporte de membrana. Por lo tanto, los miembros del linaje Roseobacter presentaban algunos de los mecanismos implicados en la tolerancia a compuestos aromáticos descritos en la literatura en otros grupos bacterianos. Por último, se estudiaron las diferencias a nivel taxonómico y funcional entre las comunidades microbianas de aguas costeras no contaminadas (control) y aguas contaminadas con hidrocarburos en base a la generación de ocho metagenomas: dos de muestras ambientales y seis a partir de experimentos de microcosmos. A nivel de los metagenomas ambientales se pudo observar que en el metagenoma de puerto había una menor detección de secuencias asignadas a alfaproteobacterias y cianobacterias y una mayor detección de gammaproteobacterias y flavobacterias respecto al agua control, confirmando los resultados de anteriores estudios basados en ARNr. A nivel de metagenomas generados a partir de experimentos de microcosmos, también se observaron diferencias a nivel taxonómico, aunque estas fueron menos evidentes que las del metagenoma de puerto. Estos cambios a nivel taxonómico fueron acompañados de cambios a nivel funcional, por ejemplo, se observó que en los metagenomas de aguas contaminadas con hidrocarburos se solían encontrar genes codificantes para enzimas implicadas sobre todo en la degradación de compuestos derivados de combustibles fósiles y de la lignina mientras que en los metagenomas controles predominaban genes codificantes para enzimas de las que se desconoce el tipo de compuestos sobre los que actúan.

RESUM

Una de les característiques del llinatge Roseobacter és la seva capacitat per a la degradació de compostos aromàtics. En la present tesi s'ha realitzat una aproximació detallada de la capacitat per degradar compostos aromàtics en el llinatge Roseobacter determinant la presència i organització de 166 gens involucrats en 6 vies de degradació de compostos centrals que ja havien estat estudiats en aquest grup (branca del protocatecuat de la via del beta-cetoadipat, benzoil-CoA, gentisat, fenilacetil-CoA, homoprotocatecuat i homogentisat) així com en les dues branques de la via central del catecol, en la branca del trencament *meta* de l'anell aromàtic del protocatecuat i en totes les vies perifèriques que convergeixen en aquestes rutes centrals en 61 genomes de Roseobacter. Es va determinar que les vies de degradació més detectades eren la branca del protocatecuat de la via del beta-cetoadipat i les vies de degradació de l'homogentisat i el fenilacetil-CoA. A més, es va observar una gran variabilitat en l'organització dels gens per a aquestes rutes catabòliques en els diferents genomes. També es va analitzar la capacitat de diversos aïllats per créixer a costa dels composts aromàtics estudiats com a única font de carboni i energia i no sempre es van complir els resultats esperats a partir de la predicció genòmica. D'altra banda, no sols s'ha estudiat el potencial per al catabolisme de compostos aromàtics sinó també els mecanismes de tolerància a dos compostos aromàtics (naftalè i toluè) que presenten membres del llinatge Roseobacter. En els cultius exposats a naftalè es van detectar canvis a nivell de proteïnes implicades en el manteniment de l'estabilitat de la membrana, la resposta a l'estrès oxidatiu i el transport de membrana. Per tant, els membres del llinatge Roseobacter presentaven alguns dels mecanismes implicats en la tolerància a compostos aromàtics descrits en la literatura en altres grups bacterians. Finalment, es van estudiar les diferències a nivell taxonòmic i funcional entre les comunitats microbianes d'aigües costaneres no contaminades (control) i aigües contaminades amb hidrocarburs a partir de la generació de vuit metagenomes: dos de mostres ambientals i sis a partir d'experiments de microcosmos. A nivell dels metagenomes ambientals es va poder observar que en el metagenoma de port hi havia una menor detecció de seqüències assignades a alfaproteobacteris i cianobacteris i una major detecció de gammaproteobacteris i flavobacteris respecte a l'aigua control, confirmant els resultats d'estudis anteriors basats en ARNr. A nivell de metagenomes generats a partir d'experiments de microcosmos, també es van observar diferències a nivel taxonòmic, encara que aquestes van ser menys evidents que les del metagenoma de port. Aquests canvis a nivell taxonòmic van ser acompanyats de canvis a nivell funcional, per exemple, es va observar que en els metagenomes d'aigües contaminades amb hidrocarburs es solien trobar gens codificants per a enzims implicats sobretot en la degradació de compostos derivats de combustibles fòssils i de la lignina mentre que en els metagenomes control predominàven gens codificants per a enzims dels quals es desconeix el tipus de substrat sobre el qual actuen.

ABSTRACT

One characteristic of the Roseobacter lineage is its capacity for the degradation of aromatic compounds. In this thesis a detailed analysis of the capacity to degrade aromatic compounds in the Roseobacter lineage has been done by determining the presence and organization of 166 genes involved in 6 degradation pathways of central compounds previously studied in this group (protocatechuate branch of the beta-ketoadipate pathway, benzoyl-CoA, gentisate, phenylacetyl-CoA, homoprotocatechuate, and homogentisate) as well as in the two branches of the central catechol pathway, in the branch of *meta*-cleavage of the protocatechuate ring and in all peripheral pathways that converge in these central routes in 61 Roseobacter genomes. The most detected degradation pathways were the protocatechuate branch of the beta-ketoadipate pathway and the degradation pathways of homogentisate and phenylacetyl-CoA. In addition, a great variability in the organization of the genes for these catabolic routes was observed in the different genomes. The ability of several isolates to grow at the expense of the aromatic compounds studied as the sole source of carbon and energy was also analyzed and the expected results based on genomic prediction were not always fulfilled. On the other hand, we have also analysed the mechanisms of tolerance to two aromatic compounds (naphthalene and toluene) of members of the Roseobacter lineage. In cultures exposed to naphthalene, changes in proteins involved in the maintenance of membrane stability, response to oxidative stress and membrane transport were detected. Therefore, members of the Roseobacter lineage presented some of the mechanisms involved in tolerance to aromatic compounds described in the literature for other bacterial groups. Finally, differences at taxonomic and functional level between microbial communities of unpolluted coastal waters (control) and waters contaminated with hydrocarbons were studied based on the generation of eight metagenomes: two of environmental samples and six from microcosm experiments. In the harbor metagenome there was less detection of sequences assigned to alphaproteobacteria and cyanobacteria and a greater detection of gammaproteobacteria and flavobacteria with respect to the control water, confirming the results of previous studies based on rRNA. In the metagenomes generated from microcosm experiments, differences at taxonomic level were also observed, although these were less evident than those of the harbor metagenome. The changes at taxonomic level were accompanied by changes at functional level. In metagenomes from contaminated waters sequences for genes coding for enzymes involved mainly in degradation of compounds derived from fossil fuels and lignin were observed. In control metagenomes genes encoding enzymes whose substrates are unknown were more prevalent.

INTRODUCCIÓN

1. El linaje Roseobacter

El linaje Roseobacter pertenece a la familia *Rhodobacteraceae* en la clase *Alphaproteobacteria*. Es un grupo filogenéticamente coherente cuyos miembros comparten al menos un 89 % de identidad en secuencia de 16S ARNr (Buchan *et al*., 2005). A pesar de que el linaje Roseobacter es robusto, la filogenia interna del linaje basada en 16S ARNr muestra grupos polifiléticos (Buchan *et al*., 2005). En los últimos años el desarrollo de las técnicas de secuenciación de genomas y la disponibilidad de genomas del linaje Roseobacter en bases de datos públicas ha permitido analizar la filogenia del grupo en base a datos genómicos. El primer estudio filogenómico fue realizado por Newton y colaboradores (2010) utilizando la secuencia concatenada de 70 genes ortólogos que se encontraban en copia única en 32 genomas del linaje Roseobacter. Estos autores definieron la presencia de, al menos, 5 grupos filogenómicos en el linaje Roseobacter (figura I1). Estudios filogenómicos posteriores con un mayor número de genomas han confirmado esta clasificación (Luo y Moran, 2014; Mas-Lladó, 2017). A pesar de obtener filogenias más robustas con los estudios filogenómicos, en estas se puede observar como microorganismos de géneros como *Ruegeria*, *Roseobacter*, *Phaeobacter* y *Oceanicola* no siempre se agrupan con los miembros de su género. Lo más probables es que esto se deba a que algunas de las cepas incluidas en estos géneros no han sido bien caracterizadas taxonómicamente. Por ello la taxonomía actual del grupo está en desacuerdo con la filogenia (Newton *et al*., 2010; Luo *et al*., 2012; 2013). Esta asignación de categorías taxonómicas no coherente con la filogenia causa confusión al interpretar la ecología y la biología evolutiva del grupo (Luo y Moran, 2014).

En relación a su ecología, el linaje Roseobacter se encuentra predominantemente en ambientes marinos, llegando a representar más del 20 % de la comunidad en aguas costeras y del 3 al 5 % en aguas oceánicas superficiales (Buchan *et al*., 2005; Moran *et al*., 2007). Los miembros del linaje Roseobacter son ubicuos en ambientes acuáticos marinos y también se encuentran en sedimentos marinos, hielo polar, nieve marina, tapetes microbianos hipersalinos y asociados a microalgas, macroalgas y animales como esponjas, corales y cefalópodos (Buchan *et al*., 2005). Algunas poblaciones del linaje Roseobacter se ven favorecidas durante las eflorescencias de fitoplancton, con el que establecen relaciones de mutualismo que pueden derivar en patogénesis (Buchan *et al*., 2014). Las relaciones de mutualismo entre los miembros del linaje Roseobacter y el fitoplancton se ven favorecidas por varias características de los roseobacters, como la quimiotaxis hacia compuestos liberados por el fitoplancton (por ejemplo, el osmolito dimetilsulfonio propionato, DMSP), y la utilización de componentes de la materia orgánica disuelta (MOD) como fuente de carbono, nitrógeno, azufre y fósforo (Buchan *et al*., 2014). En este sentido, en los genomas del linaje Roseobacter son abundantes los genes para transportadores de tipo TRAP (del inglés *tripartite ATP-independent periplasmic*), MFS (del inglés *major facilitator superfamily*) y ABC (del inglés ATP *adenosine triphosphate-binding cassette*), que probablemente sean específicos para este tipo de sustratos (Buchan *et al*., 2014). Otra característica relevante de bacterias de este linaje es la producción de metabolitos secundarios, base de la señalización química para la defensa y las interacciones entre microorganismos. Así, muchos roseobacters producen una amplia gama de compuestos bioactivos, incluyendo moléculas de señalización y compuestos antimicrobianos (Newton *et al*., 2010; Buchan *et al*., 2014). Entre las moléculas de señalización producidas por

los miembros del linaje Roseobacter se encuentran las *N*-acil homoserina lactonas, moléculas que participan en *quorum sensing,* regulando por ejemplo la producción de compuestos antimicrobianos dependiendo de la densidad celular (Buchan *et al*., 2014). En cuanto a los compuestos antimicrobianos fabricados por miembros del linaje Roseobacter, se encuentran el ácido tropoditiético (TDA) y la indigoidina (Brinkhoff *et al*., 2004; Newton *et al*., 2010; Buchan *et al*., 2014). También se ha descrito que los miembros del linaje Roseobacter producen compuestos promotores del crecimiento como auxinas y vitaminas, que facilitarían las relaciones de mutualismo con el fitoplancton y que se exportarían a través de transportadores de tipo DMT (del inglés *drug–metabolite superfamily*) (Buchan *et al*., 2005; Wagner-Döbler y Biebl, 2006; Buchan *et al*., 2014). Los miembros del linaje Roseobacter también suelen presentar estructuras para facilitar la unión a superficies, ya que muchos de ellos viven adheridos a partículas (Buchan *et al*., 2005; Newton *et al*., 2010; Buchan *et al*., 2014; Luo y Moran, 2015) e incluso sobre otros organismos, como sucede con *Ruegeria* sp. TM1040 que forma biopelículas sobre dinoflagelados del género *Pfiesteria* (Buchan *et al*., 2005).

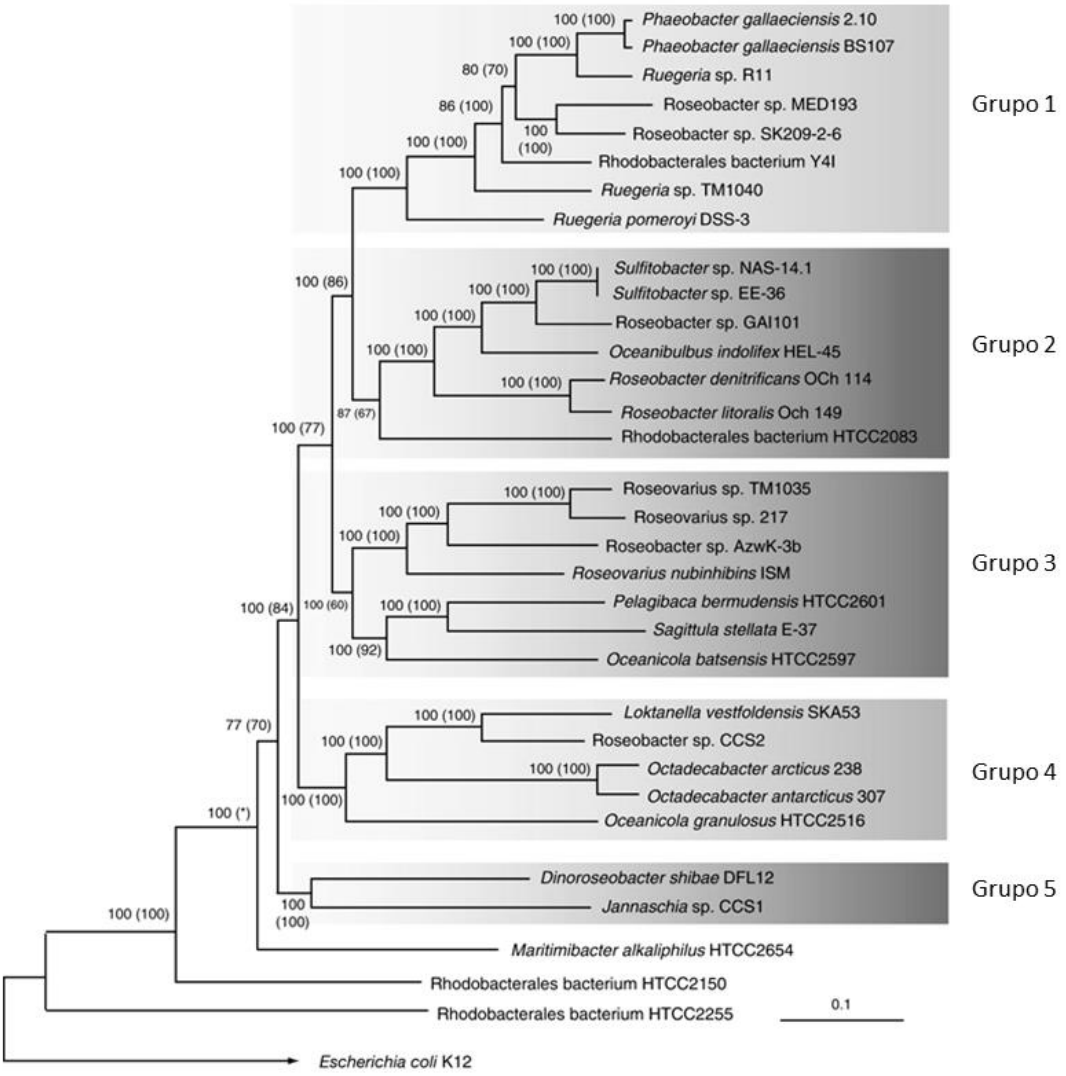

Figura I1. Grupos filogenómicos descritos en el linaje Roseobacter. Se muestra un árbol consenso elaborado por máxima verosimilitud con 32 genomas del linaje Roseobacter en base al concatenado de 70 genes ortólogos en copia única. Se muestran los valores de *bootstrap* superiores al 50 % calculados usando tanto el algoritmo de máxima verosimilitud (*maximum likelihood*) como el de *neighbor-joining*. La barra de escala representa el 10 % de divergencia de secuencia. Imagen modificada de Newton y colaboradores (2010).

Desde el punto de vista genómico, la mayoría de los miembros del linaje Roseobacter cultivados tienen genomas grandes (entre 3,1 y 4,7 Mb) y con un alto contenido mol % G+C (60 ± 4 %) en comparación con otros componentes del bacterioplancton marino (Luo y Moran, 2014). Hay algunas excepciones dentro del linaje Roseobacter con genomas más pequeños y con un menor contenido mol % G+C, como la cepa cultivada HTCC2255 (2,3 Mb, 37 % mol G+C) y algunos genomas obtenidos de organismos no cultivados mediante secuenciación de una única célula (entre 2,6 y 4,4 Mb, 39 ± 1 % mol G+C) (Luo *et al*., 2013; Zhang *et al*., 2016).

Otra característica genética importante en los miembros cultivados del linaje Roseobacter es la presencia de agentes de transferencia génica (GTA) (Luo y Moran, 2014). Los GTA fueron descubiertos en *Rhodobacter capsulatus* y son agrupaciones de genes que se cotranscriben, algunos de los cuales presentan homología con genes de fagos (Lang y Beatty, 2000). Las partículas GTA solo empaquetan ADN bacteriano y son más importantes para la transferencia génica entre roseobacters que la transformación y la transducción (Luo y Moran, 2014). Sin embargo, no está clara la abundancia y distribución de los GTA en los ecosistemas marinos, ni el papel que han tenido en la evolución del linaje Roseobacter (Luo y Moran, 2014).

Respecto al metabolismo, las bacterias cultivadas de este grupo son metabólicamente versátiles. La mayoría son organismos quimiorganoheterótrofos aunque también pueden obtener energía mediante fototrofía o litotrofía (Wagner-Döbler y Biebl, 2006; Moran y Miller, 2007). Se han descrito aislados de Roseobacter filogenéticamente diversos que presentan genes para realizar fototrofía anoxigénica aeróbica (AAnP) o, con menor frecuencia, para aprovechamiento de luz vía proteorodopsina (cepa HTCC2255) o xantorodopsinas (*Octadecabacter arcticus* 238, *Octadecabacter antarcticus* 307 y *Loktanella vestfoldensis* R-9477) (Newton *et al*., 2010; Luo y Moran*,* 2014). Estos mecanismos permiten obtener energía a partir de la luz solar, aunque desacoplados de la fijación de CO_2, puesto que los genes para la fijación del carbono del ciclo de Calvin no se encuentran en los genomas de Roseobacter estudiados (Luo y Moran, 2014). Sin embargo, se ha observado que *Roseobacter denitrificans* Och114 puede fijar CO_2 a niveles bajos a través de vías anapleróticas (Tang *et al.,* 2009), por lo que se ha especulado que los roseobacters fotótrofos podrían fijar carbono mediante vías anapleróticas en presencia de luz (Luo y Moran, 2014).

La quimiolitotrofía facultativa también es una característica común entre los miembros cultivados del linaje Roseobacter (Newton *et al*., 2010). Varios aislados del linaje presentan las formas I y II de la deshidrogenasa de monóxido de carbono (CO) codificada por los genes *cox* (Cunliffe, 2011; Luo y Moran, 2014). Se ha demostrado experimentalmente la utilización de monóxido de carbono por ejemplo en *Ruegeria pomeroyi* DSS-3^T. Esta bacteria fija CO a concentraciones típicas de aguas marinas superficiales (10 nM en aguas costeras y 2 nM en regiones oceánicas) (Moran *et al*., 2004). La oxidación del CO es una adaptación al ecosistema marino, considerado en general como oligotrófico, que permite obtener energía a partir de un compuesto inorgánico (Moran y Miller, 2007). Además, junto a la AAnP, también afecta a la eficiencia en la incorporación del carbono orgánico en la cadena trófica, y a la producción y liberación de CO y CO_2, gases importantes desde el punto de vista climático (Moran y Miller, 2007). En genomas de Roseobacter también se ha detectado la presencia de genes *sox,* para la oxidación de sulfuro o tiosulfato, que junto a la oxidación del monóxido de carbono puede proporcionar una fuente de energía adicional (Moran *et al*., 2007; Luo y Moran, 2014). En

concreto, la oxidación del tiosulfato se ha demostrado experimentalmente en *R. pomeroyi* DSS-3^T, organismo en el que se observó un aumento del 45 % en la densidad óptica de los cultivos suplementados con tiosulfato 10 mM respecto a los cultivos no suplementados con tiosulfato (Moran *et al*., 2004). Cabe destacar que la distribución taxonómica de los genes *cox* y *sox* no concuerda con la filogenia del grupo (Luo y Moran, 2014; Zhang *et al*., 2016). Este resultado sugiriere que la adquisición o pérdida de estos genes ha ocurrido en varios episodios durante la evolución del linaje Roseobacter, cuyos miembros han ido explorando y colonizando nuevos hábitats a lo largo de su historia evolutiva (Luo y Moran, 2014).

Entre la variedad de compuestos orgánicos utilizados por los miembros del linaje Roseobacter en su metabolismo quimioorganoheterótrofo vale la pena destacar algunos de ellos por su relevancia en el funcionamiento del ecosistema marino y por el interés particular de esta tesis. Una fuente de carbono y energía importante para los miembros del linaje Roseobacter es el dimetil sulfoniopropionato (DMSP) (González *et al*., 2000; Reisch *et al*., 2011). El DMSP es un compuesto osmoprotector que es producido en grandes cantidades por micro y macro algas marinas y por plantas halófitas (Yoch, 2002). Este compuesto es liberado al medio por lisis celular, senescencia e infecciones víricas, y constituye uno de los componentes principales de la materia orgánica disuelta del océano (Reisch *et al*., 2011). El DMSP se puede degradar de dos maneras: mediante la vía de la rotura y mediante la vía de la desmetilación/desmetiolación (Moran *et al*., 2003). La vía de la rotura da lugar a dimetilsulfuro (DMS), un compuesto volátil del azufre que es importante para el clima global al participar en la formación de nubes y contrarrestar el calentamiento global (Charlson *et al*., 1987). En esta vía de degradación del DMSP se genera DMS y acrilato por acción de una DMSP liasa, y se da tanto en miembros del linaje Roseobacter como en otras bacterias marinas cultivables y en algas eucariotas (Moran *et al.* 2003). En cuanto a la vía de la desmetilación/desmetiolación, el DMSP es desmetilado a 3-metil-mercaptopropionato (MMPA) y luego desmetiolado para dar lugar a metanotiol (MeSH) y acrilato (Moran *et al*., 2003). El MMPA a su vez puede volver ser desmetilado dando lugar a 3-mercaptopropionato (MPA), aunque esta última desmetilación no se ha detectado en los miembros del linaje Roseobacter (Moran *et al*., 2003). La vía de la desmetilación/desmetiolación solo se encuentra en bacterias y es la vía de degradación del DMSP más detectada. Ahora bien, las únicas bacterias en las que se da tanto la vía de degradación como la vía de la desmetilación/desmetiolación son las del linaje Roseobacter (Moran *et al.*, 2003). Esto indica una adaptación óptima del linaje Roseobacter a utilización de compuestos derivados de las algas (Geng y Belas, 2010; Buchan *et al*., 2014) tal y como ya se ha mencionado anteriormente.

Un segundo grupo de compuestos que pueden ser metabolizados por miembros cultivados del linaje Roseobacter, y que constituyen el objeto principal de esta tesis, son compuestos monoaromáticos. Como se desarrolla a continuación en el siguiente apartado existen evidencias de que los miembros del linaje Roseobacter participan en la degradación de compuestos aromáticos en los ecosistemas marinos (Buchan y González, 2010).

En los últimos años, con la obtención del genoma de la cepa HTCC2255 y de representantes no cultivados del linaje Roseobacter mediante secuenciación de una única célula (del inglés *single-cell genomics*, SCG), se han detectado diferencias importantes entre los genomas de los roseobacters cultivados (a excepción de HTCC2255) y los no cultivados. La rama evolutiva que

está representada por la cepa HTCC2255, a diferencia del resto de roseobacters cultivados, correspondería a un microorganismo del bacterioplancton, de vida libre, que se separó del ancestro común y sufrió una drástica reducción génica (Luo *et al*., 2013). Entre los genes que se perdieron se encuentran genes conservados en la mayoría de los roseobacters cultivados, como por ejemplo genes involucrados en la motilidad, quimiotaxis, metabolismo y síntesis de metabolitos secundarios, transducción de señales y varias funciones regulatorias (Luo *et al*., 2013). Con esta pérdida de genes, no esenciales para un estilo de vida libre, el genoma de roseobacters como la cepa HTCC2255 pasó a tener un perfil funcional más similar al de SAR11, el ejemplo típico de bacteria oligotrófica marina de vida libre, que al de otros roseobacters (Luo *et al*., 2013). En cuanto a los genomas SCG de grupos de Roseobacter no cultivados (los grupos CHAB-I-5, SAG-O19, DC5-80-3 y NAC11-7; este último incluye un representante cultivado, la cepa HTCC2255), estos no solo son más pequeños y presentan un menor contenido de mol % G+C sino que, además, tienen una mayor proporción de genes característicos de bacterias de vida libre (por ejemplo, de sistemas de secreción Sec) y una disminución de genes que son comunes entre los miembros cultivados del linaje Roseobacter (por ejemplo, genes de producción de antibióticos, de quimiotaxis y genes de modificación de la superficie celular). También tienen un menor porcentaje de ADN no codificante (Luo *et al*., 2014; Zhang *et al*., 2016). Estas características indican que los roseobacters no cultivados estudiados hasta el momento tienen genomas más parecidos a los modelos de bacterioplancton de vida libre con genomas reducidos (p.e. SAR11 y SAR86) que a los de los roseobacters cultivados (Luo *et al*., 2014). La excepción serían los genomas del grupo CHAB-I-5, que se parecen más a los de los genomas cultivados ya que son relativamente grandes (4,1–4,4 Mb) y contienen una elevada fracción de ADN no codificante (10–12 %) (Zhang *et al*., 2016). Respecto a la capacidad de fototrofía presente en los genomas de las bacterias no cultivadas, se ha detectado la presencia de la ruta completa de genes implicados en AAnP en los grupos SAG-O19 y DC5-80-3 y la ruta incompleta en el grupo CHAB-I-5 (Zhang *et al*., 2016). En cuanto a la capacidad de quimiolitotrofía facultativa presente en los genomas de las bacterias no cultivadas, se ha detectado la presencia de la formas I de la deshidrogenasa de monóxido de carbono en los grupos SAG-O19 y DC5-80-3 y de la forma II en el grupo CHAB-I-5 (Zhang *et al*., 2016). Además, se ha detectado la presencia de los genes *sox* en todos los grupos de roseobacters no cultivados (Zhang *et al*., 2016). Por lo tanto la mixotrofía parece ser una característica común en todos los miembros del linaje Roseobacter.

2. Degradación aeróbica de compuestos aromáticos

Los compuestos aromáticos se encuentran ampliamente distribuidos en la naturaleza puesto que se producen de forma natural por una gran variedad de procesos biológicos y biogeoquímicos (Pérez-Pantoja *et al*., 2010a). La mayor parte de los compuestos aromáticos derivan de la degradación de la lignina, que es el segundo polímero más abundante en la naturaleza tras la celulosa, y que comprende hasta el 25 % de la biomasa vegetal terrestre (Kirk y Farrell, 1987). Los productos de degradación de la lignina, junto a otros compuestos aromáticos derivados de las plantas y de las algas, constituyen componentes importantes de la materia orgánica disuelta en ambientes marinos, tales como marismas y estuarios costeros (Moran y Hodson, 1994; Sleighter y Hatcher, 2008). Los compuestos aromáticos también proceden de los aminoácidos aromáticos (Pérez-Pantoja *et al*., 2010a). Finalmente, otras

fuentes de compuestos aromáticos derivan de la extracción y uso de combustibles fósiles como por ejemplo el tráfico marítimo (GESAMP, 2007), derrames producidos en el transporte marítimo de mercancías peligrosas (como el petróleo, sus derivados refinados y otros productos químicos), el vertido de desechos industriales y aguas residuales o el uso de pinturas para barcos (Pérez-Pantoja *et al*., 2010a; Psarros *et al*., 2011).

La degradación de compuestos aromáticos es importante a nivel de los ciclos biogeoquímicos globales y la llevan a cabo casi exclusivamente microorganismos, que han desarrollado diversas estrategias para degradar estos compuestos y obtener carbono y energía (Pérez-Pantoja *et al*., 2010a). Existe una gran variedad de compuestos aromáticos y la estrategia para degradarlos consiste en transformar estos compuestos mediante una amplia variedad de vías de degradación (vías periféricas), que confluyen dando lugar a un menor número de compuestos centrales (Pérez-Pantoja *et al*., 2010a). Estos compuestos monoaromáticos centrales son degradados hasta dar lugar a intermediarios del ciclo de Krebs mediante las vías centrales para la degradación de compuestos aromáticos (Pérez-Pantoja *et al*., 2010a).

En todos los procesos de degradación aeróbica de compuestos monoaromáticos hay dos pasos clave: 1) la activación del anillo bencénico, en general mediante su oxigenación y 2) la posterior rotura del anillo. Esta rotura puede ser intradiol, es decir que se produce entre dos grupos hidroxilo (también llamada rotura *orto*), o puede ser una rotura extradiol, es decir, adyacente a un grupo hidroxilo (también llamada rotura *meta*). Una tercera posibilidad es que la rotura se produzca entre un carbono portador de un grupo carboxilo y el carbono adyacente, el cual se encuentra hidroxilado y en posición *para* respecto a otro carbono hidroxilado (rotura *para*) (Pérez-Pantoja *et al*., 2010a). Sin embargo, en condiciones aeróbicas también se produce un tipo de catabolismo que no se basa en las etapas clásicas descritas, sino en la formación de tioésteres de Coenzima A (CoA) y en una rotura no oxigenolítica (Nogales *et al*., 2007; Pérez-Pantoja *et al*., 2010a). Clásicamente, se pensaba que la rotura de anillos aromáticos con sustitutos de CoA solo se daba en vías anaeróbicas (Pérez-Pantoja *et al*., 2010a). Sin embargo, se describió una nueva vía para la degradación de compuestos aromáticos que no implicaba intermediarios aromáticos dihidroxilados para la degradación de fenilacetato en *Escherichia coli* W (Ferrández *et al*., 1998) y *Pseudomonas putida* U (Olivera *et al*., 1998). El compuesto alifático resultante sufre una serie de reacciones parecidas a las de la beta-oxidación de los ácidos grasos que implican la formación de tioésteres de CoA y al final acabará dando lugar a intermediarios del ciclo de Krebs (Nogales *et al*., 2007, Pérez-Pantoja *et al*., 2010a). A este tipo de vías como la del fenilacetato se las conoce como vías aeróbicas híbridas, puesto que combinan características de vías anaeróbicas, como la formación inicial de derivados de CoA, con reacciones típicas de vías aeróbicas como la oxigenación del anillo aromático (Díaz, 2004).

En la activación del anillo bencénico suelen estar involucradas enzimas pertenecientes a tres superfamilias: las oxigenasas Rieske con hierro no hemo (*Rieske non-heme iron oxygenases*, Rieske-NH); las flavoproteínas monooxigenasas (*flavoprotein monooxygenases*) y las oxigenasas multicomponente de dihierro soluble (*soluble diiron multicomponent oxygenases*) (Pérez-Pantoja *et al*., 2010b). Las oxigenasas Rieske-NH normalmente catalizan la incorporación de dos atomos de oxígeno al anillo bencénico (aunque algunas solo catalizan una monooxigenación) y son enzimas importantes para la activación y posterior degradación

aeróbica de compuestos aromáticos tales como benzoato, benceno, tolueno, ftalato, naftaleno o bifenilo (Pérez-Pantoja *et al*., 2010b). Son enzimas formadas por varias subunidades con un componente oxigenasa terminal (proteína hierro-azufre) y proteínas de transporte de electrones (una ferredoxina y una reductasa o una combinación de ferredoxina-NADH reductasa) (Pérez-Pantoja *et al*., 2010b). Las proteínas hierro-azufre del componente oxigenasa suelen ser heterodímeros compuestos por una subunidad alfa mayor y una subunidad beta pequeña (Pérez-Pantoja *et al*., 2010b). La subunidad alfa contiene un grupo de tipo Rieske [2Fe-2S], un hierro no hemo en el centro de activación del oxígeno y un sitio de unión del sustrato que determina la especificidad de la enzima (Pérez-Pantoja *et al*., 2010b). En cuanto a la subunidad beta, esta no siempre esta presente puesto que algunas enzimas como las ftalato 4,5-dioxigenasas contienen solo subunidades alfa (Pérez-Pantoja *et al*., 2010b). En base a análisis filogenéticos de las subunidades alfa, se han descrito dos grandes linajes de Rieske-NH, uno al que pertenecen las enzimas de la familia de la ftalato oxigenasa y otro al que pertenecen el resto de familias descritas de Rieske-NH (Duarte *et al*., 2014).

Las flavoproteínas monooxigenasas están involucradas la degradación de compuestos aromáticos y también en otros procesos biológicos como por ejemplo la síntesis de antibióticos y sideróforos (Pérez-Pantoja *et al*., 2010b). Muchas de las flavoproteínas monooxigenasas contienen una flavina como grupo prostético unida de forma no covalente, pero hay otras que tienen un grupo FAD o FMN unido de forma covalente (van Berkel *et al*., 2006). Las flavoproteínas monooxigenasas se han clasificado en base a su secuencia en seis clases distintas: A-F (van Berkel *et al*., 2006), siendo las clases A, D y F especialmente relevantes para la degradación de compuestos aromáticos (Pérez-Pantoja *et al*., 2010b). Las flavoproteínas monooxigenasas son enzimas importantes para la degradación de compuestos aromáticos tales como salicilato, 3-hidroxibenzoato, aminobenzoil-CoA o 2-aminobenzoil-CoA (Pérez-Pantoja *et al*., 2010b).

Las oxigenasas multicomponente de dihierro soluble están formadas por una hidroxilasa, que a su vez puede estar formada hasta por tres subunidades distintas (α, β y γ), junto a una oxidoreductasa y una ferredoxina (Leahy *et al*., 2003). Las oxigenasas multicomponente de dihierro soluble pueden ser divididas en cuatro grupos de enzimas en base a su estructura, y en menor medida a su especificidad de sustrato (Leahy *et al*., 2003). Estos cuatro grupos son: las monooxigenasas solubles del metano, la alqueno monooxigenasa de *Rhodococcus corallinus* B-276, las fenol hidroxilasas y las monooxigenasas de alquenos/aromáticos de 4 componentes (Leahy *et al*., 2003). Las oxigenasas multicomponente de dihierro soluble son enzimas capaces de monooxigenar benceno, tolueno y fenol para generar fenol, metilfenol y catecol, respectivamente (Pérez-Pantoja *et al*., 2010b). Cabe destacar que entre las enzimas que pertenecen a las oxigenasas multicomponente de dihierro soluble se encuentran las enzimas que se encargan de la hidroxilación de los compuestos aromáticos que deben activarse mediante la unión de un grupo CoA, como el fenilacetato y el benzoato (Pérez-Pantoja *et al*., 2010b).

En cuanto a la rotura del anillo bencénico tras la hidroxilación, como ya se ha comentado anteriormente, esta puede ser intradiólica o extradiólica (Pérez-Pantoja *et al*., 2010a). Las intradiol dioxigenasas tienen un centro mononuclear de Fe^{3+} coordinado con dos tirosinas y dos histidinas como ligandos (Fetzner, 2012). Las intradiol dioxigenasas son prácticamente

exclusivas de actinobacterias y proteobacterias, aunque se han detectado algunas excepciones como la presencia de una protocatecuato 3,4-dioxigenasa en *Deinococcus geothermalis* DSM 11300 (*Deinococci*) y en *Solibacter usitatus* Ellin6076 (*Acidobacteria*) (Pérez-Pantoja *et al*., 2010b). Todas las dioxigenasas intradiólicas pertenecen a la misma superfamilia (Duarte *et al*., 2014) en la que se distinguen siete grupos: el de las hidroxibenzoquinol dioxigenasas (grupo 1), catecol 1,2-dioxigenasas de proteobacterias (grupo 2), catecol 1,2-dioxigenasas de actinobacterias (grupo 3), subunidades alfa de las protocatecuato dioxigenasas (grupo 5), subunidades beta de las protocatecuato dioxigenasas (grupo 7) y finalmente, los grupos 4 y 6 cuyas enzimas todavía no tienen una función asignada (Pérez-Pantoja *et al*., 2010b). En cuanto a las extradiol dioxigenasas, estas suelen presentar un centro activo para la catálisis con un Fe^{2+} o menos frecuentemente con un Mn^{2+} (Fetzner, 2012). Las extradiol dioxigenasas pertenecen al menos a 3 superfamilias distintas: extradiol dioxigenasas del oxígeno quelado adyacente (*extradiol dioxygenases of the vicinal oxygen chelate superfamily)*, Lig B y cupina (Pérez-Pantoja *et al*., 2010b). Las extradiol dioxigenasas del oxígeno quelado adyacente (tipo I, EXDO I) se encargan de la rotura dioxigenolítica de los derivados del catecol en varias vías de degradación de compuestos mono y poliaromáticos y participan en la degradación de compuestos como benceno, tolueno, bifenilo, naftaleno, catecol y bibenzofurano, entre otros (Pérez-Pantoja *et al*., 2010b). La superfamilia de las extradiol dioxigenasas de tipo LigB (tipo II) estan formadas por dos tipos de subunidades, α y β, siendo la subunidad beta la que contiene el centro activo en una hendidura profunda que queda cubierta por la subunidad alfa (Sugimoto *et al*., 1999). Las enzimas de la superfamilia LigB se encargan de la degradación de mono y poliaromáticos como protocatecuato (rotura *meta*), homoprotocatecuato, galato o carbazol (Duarte *et al*., 2014). Finalmente, las enzimas de la superfamilia cupina (extradiol dioxigenasas tipo III) se caracterizan por la presencia del dominio cupina que consiste en un barril beta formado por seis láminas beta y que puede encontrarse en copia única (monocupinas) o duplicado (bicupinas) (Fetzner, 2012). Entre los sustratos sobre los que actúan las enzimas de la superfamilia cupina se encuentran compuestos aromáticos como gentisato, homogentisato y 3-hidroxiantranilato (Pérez-Pantoja *et al*., 2010b). Recientemente se ha descubierto otro tipo de dioxigenasas extradiólicas, las benzoquinol 1,2-dioxigenasas, cuyo origen no está claro y que también son dependientes de Fe^{2+} (Pérez-Pantoja *et al*., 2010b). Las benzoquinol 1,2-dioxigenasas son heterotetrámeros $\alpha_2\beta_2$, en los que las subunidades α y β no presentan identidad de secuencia con otras dioxigenasas. Estas enzimas catalizan la rotura de un amplio rango de benzoquinoles hasta sus correspondientes 4-hidroximucónico semialdehídos (Pérez-Pantoja *et al*., 2010b).

Por lo tanto, hay una extensa variedad de oxigenasas involucradas tanto en la activación como en la rotura de compuestos aromáticos. Una de las dificultades para estudiar estos enzimas es que su clasificación en las bases de datos no siempre es correcta. Estos problemas de clasificación se producen porque la mayoría de las enzimas que se encuentran en las bases de datos no se han caracterizado experimentalmente, solo se les ha asignado una determinada función en base a estudios de predicción por homología de secuencia (Duarte *et al*., 2014). Para corregir los errores de las bases de datos de enzimas y mejorar la clasificación funcional de estas, Duarte y colaboradores (2014) crearon una nueva base de datos de degradación, AromaDeg, que integra datos experimentales con análisis filogenómicos.

3. Evidencias de la degradación de compuestos aromáticos por bacterias del linaje Roseobacter

La primera evidencia de la capacidad de los miembros del linaje *Roseobacter* para degradar compuestos aromáticos vino del aislamiento de la bacteria *Sagittula stellata* E-37^T de un cultivo de enriquecimiento preparado con agua de mar de la costa de Georgia (EEUU) y pasta de papel (González *et al*., 1997). En concreto se observó que *S. stellata* E-37^T se unía a las partículas de celulosa y lignocelulosa, y que podía hidrolizar la celulosa y solubilizar y mineralizar la lignina (González *et al*., 1997). En el año 2000 Buchan y colaboradores determinaron la presencia de una enzima clave de la vía de degradación del beta-cetoadipato, la protocatecuato 3,4-dioxigenasa, en seis aislados del linaje que fueron aislados a partir de agua de mar y sedimentos de estuarios y zonas costeras del sudeste de Estados unidos. En concreto cinco de ellos se aislaron a partir de enriquecimientos con lignina y otros compuestos aromáticos (*S. stellata* E-37^T y las cepas de *Sulfitobacter* sp. EE-36, Y3F, IC4 y S25com04). El sexto aislado, GAI-16, se aisló directamente de aguas costeras usando un medio no selectivo y pobre en nutrientes (Buchan *et al*., 2000). A su vez se demostró que los seis aislados eran capaces de crecer a expensas de al menos tres de los siguientes compuestos aromáticos: antranilato, benzoato, *p*-hidroxibenzoato, salicilato, vanillato, ferulato, protocatecuato y cumarato. Además se vio que en cuatro de los aislados la actividad de la enzima protocatecuato 3,4-dioxigenasa se inducía al crecer en *p*-hidroxibenzoato (Buchan *et al*., 2000). En estudios posteriores estos autores ampliaron la lista de aislados marinos de Roseobacter que tenían el gen para protecuato 3,4-dioxigenasa y/o eran capaces de crecer con compuestos aromáticos (Buchan *et al*., 2001; 2004).

Posteriormente, Brito y colaboradores (2006) aislaron ocho cepas de Roseobacter de sedimentos de un manglar en Brasil (bahía de Guanabara), usando agua de mar artificial suplementada con pireno, naftaleno o fluorantreno como únicas fuentes de carbono. En cultivo puro todos los aislados pudieron transformar estos hidrocarburos poliaromáticos, con una capacidad transformación de entre el 10 y el 100 % (Brito *et al*., 2006). Un año después Harwati y colaboradores obtuvieron 98 aislados del linaje Roseobacter de aguas procedentes del puerto Semarang (Indonesia) y demostraron la capacidad de 43 de ellos para degradar componentes del crudo (Harwati *et al*., 2007). Estos autores además describieron dos nuevas especies, *Tranquillimonas alkanivorans* y *Tropicibacter naphthalenivorans*, como degradadores de alcanos y compuestos aromáticos policíclicos, respectivamente (Harwati *et al*., 2008; 2009). En los últimos años se han seguido aislando miembros del linaje Roseobacter implicados en la degradación de compuestos monoaromáticos y compuestos aromáticos policíclicos. Por ejemplo, se han obtenido nuevos aislados relacionados con los géneros *Citreicella*, *Marinovum*, *Pelagibaca*, *Sulfitobacter*, *Thalassobius*, *Thalassococcus* y *Tropicibacter* capaces de degradar ftalato (Iwaki *et al*., 2012a; 2012b). En cuanto a los nuevos aislados implicados en la degradación de compuestos aromáticos policíclicos se encuentran, por ejemplo, los degradadores de naftaleno *Citreicella aestuarii* 357 (Suárez-Suárez *et al*., 2012) y *Confluentimicrobium naphthalenivorans* NS6^T (Jeong *et al*., 2015), *Celeribacter indicus* P73^T capaz de degradar bifenilo, naftaleno, 2-metilnaftaleno, 2,6-dimetilnaftaleno, acenafteno, antraceno, fenantreno, dibenzotiofeno, dibenzofurano, fluoreno, 4-metildibenzotiofeno y

fluoranteno (Lai *et al*., 2014), y el degradador de fenantreno y fluoreno *Celeribacter persicus* SBU1^{T} (Jami *et al*., 2016).

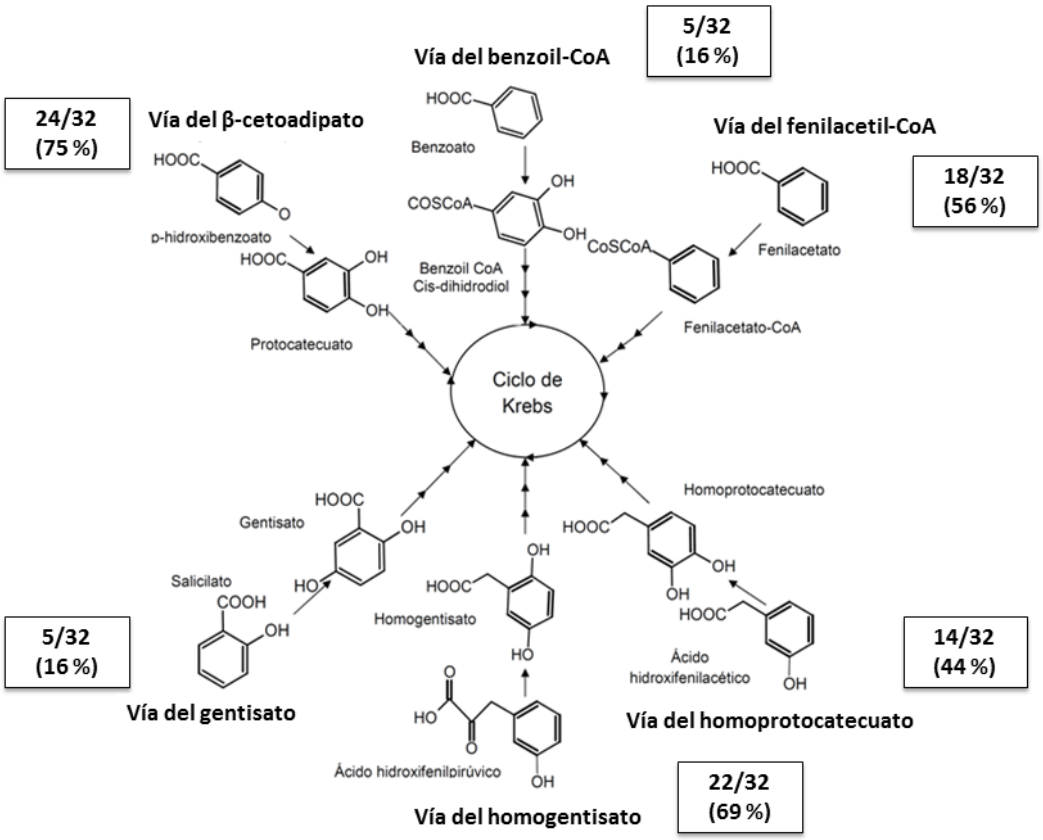

Figura I2. Vías de degradación de compuestos monoaromáticos detectadas en genomas del linaje Roseobacter. Imagen modificada de Moran *et al*., 2007 con los datos de Newton *et al*., 2010. Se indica el número de genomas respecto al total analizado (32) que presentan el gen codificante para la enzima principal en la degradación de estos seis compuestos, y entre paréntesis, el porcentaje que representan.

Mucha de la información de que se dipone sobre la capacidad de los miembros del linaje Roseobacter para degradar hidrocarburos y compuestos aromáticos se basa en análisis de genomas (Buchan y González, 2010) porque hay pocos estudios que incluyan análisis químicos o bioquímicos de degradación. Algunos ejemplos serían los realizados recientemente con algunos de los aislados degradadores de compuestos poliaromáticos (Suárez-Suárez, 2013; Lai *et al*., 2014 y Jeong *et al*., 2015). En estudios basados en homología de secuencia se describió que los miembros del linaje Roseobacter presentaban hasta seis vías para la degradación aeróbica de compuestos monoaromáticos: benzoil-CoA, fenilacetil-CoA, homoprotocatecuato, homogentisato, gentisato y rama del protocatecuato de la vía del β-cetoadipato (Moran *et al*., 2007; Newton *et al*., 2010; Buchan y González, 2010). Como se puede observar en la figura I2 las vías de degradación más frecuentemente detectadas en el estudio de Newton y colaboradores (2010) fueron: la rama del protocatecuato de la vía del β-cetodipato (75 % de los genomas analizados), la del homogentisato (69 % de los genomas) y la del fenilacetil-CoA (56 % de los genomas). Por lo tanto, la presencia de genes implicados en la degradación de compuestos monoaromáticos se describió como una característica común de los miembros cultivados del linaje Roseobacter. La gran variedad de vías implicadas en la degradación de compuestos monoaromáticos en el linaje Roseobacter es una característica más común en

bacterias de suelo (p.e. *Pseudomonas*) que en las bacterias marinas (Buchan y González, 2010). A pesar de que se detecta una gran variedad de vías de degradación de compuestos aromáticos en el conjunto del linaje Roseobacter, en algunas especies la presencia de estas rutas es mayor. Por ejemplo, *Sagittula stellata* E-37^T, *Jannaschia* sp. CCS1, *Ruegeria pomeroyi* DSS-3^T y *Maritimibacter alkaliphilus* HTCC2654^T presentan genes implicados en la degradación de protocatecuato, gentisato, homoprotocatecuato, benzoato y fenilacetato (Buchan y González, 2010). Además, estos aislados también tienen genes que presentan homología con alcano hidroxilasas y con otras hidroxilasas y dioxigenasas implicadas en la rotura de anillos aromáticos (Buchan y González, 2010). Desde la publicación en 2010 de los artículos de Newton y colaboradores y de Buchan y González, se han secuenciado una gran cantidad de genomas, sin embargo, en la mayoría de estos aislados no se estudiado su capacidad para degradar hidrocarburos. En los últimos años, uno de los estudios en los que se ha analizado la capacidad degradadora de hidrocarburos en nuevos genomas del linaje Roseobacter es el de Más-Lladó (2017). En concreto, en este estudio se trabajó con nueve genomas de aislados de puertos deportivos de Mallorca y en siete de ellos se predijo la presencia para alcano monooxigenasas, (Más-Lladó, 2017). Por lo tanto, en base a los resultados de los estudios publicados, sería interesante estudiar la presencia de genes implicados en la degradación de compuestos aromáticos en los genomas de Roseobacter cuyo potencial cabatólico aún no ha sido analizado. De esta forma, se podría comprobar si la presencia de genes catabólicos es una característica común de los genomas del linaje Roseobacter, y por ello, es uno de los objectivos de la presente tesis.

Por otra parte, se han realizado varios estudios sobre la respuesta a hidrocarburos que presentan los miembros del linaje Roseobacter. La mayor parte de estos estudios demuestran que los miembros del linaje Roseobacter responden favorablemente a la presencia de hidrocarburos. Sin embargo, no se sabe si presentan preferencia para degradar ciertos hidrocarburos, ni si la degradación de estos compuestos se ve favorecida por determinadas condiciones ambientales (Buchan y González, 2010). Un ejemplo de estos estudios es el de Brakstad y Lødeng (2005) en el que se observaba que la presencia del linaje Roseobacter, en base a 16S ARNr, aumentaba de un 21 a un 89 % tras 21 días de exposición al petróleo en aguas costeras y de <1 a un 86 % en aguas marinas cercanas a una plataforma petrolífera. Además vieron que los miembros del linaje Roseobacter que se veían favorecidos en estas condiciones estaban relacionados con los géneros *Sulfitobacter* y *Roseobacter* (Brakstad y Lødeng, 2005). Secuencias afiliadas al género *Sulfitobacter* fueron también abundantes en las librerías de pirosecuenciación generadas a partir de muestras de arena del Golfo de Méjico contaminadas por el derrame de petróleo de la plataforma Deepwater Horizon (DWH) (Kostka *et al*., 2011). Por otra parte, no todos los miembros del linaje Roseobacter proliferan en presencia de hidrocarburos (Buchan y González, 2010). Por ejemplo, en el estudio de Prabagaran y colaboradores (2007), realizado en microcosmos de agua de mar de la costa de Ushuaia suplementados con la fracción soluble en agua del petróleo, se observó una disminución de poblaciones de Roseobacter tras la exposición a petróleo. Mckew y colaboradores (2007) realizaron experimentos de microcosmos con agua de mar de una marisma costera en Stanford-le-Hope, cercana a dos refinerías de petróleo en el estuario del Támesis. Utilizaron varios hidrocarburos entre los que se encontraban alcanos e hidrocarburos aromáticos policíclicos. Este estudio sugería que ciertas poblaciones de Roseobacter podían ser

estimuladas por la presencia de alcanos e inhibidas por la presencia de determinados componentes del petróleo, como los hidrocarburos aromáticos policíclicos (Mckew *et al*., 2007). Además, en el estudio de Piña-Villalonga (2012), con agua de mar de la costa de Mallorca suplementada con diésel, también se observó cómo algunas poblaciones de Roseobacter proliferaban en presencia de diésel mientras otras no.

Finalmente, cabe destacar que se han realizado varios estudios de diversidad microbiana en ambientes crónicamente contaminados con hidrocarburos (puertos), entre los que encontramos los de Nogales y colaboradores (2007) y Piña-Villonga (2012) en varios puertos de Mallorca, Zhang y colaboradores (2007) en Puerto Victoria (Hong Kong) y el de Ma y colaboradores (2009) en Puerto Xiamen (Singapur). En todos ellos, se detectó la presencia de miembros del linaje Roseobacter. Por lo tanto, el potencial para el catabolismo de compuestos aromáticos en el linaje Roseobacter, junto a la respuesta favorable de algunas poblaciones a la exposición a hidrocarburos, convierten al linaje Roseobacter en uno de los más interesantes para el estudio de la degradación microbiana de compuestos aromáticos en ambientes marinos. Además, la elevada abundancia de microorganismos de este linaje en ambientes contaminados con hidrocarburos (p.e. puertos) convierte este tipo de ambientes en idóneos para estudiar la contribución de este linaje en la degradación de hidrocarburos.

4. Potencial para la degradación de hidrocarburos en el ambiente

Se han realizado numerosos estudios para determinar la composición de las comunidades microbianas en ambientes crónicamente contaminados y los cambios que se producen en las comunidades en respuesta a la contaminación (Cappello *et al*., 2007; Yakimov *et al*., 2007; Nogales, 2010; Berthe-Corti y Nachtkamp, 2010). La mayoría de estos estudios se basaron en el análisis de secuencias de 16S ARNr, por lo tanto permitían conocer la estructura de las comunidades microbianas existentes en estos ambientes pero no permitían realizar una aproximación al potencial catabólico de las comunidades. También se han realizado estudios para estudiar el potencial catabólico de estas comunidades a partir de la amplificación mediante PCR de determinados genes catabólicos para alcano monooxigenasas (AlkB) o dioxigenasas de compuestos aromáticos y la comparación de los productos resultantes entre diversas comunidades (Iwai *et al*., 2010; Guibert *et al*., 2012; Marcos *et al*., 2012).

Con la aparición de las técnicas de metagenómica se pudieron estudiar de forma más detallada tanto los variaciones estructurales de la comunidad, como los cambios funcionales que se producen cuando esta se expone a un cambio en las condiciones ambientales (Gilbert y Dupont, 2011; Simon y Daniel, 2011). Se han realizado varios estudios metagenómicos para analizar la respuesta de comunidades microbianas marinas a la exposición a hidrocarburos tanto en ambientes crónicamente contaminados como tras vertidos accidentales de petróleo (Mason *et al*., 2012; Gutierrez *et al*., 2013; Kimes *et al*., 2013; Hawley *et al*., 2014; Kim *et al*., 2014, Bargiela *et al*., 2015). Se ha observado como la diversidad química de los componentes del crudo de petróleo junto con la profundidad, la concentración de oxígeno, la temperatura y el aporte de nutrientes influyen en las poblaciones microbianas y los procesos de biodegradación que se producen en respuesta a los derrames accidentales de petróleo en agua de mar y sedimentos (Mason *et al*., 2012; Gutierrez *et al*., 2013; Kimes *et al*., 2013).

En base a estudios de secuencias de 16S ARNr, se había observado una sucesión de poblaciones bacterianas tras los vertidos de crudo, con una predominancia inicial de poblaciones especializadas en la degradación de hidrocarburos alifáticos, como los miembros del orden *Oceanospirillales* de géneros como *Alcanivorax*, *Oleispira*, *Thalassolituus, Neptunomonas y Oleiphilus*, y otras *Gammaproteobacteria* como *Marinobacter* (Yakimov *et al*., 2007; Bargiela *et al*., 2015). Con el tiempo se producía una sustitución de estas poblaciones por otras especializadas en la degradación de hidrocarburos aromáticos como las de los géneros *Cycloclasticus* y *Colwellia*, acompañado de un enriquecimiento de los géneros *Halomonas*, *Alteromonas* y *Pseudoalteromonas* (Yakimov *et al*., 2007; Gutierrez *et al*., 2013). Con los estudios metagenómicos realizados a partir del vertido de crudo de la plataforma petrolífera DWH en el Golfo de Méjico en 2010 (McNutt *et al*., 2012) se confirmó también esta dinámica de poblaciones (Mason *et al*., 2012; Gutierrez *et al*., 2013). Estos cambios a nivel taxonómico también fueron acompañados de cambios a nivel de potencial para el catabolismo de hidrocarburos en estas aguas (Lu *et al*., 2012; Mason *et al*., 2012). En el estudio de Mason y colaboradores (2012) se observó un aumento de la abundancia relativa de los genes involucrados en la degradación de alcanos tras el vertido del DWH. En cambio, la abundancia relativa de los genes involucrados en la degradación de compuestos aromáticos, como el etilbenceno y el tolueno, era menor que en las muestras no contaminadas (Mason *et al*., 2012). La mayor abundancia de genes involucrados en la degradación de alcanos, en comparación con aquellos implicados en la degradación de compuestos aromáticos del estudio de Mason y colaboradores (2012), es consistente con que los alcanos son más fácilmente degradables que los compuestos aromáticos, tal y como recogieron Das y Chandran (2011). En cambio, en el estudio de Lu y colaboradores (2012) en la estela del vertido de la DWH, no solo se observó una mayor abundancia de genes involucrados en la degradación de alcanos, sino también de genes involucrados en la degradación de compuestos aromáticos, entre ellos de compuestos aromáticos heterocíclicos. Además, estos cambios a nivel de potencial para el catabolismo de hidrocarburos también fueron acompañado de otros cambios funcionales, como el aumento en la abundancia relativa de genes involucrados en el metabolismo del carbono, del nitrógeno y del fósforo (Lu *et al*., 2012). Por lo tanto, en la estela del vertido de la DWH, en base a los genes detectados por Lu y colaboradores (2012), se veía favorecida la degradación de varias fuentes de carbono, entre ellas fuentes de carbono complejas, como podrían ser varios componentes del petróleo y sus derivados. Además, también se veía favorecida la liberación de nitrógeno y fosforo al medio y su asimilación, con lo que se podría aumentar la biomasa microbiana, ya que estos dos elementos son elementos limitantes del crecimiento en el medio marino (Lu *et al*., 2012).

En cuanto a los ambientes crónicamente contaminados con hidrocarburos, en base a estudios anteriores se sabía que la diversidad microbiana en estos ambientes era alta y que además predominaban bacterias del fílum *Bacteroidetes*, gammaproteobacterias marinas oligotróficas (OMG) y miembros del linaje Roseobacter (Nogales, 2010). Además, en estos ambientes las típicas bacterias marinas degradadoras de hidrocarburos (Yakimov *et al*., 2007) eran componentes minoritarios de la comunidad (Berthe-Corti y Nachtkamp, 2010; Nogales, 2010). En el estudio metagenómico de Kisand y colaboradores (2012) donde compararon comunidades de agua del puerto de Génova y una zona control cercana a la isla de Montecristo observaron que en ambos casos predominaban las clases *Alphaproteobacteria*,

Gammaproteobacteria y *Flavobacteria*. Sin embargo, se observaban diferencias en estas clases y en la clase *Actinobacteria* entre el agua contaminada y el control. En concreto, en el agua control destacaba la clase *Actinobacteria* como predominante, mientras que, en el agua contaminada predominaban *Gammaproteobacteria* y *Flavobacteria,* y prácticamente no se detectaba *Actinobacteria* (Kisand *et al*., 2012). Estos cambios a nivel taxonómico también iban acompañados de cambios a nivel funcional, puesto que en las aguas control había un mayor potencial genético para la captación nutrientes a bajas concentraciones, con una elevada presencia de transportadores ABC, mientras que en las aguas contaminadas se detectaban más genes para factores implicados en la respuesta a estrés, la motilidad celular, en la captación de nutrientes típicas de organismos copiotróficos y en el metabolismo de compuestos aromáticos (Kisand *et al*., 2012).

Para poder profundizar el conocimiento del potencial catabólico de las aguas de puertos a partir de estudios metagenómicos en la presente tesis se planteó realizar un estudio metagenómico de las aguas de puerto y su comparación con un metagenoma de aguas control en una zona adyacente. En base a estudios de 16S ARNr de nuestro grupo de investigación (Nogales *et al*., 2007; Aguiló-Ferretjans *et al*., 2008) se sabe que el linaje Roseobacter es un grupo importante en las aguas crónicamente contaminadas de puertos, aunque se desconoce la contribución de bacterias de este linaje al potencial catabólico de estas aguas. Por lo tanto, con el presente estudio no solo se pretende establecer el potencial catabólico presente en estas aguas, sino también su posible relación con los miembros del linaje Roseobacter.

5. Mecanismos de tolerancia a compuestos aromáticos

Los compuestos aromáticos pueden ser potencialmente tóxicos para los microorganismos, incluso para aquellos que pueden degradarlos. Por su naturaleza lipofílica se acumulan en las membranas citoplasmáticas, aumentando su fluidez y pudiendo llegar a lisar la membrana causando muerte celular (Ramos *et al*., 2002; Bernal *et al*., 2007b). De hecho, la degradación de compuestos aromáticos además de ser utilizada como fuente de carbono y energía para las células también podría ser utilizada para reducir la toxicidad de los compuestos aromáticos cuando estos se encuentran presentes el medio (Ramos *et al*., 2015). Ahora bien, no todos los microorganismos tienen la capacidad para degradar compuestos aromáticos y además, no hay correlación entre la capacidad para degradar un compuesto aromático y la capacidad para tolerarlo cuando este se encuentra en elevadas concentraciones. Por lo tanto, la capacidad para degradar un compuesto aromático no implica que este no sea tóxico para la célula cuando se encuentra en elevadas concentraciones (Ramos *et al*., 2015).

Una estrategia para sobrevivir en presencia de compuestos aromáticos potencialmente tóxicos para la célula, independientemente de su capacidad para degradar estos compuestos, consiste en desarrollar mecanismos de tolerancia a estos compuestos de forma que pueda mantener su viabilidad (Ramos *et al*., 2015). Los mecanismos implicados en la tolerancia a compuestos aromáticos han sido bien estudiados en el género *Pseudomonas*, y especialmente en la cepa *Pseudomonas putida* DOT-T1E (figura I3), que es un degradador eficiente de benceno, etilbenceno y tolueno entre otros compuestos aromáticos (Udaondo *et al*., 2012). Además, esta cepa presenta una alta resistencia a solventes orgánicos como por ejemplo el octanol, decanol, benceno y tolueno entre otros (Udaondo *et al*., 2012). Entre los mecanismos

implicados en la tolerancia a compuestos aromáticos documentados en la cepa *P. putida* DOT-T1E se encuentran: 1) alteraciones a nivel de permeabilidad de la membrana; 2) expulsión de los compuestos tóxicos mediante bombas; 3) activación de mecanismos de reparación del ADN, chaperonas y respuesta a estrés oxidativo, y 4) cambios en el metabolismo energético (Udaondo *et al*., 2012; Ramos *et al*., 2015).

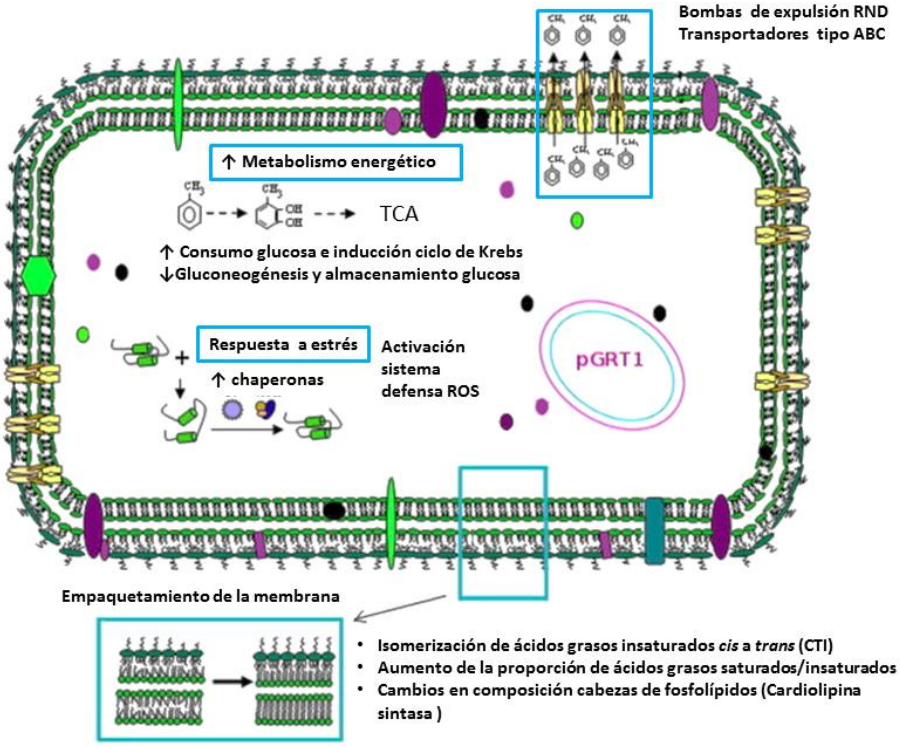

Figura I3. Mecanismos implicados en la tolerancia a tolueno en *Pseudomonas putida* DOT-T1E. Imagen modificada de Udaondo *et al*., 2012.

Por lo que respecta a los mecanismos implicados en la alteración de la permeabilidad de la membrana estos consisten en cambios en la isomerización y composición de los ácidos grasos, y cambios en los grupos de las cabezas polares de los fosfolípidos (Ramos *et al*., 2015). La isomerización de los ácidos grasos insaturados es una respuesta a corto plazo tras la exposición a compuestos aromáticos y consiste en un cambio de configuración *cis* a *trans* de los ácidos grasos insaturados por acción de la isomerasa *cis-to-trans* (CTI) descrita en varias cepas de *Pseudomonas putida* (Ramos *et al*., 2015). En cuanto a los cambios en la composición de los ácidos grasos, esta es una respuesta a más largo plazo y consiste en incrementar la proporción de ácidos grasos saturados frente a los insaturados de los fosfolípidos. Con estos cambios, aumenta el empaquetamiento de la membrana y disminuye su fluidez al disminuir la repulsión estérica (Ramos *et al*., 2015). Otro cambio en los ácidos grasos insaturados con configuración *cis* consiste en la adición de un grupo metileno en la posición del doble enlace *cis* dando lugar a ácidos grasos de ciclopropano (CFAs), que parecen tener un papel importante a la hora de tolerar solventes orgánicos (Ramos *et al*., 2015). Finalmente, se producen cambios en las cabezas polares de los fosfolípidos que también afectan a la fluidez de membrana (Ramos *et al*., 2002). En concreto en las cepas de *P. putida* S12 y DOT-T1E se observó un aumento del contenido en cardiolipina tras la exposición a tolueno (Ramos *et al*., 2015). Estudios realizados con mutantes de *P. putida* DOT-T1E deficientes en cardiolipina sintasa sugirieron que esta

enzima provocaba cambios en la arquitectura de la membrana que mejoraban la eficiencia de las bombas de expulsión (Bernal *et al*., 2007a).

Cabe destacar que los cambios a nivel de membrana asociados a la tolerancia a compuestos aromáticos no son suficientes para prevenir la entrada de estos al interior celular cuando las células están expuestas a elevadas concentraciones de compuestos aromáticos. Por eso las células disponen de mecanismos adicionales para la tolerancia a estos compuestos (Ramos *et al*., 2015). Uno de ellos es la expulsión de los compuestos aromáticos desde el citoplasma o el periplasma al exterior de la célula mediante bombas de expulsión. Este es un mecanismo ampliamente estudiado puesto que se considera el más eficiente para la tolerancia a compuestos aromáticos ya que evitan que estos se acumulen a concentraciones letales en el interior celular (Ramos *et al*., 2015). Se conocen cinco familias de bombas de expulsión en bacterias: familia de transportadores ABC (del inglés *ATP-binding cassette transporters*), RND (del inglés *resistance-nodulation-cell division*), MATE (del inglés *multi-antimicrobial extrusion proteins*), MFS (del inglés *major facilitator superfamily*) y SMR (del inglés *small multidrug resistance proteins*). En la tolerancia a compuestos aromáticos están implicados transportadores de las familias ABC y RND, siendo la familia RND la más importante para la tolerancia a estos compuestos en bacterias Gram negativas (Udaondo *et al*., 2012; Ramos *et al*., 2015). Los transportadores de tipo RND utilizan un gradiente protónico para transportar sus sustratos desde el citoplasma o la membrana interna al exterior celular. Están formados por tres componentes: un sistema de expulsión en la membrana citoplasmática que actúa como bomba de expulsión dependiente de la fuerza motriz de protones, una proteína de fusión a la membrana (MFP) también llamada adaptador periplásmico y un factor de la membrana externa (OMP) también llamado canal de la membrana externa (Nikaido y Takatsuka, 2009).

Los compuestos aromáticos también provocan cambios en la cadena de transporte electrónico (Udaondo *et al*., 2012; Ramos *et al*., 2015). En concreto, en *P. putida* DOT-T1E se ha detectado una sobreexpresión de varias proteínas de la cadena de transporte electrónico como las citocromo *c* oxidasas de tipo aa_3 y cbb_3 y la citocromo de tipo *bd* quinol oxidasa tras la exposición a compuestos aromáticos (Ramos *et al*., 2015). El funcionamiento de la cadena de transporte electrónico se ve afectado por la acumulación de compuestos aromáticos en la membrana y, además, en estas condiciones hay una mayor demanda de energía por parte de la célula. En respuesta a esta situación, parece que se produce una sobreexpressión de determinadas proteínas de la cadena de transporte electrónico (Ramos *et al*., 2015).

Por otra parte, debido a los daños provocados por la acumulación de compuestos aromáticos y al aumento de la demanda energética para poder hacer frente a estos compuestos, se produce un aumento de la producción de peróxido de hidrógeno y de especies reactivas del oxígeno que provocan daños a nivel de los ácidos nucleicos y las proteínas (Udaondo *et al*., 2012; Ramos *et al*., 2015). En respuesta a estas condiciones de estrés oxidativo, las células activan sistemas de reparación del ADN, chaperonas para poder plegar todas aquellas proteínas desnaturalizadas por la presencia de los compuestos aromáticos y mecanismos que evitan el estrés oxidativo, como el sistema de defensa a las especies reactivas del oxígeno (ROS), catalasas, peroxidasas, glutaredoxinas o superóxido dismutasa (SOD) (Udaondo *et al.*, 2012; Ramos *et al*., 2015).Todos estos mecanismos implicados en la tolerancia a solventes orgánicos

son muy costosos para la célula a nivel energético. Por ello, tras la exposición a solventes orgánicos en *P. putida* DOT-T1E además de cambios en la cadena de transporte electrónico se observó un aumento del consumo de glucosa y una disminución en la expresión de proteínas implicadas en el almacenamiento de azúcares y la gluconeogénesis (Udaondo *et al*., 2012; Ramos *et al*., 2015).

Los mecanismos de respuesta a solventes expuestos anteriormente, o al menos algunos de ellos, parecen ser comunes a otras bacterias Gram negativas como *Vibrio* o *Escherichia coli*, así como a Gram positivas como por ejemplo *Clostridium acetobutylicum* (Segura *et al*., 2012). Sin embargo, la tolerancia a hidrocarburos de miembros del linaje Roseobacter prácticamente no se ha explorado. Solo contamos con algunos resultados de tolerancia, como los cambios en el proteoma de *Ruegeria pomeroyi* DSS-3^T tras la exposición a naftaleno y diésel (Christie-Oleza *et al*., 2012), y los cambios en el proteoma de siete aislados de aguas de Mallorca tras la exposición a diésel (Mas-Lladó, 2017). En el experimento de Christie-Oleza y colaboradores (2012), observaron que tras la exposición a naftaleno de *Ruegeria pomeroyi* DSS-3^T se detectaba una mayor abundancia de ciertas proteínas que podían estar implicadas en la estabilización de la membrana, así como de bombas de expulsión y de una proteína que podía estar implicada en la destoxicación celular. En general, en el experimento de Mas-Lladó (2017), se observó la respuesta descrita en la literatura, que implicaba la expresión de proteínas para el mantenimiento de la estabilidad celular y el transporte de la membrana, proteínas implicadas en la reducción del estrés oxidativo y chaperonas. Sin embargo, a pesar de detectar la respuesta general de tolerancia a compuestos orgánicos, también se observaron diferencias en la respuesta de diferentes aislados del linaje Roseobacter (Mas-Lladó, 2017). En el contexto de esta tesis, y teniendo en cuenta la capacidad para el catabolismo de compuestos aromáticos (ver apartado 3) de miembros del linaje Roseobacter, resultaba de interés explorar posibles mecanismos de tolerancia a compuestos aromáticos.

HIPÓTESIS Y OBJETIVOS

El linaje Roseobacter es un grupo importante para el funcionamiento de los ecosistemas marinos por sus características fisiológicas y metabólicas. Uno de los ambientes donde predominan bacterias de este linaje son ambientes crónicamente contaminados por hidrocarburos. Esto indicaría, como mínimo, su capacidad para tolerar la presencia de estos contaminantes. A esto se añade la capacidad para la degradación de compuestos monoaromáticos extendida en los miembros cultivados del linaje (basada en predicción génica en la mayoría de los casos) y al reciente aislamiento de bacterias de este linaje degradadoras de hidrocarburos poliarómaticos.

En base a esta información se planteó la hipótesis de que las bacterias del linaje Roseobacter participarían en la degradación de compuestos aromáticos en ambientes crónicamente contaminados como por ejemplo aguas de puertos deportivos. Sin embargo, el hecho de que no hayan sido reconocidas como bacterias degradadoras de hidrocarburos podría ser debido a que tendrían estrategias de degradación, estructuras de genes catabólicos o de resistencia al estrés diferentes de los de las bacterias modelo de degradación estudiadas hasta el momento. Por lo tanto, en esta tesis se planteó analizar el potencial para el catabolismo de compuestos aromáticos de miembros del linaje Roseobacter como base para entender el potencial catabólico de comunidades microbianas de ambientes contaminados por hidrocarburos donde proliferan estas bacterias.

Para comprobar esta hipótesis se propusieron los siguientes objetivos:

1. Determinar la presencia y organización de los genes para rutas centrales para la degradación aeróbica de compuestos aromáticos, así como de rutas periféricas que convergen en esas rutas centrales, en genomas de bacterias del linaje Roseobacter.

2. Analizar los mecanismos implicados en la tolerancia a compuestos aromáticos que presentan aislados del linaje Roseobacter desde una aproximación proteómica.

3. Analizar mediante metagenómica el potencial catabólico para compuestos aromáticos en muestras ambientales contaminadas crónicamente con hidrocarburos o artificialmente en el laboratorio y relacionar este potencial catabólico con la presencia de miembros del linaje Roseobacter.

MATERIAL Y MÉTODOS

1. Predicción de genes para la degradación y tolerancia de compuestos aromáticos en genomas del linaje Roseobacter

Este estudio se ha realizado con información de 61 genomas del linaje Roseobacter (tabla S1.1). Catorce de estos genomas se encontraban cerrados mientras los 47 restantes se encontraban en boceto. Para determinar el potencial para el catabolismo de compuestos aromáticos se analizó la presencia y organización de 166 genes involucrados en rutas catabólicas centrales de compuestos monoaromáticos y de vías periféricas que confluyen en estos metabolitos centrales (tablas M1 y M2, ver detalle en la tabla S1.2). Para determinar posibles mecanismos de tolerancia a compuestos aromáticos (Ramos *et al*., 2002) se analizaron genes para 24 proteínas codificantes para las subunidades de seis bombas de tipo RND (tabla M3).

Tabla M1. Categorías funcionales de los genes involucrados en la degradación de los compuestos aromáticos estudiados. Las categorías se muestran organizadas según los compuestos aromáticos de rutas catabólicas centrales o en los que confluyen las rutas periféricas analizadas.

Compuesto	Genes catabólicos		Genes reguladores	Genes para transporte	Genes conservados/ accesorios[a]	Total
	Rutas centrales	Rutas periféricas				
Protocatecuato	16	26	9	3	3	57
Catecol	12	6	5	0	1	24
Homogentisato	3	3	3	0	5	14
Fenilacetil-CoA	11	5	4	2	1	23
Homoprotocatecuato	8	2	2	1	1	14
Gentisato	3	14	1	1	2	21
Benzoil-CoA	6	5	1	1	0	13
Total	59	61	25	8	13	166

a: Los genes conservados/accesorios son genes que se encuentran conservados junto a los genes implicados en la degradación de un compuesto aromático y cuya relación con la degradación de ese compuesto no se ha estudiado.

La predicción génica se hizo en base a la identidad de secuencia de proteínas con modelos de degradación y tolerancia a compuestos orgánicos descritos en la literatura. Los modelos usados para el estudio de la degradación y tolerancia de compuestos aromáticos fueron los que se muestran en las tablas M2 (para más detalle ver también la tabla S1.2.) y M3. Se obtuvieron las secuencias de las proteínas modelo de la base de datos del National Center for Biotechnology Information, NCBI (NCBI Resource Coordinators, 2017). Se realizaron análisis de identidad de secuencia de estas proteínas con las proteínas codificadas en los genomas de roseobacters estudiados usando el programa BLASTp (Basic Local Alignment Search Tool) (Altschul *et al*., 1997). De esta manera se localizaron en los genomas de roseobacters las secuencias génicas codificantes para las proteínas analizadas. El criterio para considerar un resultado como válido era que al menos presentara un 50 % de identidad con la proteína modelo en al menos un 50 % de la secuencia y con un *E-value* menor de 10^{-5}.

En algunos casos con este criterio de identidad de secuencia no se pudo detectar ninguna proteína. En estos casos se utilizó un criterio más relajado, que las proteínas presentaran al menos un 30 % de identidad con un *E-value* menor de 10^{-5} con la proteína modelo. Sin embargo, en este caso solo se dieron como positivos aquellos resultados en que la secuencia detectada en los genomas de roseobacters estuviera en un conjunto génico en el que la mayor

Tabla M2. Organismos modelo para la detección en genomas del linaje Roseobacter de los genes implicados en las rutas centrales y periféricas del catabolismo de compuestos aromáticos.

Vías catabólicas	**Organismo modelo**	**Referencias**
Beta-cetoadipato (Rotura *orto*)		
Protocatecuato	*Citreicella* sp. SE45	Buchan *et al*., 2004
	Pseudomonas putida KT2440	Jiménez *et al*., 2002
	Rhodococcus opacus 1CP	Eulberg *et al*., 1998
Catecol	*Pseudomonas putida* BIRD-1	Matilla *et al*., 2011
Protocatecuato (Rotura *meta*)	*Sphingomonas* sp. KA1	Shintani *et al*., 2007
Catecol (Rotura *meta*)	*Rhodobacter sphaeroides* 2.4.1.^{T}	Kontur *et al*., 2012
	Pseudomonas putida mt-2	Velázquez *et al*., 2005
Homogentisato	*Ruegeria pomeroyi* DSS-3^{T}	Moran *et al*., 2007
	Pseudomonas putida KT2440	Arias-Barrau *et al*., 2004
	Citreicella aestuarii 357	Suárez-Suárez, 2013
Fenilacetil-CoA	*Escherichia coli* W	Ferrández *et al*., 1998
	Pseudomonas putida U	Olivera *et al*., 1998
Homoprotocatecuato	*Escherichia coli* W	Díaz *et al*., 2001
	Brevibacterium fuscum var. *dextranlyticum*	Wang y Lipscomb, 1997
Gentisato	*Citreicella aestuarii* 357	Suárez-Suárez, 2013
	Ralstonia sp. U2	Zhou *et al*., 2001
Benzoil-CoA	*Azoarcus evansii* KB 740^{T}	Gescher *et al*., 2005
3-hidroxibenzoato (Vía gentisato)	*Polaromonas naphthalenivorans* CJ2	Park *et al*., 2007
3-hidroxibenzoato (Vía protocatecuato)	*Comamonas testosteroni* KH122-3s	Hiromoto *et al*., 2006
3-hidroxifenilacetato (Vía homogentisato)	*Aspergillus nidulans* FGSC A4	Ferrer-Sevillano y Fernández-Cañón, 2007
4-hidroxibenzoato (Vía benzoil-CoA)	*Thauera aromatica* K172	Breese *et al.,* 1998
4-hidroxibenzoato (Vía protocatecuato)	*Citreicella sp.* SE45	Buchan *et al*., 2004
4-hidroxibenzoil-CoA (Vía protocatecuato)	*Pseudomonas* sp. CBS3	Zhuang *et al*., 2002
4-hidroxifenilacetato (Vía homogentisato)	*Citreicella aestuarii* 357	Suárez-Suárez, 2013
4-hidroxifenilacetato (Vía homoprotocatecuato)	*Escherichia coli* W	Díaz *et al*., 2001

Tabla M2. Organismos modelo para la detección en genomas del linaje Roseobacter de los genes implicados en las rutas centrales y periféricas del catabolismo de compuestos aromáticos (continuación).

Vías catabólicas	Organismo modelo	Referencias
4-hidroxifenilpiruvato (Vía homogentisato)	*Pseudomonas fluorescens* F113	Redondo-Nieto *et al*., 2012
Benzoato (Vía benzoil-CoA)	*Azoarcus evansii* KB740^{T}	Gescher *et al*., 2005
Benzoato (Vía catecol)	*Acinetobacter* sp. ADP1	Collier *et al*., 1998
	Pseudomonas putida (plásmido TOL pWW0)	Ramos *et al*., 1997
Benzoato (Vía protocatecuato)	*Aspergillus niger* CBS 513.88	van Gorcom *et al*., 1990
Estireno (Vía fenilacetil-CoA)	*Pseudomonas* sp. VLB120	Panke et al., 1998
Fenilacetato (Vía fenilacetil-CoA)	*Escherichia coli* W	Díaz *et al*., 2001
Ferulato (Vía protocatecuato)	*Agrobacterium tumefaciens* C58	Campillo *et al*., 2014
	Pseudomonas fluorescens 2-79	Nesemann *et al*., 2015
	Pseudomonas putida KT2440	Belda *et al*., 2016
	Pseudomonas putida NBRC 14164^{T}	Ohji *et al*., 2014
Ftalato (Vía protocatecuato)	*Burkholderia cepacia* DB01	Chang y Zylstra, 1998
Isoftalato (Vía protocatecuato)	*Comamonas testosteroni* YZW-D	Wang *et al*., 1995
Naftaleno (Vía gentisato)	*Ralstonia* sp. U2	Zhou *et al*., 2001
Salicilato (Vía catecol)	*Pseudomonas stutzeri* AN10	Bosch *et al*., 1999a
Salicilato (Vía gentisato)	*Ralstonia* sp. U2	Zhou *et al*., 2001
	Streptomyces sp. WA46	Ishiyama *et al*., 2004
Tereftalato (Vía protocatecuato)	*Comamonas testosteroni* YZW-D	Wang *et al*., 1995
Vanillina (Vía protocatecuato)	*Pseudomonas putida* KT2440	Nelson *et al*., 2002
Vanillato (Vía protocatecuato)	*Pseudomonas putida* NBRC 14164^{T}	Ohji *et al*., 2014
	Agrobacterium tumefaciens C58	Campillo *et al*., 2014
Protocatecuato-catecol (interconversión)	*Micromonospora lupini* Lupac 08	Alonso-Vega *et al*., 2012

Tabla M3. Organismos modelo para la detección en genomas de Roseobacter de genes para bombas de tipo RND implicadas en tolerancia a disolventes orgánicos.

Tipo de bomba	Organismo modelo	Referencia
AcrAB-TolC	*Escherichia coli* str. K-12 substr. MG1655	Aono *et al*., 1998
SrpABC	*Pseudomonas putida* S12	Kieboom *et al*., 1998
TtgABC	*Pseudomonas putida* DOT-T1E	Ramos *et al*., 1998
TtgDEF	*Pseudomonas putida* DOT-T1E	Mosqueda y Ramos, 2000
TtgGHI	*Pseudomonas putida* DOT-T1E	Rojas *et al.,* 2001
MexAB-OprM	*Pseudomonas aeruginosa* PAO1	Dreier y Ruggerone, 2015

parte de los genes se hubieran detectado con el criterio más estricto. También se consideraron aquellos casos en que el gen detectado fuera el único o uno de los pocos genes implicados en una vía de degradación de compuestos aromáticos que no hubiera sido detectado con el criterio más estricto, siempre y cuando el resto de genes de la ruta hubiesen sido detectados con fiabilidad. Una vez detectados los genes con confianza en algún miembro del linaje Roseobacter, se utilizó esta secuencia proteica para buscar los genes en el resto de genomas del linaje Roseobacter (ver tabla S1.2). En caso de no poder detectar con confianza alguno de los genes buscados en ningún miembro del linaje Roseobacter se siguió utilizando el gen del modelo externo a Roseobacter para detectar esta proteína en los miembros del linaje Roseobacter.

2. Cultivos de aislados del linaje Roseobacter

2.1. Determinación de la capacidad de crecimiento con compuestos aromáticos como fuente de carbono y energía

Para varios aislados de Roseobacter cuyos genomas estaban incluidos en el estudio se analizó la capacidad de crecimiento en medio mineral con varios compuestos monoaromáticos como única fuente de carbono y energía. Los cultivos se realizaron en medio mineral marino (MMMb) con la siguiente composición final: sales marinas (Sigma-Aldrich) 4 % (p/v); Tris-HCl 0,1 M pH 7,4; Na_2HPO_4 6,33 µM, NH_4Cl 56,09 µM; citrato férrico amónico 0,03 µM; extracto de levadura 0,05 % (p/v). Se utilizaron los siguientes compuestos aromáticos: protocatecuato, 4-hidroxibenzoato, salicilato, gentisato, homogentisato, benzoato, fenilacetato, 4-hidroxifenilacetato, homoprotocatecuato y catecol a una concentración final de 3 mM; y dibenzotiofeno y naftaleno a una concentración final de 0,1 % (p/v). Las disoluciones concentradas de los compuestos solubles en agua (gentisato, protocatecuato, 4-hidroxibenzoato, homogentisato, benzoato, salicilato, catecol y succinato) se prepararon con agua milliQ (Millipore) y se esterilizaron por filtración con filtros de celulosa de 0,22 µm de tamaño de poro (Nalgene). Las disoluciones concentradas de fenilacetato, 4-hidroxifenilacetato, naftaleno, dibenzotiofeno y homoprotocatecuato se prepararon con acetona y se filtraron con filtros de teflón de 0,22 µm (Millipore). En el caso de los compuestos aromáticos disueltos en acetona se añadió la cantidad necesaria a tubos estériles y se evaporó el disolvente antes de añadir el medio de cultivo y el inóculo. En todos los casos se realizaron cultivos previos de las bacterias en medio MMMb con succinato al 0,5 % (p/v) como fuente de carbono, que se incubaron a temperatura ambiente con agitación constante (180 rpm) durante 72 h. Una vez se alcanzaba la fase exponencial de crecimiento (DO_{600} 0,4–0,8) se hacía un inóculo 1/100 en tubos que contenían medio MMMb y cada uno de los compuestos

aromáticos ensayados. También se prepararon controles negativos con medio MMMb y cada uno de los compuestos ensayados sin inocularlos con bacterias y tubos con medio MMMb inoculados pero sin fuente de carbono añadida. Todos los experimentos se realizaron por duplicado. Los cultivos se incubaron 3 semanas a temperatura ambiente y en agitación (180 rpm) tiempo durante el cual se realizaron medidas de densidad óptica a una longitud de onda de 600 nm.

2.2. Cultivos para el análisis proteómico preliminar de tolerancia a naftaleno y tolueno

Los análisis preliminares de tolerancia a tolueno se realizaron con 9 aislados procedentes de aguas de Mallorca (*Sulfitobacter mediterraneus* 1FIGIMAR09, *Sulfitobacter* sp. 3SOLIMAR09, *Loktanella* sp. 1ANDIMAR09, *Ruegeria* sp. 6PALISEP08, *Thalassobacter stenotrophicus* 16PALIMAR09 y 1CONIMAR09, *Loktanella* sp. 5RATIMAR09 y 3ANDIMAR09, *Pseudophaeobacter* sp. 11ANDIMAR09). También se incluyeron las cepas *Ruegeria pomeroyi* DSS-3^T y *Citreicella aestuarii* 357, esta última descrita como degradadora de naftaleno y dibenzotiofeno (Suárez-Suárez, 2013). Los análisis preliminares de tolerancia a naftaleno se realizaron con *Citreicella aestuarii* 357 y *Sulfitobacter* sp. 3SOLIMAR09. No se incluyó la cepa DSS-3^T porque ya había evidencias previas de que se producían cambios en su proteoma tras la exposición a naftaleno (Christie-Oleza *et al*., 2012). Estos análisis preliminares de cambios en el proteoma se realizaron por electroforesis SDS-PAGE (ver apartado 3.1).

Para el estudio de tolerancia a tolueno se crecieron las bacterias por triplicado en 50 ml de caldo marino (MB) (Cultimed) en Erlenmeyers de 100 ml a 30 °C y en agitación (180 rpm) hasta alcanzar la fase exponencial (DO_{600} 0,4–0,8). A continuación los cultivos se dividieron en alícuotas de 6 ml en tubos de cristal con tapón de teflón y se hicieron tratamientos con dos concentraciones distintas de tolueno (0,1 % y 0,3 % v/v, concentración final). El tolueno fue previamente esterilizado mediante filtración con filtros de teflón de 0,22 µm (Millipore). Estas concentraciones fueron las utilizadas por Segura y colaboradores (2007) para analizar la tolerancia a tolueno de varias cepas marinas. Tras 30 minutos, 1 y 3 h de incubación con tolueno se tomaron 3 ml de los cultivos, se recogieron las células por centrifugación a 15.871 x g durante 5 minutos usando una centrífuga 5424 R (Eppendorf) y se congelaron a –20 °C hasta su posterior análisis.

En los tratamientos con naftaleno las bacterias se crecieron por triplicado en 100 ml de MB en Erlenmeyers de 300 ml como se ha indicado anteriormente. A continuación, se trasvasaron 30 ml del cultivo con MB a Erlenmeyers de 100 ml estériles que contenían naftaleno. De esta forma, tras el pase de recipiente, cada cultivo con MB tenia una concentración de 0,1 % (p/v) de naftaleno. Las células se recogieron tras 30 minutos, 1, 3, 7 y 24 h de exposición al naftaleno como se ha mencionado anteriormente.

2.3. Cultivos para el análisis de tolerancia a tolueno y naftaleno por proteómica masiva (*shot-gun*)

En base a los resultados obtenidos en los perfiles proteicos con SDS-PAGE (ver apartado 3.1) se seleccionaron las bacterias *Ruegeria pomeroyi* DSS-3^T, *Sulfitobacter* sp. 3SOLIMAR09 y *Citreicella aestuarii* 357 para los análisis de tolerancia a naftaleno y tolueno mediante proteómica masiva. En ambos casos, se crecieron estas bacterias por duplicado en 120 ml de

MB en Erlenmeyers de 300 ml a 30 °C en agitación (180 rpm) hasta alcanzar la fase exponencial (DO_{600} 0,4–0,8). Se mezclaron los dos cultivos, se dividió el volumen en 6 alícuotas de 30 ml en botellas de 100 ml con tapón de teflón en el caso del experimento de tolerancia a tolueno y en Erlenmeyers de 100 ml en el caso del experimento de tolerancia a naftaleno. A tres de los cultivos se les añadió el tratamiento [tolueno 0,3 % (v/v) o naftaleno 0,1 % (p/v)] tal y como se ha descrito anteriormente. Se realizaron incubaciones de 3 h en el experimento con tolueno y 24 h en el experimento con naftaleno. Pasado este tiempo se tomaron muestras de 10 ml de cultivo y se recogieron las células por centrifugación a 15.871 x g durante 5 minutos usando una centrífuga 5810 R (Eppendorf). Las muestras se congelaron a –80 °C y se enviaron congeladas en hielo seco a la *School of Life Sciences* de la Universidad de Warwick (Reino Unido) para su posterior análisis.

Durante la incubación con tolueno y naftaleno se determinó la densidad óptica de los cultivos y se sembró cada cultivo por estría en agar marino (Cultimed) y en agar Luria Bertani (Conda) para determinar su capacidad de crecimiento tras la exposición a los hidrocarburos y la pureza de los cultivos. En el caso de *R. pomeroyi* DSS-3^T tras 24 h de exposición a naftaleno también se sembraron diluciones de los cultivos en placas de agar marino para determinar el número de viables tras la exposición a naftaleno. Todas las placas se incubaron de 72 a 96 h a 30 °C.

2.4. Cultivos del mutante de *R. pomeroyi* DSS-3^T defectivo para el gen codificante de la proteína SPO2567

Se realizaron cultivos líquidos del mutante defectivo para SPO2567, SPO2567$^-$ (ver apartado 5) y de *R. pomeroyi* DSS-3^T silvestre (wt) en MB de forma simultánea para determinar posibles diferencias en el crecimiento en este medio. Las dos cepas se crecieron por duplicado en 120 ml de MB en Erlenmeyers de 300 ml a 30 °C en agitación (180 rpm) y se realizaron lecturas de su densidad óptica hasta alcanzar la fase estacionaria. También se analizó la viabilidad del mutante SPO2567$^-$ en comparación con la cepa de *R. pomeroyi* DSS-3^T silvestre tras exposición a naftaleno (0,1 % p/v). Para ello se crecieron las dos cepas por duplicado tal como se detalla en el apartado 2.3 durante 48 h. A las 24 y 48 h de incubación se realizaron medidas de densidad óptica de los cultivos y se recogieron células por centrifugación como se ha descrito anteriormente (3 ml de cultivo). Se determinó la biomasa recogida en cada tubo (peso húmedo) y se hicieron recuentos de viables en MB.

3. Procesamiento de muestras para análisis proteómico

3.1. Fraccionamiento y electroforesis SDS-PAGE en los experimentos preliminares

Para los análisis de proteómica se realizó un experimento preliminar en el que se utilizó un método de baja resolución (SDS-PAGE) para ver cambios en las proteínas expresadas por los diferentes aislados tras la exposición a tolueno y naftaleno. El objetivo de estos experimentos era seleccionar las bacterias que mostrasen cambios más relevantes para un posterior análisis de alta resolución por proteómica masiva. Para simplificar los patrones de proteínas, se estudió el proteoma de los aislados en dos fracciones: una fracción enriquecida en proteínas citoplasmáticas y una enriquecida en proteínas de membrana. Para obtener las distintas fracciones se resuspendieron los precipitados celulares de cultivos realizados como se detalla en apartado 2.2 con 1 ml Tris-HCl 50 mM pH 8,0 y se recogieron las células por centrifugación a

15.871 x g durante 5 minutos usando una centrífuga 5424 R (Eppendorf). A continuación, se resuspendendieron los precipitados en 800 µl Tris-HCl 50 mM pH 8,0 y se sonicaron las células durante 2 minutos usando pulsos de ultrasonidos de 5 segundos con intervalos de 10 segundos (MSE Soniprep 150). Una vez obtenidos los lisados celulares, se centrifugaron a 3.381 x g usando una centrífuga 5424 R (Eppendorf) durante 2 minutos para descartar restos celulares. Los sobrenadantes se recogieron en tubos limpios. Finalmente, se centrifugaron estos sobrenadantes a 15.871 x g durante 30 minutos usando una centrífuga 5424 R (Eppendorf). Los precipitados obtenidos contenían la fracción enriquecida en proteínas de membrana mientras los sobrenadantes contenían la fracción enriquecida en proteínas citoplasmáticas. Todas las fracciones se conservaron a –20 °C hasta su procesamiento.

En el caso del estudio de la fracción enriquecida en proteínas de membrana, se resuspendió cada precipitado en 15 µl de H_2O milliQ estéril y 15 µl de tampón de carga 2X [0,125 M Tris-HCl, 4 % SDS (p/v), 20 % glicerol (v/v), 100 mM ditiotreitol (DTT), pH 6.8]. En la fracción enriquecida en proteínas citoplasmáticas se mezclaron 15 µl del sobrenadante recogido con 15 µl de tampón de carga 2X. Se incubaron las muestras a 100 °C durante 5 minutos y se centrifugaron a 15.871 x g durante 5 minutos usando una centrífuga 5424 R (Eppendorf). Los sobrenadantes se cargaron en geles de SDS-PAGE con una zona de apilamiento al 5 % y una zona de resolución al 10 % de acrilamida/bisacrilamida (BioRad). Se utilizó el marcador de peso molecular Precision Plus Protein™ Dual Color Standards (BioRad). Las electroforesis se corrieron a una intensidad de corriente de 10 mA por gel durante aproximadamente 3 h en el sistema Mini Protean Tetra Cell (Bio-Rad) con tampón de electroforesis [0,025 M Tris-HCl, 0,192 M glicina, 0,1 % SDS (p/v) pH 8,3]. Los geles se tiñeron con azul Coomassie [0,125 % (p/v) Coomassie Brilliant Blue R250 (Fluka), 45 % metanol (v/v) y 10 % ácido acético (v/v) en agua milliQ] *overnight* (o/n) y se destiñeron mediante varios lavados con decolorante [45 % metanol (v/v) y 10 % ácido acético (v/v) en agua destilada] y finalmente en una disolución decolorante al 25 % metanol (v/v) y 7 % ácido acético (v/v) en agua destilada.

En el caso del mutante SPO2567$^-$de *R. pomeroyi* DSS-3^T, los precipitados celulares se resuspendieron en 2 volúmenes (2 µl por cada mg de biomasa, peso húmedo) de tampón de carga 2X. Se lisaron las células con 3 ciclos de incubación a 90 °C durante 5 minutos seguidos de 5 minutos de sonicación en un baño de ultrasonidos (Selecta P 3000513 Ultrasonic Cleaning Bath). Los lisados se cargaron directamente en geles de poliacrilamida como se ha indicado anteriormente y se corrieron a 10 mA por gel durante aproximadamente 2 h.

3.2. Procesamiento de muestras para proteómica masiva

Para el estudio de tolerancia al naftaleno se analizó el proteoma total y la fracción enriquecida en proteínas de membrana preparada como se ha descrito anteriormente. En el caso de los cultivos tratados con tolueno se analizó solo la fracción enriquecida en proteínas de membrana debido a la disponibilidad de tiempo de análisis proteómico en la *School of Life Sciences* de la Universidad de Warwick.

Todas las fracciones proteicas se resuspendieron en 2 volúmenes (2 µl por cada mg de biomasa, peso húmedo) de NuPAGE® LDS Sample Buffer (Invitrogen). Se realizaron 3 ciclos de incubación a 95 °C durante 5 minutos seguidos de 5 minutos de sonicación en un baño de ultrasonidos (Branson 1210E-MTH Ultrasonic Cleaner) para asegurar la disolución de las

fracciones proteicas. Los lisados se cargaron directamente en geles NuPAGE® 10 % Bis-Tris Gel (Invitrogen) y se corrieron a 200 V en un equipo Mini Protean Tetra Cell (Bio-Rad) durante 10 minutos para una migración corta. La electroforesis se realizó usando el tampón NuPAGE® MOPS SDS Running Buffer (Invitrogen). Se tiñeron los geles con Coomassie Blue G-250 SimplyBlue SafeStain™ (Invitrogen) para observar las bandas en las que se encontraban las fracciones proteicas. A continuación, se cortaron las bandas de los genes, y a su vez, cada banda fue cortada en fragmentos pequeños que se guardaron a –20 °C hasta su posterior procesado.

Los fragmentos de gel se descongelaron y se lavaron dos veces durante 20 minutos en agitación (600 rpm) con bicarbonato de amonio 50 mM en etanol 50 % (v/v) a 55 °C. A continuación se deshidrataron incubándolos en etanol absoluto a 55 °C en agitación (600 rpm) durante 5 minutos, repitiendo este proceso varias veces hasta la completa deshidratación. A continuación las muestras se prepararon para la digestión proteica con tripsina. En primer lugar, para reducir los puentes disulfuro de las cisteínas se añadió una disolución de DTT 10 mM en bicarbonato de amonio 50 mM a los fragmentos de gel deshidratados y se incubaron entre 30 y 45 minutos a 56 °C con agitación (600 rpm). A continuación se alquilaron los residuos de cisteína mediante incubación en iodoacetamida 55 mM disuelta en bicarbonato de amonio 50 mM durante 20–30 minutos a temperatura ambiente y en oscuridad. Se lavaron y deshidrataron los fragmentos de los geles tal como ya se ha descrito anteriormente y después de ello se realizó la digestión de las proteínas añadiendo 100 ng de tripsina (Sigma-Aldrich) disuelta en bicarbonato de amonio 50 mM e incubando a 37° C o/n en agitación (600 rpm). La digestión se paró añadiendo 5 % (v/v) de ácido fórmico en acetonitrilo 25 % (v/v) y sonicando las muestras durante 10 minutos (Fisherbrand FB15062 Ultrasonic Bath) para extraer los péptidos digeridos de los fragmentos de gel. Se transfirió el sobrenadante a nuevos tubos y se repitió este procedimiento tres veces para poder extraer el mayor número de péptidos de los fragmentos del gel. Los sobrenadantes que contenían los péptidos digeridos se secaron mediante centrifugación en vacío (miVac Duo Concentrator, GeneVac) durante al menos 5 h a 40 °C. Los péptidos se conservaron a –20 °C hasta su posterior análisis.

Para el análisis por espectrometría de masas se descongelaron los péptidos, se resuspendieron en una disolución de acetonitrilo 2,5 % v/v con ácido trifluoroacético 0,05 % v/v, se sonicaron en un baño de ultrasonidos (Fisherbrand FB15062 Ultrasonic Bath) durante 30 minutos para solubilizar los péptidos y se centrifugaron a 17.000 x g durante 5 minutos (Heraus Fresco 21, Thermo Scientific). Se analizaron alícuotas de 20 µl de los sobrenadantes, que contenían los péptidos digeridos, mediante cromatografía líquida acoplada a espectrometría de masas (LC-MS/MS) por los técnicos del servicio de proteómica de la *School of Life Sciencies* de la Universidad de Warwick (Reino Unido). Se realizó una cromatografía en fase inversa para separar los péptidos antes de su análisis en el espectrómetro de masas utilizando dos columnas: una precolumna PepMap µ-precolumn cartridge 300 µm i.d. x 5 mm 5 µm 100 Å y una columna Acclaim PepMap RSLC 75 µm x 25 cm 2 µm 100 Å (Thermo Scientific). Las columnas se instalaron en un sistema Ultimate 3000 RSLCnano system (Dionex). Se utilizaron dos fases móviles: el tampón de la fase A estaba compuesto por ácido fórmico al 0,1 % v/v en agua mientras que el tampón de la fase B estaba compuesto por ácido fórmico al 0,1 % v/v en acetonitrilo. Las muestras se cargaron en la precolumna equilibrada en una disolución acuosa de acetonitrilo (2 % v/v) que contenía ácido trifluoroacético al 0,1 % (v/v) durante 8 minutos y

un flujo de 10 µl/min. A continuación se eluyeron los péptidos en la columna analítica utilizando un flujo 300 nl/min incrementando la concentración de la fase B del 3 al 25 % durante 90 minutos, a continuación pasando de 25 % a 35 % de B durante 10 minutos, luego se incrementó hasta el 90 % de fase B durante 5 minutos y finalmente se reequilibró durante 15 minutos hasta llegar a un 3 % de fase B. Una vez eluídos los péptidos estos se ionizaron [Triversa Nanomate nanospray source (Advion Biosciences)] y se resolvieron en un espectrómetro de masas Thermo Orbitrap Fusion (Q-OT-qIT, Thermo Scientific). Las muestras de proteoma total se resolvieron durante 2 h mientras las muestras de fraccionamientos se analizaron durante 1 hora.

4. Análisis de los datos de proteómica masiva

La identificación y cuantificación relativa de las proteínas de cada muestra se realizó con el programa Progenesis QI for proteomics 2.0 (Nonlinear Dynamics). El software designó uno de los espectros de masas de cada experimento como espectro de referencia y el resto de espectros de cada experimento se alinearon en referencia a ese espectro. La intensidad de todos los picos fue normalizada en base a la intensidad total. A continuación se obtuvo una lista de picos de péptidos con sus valores de masa/carga y tiempos de retención. Solo se seleccionaron aquellos picos que tuvieran una carga entre +2 y +4 descartando así todos los péptidos que no habían sido bien digeridos por la tripsina. Todos los péptidos seleccionados se identificaron mediante el programa Mascot versión 2.5.0 (Matrix Science) en base al espectro teórico del proteoma de cada microorganismo obtenido a partir de la digestión virtual con tripsina de las proteínas anotadas en su genoma. Los péptidos identificados se volvieron a importar a Progenesis, se realizó una cuantificación relativa usando todos los péptidos identificados con Mascot y se exportaron los resultados para su análisis. A continuación, se cribaron los resultados para conservar solo aquellas proteínas de las que al menos se habían detectado dos péptidos no redundantes, con un valor de probabilidad ANOVA (p) menor de 0,05 y que al menos presentaban una variación de abundancia relativa en un factor igual o superior a 2x (mayor o menor) entre la condición control y el tratamiento con hidrocarburos. Para poder comparar los valores de abundancia relativa para cada proteína se dividió el valor de abundancia normalizada que proporcionaba el programa Progenesis entre el tamaño de la proteína a la que correspondia cada péptido detectado. Con esta normalización adicional se obtuvo el porcentaje de abundancia relativa normalizada de cada péptido en función de la longitud de la proteína a la que se había asignado.

Para poder determinar si se producía una respuesta común de *R. pomeroyi* DSS-3^T, *Sulfitobacter* sp. 3SOLIMAR09 y *C. aestuarii* 357 a la exposición a naftaleno y tolueno se realizaron agrupaciones de las proteínas consideradas como significativas con los criterios mencionados anteriormente. Estas agrupaciones se realizaron con CD-HIT (Huang *et al*., 2010) en base a la identidad de secuencia, generando así agrupaciones de proteínas que al menos presentaran un 50 % de identidad en al menos el 50 % de su secuencia. Una vez agrupadas las proteínas se construyó una matriz con la suma de las abundancias relativas normalizadas de cada grupo de proteínas en cada condición. Se realizó un análisis de componentes principales (PCA) utilizando el programa PAST v. 3.08 (Hammer *et al*., 2001) para comparar globalmente los proteomas y determinar grupos de proteínas que diferenciaban la respuesta a naftaleno o tolueno de los 3 organismos estudiados de las condiciones control.

Por otra parte, todas las proteínas consideradas estadísticamente significativas en cada experimento se clasificaron en categorías funcionales usando el servidor de anotación automático BlastKOALA de KEGG (Kanehisa *et al*., 2016b). Se compararon las abundancias relativas normalizadas de las proteínas de cada categoría para determinar la respuesta funcional al naftaleno o tolueno. Además de estudiar los cambios funcionales que se daban en los proteomas de los tres organismos tras la exposición a naftaleno también se estudiaron en detalle las proteínas individuales que formaban parte de la respuesta de cada microorganismo a la exposición a naftaleno. Para poder entender la función de cada una de las proteínas detectadas que habían pasado los criterios de cribado se utilizó la asignación funcional de KEGG (Kanehisa *et al*., 2016a). En el caso de todas aquellas proteínas que no pudieron ser asignadas funcionalmente por KEGG o bien habían sido asignadas a categorías generales, se realizó una búsqueda manual de la función de estas proteínas en base a a su anotación en los genomas.

5. Generación de un mutante de *R. pomeroyi* DSS-3^T defectivo para el gen codificante de la proteína SPO2567

Christie-Oleza y colaboradores (2012) propusieron que la proteína SPO2567 de *R. pomeroyi* DSS-3^T podía estar relacionada con la estabilización de la membrana tras la exposición a naftaleno, diésel y elevadas temperaturas. Para estudiar el posible papel de esta proteína se generó un mutante defectivo para el gen codificante de esta proteína mediante el método Gibson (Gibson *et al*., 2009).

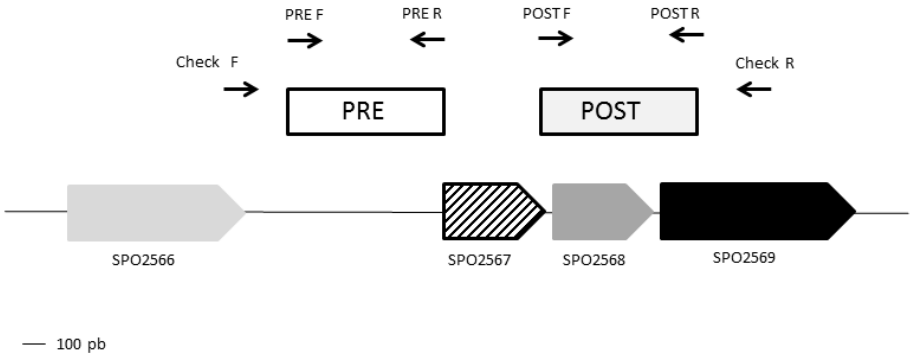

Figura M1. Esquema del fragmento de cromosoma de *R. pomeroyi* DSS-3^T que contiene el gen codificante para la proteína SPO2567. Se muestran también los fragmentos PRE y POST usados en la clonación y la posición donde hibridan los diferentes cebadores, indicados con flechas.

La clonación se realizó en el vector pk18SmobSacB (Schäfer *et al*., 1994), portador de un casete de resistencia a kanamicina (Km). Para generar el mutante defectivo en SPO2567 de *R. pomeroyi* DSS-3^T se amplificaron por PCR dos fragmentos de aproximadamente 700 pb corriente arriba (fragmento PRE) y corriente abajo del gen de interés (fragmento POST), que solapaban 5 pb con el principio y 44 pb con el final del gen SPO2567 (ver figura M1). Se amplificó también un gen de resistencia a gentamicina (Gm) del plásmido p34S-Gm (Dennis y Zylstra, 1998) para posteriormente poder seleccionar el mutante. En cuanto a los cebadores necesarios para la construcción del mutante (tabla M4), estos se diseñaron con la herramienta NEBuilder Assembly Tool de New England BioLabs (http://nebuilder.neb.com).

Tabla M4. Cebadores utilizados para la construcción del mutante SPO2567^{-} de *R. pomeroyi* DSS-3^{T}.

Cebador	Tm (°C)	mol % G+C	Secuencia 5'→ 3'
PRE F	88,2	49,0	TTTCACACAGGAAACAGCTATGACATGATTACGCGTCGTCAGAGCTGTTGGGC
PRE R	77,9	41,0	AACAGTTTTTATGCAGGTCATTTTGGGTCTCCAGATCAG
Gm F	76,5	34,0	AGACCCAAAATGACCTGCATAAAAACTGTTGTAATTCATTAAGC
Gm R	87,2	57,1	CTTGCTGCTGCGCTCGGCGGCGTTGTGACAATTTA
POST F	93,0	68,5	TGTCACAACGCCGCCGAGCGCAGCAGGCAAGGGTCG
POST R	92,8	66,0	ACTCTAGAGGATCCCCGGGTACCGAGCTCGTTCCGCTCCATCCTGCCCGC
Check F	69,5	65,0	GGGTTCAGTGCTGACCTCGC
Check R	81,8	75,0	AGGCGCGGCAATCGGCGCAG
pk18 F	40,0	50,0	GTTGTGTGGAATTGTGAGCG
pk18 R	42,0	58,0	AGTGCCAAGCTTGCATGCC

Los fragmentos PRE, POST y Gm se amplificaron en tres reacciones de PCR independientes. Las reacciones de PCR contenían: MyTaq^{TM} Mix (Bioline) y los correspondientes cebadores *forward* y *reverse* (ver tabla M4) a una concentración de 0,4 µM. El ciclo de PCR usado fue: 95 °C 1 min, seguido de 2 ciclos a 95 °C 15 s, 58 °C 15 s, 72 °C 1 min, y a continuación 30 ciclos a 95 °C 15 s, 70 °C 15 s y 72 °C 1 min. Finalmente se realizó una extensión durante 3 min a 72 °C. Para comprobar que los fragmentos amplificados se correspondían con los tamaños esperados estos se resolvieron mediante electroforesis en gel de agarosa al 1 % p/v (Hi-Res Standard agarose, Geneflow) a 100 V en tampón TBE 1X (concentración final Tris 0,1M, ácido bórico 0,1M y 0,002M de Na_2EDTA en agua destilada) usando el marcador 1 kb DNA Ladder (New England Biolabs) como referencia. Los productos de PCR se purificaron con el QIAquick PCR Purification Kit (Qiagen) y se determinó la concentración de ADN con un espectrofotómetro Nanodrop 2000c (Thermo Scientific). A continuación se siguió el protocolo NEBuilder HiFi DNA Assembly Reaction (New England Biolabs) para una realizar una construcción con 4 fragmentos de ADN: el plásmido pk18SmobSacB (Schäfer *et al*., 1994), los fragmentos PRE y POST, y el gen de resistencia a gentamicina, que se insertaría entre los fragmentos PRE y POST. Esta construcción se denominó pk18mobSacB-PRE-Gm-POST y se clonó en células competentes NEB® 5-alpha competent *Escherichia coli* cells (New England BioLabs) mediante choque térmico siguiendo las especificaciones del fabricante. La selección de los transformantes que habían adquirido el plásmido pk18mobSacB-PRE-Gm-POST se realizó en placas de agar Luria Bertani (LB) (Sigma) con gentamicina (concentración final en placa de 0,01 mg/ml) y kanamicina (concentración final en placa de 0,05 mg/ml). Además se comprobó que los transformantes tuvieran el plásmido pk18mobSacB-PRE-Gm-POST amplificando por PCR con los cebadores PRE F y POST R (figura M1 y tabla M4). La reacción de PCR se realizó como se ha descrito anteriormente. El programa de PCR utilizado fue: 95 °C 1 min; 30 ciclos de 95 °C 15 s, 70 °C 15 s, 72 °C 2 min, seguido de una extensión final de 3 min a 72 °C.

Una vez comprobado que los clones de *E. coli* tenían la construcción correcta se purificó el plásmido pk18mobSacB-PRE-Gm-POST con el kit QIAprep Spin Miniprep Kit (Qiagen) y se transfirió por transformación a células competentes de *E. coli* S17 λ_{pir} (Simon *et al*., 1983). Las células competentes se prepararon a partir del inóculo de una colonia en 30 ml de medio SOB (triptona 20 g/l, extracto de levadura 5 g/l, NaCl 10 mM, KCl 6,7 mM, $MgCl_2$ 10 mM, $MgSO_4$ 10 mM) en un Erlenmeyer de 250 ml incubado o/n a 30 °C en agitación (180 rpm). A continuación,

se añadieron 8 ml de este cultivo a un Erlenmeyer de 2 l que contenía 200 ml de medio SOB y se incubó a 30 °C en agitación (180 rpm) hasta alcanzar una DO_{600} de 0,3. Se dividió el cultivo en 4 tubos estériles de 50 ml y se incubaron en hielo durante 15 minutos. Se centrifugaron los tubos a 3.000 x g durante 15 minutos a 4 °C en una centrífuga 5810 R (Eppendorf) y se eliminó el sobrenadante. Se resuspendió el precipitado de cada tubo en 16 ml del tampón de transformación 1 (RbCl 100 mM, $MnCl_2$ x $4H_2O$ 60 mM, KCH_3CO_2 30 mM, $CaCl_2$ x $2H_2O$ 10,2 mM, glicerol 150 ml/l) que se encontraba refrigerado, se agitó de forma suave cada tubo y se incubaron en hielo durante 15 minutos. Se volvieron a centrifugar los tubos a 3.000 x g durante 15 minutos a 4 °C en una centrífuga 5810 R (Eppendorf), se eliminó el sobrenadante y se resuspendió el precipitado en 4 ml del tampón de transformación 2 (RbCl 10 mM, MOPS 10 mM, $CaCl_2$ x $2H_2O$ 74,8 mM, glicerol 150 ml/l) el cual también se encontraba refrigerado y finalmente las alícuotas se guardaron a –80 °C.

La transferencia del plásmido pk18mobSacB-PRE-Gm-POST a las células competentes *E. coli* S17 λ_{pir} se llevó a cabo siguiendo un protocolo de transformación por choque térmico. Para ello, se descongeló una alícuota con 300 µl de células competentes de *E. coli* S17 λ_{pir} en hielo y se añadieron 2 µl del plásmido pk18mobSacB-PRE-Gm-POST purificado. Tras 40 minutos de incubación las células se expusieron a un choque térmico de 42 °C durante 45 segundos y se incubaron en hielo durante 20 minutos. A continuación, se añadió 1 ml de caldo LB y se incubaron las células a 30 °C en agitación (180 rpm) durante una hora. Los transformantes se seleccionaron en placas de agar LB con gentamicina (concentración final en placa de 0,01 mg/ml) y kanamicina (concentración final en placa de 0,05 mg/ml) que se incubaron a 30 °C o/n. Se seleccionaron 6 colonias de transformantes y se comprobó por PCR que hubieran adquirido el plásmido pk18mobSacB-PRE-Gm-POST, para lo cual se utilizaron los cebadores pk18F y pK18R (ver tabla M4). Los cebadores pk18F y pK18R hibridan con dos zonas del plásmido pk18mobSacB, y por tanto, solo hibridan en los transformantes, que son los que han incorporado el plásmido pk18mobSacB-PRE-Gm-POST. Las condiciones de reacción fueron las mencionadas anteriormente y el ciclo de amplificación fue: 95 °C 3 min; 30 ciclos de 95 °C 30 s, 60 °C 30 s y 72 °C 1 min, con una extensión final de 5 min a 72 °C. Para confirmar que el plásmido pk18mobSacB-PRE-Gm-POST se encontrase en las células de *E. coli* S17 λ_{pir} se purificó el plásmido con el kit QIAprep Spin Miniprep Kit (Qiagen) y se secuenció por el método Sanger en los servicios de secuenciación de la Universidad de Warwick.

Finalmente, se transfirió el plásmido pk18mobSacB-PRE-Gm-POST por conjugación desde la cep*a E. coli* S17 λ_{pir} a *R. pomeroyi* DSS-3^T. La cepa receptora *Ruegeria pomeroyi* DSS-3^T se creció en MB (Difco) a 30 °C durante 2 días, mientras que la cepa donadora *E. coli* S17λ_{pir} con el plásmido pk18mobSacB-PRE-Gm-POST se creció en caldo LB a 30 °C o/n. Una vez crecidos los dos cultivos se mezclaron 5 ml del cultivo de la cepa donadora con 10 ml del cultivo de la cepa receptora en un tubo de 50 ml estéril y se centrifugó a 3.220 x g durante 15 minutos en una centrífuga 5810 R (Eppendorf). Se descartó el sobrenadante por decantación, y se resuspendió el precipitado de células con el residuo de sobrenadante que quedaba en el tubo (aproximadamente 1 ml). Se depositaron las células en el centro de una placa de agar marino (MA), se dejó secar la placa y se incubó o/n a 30 °C. El siguiente día se recogieron las células del centro de la placa con un asa de siembra, se resuspendieron en 320 µl de medio mineral marino (MMM) (Lanfranconi *et al*., 2010) y se sembraron varios volúmenes de muestra en placas de MMM con succinato al 0,5 % (p/v) como fuente de carbono y Gm (concentración

final en placa de 0,01 mg/ml). Las placas se incubaron a 30 °C durante 4 días. Se seleccionaron 50 colonias crecidas en estas placas y se sembraron en MA con Gm (concentración final en placa de 0,01 mg/ml) o con Km (concentración final en placa de 0,05 mg/ml), incubándose a 30 °C durante 3 días. En las placas de agar marino con Gm podía crecer tanto el mutante de *R. pomeroyi* DSS-3^T como la cepa donadora mientras que en las placas con Km solo podía crecer la cepa donadora *E. coli* S17λ_{pir} con el plásmido pk18mobSacB-PRE-Gm-POST. Por ello seleccionamos todas aquellas colonias que crecían con Gm pero no con Km. Para comprobar estos mutantes se amplificó por PCR con los cebadores Check F y Check R (tabla 3.1) la región donde se había realizado la mutación. Las condiciones de reacción fueron las descritas anteriormente. El ciclo de PCR usado fue: 95 °C 3 min; 30 ciclos de 95 °C 30 s, 64 °C 30 s y 72 °C 3 min, con una extensión final de 5 min a 72 °C. Se comprobó mediante electroforesis en gel de agarosa (1 % p/v en TBE) que el tamaño del fragmento amplificado en los mutantes de *R. pomeroyi* DSS-3^T se correspondiese con el esperado, que debería ser de 2.381 pb. Finalmente, se secuenció por el método de Sanger el producto de PCR obtenido con los cebadores Check F y Check R para confirmar que se hubiera conseguido generar un mutante de *R. pomeroyi* DSS-3^T defectivo para el gen codificante de la proteína SPO2567.

6. Análisis *in silico* de las características de la proteína SPO2567

En primer lugar se determinó la existencia de homólogos de la proteína SPO2567 utilizando la herramienta de identificación y anotación de dominios Simple Modular Architecture Research Tool (SMART) que se basó en la búsqueda de homólogos de 1.200 dominios de proteínas curados manualmente en una base de datos de 33 millones de proteínas (Schultz *et al*., 1998; Letunic *et al*., 2015). A continuación se realizó una filogenia de la proteína SPO2567 y todas las proteínas homólogas halladas por SMART. Para ello, se alinearon las secuencias de proteínas utilizando el programa Clustal X (Thompson *et al*., 1997), se cargó el alineamiento en el programa Gblocks (Castresana, 2000) para eliminar las regiones poco conservadas y se utilizó tanto el alineamiento crudo como el pasado por Gblocks para realizar dos filogenias por el método de parsimonia (Eck y Dayhoff, 1966) utilizando el paquete PHYLYP (Felsestein, 1989).

A continuación, se realizó una predicción de estructura de la proteína SPO2567 con el programa de predicción de estructuras Phyre2 (Kelley *et al*., 2015). También se analizaron las posibles interacciones de la proteína SPO2567 con otras proteínas, utilizando la herramienta Search Tool for Recurring Instances of Neighbouring Genes, STRING v.10 (Szklarczyk *et al*., 2015, von Mering *et al*., 2003) que permite conocer las asociaciones funcionales directas e indirectas que presentan las proteínas de 2.031 organismos. Con esta herramienta se pudo establecer el tipo de relación que tenía SPO2567 con el resto de proteínas de *R. pomeroyi* DSS-3^T, el grado de confianza de esta interacción y si esta interacción también se encontraba conservada en otros microorganismos. Entre las proteínas que resultaron estar relacionadas con SPO2567 se detectaron varias proteínas de función desconocida cuya identificación se intentó refinar mediante el uso de la herramienta Conserved Domain Search del NCBI (Marchler-Bauer *et al*., 2015).

7. Generación de metagenomas de muestras marinas

7.1. Toma de muestras

Se realizaron cuatro muestreos, uno en abril de 2013, uno en de marzo 2014 y dos en abril de 2014. Para todos ellos se recogieron muestras de agua de la bahía Cala Penyes Rotges (Calvià, Mallorca) a una distancia aproximada de 1 km de la costa (39°29'05" N, 2°28'16" E). Estudios previos habían mostrado que las comunidades microbianas en esta zona, considerada control, tienen la composición típica de aguas costeras oligotróficas (Nogales *et al*., 2007; Aguiló-Ferretjans *et al*., 2008). En el muestreo de marzo de 2014 también se recogieron muestras de agua desde los pantalanes del interior del puerto deportivo de Port Adriano (39°29'21" N, 2°28'44" E). En el muestreo de marzo de 2014 se recogieron alrededor de 100 l de agua de puerto y 100 l de agua superficial de la zona control para realizar estudios de metagenómica de las comunidades microbianas. En los muestreos de abril 2013 y 2014 se recogieron alrededor de 200 l de agua superficial de la zona control para realizar experimentos de microcosmos en el laboratorio. Todas las muestras se tomaron tras al menos tres días de mar en calma y con ausencia de precipitaciones. En todos los casos las muestras se tomaron de la superficie (aproximadamente en los 5 cm superiores) con garrafas de plástico previamente lavadas con HCl 0,1 N y agua destilada. Las muestras de agua fueron trasladadas de inmediato al laboratorio para su procesamiento.

En todos los muestreos se determinó la temperatura *in situ* del agua y la salinidad y conductividad con CyberScan COND 610 (Eutech Instruments). Para determinar la concentración de clorofilas de las aguas muestreadas se filtraron 4 l de agua a través de filtros GF/F (Whatman) para recoger la biomasa y se guardaron a –20°C hasta su extracción. La extracción se realizó con una solución del 90 % acetona (v/v) en una disolución de $MgCO_3$ saturada. Se realizó la determinación de las clorofilas mediante espectofotometría en base al método tricromático de Jeffery y Humphrey (1975).

7.2. Experimentos de microcosmos

7.2.1. Preparación de los microcosmos

La preparación de microcosmos se realizó en base a los experimentos descritos por Lanfanconi y colaboradores (2010). Los microcosmos se prepararon en dos años sucesivos, en abril de 2013 y 2014. En el año 2014 se hicieron dos experimentos independientes de forma sucesiva, con muestras tomadas en un intervalo de 4 días. En los tres experimentos realizados se prepararon cuatro recipientes de plástico con 44 l de agua de la zona control cubiertos con tapas de metacrilato y oxigenados mediante bombas de acuario (Ehem Air pump 400) que expulsaban aire filtrado (filtro de jeringa de 0,2 µm, Nalgene). Los microcosmos se incubaron con luz natural y a una temperatura de 17° C, similar a la temperatura ambiental en el momento de recogida de las muestras (entre 15–17° C). Se dejaron estabilizar los microcosmos durante 2 h y se recogió la primera muestra (t_0). A continuación se aplicó el tratamiento a cada microcosmos: se utilizaron dos de ellos como controles (sin aditivos) y se añadieron 30 mg/l de diésel y 7,89 mg/l de $NaNO_3$ en los otros dos recipientes. Para compensar las posibles pérdidas de diésel en los microcosmos tratados, ya fuera por evaporación o por adsorción a las paredes de los recipientes, cada día se añadieron 10 mg/l de diésel y 2,63 mg/l de $NaNO_3$ siguiendo el

protocolo descrito por Lanfanconi y colaboradores (2010) y Piña-Villalonga (2012). Teniendo en cuenta la información previa sobre la dinámica de comunidades en este tipo de experimentos (Lanfranconi *et al*., 2010) los microcosmos se incubaron un máximo de 67 h. A las 24, 42 y 64 h de incubación de los microcosmos (67 h en el caso de los experimentos de 2013) se tomaron muestras para hacer un seguimiento rápido de los cambios en las comunidades microbianas mediante polimorfismo de longitud de fragmentos de restricción terminales (T-RFLP) de amplificados de 16S ADNr y recuentos de procariotas por DAPI y FISH (hibridación *in situ* con sondas fluorescentes).

7.2.2. Análisis de polimorfismos de longitud de fragmentos de restricción terminales (T-RFLP)

Se recogieron muestras de 2 l en los tiempos indicados en el apartado anterior. El procotolo para recoger las células por filtración y para la extracción de ácidos nucleicos se describe a continuación en el apartado 7.3. Se analizó la diversidad bacteriana en las muestras de cada microcosmos mediante la amplificación por PCR de 16S ADNr con los cebadores universales 27F (Lane, 1991) y 1389R (Osborn *et al.,* 2000) marcados en su extremo 5' con los fluorocromos 6-FAM y HEX, respectivamente. Para ello se siguieron protocolos descritos previamente (Nogales *et al*., 2007, Piña-Villalonga, 2012). Los fragmentos amplificados se digirieron separadamente con las enzimas de restricción AluI y HhaI (New England Biolabs) tal como se ha descrito (Piña-Villalonga, 2012). La separación electroforética de los fragmentos se realizó en un analizador genético ABI 3130 (Applied Biosystems) en los Servicios Científico-Técnicos de la UIB. El análisis de los perfiles electroforéticos se realizó como se describe en Piña-Villalonga (2012).

7.2.3. Recuento de procariotas totales e hibridación *in situ* con sondas fluorescentes (FISH)

En cada muestreo se tomaron 50 ml de agua, se fijaron con formaldehído al 4 % (v/v) y se conservaron a 4 °C durante un tiempo máximo de 16 h. A continuación se filtraron usando filtros de policarbonato con un tamaño de poro de 0,2 µm de diámetro (Millipore) y estos filtros se lavaron dos veces con tampón fosfato salino (PBS) (Sambrook *et al*., 2001). Se realizaron recuentos de procariotas totales mediante tinción con DAPI (4',6'-diamidino-2-fenilindol) en un microscopio de epifluorescencia Axiophot (Zeiss). También se cuantificó la abundancia de determinados grupos bacterianos mediante FISH con sondas específicas marcadas con el fluorocromo Cy3 (Interactiva Biotechnologie y Thermo Scientific). En concreto se usaron las sondas EUB338 I–III para determinar la presencia de *Bacteria* (Amann *et al*., 1990; Daims *et al*., 1999), la sonda ROS537 (Eilers *et al*., 2001) para miembros del linaje Roseobacter, y la sonda Gam42a (Manz *et al*., 1992) para *Gammaproteobacteria* (esta última utilizada con competidor). La tinción de DAPI y las hibridaciones se realizaron siguiendo el método descrito por Glöckner y colaboradores (1996). Por cada muestra se contó un mínimo de 500 células teñidas con DAPI.

7.2.4. Análisis de limitación de nitrógeno en las aguas de los microcosmos

Dado que los tres muestreos en los que se realizaron experimentos de microcosmos se trataron con diésel y nitrógeno, se analizó si se producía un incremento en el número de procariotas totales tras la adición de nitrógeno en un tiempo corto de incubación. Para ello se

prepararon 4 botellas estériles con 400 ml de agua, dos de las cuales se suplementaron con $NaNO_3$ a una concentración final de 16 µM y las otras dos se dejaron como controles. Se incubaron las botellas durante 24 h en las mismas condiciones que los microcosmos y se tomaron muestras para realizar tinciones DAPI con las que poder hacer un recuento de procariotas totales como se ha descrito en el apartado anterior.

7.3. Extracción de ácidos nucleicos totales para generar metagenomas

Para generar los metagenomas de muestras ambientales se filtraron 100 l de agua de la zona control y del puerto a través de filtros de membrana de fluoruro de polivinilideno con un tamaño de poro de 5 µm (Millipore) para eliminar los organismos eucariotas y las partículas presentes en las muestras. A continuación, se filtraron las muestras por filtros de membrana de fluoruro de polivinilideno con un tamaño de poro de 0,22 µm (Millipore) cambiando los filtros cada 2 l de agua filtrada. Los filtros se congelaron a –80° C hasta la extracción de los ácidos nucleicos.

En el caso de las muestras de microcosmos se siguió el mismo procedimiento pero el volumen total filtrado de cada muestra fueron los 44 l de cada recipiente. Para la preparación de las muestras en el experimento realizado en abril de 2013 se procesaron dos microcosmos (un control y un tratamiento) tras 42 h de incubación y los dos restantes tras 67 h de incubación, con el fin de obtener 4 metagenomas que proporcionaran información sobre variación temporal de comunidades. En los dos experimentos realizados en abril de 2014 se procesaron todos los microcosmos tras 64 h de incubación. Tras el análisis de diversidad mediante T-RFLP se determinó que en el primer experimento (M1) había mayores diferencias entre los microcosmos control y los tratados. Por esta razón se utilizaron únicamente las muestras del experimento M1 para generar metagenomas.

Para la extracción de los ácidos nucleicos de cada filtro se siguió el protocolo descrito por Nogales y colaboradores (2007) el cual fue modificado añadiendo 0,05 % (v/v) de RNase CocktailTM Enzyme Mix (Ambion, Life Technologies) en el mismo paso en el que se añadía lisozima. Además de comprobar la extracción de los ácidos nucleicos mediante electroforesis en gel de agarosa al 1 % p/v en tampón TAE (Sambrook *et al*., 2001) y su tinción con bromuro de etidio (0,5 µg/ml) se cuantificaron los extractos con un espectofotómetro Nanodrop 2000c (Thermo Scientific). Se extrajeron ácidos nucleicos de tantos filtros por muestra como fue necesario para conseguir una concentración de ácidos nucleicos en el rango recomendado por la empresa que realizó la secuenciación (Macrogen, Corea). En el caso de los metagenomas generados a partir de los microcosmos de 2014, se decidió utilizar muestras de ADN compuestas de los dos microcosmos control y tratados de M1, de manera que se generaron únicamente dos metagenomas mezclando los ácidos nucleicos extraídos de los dos microcosmos réplica de cada condición. Las secuencias de los metagenomas se generaron con la tecnología Illumina HiSeq2000 (lecturas pareadas 250–300 pb) a partir de librerías preparadas con el kit TruSeq Nano DNA library preparation (Illumina Inc.) en la empresa Macrogen (Corea).

8. Análisis de los metagenomas

Una vez obtenidas las lecturas crudas de cada metagenoma se realizó una selección de lecturas de calidad utilizando el programa Trimmomatic (Bolger *et al*., 2014). Con este paso se eliminaron los adaptadores usados para la generación de las librerías, las lecturas que tuvieran una longitud inferior a 90 pb (parámetro MINLEN:90), un índice de calidad Q medio para toda la lectura inferior a 20 (AVGQUAL:20), aquellas cuyo límite de índice de calidad Q medio para una ventana de 4 nucleótidos cayera por debajo de 20 (SLIDINGWINDOW:4:20), y aquellas cuyo último nucleótido tuviesen también un índice de calidad inferior a 20 (TRAILING:20). La calidad global de las lecturas que habían pasado la criba de Trimmomatic se analizó con el programa fastQC (Andrews, 2010). Todas las lecturas conservadas para el análisis tenían valores de fastQC en el rango de calidad bueno (Q ≥ 28) o razonable (Q entre 20 y 28).

A continuación se siguieron varias estrategias para analizar las diferencias taxonómicas y funcionales existentes entre las comunidades de aguas no contaminadas (zona control y microcosmos no tratados) y aguas contaminadas (agua del puerto y de microcosmos tratados). En primer lugar se analizaron las lecturas con el servidor de análisis de metagenomas MG-RAST (Meyer *et al*., 2008). Se analizaron las dos conjuntos de lecturas pareadas (R1 y R2), el conjunto de lecturas desapareadas (R3) y además se realizó un análisis conjunto de todas las lecturas pareadas de cada metagenoma mediante la opción *Join paired ends* que permitió unir todas las lecturas pareadas en un mismo conjunto de datos y reducir el número de lecturas del análisis. Estas fueron las lecturas que se analizaron en profundidad para la asignación taxonómica y funcional. En primer lugar se evaluó si el esfuerzo de muestreo había sido suficiente mediante curvas de rarefacción. A continuación, se determinaron las categorías taxonómicas presentes en cada metagenoma utilizando la base de datos de RefSeq (Pruitt *et al*., 2007) que es una base de datos de nucleótidos y proteínas curada y no redundante que representa una elevada diversidad taxonómica. También se recuperaron las secuencias de 16S ADNr de bacterias, arqueas y eucariotas detectadas con MG-RAST y se analizaron con SILVAngs (Pruesse *et al*., 2007) para determinar los grupos taxonómicos presentes en cada metagenoma. Se obtuvo así una segunda clasificación taxonómica que pudiera completar a la obtenida por MG-RAST. Se calcularon los índices de diversidad de Shannon (H) y la equitabilidad (J) de la asignación taxonómica realizada con MG-RAST para cada metagenoma con el programa PAST v. 3.08 (Hammer *et al*., 2001). La asignación funcional de las lecturas de cada metagenoma se realizó en base a las bases de datos de COG (Tatusov *et al*., 2000), basada en la agrupación de proteínas ortólogas; NOG (Jensen *et al*., 2008), basada en grupos de proteínas ortólogas no supervisados; KO (Kanehisa *et al*., 2016a), basada en la base de datos de proteínas ortólogas de KEGG (Kyoto Encyclopedia of Genes and Genomes) (Ogata *et al*., 1999); y de subsistemas SEED (Overbeek *et al*., 2005; 2014), basada en las agrupaciones de proteínas relacionadas funcionalmente. El criterio de identidad de secuencia seguido tanto en el análisis taxonómico como en el funcional fue que los fragmentos de ADN de los metagenomas presentaran al menos un 30 % de identidad en al menos un fragmento de 45 nucleótidos con la proteína/gen modelo y con un *E-value* menor de 10^{-5}. Además se analizaron las lecturas de cada metagenoma con la base de datos AromaDeg (Duarte *et al*., 2014) para completar el análisis funcional de los genes involucrados en el catabolismo de compuestos aromáticos. En concreto se analizaron los dos conjuntos de lecturas pareadas (R1 y R2) por

separado realizando una comparación de las lecturas de cada metagenoma contra la base de datos de AromaDeg con el programa BLASTx (Altschul *et al*., 1997). El criterio de identidad de secuencia seguido fue de al menos un 30 % de identidad en al menos un fragmento de 45 aminoácidos con la proteína modelo de AromaDeg y con un *E-value* menor de 10^{-5}. Como los resultados que se obtenían para las lecturas R1 y R2 de cada metagenoma eran similares se usó la anotación obtenida con lecturas R1 de cada metagenoma para analizar la diversidad de proteínas relacionadas con la degradación de compuestos aromáticos de cada metagenoma. Además, para estudiar la variabilidad en la presencia de genes para enzimas implicadas en la degradación de compuestos aromáticos se realizaron análisis de componentes principales (PCA) usando el programa PAST v. 3.08 (Hammer *et al*., 2001) a partir de los resultados de AromaDeg (Duarte *et al*., 2014).

Finalmente, la última aproximación seguida en este trabajo fue realizar ensamblajes de las lecturas de los metagenomas con el fin de obtener fragmentos de secuencia más largos. Se ensamblaron las lecturas con el programa MetaVelvet (Namiki *et al*., 2012) usando distintos k-mers (51, 57, 61, 67, 71, 77, 81, 87 y 91) y se determinó el mejor ensamblaje en base al número de *contigs* obtenidos, su longitud, valor N50 y la profundidad de su cobertura media (Miller *et al*., 2010). A continuación, se analizaron los *contigs* de cada metagenoma con la base de datos de AromaDeg (Duarte *et al*., 2014) para poder analizar el potencial para el catabolismo de compuestos aromáticos y para saber si estos genes se organizaban o no en estructuras supragénicas. El criterio de identidad de secuencia seguido también fue de al menos un 30 % de identidad en al menos un fragmento de 45 aminoácidos con la proteína modelo de AromaDeg y con un *E-value* menor de 10^{-5}.

CAPÍTULO 1: POTENCIAL PARA EL CATABOLISMO DE COMPUESTOS AROMÁTICOS EN MIEMBROS DEL LINAJE ROSEOBACTER

El objetivo principal de esta parte de la tesis era estudiar el potencial para el catabolismo de compuestos aromáticos en los miembros del linaje Roseobacter utilizando la información disponible en genomas de estos microoganismos. Hasta la realización de esta tesis el estudio más exhaustivo sobre la capacidad para degradar compuestos aromáticos en los miembros del linaje Roseobacter era el estudio de Newton y colaboradores (2010) basado en la búsqueda de la enzima principal de seis vías centrales para la degradación de compuestos aromáticos (rama del protocatecuato de la vía del beta-cetoadipato, benzoil-CoA, gentisato, fenilacetil-CoA, homoprotocatecuato y homogentisato) en 32 genomas de Roseobacter. En la presente tesis se ha realizado una aproximación más detallada, puesto que no solo se ha detectado la presencia de genes para las enzimas principales de las seis vías citadas anteriormente en un número mayor de genomas, sino que se ha ampliado tanto el número de genes analizados como el número de rutas estudiadas. Así, se ha determinado la presencia y organización de 166 genes involucrados en las 6 vías centrales anteriormente citadas así como en la rama del catecol de la vía del beta-cetoadipato, la degradación por rotura *meta* del anillo aromático del protocatecuato y el catecol, y en las vías periféricas que convergen en estas rutas centrales. El estudio se ha realizado en 61 genomas del linaje Roseobacter. La presencia de estos 166 genes se ha determinado mediante análisis de identidad de secuencia de las proteínas codificadas por estos genes (ver apartado 1 material y métodos). Al ampliar el estudio de las rutas centrales estudiadas por Newton y colaboradores (2010) y añadir el estudio de las rutas periféricas que convergen en las rutas centrales se ha obtenido una visión más amplia del espectro de compuestos aromáticos que podrían ser degradados por los miembros del linaje Roseobacter. Además, el estudio de los genes para todos los pasos intermedios de las rutas de degradación de los compuestos aromáticos hasta la formación de intermediarios del ciclo de Krebs permite postular con más fiabilidad el potencial para degradar compuestos aromáticos en los genomas estudiados. Cabe destacar que en esta tesis también se ha analizado la capacidad de varios aislados para crecer a expensas de los compuestos aromáticos estudiados como única fuente de carbono y energía (ver apartado 9). Por todo ello, esta tesis presenta el estudio más detallado del potencial catabólico de compuestos aromáticos realizado en un mayor número de genomas del linaje Roseobacter descrito hasta la fecha.

1. Genes para la degradación de compuestos aromáticos en genomas del linaje Roseobacter

Tal y como puede observarse en la tabla M1 la mayor parte de los 166 genes estudiados (ver detalles en la tabla suplementaria S1.2) codificaban para proteínas implicadas en las vías catabólicas centrales y periféricas. Pero también se estudiaron genes codificantes para proteínas involucradas en la regulación de estas rutas, para transportadores de compuestos aromáticos y para varias proteínas cuya función no siempre está clara pero cuyos genes codificantes suelen estar conservados junto a los genes codificantes para las proteínas catabólicas estudiadas.

En primer lugar se analizó la presencia en los genomas de los genes de rutas centrales de degradación. Inicialmente se tuvo en cuenta solo la presencia de los genes que codificaban para las enzimas que catalizan la rotura del anillo aromático (considerados en este trabajo como genes/enzimas principales). En caso de que la enzima principal estuviera codificada por más de un gen, solo se consideró como resultado positivo si se detectaban todos los genes

codificantes para las diferentes subunidades. En base a este criterio, en la tabla 1.1 se muestra una visión general de la presencia de las rutas centrales de degradación en los genomas analizados. En concreto, se observó que las enzimas principales de la rama del protocatecuato de la vía del beta-cetoadipato y las vías de degradación del homogentisato y fenilacetil-CoA eran las más detectadas en genomas de los miembros del linaje Roseobacter (tabla 1.1) al igual que se había mostrado en el estudio previo de Newton y colaboradores (2010). También se determinó si la presencia de genes para las enzimas principales era homogénea entre los cinco grupos filogenómicos del linaje Roseobacter descritos por Newton y colaboradores (2010). Solo en el grupo G3 se detectaron los genes codificantes para las enzimas principales de todas las rutas estudiadas (tabla 1.1). Cabe destacar que en el grupo G3 se observaron porcentajes de detección elevados del gen para gentisato 1,2 dioxigenasa cuando esta es una de las vías de degradación de compuestos monoaromáticos menos detectadas en el linaje Roseobacter (ver tablas 1.1, 1.2 y tabla suplementaria S1.2). Esto se debió probablemente a la presencia en el grupo filogenómico G3 de cuatro cepas de *Citreicella aestuarii*, todas ellas portadoras de genes para la degradación del gentisato. Por otra parte, se pudo observar que los genes codificantes para las enzimas principales de la rama del protocatecuato de la vía del beta-cetoadipato y de las vías del homogentisato, fenilacetil-CoA y homoprotocatecuato también eran los únicos que se observaban en los cinco grupos filogenómicos (tabla 1.1). Ahora bien el porcentaje de detección en los cinco grupos variaba considerablemente y solo los genes codificantes para la enzima principal de la rama del protocatecuato de la vía del beta-cetoadipato estaban presentes en al menos el 50 % de los genomas de cada uno de los cinco grupos filogenómicos.

Tabla 1.1. Detección de los genes codificantes para las enzimas principales de las vías centrales de la degradación de compuestos aromáticos en genomas del linaje Roseobacter. Los valores entre paréntesis indican porcentajes.

		Grupos filogenómicos[a]				
Vía central	**Total**	**G1**	**G2**	**G3**	**G4**	**G5**
Beta-cetoadipato (Rotura *orto*)						
Protocatecuato	41 (67,2)	11 (78,6)	7 (77,8)	9 (60,0)	7 (63,6)	7 (58,3)
Catecol	9 (14,8)	2 (14,3)	0(0,0)	4 (26,7)	0 (0,0)	3 (25,0)
Protocatecuato (rotura *meta*)	3 (4,9)	0 (0,0)	1 (11,1)	1 (6,7)	0 (0,0)	1 (8,3)
Catecol (rotura *meta*)	3 (4,9)	0 (0,0)	1 (11,1)	2 (13,3)	0 (0,0)	0 (0,0)
Homogentisato	42 (68,9)	14 (100,0)	9 (100)	10 (66,7)	3 (27,3)	6 (50,0)
Fenilacetil-CoA	28 (45,9)	14 (100,0)	6 (66,7)	3 (20,0)	0 (0,0)	5 (41,7)
Homoprotocatecuato	22 (36,1)	12 (85,7)	3 (33,3)	3 (20,0)	1 (9,1)	3 (25,0)
Gentisato	15 (24,6)	1 (7,1)	1 (11,1)	7 (46,7)	0 (0,0)	6 (50,0)
Benzoil-CoA	7 (11,5)	2 (14,3)	2 (22,2)	2 (13,3)	0 (0,0)	1 (8,3)

a: El número de genomas incluidos en cada grupo filogenómico es: G1, 14; G2, 9; G3, 15; G4, 11; G5, 12.

Seguidamente se analizó la presencia de todos los genes implicados en las vías degrativas centrales. Para considerar la presencia de una ruta completa debían detectarse la totalidad de genes de la ruta con al menos un 30 % de identidad con un *E-value* menor de 10^{-5} respecto al modelo. En caso de no detectar uno o más genes de una ruta se consideraba que la vía estaba incompleta en el genoma (tabla 1.2 y detalle en tabla suplementaria S1.2). Este resultado indicaría que o bien esos microorganismos presentaban rutas degradativas truncadas o bien que presentaban enzimas cuya secuencia era suficientemente divergente a los modelos usados para que no se detectasen con los criterios utilizados en este estudio. En los genomas de

ninguno de los grupos filogenómicos se detectaron rutas completas para la degradación del catecol y protocatecuato por rotura en posición *meta*, del catecol por rotura en posición *orto* y del benzoil-CoA (tabla 1.2). Ahora bien, que no se considere una vía completa por la falta de detección de al menos uno de sus genes no implica que esta vía de degradación este truncada en este organismo. Por ejemplo, en ninguno de los genomas se pudo detectar el gen *boxD* de la vía de degradación de benzoil-CoA pero se sabe que *Sagittula stellata* E-37^T puede catabolizar el benzoato (Gulvik y Buchan 2013). Por lo tanto, el criterio utilizado en esta tesis para determinar la presencia de rutas completas podría resultar en ocasiones demasiado estricto. De hecho los experimentos de crecimiento en compuestos aromáticos realizados en esta tesis también muestran algunas discrepancias con la predicción génica (ver apartado 9). Ahora bien, al carecer de estudios de crecimiento de los 61 microorganismos estudiados en presencia de estos compuestos se prefirió mantener este criterio estricto para evitar afirmar la presencia de rutas completas sin al menos una base genética. La detección de las rutas completas para la degradación del protocatecuato (rotura *orto*), homogentisato, homoprotocatecuato y fenilacetil-CoA era más alta en genomas de los grupos G1 y G2 que en el resto de grupos filogenómicos, sobre todo en el caso de G1. Además, en el caso del G3, la elevada detección del gen codificante para la enzima principal para la degradación del gentisato iba siempre acompañada de la detección de la ruta completa para la degradación de este compuesto (tablas 1.1 y 1.2). Por otra parte cabe destacar que el grupo filogenómico en el que se detectó una menor cantidad de vías degradativas y con unos porcentajes de detección más bajos fue el grupo G4, en el que solo se detectó la rama del protocatecuato de la vía del beta-cetoadipato y las vías degradativas del homogentisato y homoprotocatecuato con unos porcentajes de detección claramente inferiores a los que se observaban en los demás grupos filogenómicos (tablas 1.1 y 1.2).

Tabla 1.2. Detección de vías centrales completas para la degradación de compuestos aromáticos en genomas del linaje Roseobacter. Los valores entre paréntesis indican porcentajes.

Vía central	Total	Grupos filogenómicos[a]					Nº estructuras génicas
		G1	G2	G3	G4	G5	
Beta-cetoadipato (Rotura *orto*)							
Protocatecuato	39 (63,9)	11 (78,6)	7 (77,8)	9 (60,0)	5 (45,4)	7 (58,3)	25 (24,3)
Catecol	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	8 (7,8)
Protocatecuato (rotura *meta*)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	3 (2,9)
Catecol (rotura *meta*)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	1 (1,0)
Homogentisato	38 (62,3)	14 (100,0)	7 (77,8)	9 (60,0)	3 (27,3)	5 (41,7)	24 (23,3)
Fenilacetil-CoA	28 (45,9)	14 (100,0)	6 (66,7)	3 (20,0)	0 (0,0)	5 (41,7)	13 (12,6)
Homoprotocatecuato	22 (36,1)	12 (85,7)	3 (33,3)	3 (20,0)	1 (9,1)	3 (25,0)	13 (12,6)
Gentisato	15 (24,6)	1 (7,1)	1 (11,1)	7 (46,7)	0 (0,0)	6 (50,0)	14 (13,6)
Benzoil-CoA	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	2 (1,9)

a: El número de genomas incluidos en cada grupo filogenómico es: G1, 14; G2, 9; G3, 15; G4, 11; G5, 12

A continuación se analizó el tipo de organización que presentaban estos genes en los genomas estudiados. Este análisis permitió determinar si había sintenia entre los miembros del linaje Roseobacter y si estos genes se encontraban organizados en operones o en estructuras supraoperónicas como en algunos de los modelos de degradación descritos (Harwood y

Parales, 1996; Parke, 1997). Cabe destacar que se estudió la organización de estos genes en todos los genomas en los que se había detectado la enzima principal, independientemente de que se hubiera podido detectar la via de degradación completa o no. Para todas las vías centrales estudiadas se detectó más de un tipo de estructura génica excepto en la rotura *meta* del catecol (tabla 1.2). La vía degradativa en la que se observaba una mayor variabilidad en cuanto a la organización génica fue en los genes de la rama del protocatecuato de la vía del beta-cetoadipato (tabla 1.2). En los siguientes apartados se detalla la organización de los genes codificantes para las enzimas de las diferentes vías de degradación de compuestos aromáticos en los genomas de bacterias del linaje Roseobacter analizados.

2. Vías de degradación de protocatecuato y catecol

Como se comentó en la introducción (apartado 2) el protocatecuato y el catecol pueden degradarse tanto a partir de la rotura del anillo aromático en posición *orto* como a partir de la rotura en posición *meta* (figura 1.1). En el primer caso la ruta de degradación recibe el nombre de vía del beta-cetoadipato.

2.1. Rotura *orto* (Vía del beta-cetoadipato)

La vía de degradación del beta-cetoadipato está presente en muchos grupos bacterianos, en concreto ha sido principalmente estudiada en bacterias de suelo, y es una vía central convergente para degradación de compuestos aromáticos (Harwood y Parales, 1996). La vía del beta-cetoadipato tiene dos ramas: la rama del protocatecuato y la rama del catecol (figura 1.1). Ambos compuestos son metabolitos centrales derivados de la oxigenación de diversos compuestos fenólicos, de monómeros de lignina y de hidrocarburos aromáticos (Harwood y Parales, 1996). En la vía del beta-cetoadipato la rotura del anillo bencénico se produce en posición *orto* y es catalizada por intradiol dioxigenasas que contienen un grupo hierro III no hemo (Harwood y Parales, 1996; Pérez-Pantoja *et al*., 2010a). Una vez roto el anillo bencénico las siguientes reacciones de cada rama conducen a la generación de intermediarios del ciclo de Krebs (Harwood y Parales, 1996). La vía del beta-cetoadipato suele estar codificada en el cromosoma y se encuentra muy conservada en diversos grupos bacterianos aunque hay una gran diversidad en cuanto a su regulación y a la forma en que se encuentran organizados y agrupados los genes (Harwood y Parales, 1996).

2.1.1. Rama del protocatecuato de la vía del beta-cetoadipato

Como se ha mencionado anteriormente la rama del protocatecuato de la vía del beta-cetoadipato fue la más detectada en los genomas del linaje Roseobacter (tablas 1.1 y 1.2). En esta ruta la degradación de protocatecuato se lleva a cabo por la acción de seis enzimas codificadas por ocho genes *pca* (figura 1.1). La enzima principal que se encarga de la rotura del protocatecuato en posición *orto* es la protocatecuato 3,4 dioxigenasa (EC 1.13.11.3), una enzima con dos subunidades que se encuentran codificadas por los genes *pcaG* y *pcaH*. Además se han documentado otras proteínas relacionadas con esta ruta que actúan como reguladores (PcaQ y PcaR). Los genes *pca* normalmente se encuentran agrupados, pero la organización génica en las agrupaciones no se mantiene entre los organismos que la presentan

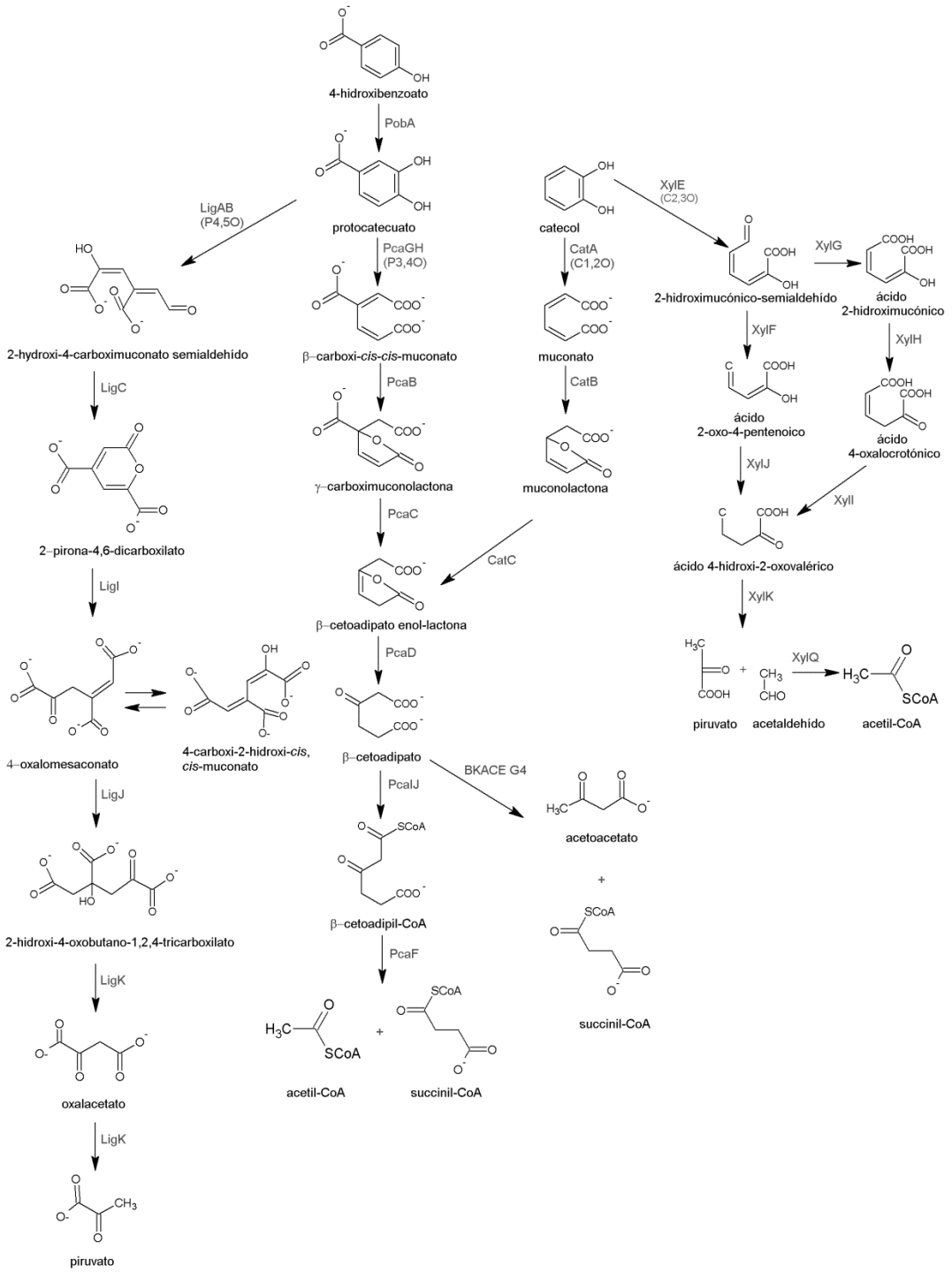

Figura 1.1. Rutas centrales de degradación de protocatecuato y catecol: vía del beta-cetoadipato (rotura *orto*) y rutas de degradación por rotura *meta*. Abreviaturas: PobA, 4-hidroxibenzoato hidroxilasa; PcaGH (P3,4O), protocatecuato 3,4-dioxigenasa; PcaB, 3-carboxi-*cis*,*cis*-muconato cicloisomerasa; PcaC, 4-carboximuconolactona descarboxilasa; PcaD, 3-oxoadipato enol-lactona hidrolasa; PcaIJ, 3-oxoadipato CoA-transferasa; PcaF, β-cetoadipil-CoA tiolasa; BKACE G4, enzima de rotura de β-cetoácidos del grupo G4; CatA (C1,2O), catecol 1,2-dioxigenasa; CatB, muconato cicloisomerasa; CatC, muconolactona δ-isomerasa; XylE (C2,3O), catecol 2,3-dioxigenasa; XylF, hidroximucónico-semialdehído hidrolasa; XylG, hidroximucónico-semialdehído deshidrogenasa; XylH, 4-oxalocrotonato isomerasa; XylI, 4-oxalocrotonato descarboxilasa; XylJ, 2-oxo-4-pentanoato hidratasa; XylK, 4-hidroxi-2-oxovalerato aldolasa; XylQ, acetaldehído deshidrogenasa; LigAB (P4,5O), protocatecuato 4,5-dioxigenasa; LigC, 2-hidroxi-4-carboximuconato-6-semialdehído deshidrogenasa; LigI, 2-pirona-4,6-dicarboxilato hidrolasa; LigJ, 4-oxalomesaconato hidratasa; LigK, 4-carboxi-4-hidroxi-2-oxoadipato aldolasa. Figura elaborada a partir de información de Harwood y Parales, 1996; Masai *et al*., 1999; Hara *et al*., 2003; Peng *et al*., 2008 y Bastard *et al*., 2014.

a excepción de las parejas de genes *pcaGH* y *pcaIJ*, codificantes para subunidades de las mismas enzimas, y que siempre se encuentran juntos (Harwood y Parales, 1996). Por otra parte, se ha observado que en algunos microorganismos como *Acinetobacter calcoaceticus* y *Agrobacterium tumefaciens* los genes *pca* se encuentran formando parte de estructuras supraoperónicas (Harwood y Parales, 1996), como por ejemplo con los genes *pob*, como *pobA,* codificante para la enzima 4-hidroxibenzoato hidroxilasa encargado de la transformación del 4-hidroxibenzoato a protocatecuato (Harwood y Parales, 1996).

El primer trabajo en el que se estudió la organización de los genes para la rotura *orto* del protocatecuato en el linaje Roseobacter fue el de Buchan y colaboradores (2004) en el cual analizaron las secuencias que flanqueaban a los genes *pcaG* y *pcaH* en ocho miembros del linaje Roseobacter. En ese estudio observaron hasta 5 estructuras distintas para los genes *pca,* en todas las cuales se encontraban los genes *pca* junto a un gen conservado, que se cotranscribía con los genes *pca* en *Ruegeria pomeroyi* DSS-3^T, y al que denominaron *chp*, del inglés *conserved hypothetical protein* (Buchan *et al*., 2004). Buchan y colaboradores (2004) postularon que la presencia de este gen era una característica del linaje Roseobacter ya que la presencia de genes homólogos a *chp* en asociación a los genes *pca* no había sido descrita en otros organismos (Harwood y Parales, 1996; Jiménez *et al*., 2002). Se observó también que cepas que podían pertenecer a la misma especie tenían la misma organización génica (Buchan *et al*., 2004). Por otra parte, el gen *pobA*, se encontró adyacente a los genes *pca* en algunas cepas (Buchan *et al*., 2004). En un estudio posterior (Gulvick y Buchan, 2013) se estudió la organización de los genes *pca* en 3 cepas más (*Roseobacter* sp. GAI101, *Jannaschia* sp. CCS1 y *Ruegeria* sp. KLH11), observando resultados similares.

En la figura 1.2 se muestran los resultados de la detección de los diferentes genes de la rama del protocatecuato de la ruta del beta-cetoadipato en los genomas analizados en estudio usando el criterio estándar del 50 % de identidad con un *E-value* menor de 10^{-5}. Como se observa en la figura se pudieron detectar los genes *pcaG*, *pcaH*, *pcaC*, *pcaD*, *pcaJ*, *pcaF* y *chp* en la mayoría de los genomas estudiados. En el caso de *pcaB*, *pcaI,* y los reguladores *pcaQ* y *pobR,* la mayoría de genes fueron detectados usando el criterio de 30 % de identidad con un *E-value* menor de 10^{-5}. Ahora bien, solo se dieron como válidas las identificaciones de estos genes cuando se encontraban adyacentes a los otros genes *pca* o al gen *pobA*, o bien cuando eran el único gen de la ruta que no había sido detectado mediante el criterio más estricto.

Como se puede observar en la figura 1.2, en dos de los genomas, *Roseobacter* sp. MED193 y *Loktanella vestfoldensis* SKA53, se pudieron identificar dos copias distintas de genes *pcaGH* estando ambas flanqueadas por distintos genes *pca*. Como se ha descrito generalmente en otras bacterias (Harwood y Parales, 1996), los genes *pcaG* y *pcaH* se encontraron juntos en 41 genomas (67 %) de los microorganismos estudiados, en 39 de los cuales también se pudieron detectar los otros 6 genes codificantes para las enzimas involucradas en la degradación de protocatecuato (figura 1.2). En *Roseobacter* sp. CCS2 se observó un posible gen de fusión entre de *pcaC* y *pcaD* que se ha descrito como *pcaL* en *Rhodococcus opacus* 1CP (Eulberg *et al*., 1998) y en la alfaproteobacteria *Caulobacter crescentus* (Jiménez *et al*., 2002). En cuanto a la presencia de *pcaB,* este gen se pudo detectar en todos los microorganismos en los que se había detectadado *pcaGH* aunque en la mayoría de casos aplicando el criterio de detección más relajado (30 % de identidad de secuencia con un *E-value* < 10^5). Buchan y colaboradores

(2004) comentaron la posible ausencia de *pcaB* en los miembros del linaje Roseobacter aunque en el estudio de Gulvick y Buchan (2013) se mostró la presencia del gen *pcaB* en el genoma de *Ruegeria* sp. KLH11. La escasa detección de *pcaB* era extraña si se tenía en cuenta que el sustrato para PcaB, el β-carboxi-*cis*-*cis*-muconato es muy tóxico y que, por tanto debía ser procesado para permitir el crecimiento en protocatecuato (Parke *et al*., 2000). Con el presente estudio se confirmó la presencia generalizada de *pcaB* en los miembros del linaje Roseobacter (figura 1.2). En cuanto a los genes *pcaI*, *pcaJ* y *pcaF* estos fueron detectados en la mayoría de los genomas estudiados independientemente o no de la presencia de *pcaGH* en ellos.

De acuerdo con los resultados de esta tesis veinte miembros del linaje Roseobacter estudiados podrían presentar la rama del protocatecuato de la vía del beta-cetoadipato incompleta puesto que en la mayoría de ellos el primer gen detectado fue *pcaD*, *pcaI* o *pcaJ* (figura 1.2). Esto significaría que solo podrían realizar las últimas reacciones de esta vía de degradación. En el caso de *K. vulgare* WSH-001 la vía estaría incompleta porque solo se encontraron los genes *pcaGHBCD* para convertir el protocatecuato en beta-cetoadipato pero no los genes *pcaIJF* para canalizar el beta-cetoadipato hacia el ciclo de Krebs (figura 1.2). También estaría incompleta la vía en *K. vulgare* Y25 al no detectar *pcaD* (figura 1.2).

En cuarenta y cuatro de los genomas analizados se detectó el gen *chp* (figura 1.2). Estos resultados confirmaban la evidencia de Buchan y colaboradores (2004) de que la presencia de este gen en asociación con los genes *pca* es característica del linaje Roseobacter. Recientemente Bastard y colaboradores (2014) demostraron que las proteínas de la familia a la que pertenece la proteína codificada por el gen *chp* (DUF849) catalizan una reacción genérica de condensación de β-cetoácidos con acetil-CoA para producir un éster de CoA y acetoacetato. Estos autores han denominado a esta familia como enzimas para la rotura de β-cetoácidos (BKACE). Estos autores diferenciaron hasta siete grupos dentro de la familia BKACE (G1-G7) y determinaron la presencia de genes codificantes para BKACE G4 en genomas de alfaproteobacterias en asociación con los genes *pca*. Además, probaron que las proteínas BKACE G4, incluyendo proteínas clonadas de bacterias del linaje Roseobacter, transformaban preferentemente beta-cetoadipato y beta-cetoglutarato (Bastard *et al*., 2014). En base a esta información se confirmó que los genes *chp* que se habían predicho en los 31 genomas de Roseobacter comunes entre nuestro estudio y el estudio de Bastard y colaboradores (2014) habían sido clasificados como BKACE G4 por estos autores.Teniendo en cuenta la identidad de secuencia inter e intragrupo para las BKACE (Bastard *et al*., 2014) se consideró que todos los genes *chp* que se habían detectado en nuestro estudio (figura 1.2) codificaban para proteínas del grupo G4 de BKACE. Por lo tanto, los miembros del linaje Roseobacter tendrían dos maneras distintas de transformar el beta-cetoadipato: mediante la beta-cetoadipato succinil-CoA transferasa (PcaIJ) y mediante la proteína conservada codificada por el gen *chp*.

En cuanto al gen *pobA*, que codifica para la enzima que transforma 4-hidroxibenzoato a protocatecuato, (figura 1.1.), este se detectó en 42 genomas (69 % de los genomas estudiados). En ellos estaban incluidos los genomas de *Rhodobacteraceae* bacterium HTCC2083 y *Maritimibacter alkaliphilus* HTCC2654^{T}, en los que no se había podido detectar *pcaGH,* pero sí se había podido detectar la protocatecuato 4,5-dioxigenasa que cataliza para la

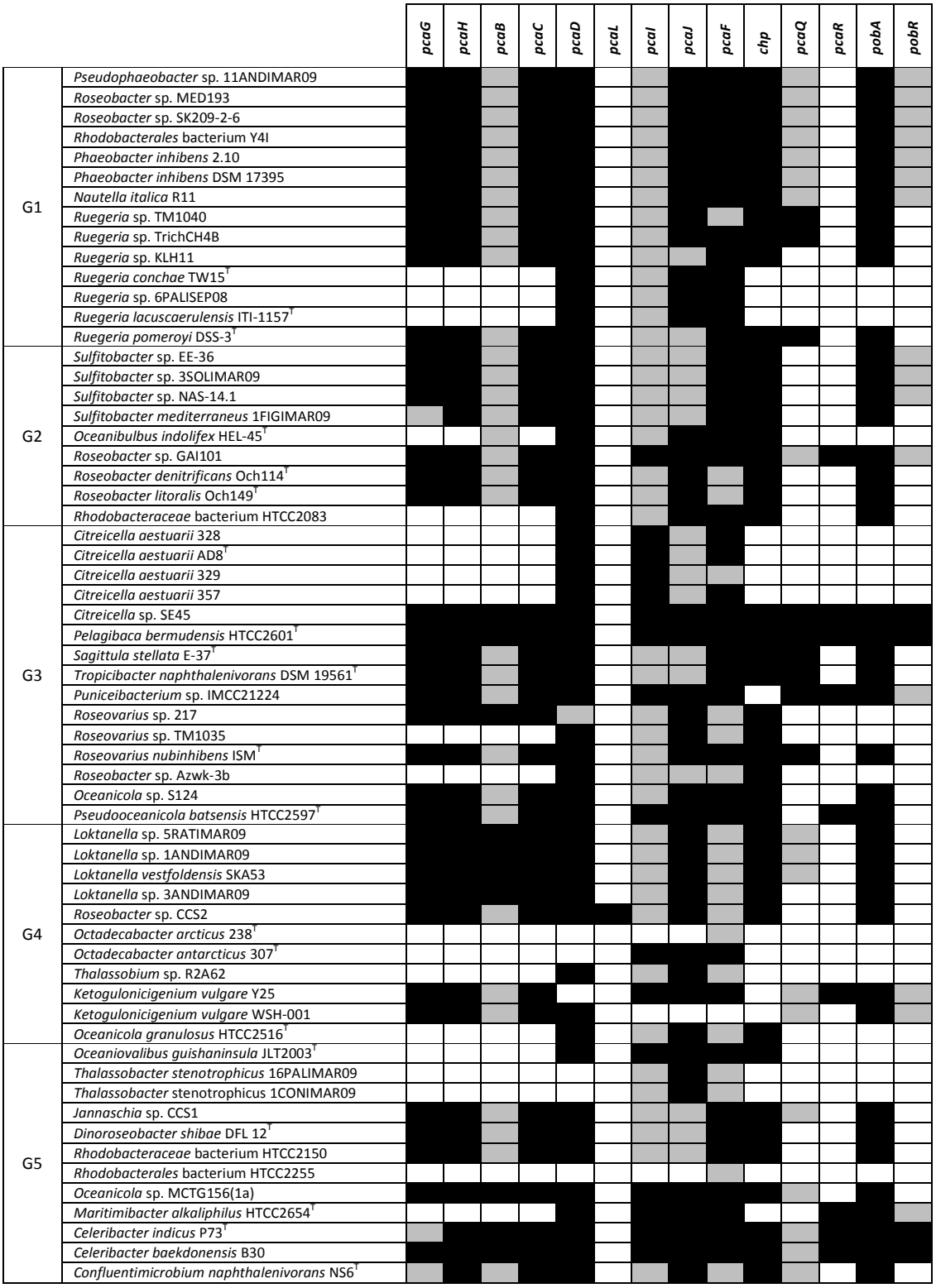

Figura 1.2. Detección de genes codificantes para las proteínas involucradas en la rama del protocatecuato de la vía del beta-cetoadipato en genomas del linaje Roseobacter. Se incluyen también los genes implicados en la transformación de 4-hidroxibenzoato a protocatecuato. Los genomas están ordenados en base a su filogenia (Luo y Moran, 2014; Mas-Lladó, 2017). Código de color: Negro 50 % de identidad de secuencia mínima con la proteína modelo, *E-value* $<10^{-5}$; gris 30 % de identidad de secuencia mínima con la proteína modelo, *E-value* $<10^{-5}$; blanco, no detectados.

rotura *meta* del protocatecuato (ver apartado 2.2.1). El resto de genes *pobA* se detectaron en 40 de los 41 genomas en los que se había detectado los genes *pcaGH* (no se detectó en *Roseovarius* sp. 217) (figura 1.2).

Por otra parte, cabe destacar que la mayoría de genomas estudiados se encontraban en estado de boceto, pero si se tienen en cuenta los 14 genomas que estaban completos (ver tabla S1.1) se pudo observar que los genes implicados en la rama del protocatecuato de la vía del beta-cetoadipato se encontraban generalmente en el cromosoma. Las excepciones fueron *Ruegeria pomeroyi* DSS-3^T, en el que los genes *chp*, *pcaCHG*, *pobA* y *pcaQ* se encontraban en un megaplásmido, tal y como ya había descrito anteriormente Buchan y colaboradores (2004), y en las cepas industriales *K. vulgare* Y25 y *K. vulgare* WSH-001, en las que la totalidad de genes *pca* detectados se encontraban en un plásmido. Respecto a la organización de los genes se observó una gran variabilidad de tipos de agrupaciones génicas (figura 1.3). Tal y como ya habían descrito Harwood y Parales (1996) para microorganismos aislados del suelo, los genes *pcaGH* codificantes para la enzima principal de la vía se encontraban siempre juntos y orientados en el mismo sentido. Además, junto a *pcaGH* se solía encontrar *pcaC* formando la estructura *core pcaCHG* (figura 1.3) que ya había sido observada por Buchan y colaboradores (2004) en ocho roseobacters. Los dos únicos genomas en los que no se encontró esta estructura fueron *Roseobacter* sp. CCS2 (en el que en cambio se encontró el gen *pcaL* adyacente a *pcaGH*) y *Sulfitobacter mediterraneus* 1FIGIMAR09 en el que el gen *pcaC* se encontró agrupado con *pcaB* y *pcaD* (figura 1.4). Partiendo de este *core pcaCHG* se pudieron establecer hasta 25 estructuras génicas únicas con diversos grados de complejidad según la disposición del resto de genes. Tan solo 7 de estas estructuras eran comunes en varios genomas, que generalmente correspondían a bacterias que pertenecían al mismo género o a géneros cercanos. Un ejemplo, sería la estructura que presentan los miembros del género *Phaeobacter* (figura 1.3). Aunque también se encontraron excepciones como bacterias filogenéticamente distantes que tenían la misma estructura génica. Un ejemplo sería *R. pomeroyi* DSS-3^T del grupo G1 de Roseobacter que compartía la misma organización de los genes *pca* que *S. stellata* E-37^T y *Tropicibacter naphthalenivorans* DSM 19561^T del grupo G3 y que *Jannaschia* sp. CSS1 del grupo G5 (figura 1.3). Cabe destacar que también se observó diferente organización génica en las dos agrupaciones de genes *pcaCHG* que presentaban *Roseobacter* sp. MED193 y *L. vestfoldensis* SKA53 (figura 1.3).

Por lo que respecta a la localización de otros genes de la ruta en relación al grupo génico *pcaCHG*, en 29 genomas, incluyendo las dos agrupaciones de genes alrededor de *pcaCHG* de *Roseobacter* sp. MED193 y *L. vestfoldensis* SKA53, se observó el gen regulador *pcaQ* junto a los genes *pcaCHG*. PcaQ es un regulador de la familia LysR que activa la transcripción de los genes *pca* en alfaproteobacterias como *Agrobacterium tumefaciens* A348 y varias especies de *Rhizobium* (Parke, 1996). En 25 estructuras génicas *pcaQ* se encontraba corriente arriba y en sentido contrario a los genes *pcaCHG,* en concordancia con los resultados descritos anteriormente (Buchan e*t al.*, 2004; Gulvik y Buchan, 2013).

El gen *chp* se detectó junto a los genes *pcaGH*, en concreto junto a *pcaG*, y en la misma orientación en 30 genomas (figura 1.3). Esto sugería que en todos estos roseobacters este gen

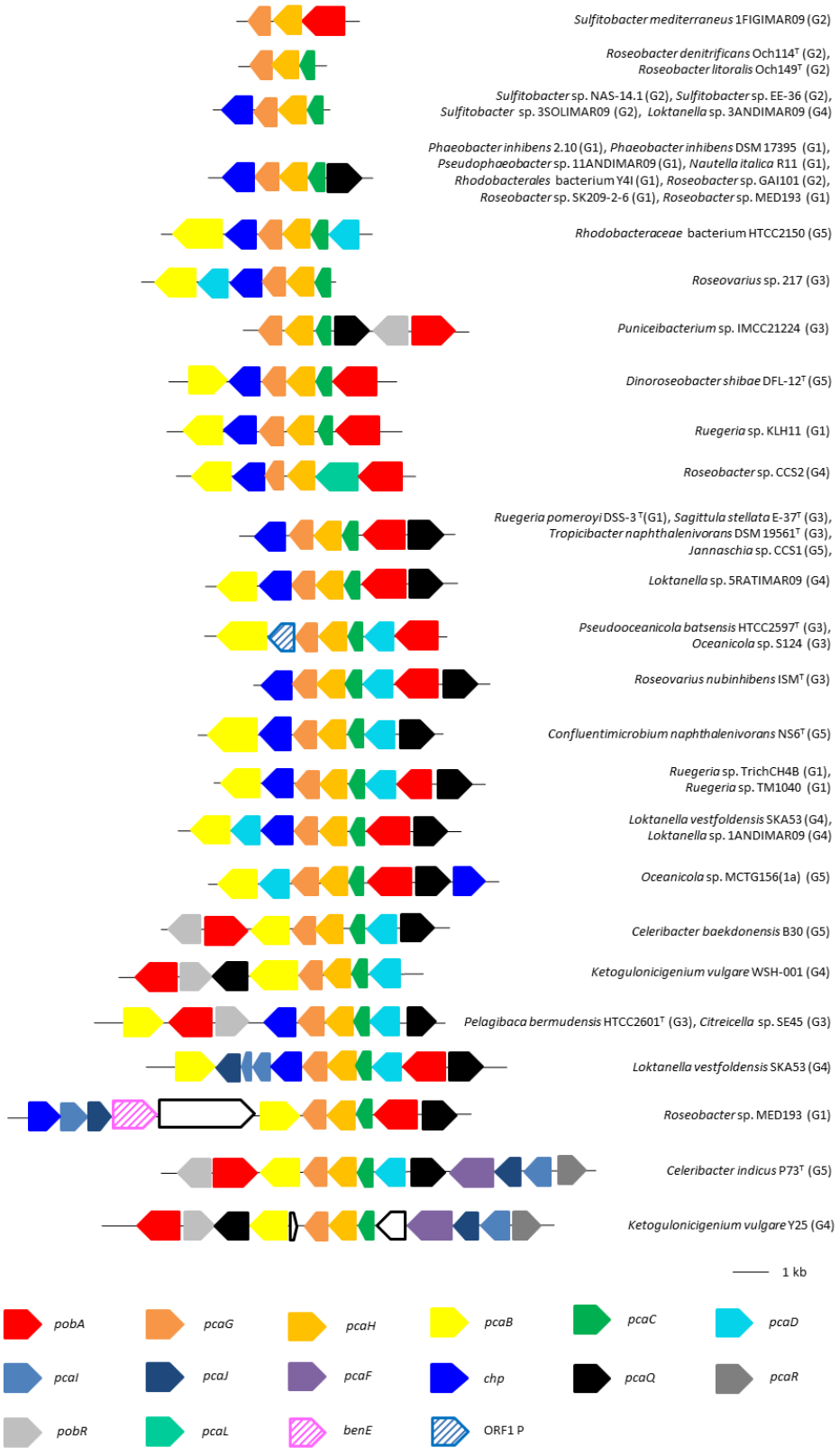

Figura 1.3. Organización de los genes de la ruta del protocatecuato de la vía del beta-cetoadipato en los genomas del linaje Roseobacter. Las abreviaturas son las mismas que las de la figura 1.1 junto a: *pcaL,* gen para 3-oxodadipato enol-lactona hidrolasa/4-carboximuconolactona descarboxilasa; *chp,* gen para proteína hipotética conservada (BKACE G4); *pcaQ*, *pcaR*, *pobR* y *pcaR,* genes para reguladores transcripcionales; *benE,* gen para un transportador de benzoato y ORF1 P, proteína codificante para una esterasa C. Los genes de color blanco y contorno negro codifican para proteínas hipotéticas. El texto entre paréntesis indica los grupos genómicos.

probablemente se transcribía a la vez que los genes *pca*, tal y como ya habían descrito Buchan y colaboradores (2004) en *R. pomeroyi* DSS-3^T. En otras ocasiones el gen *chp* se encontró separado de los genes *pcaCHG,* como por ejemplo en en *Oceanicola* sp. MCTG156(1a) y en una de las dos agrupaciones de genes de *Roseobacter* sp. MED193 (figura 1.3). También se detectó el gen *chp* en regiones diferentes del genoma, sobre todo en aquellos genomas que presentaban la rama del protocatecuato de la vía del beta-cetoadipato incompleta, como por ejemplo en *O. guishaninsula* JLT2003^T.

En cuanto a *pcaD*, este se encontró en la misma agrupación génica que el *core pcaCHG* en 14 de las estructuras génicas detectadas en 16 genomas de roseobacters, y en 13 de estos genomas se encontró corriente arriba del *core pcaCHG* y en la misma orientación (figura 1.3). Respecto a *pcaB*, se encontró junto en la misma agrupación que *pcaCHG* en 19 de las estructuras génicas detectadas en 21 genomas de roseobacters. En concreto, *pcaB* se localizó corriente abajo de los genes *pcaCHG* en 20 genomas y, en la mayoría de casos, orientado en el mismo sentido que los genes del *core pcaCHG* y junto a *chp* (en caso de que estuviera presente). En algunos genomas se encontraron uno o varios genes entre *pcaB* y *chp,* como por ejemplo *pcaD* o los genes *pob* (figura 1.3). Por otra parte, los genes *pcaD* y *pcaB* también se podían encontrar separados de los genes *pcaCHG*, bien juntos o con un gen entre ellos que podía ser un gen para una esterasa, o incluso por *pcaC,* como en *Sulfitobacter mediterraneus* 1FIGIMAR09 (figura 1.4).

Otra estructura génica detectada en los genomas de Roseobacter fue la formada por los genes *pcaI* y *pcaJ* codificantes para la beta-cetoadipato succinil-CoA transferasa, que casi siempre se encontraban separados del *core pcaCHG* a excepción de *K. vulgare* Y25, *Celeribacter indicus* P73^T y en una de las dos copias de genes *pca* de *L. vestfoldensis* SKA53 y *Roseobacter* sp. MED193 (figura 1.3). La presencia de la agrupación *pcaIJ* en una región distinta a *pcaCHG* en los genomas de nuestro estudio ya había sido descrita por Jiménez y colaboradores (2002) en genomas de bacterias no marinas. En los genomas analizados esta agrupación *pcaIJ* podía encontrarse en una región del genoma separada del resto de genes *pca* o junto a los genes *pcaF* y *pcaR* formando las agrupaciones *pcaRIJF*, *pcaRIJ, pcaIJF y pcaFIJ* (figura 1.4). La agrupación *pcaRIJF* no había sido descrita anteriormente en los miembros del linaje Roseobacter pero sí se había encontrado en otras alfaproteobacterias como *Agrobacterium tumefaciens* C58 y en el megaplásmido pSymB de *Sinorhizobium meliloti* (Jiménez *et al*., 2002; MacLean *et al*., 2006). PcaR es un regulador de la familia IclR que puede actuar como activador de la transcripción de varios genes *pca* en *Pseudomonas* (Jiménez *et al*., 2002; Tropel y van der Meer, 2004). Por lo tanto, al encontrar la agrupación *pcaRIJF* en algunos genomas del linaje Roseobacter se puede hipotetizar que los genes *pcaIJF* se encontrarían regulados por PcaR en dichas bacterias. En concreto, se detectó la agrupación *pcaIJF* precedida por el gen regulador *pcaR* orientado en sentido contrario en cuatro genomas de roseobacters (figura 1.4). Cabe destacar que la agrupación *pcaRIJF* también pudo ser identificada en el plásmido de *K. vulgare* Y25, aunque en este caso se encontraba junto a la agupación génica que contenía el *core pcaCHG* (figura 1.3). Por lo tanto, en base a los resultados obtenidos, las dos estrategias que presentarían los miembros del linaje Roseobacter para degradar beta-cetoadipato, mediante PcaIJ y mediante el producto del gen *chp*, no estarían bajo el mismo control regulatorio. El gen

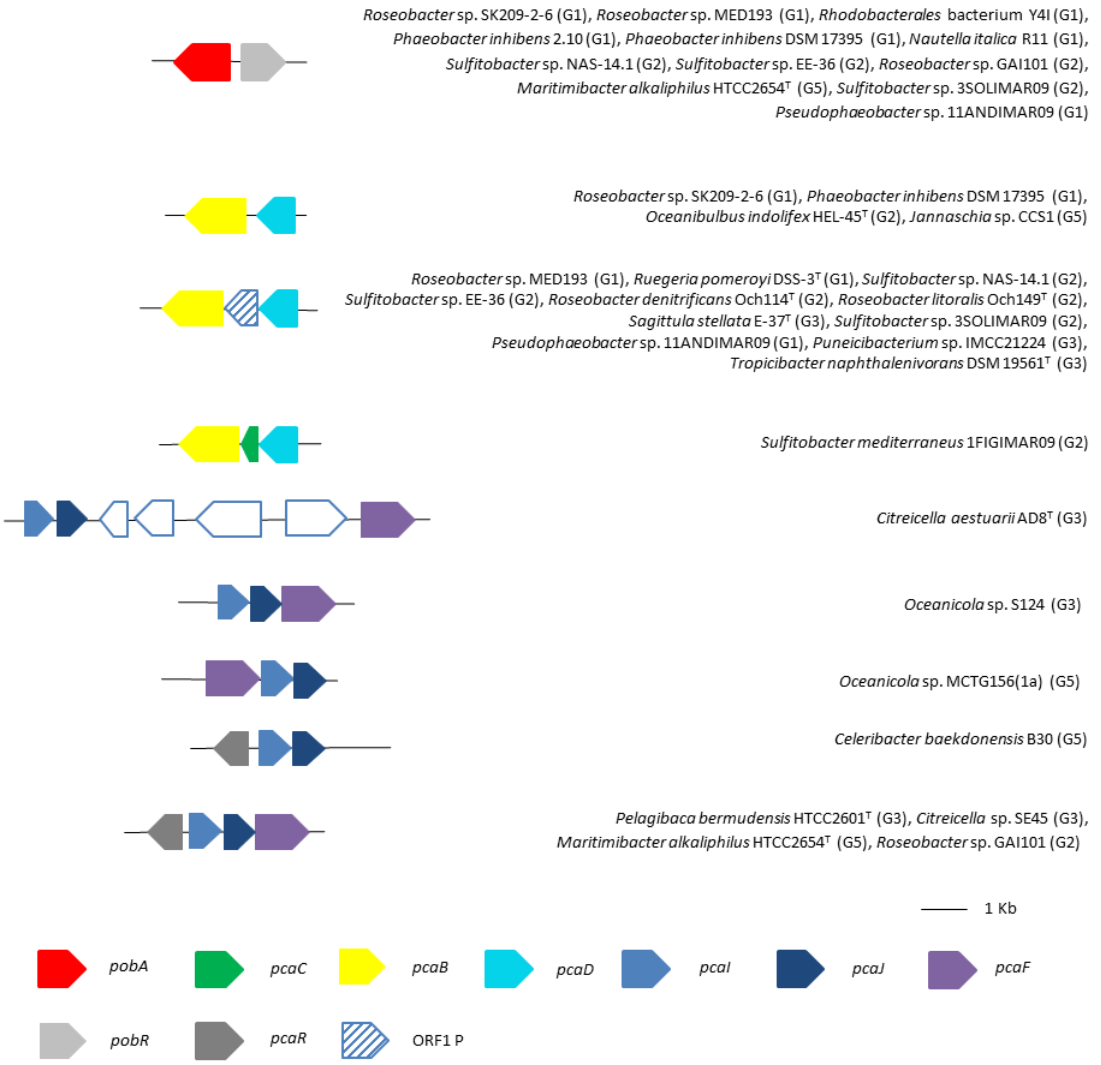

Figura 1.4. Organización de los genes *pca* en regiones separadas del *core pcaCHG* en los genomas del linaje Roseobacter. Las abreviaturas son las mismas que las de la figura 1.1. y 1.3. Los genes de color de blanco con contorno azul hacen referencia a genes que se observan ocasionalmente, pero que no se encuentran conservados ni codifican para proteínas implicadas en la degradación de compuestos aromáticos. El texto entre paréntesis indica los grupos genómicos.

chp suele encontrarse junto al *core pcaCHG* y bajo el control de PcaQ, mientras que, los genes *pcaIJ* suelen encontrarse en una región distinta del genoma, en muchas ocasiones sin un regulador cercano y en otras bajo el control de PcaR. La agrupación de los genes *pcaIJ junto a pcaF* se encontró en tres genomas de Roseobacter, en cada uno de los cuales la organización génica fue diferente. En *C. aestuarii* AD8^T (figura 1.4) los genes *pcaIJ* se encontraban corriente arriba de *pcaF* y separados de este por cuatro genes implicados en la síntesis de aminoácidos y en el transporte de membrana (figura 1.4). En *Oceanicola* sp. S124 y *Oceanicola* sp. MCTG156(1a) estos tres genes se encontraban juntos aunque la posición de *pcaF* difería respecto a *pcaIJ* (figura 1.4). En cuanto a la presencia de la agrupación *pcaRIJ* de *C. baekdonensis* B30 y de la agrupación *pcaIJF* en *Oceanicola* sp. S124, cabe destacar que se encontraban al final de un *contig* con lo cual no se puede confirmar si formaban parte o no de una agrupación génica más extensa (figura 1.4). En cuanto a *pcaF*, si no estaba formando parte de agrupaciones anteriormente comentadas siempre se encontraba en otras regiones del genoma. Además, aunque la mayor parte de *pcaF* identificados en nuestro estudio tuvieron

más de un 50 % de identidad con la proteína PcaF modelo de *C. aestuarii* SE45, todos ellos también presentaron identidad suficiente con el gen *paaJ,* codificante para 3-oxoadipil-CoA tiolasa de la vía de degradación del fenilacetil-CoA (ver apartado 4). Esta situación se dio porque hay una alta identidad de secuencia entre PcaF y PaaJ (>50 %), ya que son isoenzimas (Nogales *et al*., 2007). De hecho, 24 de los genes *pcaF* separados de los genes *pca* que se han observado en este estudio se encontraban agrupados con los genes para la degradación del fenilacetil-CoA. Se ha probado experimentalmente que el producto de *paaJ* en *E. coli* puede complementar a un mutante de *pcaF* de *P. putida* KT2440 (Nogales *et al*., 2007). Es por ello que en esta tesis se ha considerado en estos 24 genomas de Roseobacter habría potencial génico para el paso final de la ruta de rotura *orto* del protocatecuato.

En cuanto a la presencia del gen *pobA*, en caso de encontrarse en la misma estructura génica que *pcaHG,* este se localizó corriente arriba de estos genes en 19 genomas del linaje Roseobacter y adyacente a *pcaQ* cuando este último estaba presente (figura 1.3). Esto sugeriría que, tanto los genes *pca* como *pobA* se encontrarían regulados por PcaQ en bacterias con este tipo de agrupaciones génicas. En seis genomas de Roseobacter se pudo observar como el gen *pobA* se encontraba corriente abajo de los genes *pcaCHG* y junto a otro gen regulador, *pobR* (figura 1.3). En estos casos podría postularse que *pobA* no se encontraría regulado por PcaQ sino por PobR. En doce genomas del linaje Roseobacter *pobA* se encontraba junto al regulador *pobR* pero alejado de la agrupación de genes alrededor del *core pcaCHG* (figura 1.4). Por lo tanto, se puede proponer que los miembros del linaje Roseobacter presentarían dos estrategias distintas para regular *pobA*. La primera consistiría en localizar el gen *pobA* entre el gen regulador *pcaQ* y los genes *pca(D)CHG* de forma que *pobA* estuviera regulado por PcaQ. De esta manera, se aseguraría una expresión coordinada de la proteína implicada en la degradación del 4-hidroxibenzoato (PobA) y las proteínas implicadas en la degradación de su producto, el protocatecuato. En la segunda estrategia se podría potencialmente ajustar la expresión de *pobA* y de los genes *pca(D)CHG* de forma diferente al estar controlada por dos reguladores distintos, PcaQ y PobR, lo que podría dar como resultado una mayor flexibilidad metabólica.

Con los resultados de este estudio ha quedado patente que en los miembros del linaje Roseobacter los genes *pca* no se organizan en una sola estructura génica tal y como se había descrito en *Acinetobacter calcoaceticus* o *Agrobacterium tumefaciens* (Harwood y Parales, 1996; Parke, 1997) sino que se organizaban en varias agrupaciones distintas. Esta variabilidad a la hora de organizar los genes *pca* es común en bacterias puesto que las distintas presiones selectivas han dado lugar a diferencias en la organización génica, en las moléculas inductoras, en los reguladores y en los sistemas de transporte involucrados en la rama del protocatecuato de la vía del beta-cetoadipato (Harwood y Parales, 1996; Jiménez *et al*., 2002). Ahora bien, no se pudo establecer una relación entre el tipo de organización de los genes *pca* y los cinco grupos filogenómicos de Roseobacter (figura 1.3). Por lo tanto, se puede presuponer que mecanismos genéticos como la transferencia génica horizontal (TGH) o reordenamientos del ADN han podido contribuir a la aparición de las diferentes agrupaciones génicas observadas (figuras 1.3 y 1.4).

Finalmente, la presencia de genes codificantes para las enzimas de la rama del protocatecuato de la vía del beta-cetoadipato también se analizó en cuatro genomas de roseobacters no

cultivados obtenidos por tecnología de secuenciación SCG: AAA076-C03, AAA015-O19, AAA298-K06, y AAA300-J04, cuyos genomas tenían una cobertura del 76, 55, 48 y 23 % respectivamente (Swan *et al*., 2013; Luo *et al*., 2014). Solo se pudieron detectar con el criterio del 30 % de identidad con un *E-value* menor de 10^{-5} los genes *pcaG*, *pcaH*, *pcaI*, *pcaJ*, *pcaF* y *chp* en AAA015-O19 y en AAA298-K06 (en el cual no se detectó *pcaJ*), y los genes *pcaF* y *chp* en AAA076-C03. Posibles explicaciones al hecho de no detectar con certeza estos genes en este tipo de genomas son por un lado que los miembros no cultivados del linaje Roseobacter podrían poseer genes divergentes respecto a las que presentaban los miembros cultivados del linaje y, por ello, no se pudieron detectar con los criterios utilizados en este estudio. Por otro lado, estos resultados también llevan a hipotetizar que la rama del protocatecuato de la vía del beta-cetoadipato no sería tan relevante en las poblaciones no cultivadas de este grupo. Por lo tanto ambas posibilidades ponen en evidencia que hay diferencias a nivel de potencial catabólico entre los miembros del linaje Roseobacter cultivados y no cultivados, tal y como ya se habían descrito anteriormente para otras características del linaje (Luo *et al*., 2014; Zhang *et al*., 2016), lo que redunda en la necesidad de realizar más estudios sobre ello.

2.1.2. Rama del catecol de la vía del beta-cetoadipato

La rama del catecol de la vía del beta-cetoadipato (figura 1.1) ha sido bien caracterizada en bacterias de los géneros *Pseudomonas*, *Burkholderia*, *Cupriavidus* y *Ralstonia* (Harwood y Parales, 1996; Jiménez *et al*., 2002; Pérez-Pantoja *et al*., 2012), pero se consideraba que no estaba presente en las bacterias del linaje Roseobacter (Buchan y González, 2010). Esta rama está formada por reacciones catalizadas por tres enzimas: catecol 1,2-dioxigenasa (CatA), una muconato cicloisomerasa (CatB) y muconolactona isomerasa (CatC). Los genes que codifican para estas enzimas en *Pseudomonas* suelen encontrarse agrupados formando la agrupación génica *catRBCA*, junto con el gen *catR* codificante para un regulador transcripcional de tipo LysR (Cámara *et al*., 2007) también llamado *catM* (Cosper *et al*., 2000). El producto de degradación que se origina, beta-cetoadipato enol-lactona, puede ser transformado por acción de CatD, CatI, CatJ y CatF hasta intermediarios del ciclo de Krebs (Cosper *et al*., 2000). Las proteínas CatD, CatI, CatJ y CatF presentan una elevada identidad de secuencia con PcaD, PcaI, PcaJ y PcaF pertenecientes a la rama del protocatecuato de la vía del beta-cetoadipato (figura 1.1) y su función puede ser substituida por la acción de estas proteínas de la rama del protocatecuato (Göbel *et al*., 2002). En varios microorganismos, como *Cupriavidus necator* JMP134 (Pérez-Pantoja *et al*., 2008) y *Acinetobacter* sp. ADP1 (Cosper *et al*., 2000), se ha detectado la presencia de los genes codificantes para las proteínas implicadas en la degradación de benzoato junto a los genes para la degradación de catecol. La degradación del benzoato a catecol se produce mediante dos reacciones: el paso de benzoato a ácido *cis*-1,6-dihidroxi-2,4-ciclohexadieno-1-carboxílico mediante la acción de una benzoato 1,2-dioxigenasa (BenABC) y el paso de ácido *cis*-1,6-dihidroxi-2,4-ciclohexadieno-1-carboxílico a catecol mediante la benzoato *cis*-diol deshidrogenasa (BenD) (Cosper *et al*., 2000).

En este estudio se pudo determinar la presencia de los genes *catA* y *catB* en nueve de los genomas del linaje Roseobacter estudiados, pero en ninguno de ellos se pudo determinar la presencia de *catC* con los criterios de búsqueda utilizados (tabla S1.2). Estos nueve genomas fueron: *Roseobacter* sp. MED193 (G1), *Rhodobacterales* bacterium Y4I (G1), *Sagittula stellata* E-37^{T} (G3), *Roseovarius* sp. 217 (G3), *Roseovarius* sp. TM1035 (G3), *Pseudooceanicola batsensis*

HTCC2597^T (G3), *Celeribacter indicus* P73^T (G5), *Celeribacter baekdonensis* B30 (G5) y *Confluentimicrobium naphthalenivorans* NS6^T (G5). Además en estos nueve genomas también se habían detectado los genes *pcaD*, *pcaI*, *pcaJ* y *pcaF* (figura 1.2), por lo que las reacciones de la parte final de la ruta desde beta-cetoadipato enol-lactona (el producto de PcaC y CatC) serían posibles. En algunas bacterias se ha descrito la confluencia de las dos ramas de la ruta del beta-cetoadipato (protocatecuato y catecol), pero esta confluencia se produce con posterioridad a la reacción catalizada por CatC (Harwood y Parales, 1996). Por lo tanto, en base a los resultados de esta tesis ninguno de los genomas de Roseobacter estudiados presentaba la vía completa para la degradación del catecol a partir de su rotura en posición *orto* debido a la falta de detección del gen codificante para la proteína CatC. Esto no solo podía ser debido a la ausencia del gen sino que también podía deberse a que la proteína fuera divergente a la proteína modelo usada y no pudiera ser detectada con los criterios de detección seguidos. Por lo tanto, se debería realizar un estudio más exhaustivo para confirmar si los miembros del linaje Roseobacter pueden o no presentar una vía completa para la degradación del catecol a partir de su rotura en posición *orto*. Tal y como se puede observar en la figura 1.5 en los nueve genomas mencionados anteriormente *catA* y *catB* se encontraban en la misma agrupación génica y en ocho de estos organismos (excepto en *S. stellata* E-37^T) también se encontró un gen codificante para el regulador transcripcional CatR. Se observaron ocho tipos de estructuras génicas que contenían el gen *catA* y solo en el caso de *Celeribacter baekdonensis* B30 y *Celeribacter indicus* P73^T se pudo considerar que presentaran la misma estructura génica *catABR* (figura 1.5). Además, en estos dos genomas se observó la presencia de genes codificantes para una proteína hipotética y una alfa/beta hidrolasa (ORF1 C) entre los genes *catR* y *catB* (figura 1.5). El gen para la alfa/beta hidrolasa (ORF1 C) también fue detectado junto a los genes *cat* en *Pseudooceanicola batsensis* HTCC2597, *Roseobacter* sp. MED193 y en *Confluentimicrobium naphthalenivorans* NS6^T (figura 1.5). La estructura más sencilla observada fue la de *Rhodobacterales* bacterium Y4I en la que solo se pudo detectar un grupo génico formado por *catABR* mientras que en las estructuras más complejas se encontraron estos genes junto a genes implicados en otras vías para la degradación de compuestos aromáticos (figura 1.5). En *Roseovarius* sp. 217, *Roseovarius* sp. TM1035, *Roseobacter* sp. MED193 y *Confluentimicrobium naphthalenivorans* NS6^T se detectaron los genes para la degradación de benzoato a catecol (*ben*) corriente abajo del gen *catA* (figura 1.5). La presencia de genes *ben* en genomas del linaje Roseobacter no había sido descrita hasta la fecha de realización de esta tesis. Por lo tanto en el linaje Roseobacter habría dos estrategias para degradar el benzoato, mediante la vía del benzoil-CoA, tal y como ya se había descrito anteriormente (Newton *et al*., 2010), y mediante la degradación a catecol. En base a los resultados obtenidos, *Roseovarius* sp. 217, *Roseovarius* sp. TM1035 y *Roseobacter* sp. MED193 podrían degradar potencialmente el benzoato hasta catecol. En el caso de *Confluentimicrobium naphthalenivorans* NS6^T no se pudo determinar si potencialmente podía degradar el benzoato hasta catecol puesto que presentaba un gen *benA* truncado entre *catB* y *benB* (figura 1.5), cuyo producto presentaba menos de un 30 % de identidad de secuencia con la proteína BenA modelo. Si esto es así, no tendría sentido que en estas bacterias se produjese la degradación de benzoato vía catecol para después tener una vía truncada de degradación del catecol por falta del gen *catC* (no detectado en este estudio). Por lo tanto, lo más probable es que estas bacterias tengan una enzima que realice la función de CatC, aunque no se haya detectado con los criterios utilizados en este estudio.

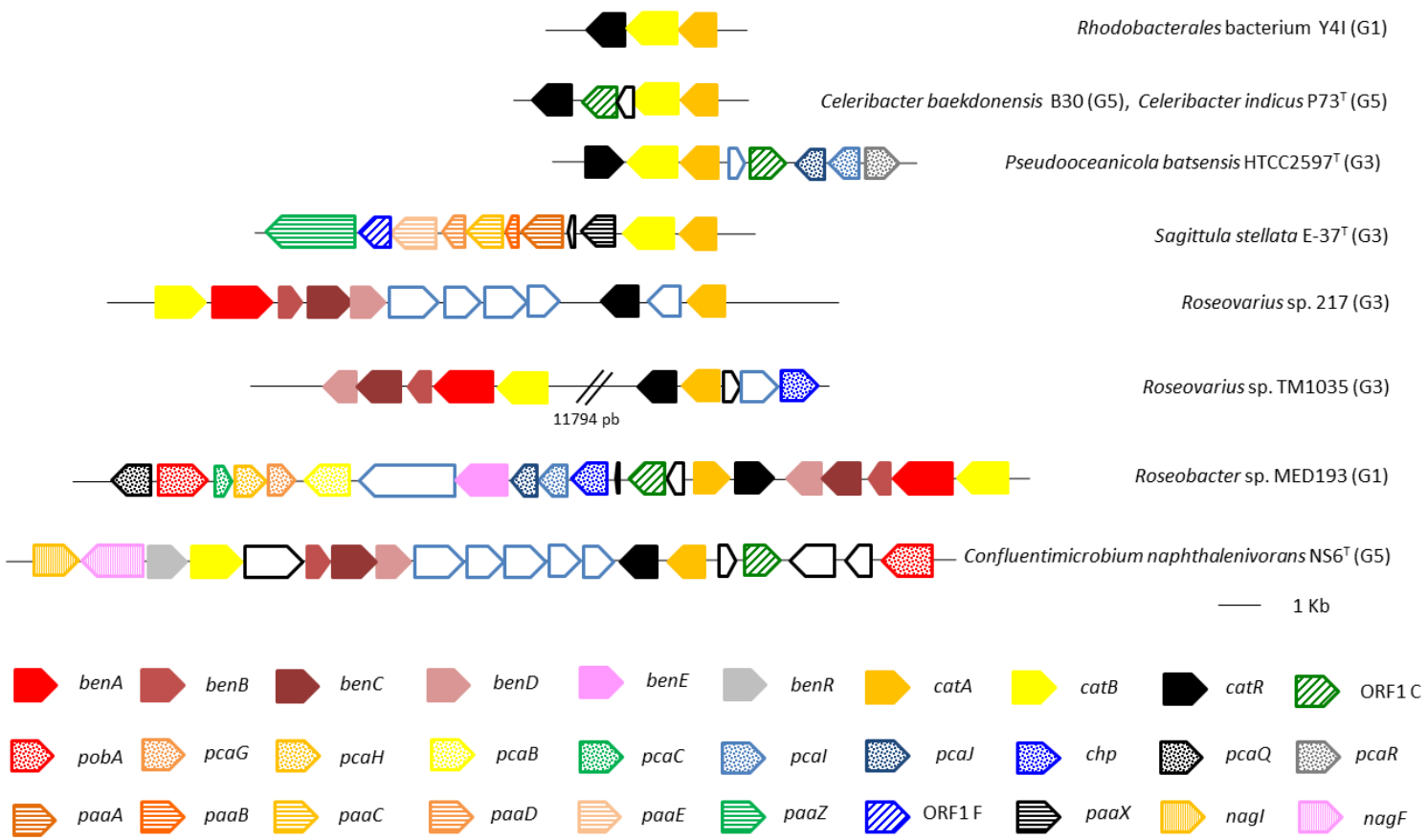

Figura 1.5. Organización de los genes de la rotura *orto* del catecol en los genomas del linaje Roseobacter. Las abreviaturas son las mismas que las de la figura 1.1 y 1.3, junto a: *catR*, regulador transcripcional del operón *cat*; *benABC,* benzoato 1,2-dioxigenasa; *benD*, 1,6-dihidroxiciclohexa-2,4-dieno-1-carboxilato deshidrogenasa; *benE*, proteína de membrana transportadora de benzoato de la superfamilia MFS; *benR*, regulador transcripcional y ORF1 C, alfa/beta hidrolasa. Las abreviaturas de los genes codificantes para las enzimas de la vía del fenilacetatil-CoA (genes *paa*) y gentisato (genes *nag*) se encuentran en las figuras 1.13 y 1.20, respectivamente. Los genes de color de blanco y contorno negro codifican para proteínas hipotéticas y los genes de color blanco con contorno azul se observaron ocasionalmente, pero no se encuentran conservados, ni codifican para proteínas implicadas en la degradación de compuestos aromáticos. El texto entre paréntesis indica los grupos genómicos.

Por otra parte, corriente arriba del gen *catA* se encontaron genes codificantes para las enzimas de la rama del protocatecuato de la vía del beta-cetoadipato en *Pseudooceanicola batsensis* HTCC2597^T, *Roseovarius* sp. TM1035 y *Roseobacter* sp. MED193, y el gen *pobA* en *C. naphthalenivorans* NS6^T (figura 1.5). Finalmente, corriente abajo del gen *catA* se detectó la presencia de los genes codificantes para las enzimas de la vía del fenilacetil-CoA (genes *paa*) en *S. stellata* E-37^T y de los genes codificantes para las enzimas de la vía del gentisato (genes *nag*) en *Confluentimicrobium naphthalenivorans* NS6^T (figura 1.5). Por lo tanto en *Roseovarius* sp. TM1035, *Roseobacter* sp. MED193 y *Confluentimicrobium naphthalenivorans* NS6^T los genes *cat* se encontraban en la misma agrupación que los genes codificantes para enzimas de al menos dos vías distintas de degradación de compuestos aromáticos, siendo la vía de degradación del benzoato (genes *ben*) siempre una de ellas.

2.2. Rotura *meta*

2.2.1. Rotura *meta* de protocatecuato

La degradación del protocatecuato a partir de la rotura del anillo bencénico en posición *meta* se ha descrito en microorganismos de varios géneros como *Pseudomonas*, *Burkholderia*, *Cupriavidus* y *Ralstonia* (Pérez-Pantoja *et al*., 2010b). En los miembros del linaje Roseobacter este tipo de degradación se consideraba poco frecuente puesto que hasta la realización de este estudio solo se había detectado la presencia de los genes implicados en la rotura *meta* del protocatecuato en *Maritimibacter alkaliphilus* HTCC2564^T (Buchan y González, 2010). Hay cinco enzimas implicadas en la degradación bacteriana del protocatecuato por rotura en posición *meta* hasta dar lugar a intermediarios del ciclo de Krebs (figura 1.1). La enzima principal es la protocatecuato 4,5-dioxigenasa (LigAB) que es una enzima dimérica que pertenece a la superfamilia LigB de las extradiol dioxigenasas (Hara *et al*., 2003; Pérez-Pantoja *et al*., 2010b). A continuación, actúan las enzimas 2-hidroxi-4-carboximuconato-6-semialdehído deshidrogenasa (LigC), 2-pirona-4,6-dicarboxilato hidrolasa (LigI), 4-oxalomesaconato hidratasa (LigJ) y finalmente 4-carboxi-4-hidroxi-2-oxoadipato aldolasa (LigK) (figura 1.1) (Hara *et al*., 2003).

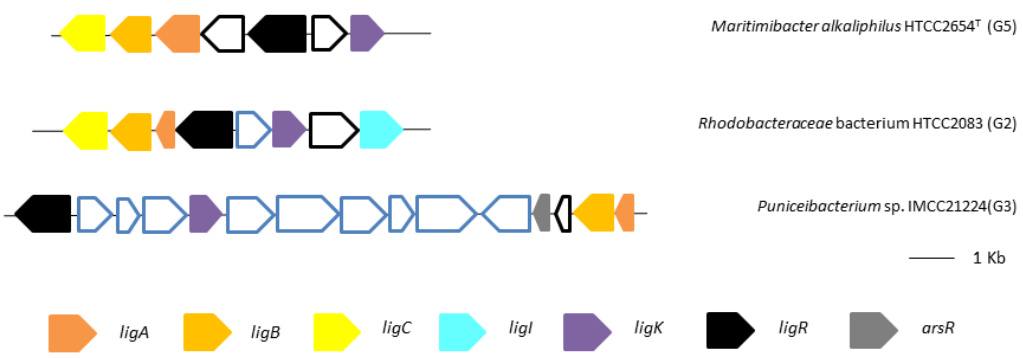

Figura 1.6. Organización de los genes de la rotura *meta* del protocatecuato en los genomas del linaje Roseobacter. Las abreviaturas son las mismas que las de la figura 1.1 junto a: *ligR,* gen para regulador de los genes *lig* y *asnR,* gen para regulador transcripcional de tipo AsnR. Los genes de color blanco y contorno negro codifican para proteínas hipotéticas y los genes de color blanco con contorno azul se observaron ocasionalmente, no se encuentran conservados, ni codifican para proteínas implicadas en la degradación de compuestos aromáticos. El texto entre paréntesis indica los grupos genómicos.

En este estudio se detectó la presencia de los genes codificantes para las proteínas LigAB en 3 miembros del linaje Roseobacter: *Maritimibacter alkaliphilus* HTCC2654^T, *Rhodobacteraceae* bacterium HTCC2083 y *Puniceibacterium* sp. IMCC21224, así como el gen *ligC* en los dos primeros. En ninguno de estos tres genomas se pudo detectar la totalidad de los genes de la ruta completa (tabla S1.2). En *Rhodobacteraceae* bacterium HTCC2083 y *M. alkiliphilus* HTCC2564^T no se detectó *ligJ* (tabla S1.2) y en *Puniceibacterium* sp. IMCC21224 no se detectó la presencia de *ligC* y *ligI* (tabla S1.2). Por lo tanto, aparentemente, ninguno de los tres presentaba esta vía degradativa completa, a excepción de que presentaran proteínas que pudieran suplir la función de las proteínas codificadas por estos genes y que no hubiera sido detectadas debido a su divergencia con las proteínas modelo usadas.

En cada uno de estos genomas se encontró una organización génica distista alrededor de los genes *ligAB*. Como se puede observar en la figura 1.6, en los tres microorganismos los genes *ligAB* aparecieron juntos y orientados en el mismo sentido y con el gen *ligR*, codificante para un regulador transcripcional, en la misma agrupación génica. Este tipo de organización génica se había documentado en otros grupos bacterianos como *Sphingomonas* (Maruyama *et al*., 2004). Respecto al resto de genes de la ruta, *ligC* se encontraba corriente abajo de *ligB* y en el mismo sentido de los genes *ligAB* en *M. alkaliphilus* HTCC2654^T y *Rhodobacteraceae* bacterium HTCC2083 (figura 1.6). Respecto a *ligK* y *ligR*, estos se encontraban corriente arriba de *ligAB* en *M. alkaliphilus* HTCC2654^T y *Rhodobacteraceae* bacterium HTCC2083, mientras que en *Puniceibacterium* sp. IMCC21224 estos se encontraban corriente abajo de *ligAB* (figura 1.6). Además, en los tres organismos *ligR* se encontraba en la misma orientación que *ligAB* mientras que *ligK* se encontraba en sentido contrario (figura 1.6). Finalmente, *ligI* se encontraba corriente arriba y en sentido contrario a *ligAB* en *Rhodobacteraceae* bacterium HTCC2083, y aislado del resto de genes *lig* en *M. alkaliphilus* HTCC2654^T (figura 1.6).

2.2.2. Rotura *meta* de catecol

La rotura del catecol en posición *meta* la realiza una catecol 2,3 dioxigenasa (XylE) que pertenece a la familia de las extradiol dioxigenasas del oxígeno quelado adyacente de tipo I, EXDO I (Pérez-Pantoja *et al*., 2010b). El producto, 2-hidroximucónico-semialdehído, es degradado hasta intermediarios del ciclo de Krebs por la acción de siete enzimas (figura 1.1). Esta vía de degradación del catecol se ha descrito en varias familias bacterianas como *Pseudomonadaceae*, *Burkholderiaceae* y *Comamonadaceae* (Aemprapa y Williams, 1998; Pérez-Pantoja *et al*., 2012) aunque hasta la fecha no se había detectado su presencia en los miembros del linaje Roseobacter (Buchan y González, 2010).

En este estudio solo se pudo detectar la presencia del gen codificante para la enzima clave, catecol 2,3 dioxigenasa (XylE), en los genomas de *Roseobacter* sp. GAI101, *Citreicella* sp. SE45 y *S. stellata* E-37^T. En otros genomas se detectaron los genes *xylG* y *xylQ* (*Pseudophaeobacter sp.* 11ANDIMAR09), y los genes *xylK* y *xylQ* en *Octadecabacter arcticus* 238^T, *C. indicus* P73^T, *Confluentimicrobium naphthalenivorans* NS6^T, y en las cuatro cepas de *C. aestuarii*: AD8^T, 357, 328 y 329 (tabla S1.2). Además, con el criterio de detección del 30 % de identidad de secuencia y un *E-value* menor de 10^{-5} con las proteínas modelo se determinó la presencia de *xylI* y *xylJ* en *Pseudophaeobacter* sp. 11ANDIMAR09 y de *xylF* y *xylJ* en *Confluentimicrobium naphthalenivorans* NS6^T. Sin embargo, en ninguno de los genomas se detectaron los genes

codificantes para todas las enzimas de esta ruta (tabla S1.2). Por lo tanto, en base a los resultados obtenidos, y con las limitaciones de los criterios utilizados en este estudio, puede decirse que los miembros del linaje Roseobacter analizados no presentaban la vía de degradación del catecol por rotura en posición *meta*.

En los genomas donde estaba presente, el gen para catecol 2,3 dioxigenasa estaba aislado del resto de genes para degradación de compuestos aromáticos. En *S. stellata* E-37^T y *Citreicella* sp. SE45 el gen codificante para XylE se encontraba junto a un gen codificante para un regulador de tipo TetR orientado en sentido contrario, lo que sugería que estaba bajo el control de este gen regulador. Este tipo de agrupación también se observó en *Rhodobacter sphaeroides* 2.4.1, el miembro del orden *Rhodobacterales* que se había tomado como modelo para la búsqueda del gen codificante para XylE (tabla S1.2), aunque no había sido descrito en otros microorganismos en los que se había estudiado esta ruta como *Comamonas* sp. JS765 o *Acidovorax* sp. JS42 (Pérez-Pantoja *et al*., 2012). En estos dos organismos se observaba un regulador de tipo LysR (XylR) cercano al gen que codifica para C23O/XylE (Pérez-Pantoja *et al*., 2012).

2.3. Rutas periféricas de degradación hacia protocatecuato

2.3.1. Degradación de ferulato y vanillato

Dado el elevado porcentaje de genomas del linaje Roseobacter en que se había propuesto la presencia de genes para la rotura *orto* del protocatecuato se decidió analizar genes para rutas periféricas que confluyen en este compuesto. Además del 4-hidroxibenzoato que se ha comentado previamente (ver apartado 2.1.1.) existen varios compuestos aromáticos precursores del protocatecuato como son el ferulato, la vanillina, el vanillato, el ftalato, el isoftalato y el tereftalato (figura 1.7). Tanto el ferulato como el vanillato son productos de la degradación de la lignina y su degradación esta relacionada puesto que el vanillato se puede formar a partir de la degradación del ferulato (Campillo *et al*., 2014).

Se detectaron los genes codificantes para las cuatro enzimas implicadas en la degradación de ferulato hasta vanillato en 17 genomas del linaje Roseobacter (figura 1.8 y tabla S1.2). En cuanto a la asignación filogenómica de estos 17 genomas, 7 pertenecían al grupo G3 (*Citreicella* sp. SE45, *P. bermudensis* HTCC2601^T, *S. stellata* E-37^T, *T. naphthalenivorans* DSM 19561^T, *Puniceibacterium* sp. IMCC21224, *Oceanicola* sp. S124 y *P. batsensis* HTCC2597^T), 4 pertenecían al G4 (*Loktanella* sp. 5RATIMAR09, 1ANDIMAR09 y 3ANDIMAR09 y *L. vestfoldensis* SKA53), 3 pertenecían al G5 (*M. alkaliphilus* HTCC2654^T, *C. indicus* P73^T y *C. baekdonensis* B30), 2 al G1 (*Pseudophaeobacter* sp. 11ANDIMAR09 y *Roseobacter* sp. MED193) y 1 al G2 (*Roseobacter* sp. GAI101). Además, también se detectó esta vía incompleta en 10 genomas más, 4 de ellos asignados al grupo filogenómico G5, 3 al grupo G3, 2 al grupo G2 y uno al G1 (tabla S1.2). Estos resultados mostraban que la degradación del ferulato, o al menos de alguno de sus productos de degradación, era una característica extendida en los genomas del linaje Roseobacter. En concreto, se detectó la presencia de los genes *atu1416* codificante para la feruloil-CoA sintasa y *atu1417* codificante para la feruloil-CoA hidratasa en 20 genomas, del gen *atu1415* codificante para la 3-hidroxiacil-CoA deshidrogenasa en 27 genomas, y del gen *atu1421* codificante para la 4-hidroxi-3-metoxifenil-β-cetopropiónico-CoA beta-cetotiolasa en 17 genomas (figura 1.8 y tabla S1.2).

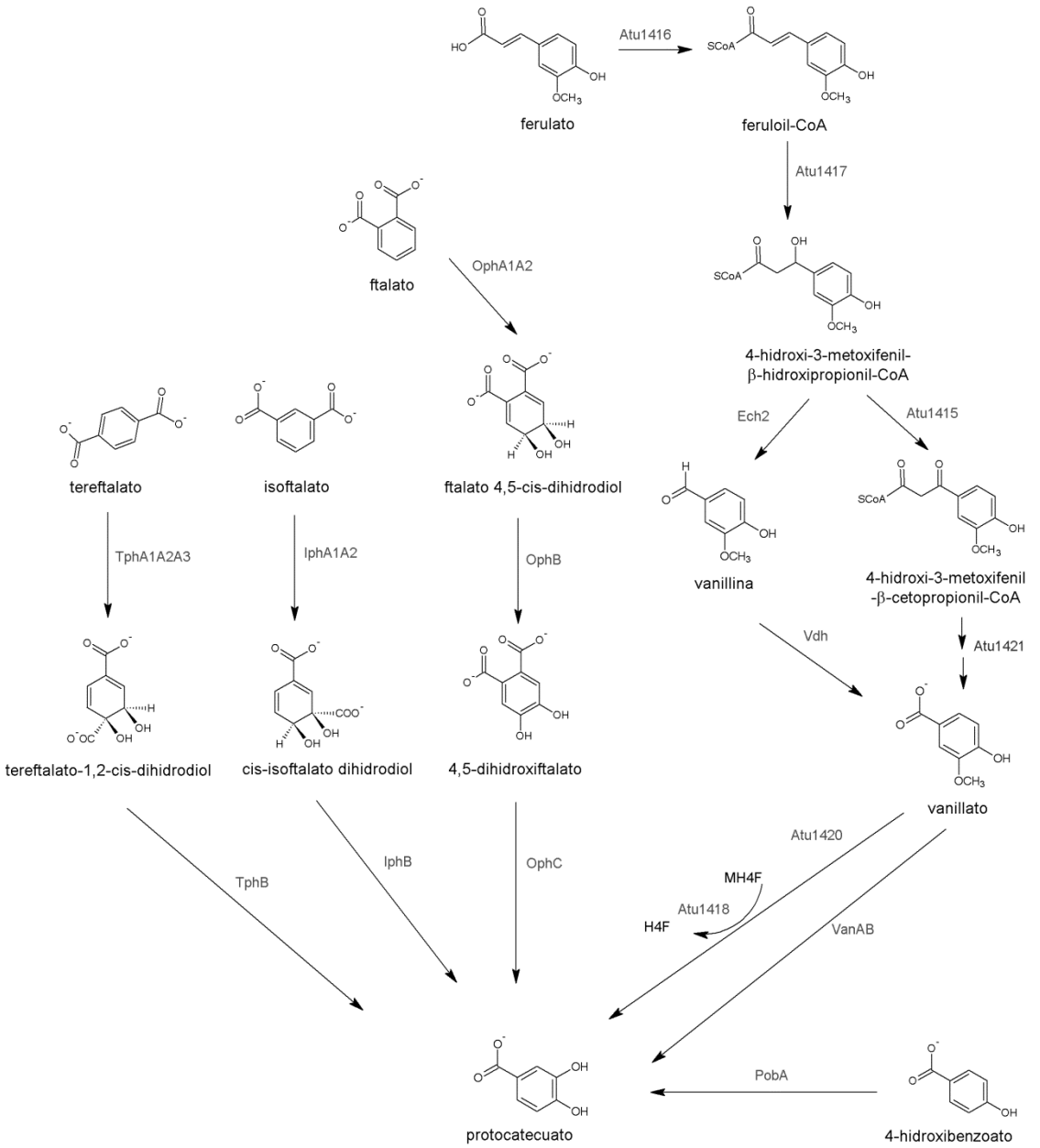

Figura 1.7. Rutas de degradación periféricas que convergen en protocatecuato. Abreviaturas: TphA1A2A3, tereftalato 1,2-dioxigenasa; TphB, tereftalato dihidrodiol deshidrogenasa; IphA1A2, isoftalato 3,4-dioxigenasa; IphB, *cis*-isoftalato dihidrodiol deshidrogenasa; OphA1A2, ftalato 4,5-dioxigenasa; OphB, ftalato 4,5-*cis*-dihidrodiol deshidrogenasa; OphC, 4,5-dihidroxiftalato descarboxilasa; Atu1416, feruloil-CoA sintasa; Atu1417, feruloil-CoA hidratasa; Atu1415, 3-hidroxiacil-CoA deshidrogenasa; Ech2, vanillina sintasa; Vdh, vanillina deshidrogenasa; Atu1421, 4-hidroxi-3-metoxifenil-β-cetopropiónico-CoA beta-cetotiolasa; Atu1420, *o*-desmetilasa; Atu1418, metilenotetrahidrofolato reductasa; VanAB, vanillato *o*-desmetilasa monooxigenasa; PobA, 4-hidroxibenzoato hidroxilasa. Figura elaborada a partir información de Wang *et al*., 1995; Harwood y Parales, 1996; Chang y Zylstra, 1998; Campillo *et al*., 2014.

Por lo que respecta a la degradación del vanillato solo se detectó la presencia del gen *vdh* codificante para vanillina deshidrogenasa en 5 genomas, del gen *atu1420* codificante para una *o*-desmetilasa que degrada el vanillato a protocatecuato en 3 genomas del linaje Roseobacter, y no se detectó la presencia del gen *ech2* codificante para la vanillina sintasa ni los genes *vanA* y *vanB* codificantes para la vanillato *o*-desmetilasa monooxigenasa (figura 1.8 y tabla S1.2). Por lo tanto, la capacidad para generar vanillina y para realizar los pasos de vanillina a vanillato, y

de vanillato a protocatecuato no parecían ser frecuentes en miembros del linaje Roseobacter (figura 1.8 y tabla S1.2).

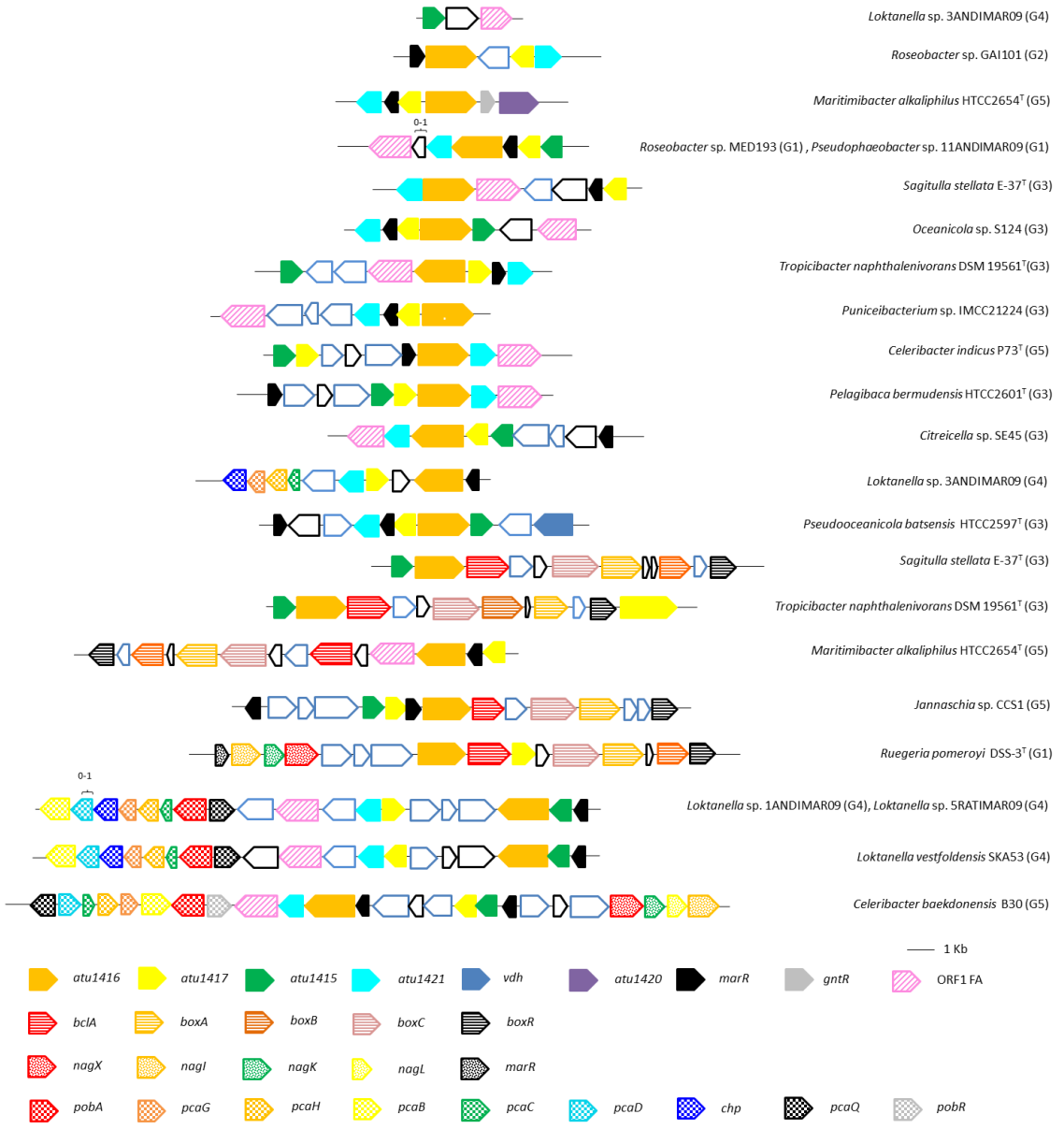

Figura 1.8. Organización de los genes de la vía del ferulato en los genomas del linaje Roseobacter. Las abreviaturas son las mismas que las de las figuras 1.4 y 1.7 junto a: *bclA,* gen para benzoato-CoA ligasa; *boxAB*, gen para benzoil-CoA dioxigenasa; *boxC,* gen para benzoil-CoA dihidrodiol liasa; *boxR,* gen para regulador de los genes *box*; *marR,* gen para regulador transcripcional de los genes *nag* de tipo MarR; *nagX*, gen para 3-hidroxibenzoato 6-hidroxilasa; *nagI,* gen para gentisato 1,2-dioxigenasa; *nagL,* gen para maleilpiruvato isomerasa; *nagK,* gen para fumarilpiruvato hidrolasa, y ORF1 FA, proteína conservada acetil-CoA deshidrogenasa. Los genes de color blanco y contorno negro codifican para proteínas hipotéticas y los genes de color blanco con contorno azul se observaron ocasionalmente, no se encuentran conservados, ni codifican para proteínas implicadas en la degradación de compuestos aromáticos. El texto entre paréntesis indica los grupos genómicos.

Respecto a la ordenación de los genes implicados en la degradación de ferulato, vanillina y vanillato en los genomas en los que se observaron se pudieron observar 21 agrupaciones génicas distintas (figura 1.8). Además, en un mismo genoma estos genes se podían encontrar en varias agrupaciones génicas (figura 1.8). En 10 genomas los genes implicados en la degradación del ferulato se encontraban junto a genes de otras vías de degradación de

compuestos monoaromáticos. En concreto en 5 genomas (*Loktanella* sp. 3ANDIMAR09, *L. vestfoldensis* SKA53, *Loktanella* sp. 1ANDIMAR09, *Loktanella* sp. 5RATIMAR09 y *C. baekdonensis* B30) se encontraban junto a los genes implicados en la degradación del protocatecuato (figura 1.8). Este resultado se corresponde con lo esperado puesto que el ferulato es un precursor del protocatecuato y es común observar los genes implicados en la degradación de un compuesto central junto a los genes implicados en la degradación de los compuestos periféricos que confluyen en ese compuesto central (Parke y Ornston, 2003). En otros 5 genomas (*S. stellata* E-37^T, *M. alkaliphilus* HTCC264^T, *R. pomeroyi* DSS-3^T, *Jannaschia* sp. CSS1 y *T. naphthalenivorans* DSM 19561^T) los genes para la degradación del ferulato se encontraban junto a los de degradación del benzoil-CoA (figura 1.8). Además, en *R. pomeroyi* DSS-3^T y *C. baekdonensis* B30 los genes para la degradación del ferulato también se encontraban junto a los genes para la degradación del gentisato (figura 1.8). De estos 10 genomas en los que los genes implicados en la degradación del ferulato se encontraban junto a genes de otras vías de degradación de compuestos monoaromáticos, 8 se encontraban en estado de boceto. En cambio, los genomas de *R. pomeroyi* DSS-3^T y *Jannaschia* sp. CCS1 estaban cerrados (ver tabla S1.1), y en ambos las agrupaciones de los genes para la degradación del ferulato de la figura 1.8 se encontraban en el cromosoma. Por otra parte, se observó la presencia conservada de un gen codificante para una proteína (denominado aquí ORF1 FA), junto a los genes implicados en la degradación del ferulato en 15 genomas del linaje Roseobacter (figura 1.8 y tabla S1.2). ORF1 FA codificaría para una acetil-CoA deshidrogenasa. En otras bacterias como *A. tumefaciens* C58 y *Pseudomonas putida* KT2440 también se había observado la presencia conservada de un gen codificante para esta enzima junto a los genes implicados en la degradación del ferulato (Nelson *et al*., 2002; Campillo *et al*., 2014;). Por lo tanto, su presencia conservada podía ser un indicador de su posible rol en esta vía de degradación en diferentes bacterias, aunque deberían realizarse más estudios para poder confirmar esta hipótesis. Por ejemplo, se podría diseñar un experimento similar al de Price y colaboradores (2018) para determinar la función de una proteína de función desconocida realizando experimentos de cultivo en diversas condiciones con mutantes defectivos para ella. De esta forma, se podría comprobar si la pérdida de su función afecta o no a la capacidad para degradar ferulato, y si afecta a otros procesos celulares.

En cuanto a la presencia de genes reguladores, se detectó la presencia de un gen codificante para un regulador de tipo MarR junto a los genes para degradación de ferulato en 18 genomas del linaje Roseobacter (figura 1.8). Además, en *M. alkaliphilus* HTCC2654^T se encontró también un gen codificante para un regulador de tipo GntR (figura 1.8). La presencia de un solo regulador de la familia MarR para los genes de degradación de ferulato se había descrito en *P. putida* KT2440 (Plaggenborg *et al*., 2003). En *A. tumefaciens* C58 también se han descrito dos tipos de reguladores transcripcionales, uno de tipo GntR y otro de tipo MarR, (Campillo *et al*., 2014). Por tanto, aunque en el linaje Roseobacter podamos encontrar los dos tipos de regulación descritos hasta el momento para los genes de degradación de ferulato, mayoritariamente esta parece depender de un solo regulador de tipo MarR.

2.3.2. Degradación de ftalato, isoftalato y tereftalato

Los tres isómeros del ftalato son compuestos aromáticos cuya presencia en el medio ambiente está ampliamente extendida debido a que son compuestos químicos con una gran variedad de

usos como la fabricación de plásticos, perfumes, lubricantes, adhesivos, materiales de construcción, muebles, fármacos, juguetes, materiales de limpieza o insecticidas (Schettler, 2006). Como se puede observar en la figura 1.7 cada uno de los isómeros del ftalato se degrada por una vía distinta hasta llegar a protocatecuato. Como ya se ha comentado en el apartado 3 de la introducción se dispone de aislados de Roseobacter capaces de degradar ftalato de los géneros *Citreicella*, *Marinovum*, *Pelagibaca*, *Sulfitobacter*, *Thalassobius*, *Thalassococcus* y *Tropicibacter* (Iwaki *et al*., 2012a).

En nuestro estudio las vías de degradación de los isómeros del ftalato resultaron poco frecuentes puesto que solo se detectaron en 5 genomas del linaje Roseobacter (tabla S1.2). La presencia de los genes *oph* implicados en la degradación aeróbica del ftalato se detectó en genomas de *C. indicus* P73^T, *Oceanicola* sp. MCTG156(1a), *M. alkaliphilus* HTCC2654^T y *P. bermudensis* HTCC2601^T (tabla S1.2). Además, cabe destacar que en ninguno de estos 4 genomas se pudo detectar la presencia del gen *ophD*, codificante para un transportador de ftalato (tabla S1.2). En dos de estos genomas, *M. alkaliphilus* HTCC2654^T y *Oceanicola* sp. MCTG156(1a), se encontraron dos agrupaciones de genes distintas (figura 1.9A). En total se pudieron observar 3 tipos de agrupaciones génicas distintas (figura 1.9A).

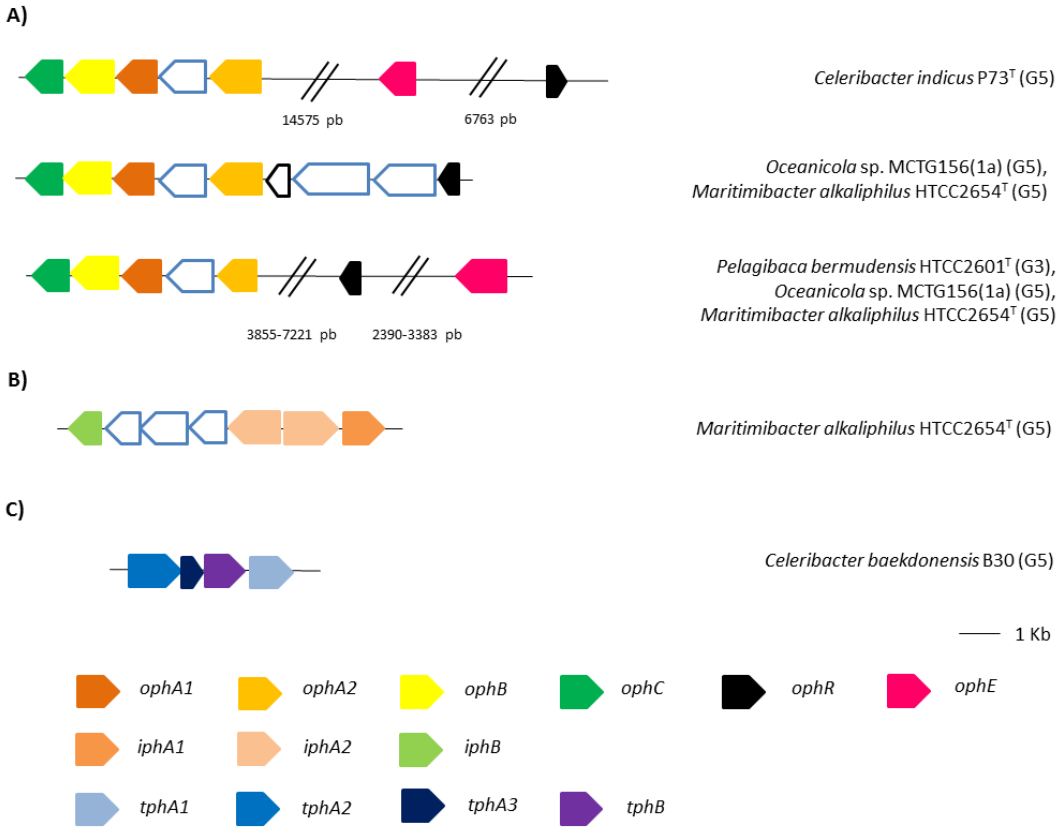

Figura 1.9. Organización de los genes de las vías del ftalato, isoftalato y tereftalato en los genomas del linaje Roseobacter. Panel A, degradación del ftalato. Panel B, degradación del isoftalato. Panel C, degradación del tereftalato. Las abreviaturas son las mismas que las de la figura 1.7 junto a: *ophR* regulador operón *oph* y *ophE* gen para posible quinolinato fosforibosil transferasa inducible por ftalato. Los genes de color blanco y contorno negro codifican para proteínas hipotéticas; los genes de color blanco con contorno azul se observaron ocasionalmente, no se encuentran conservados, ni codifican para proteínas implicadas en la degradación de compuestos aromáticos. El texto entre paréntesis indica los grupos genómicos.

En todas las estructuras observadas los genes codificantes para las dos subunidades de la ftalato 4,5-dioxigenasa (*ophA1* y *ophA2*), la deshidrogenasa (*ophB*) y la descarboxilasa (*ophC*) se encontraban juntos y orientados en el mismo sentido, solo separados por un gen codificante para una proteína aparentemente no relacionada con la degradación del ftalato (figura 1.9A). La diferencia entre las tres estructuras se basaba en la posición de los genes codificantes para el regulador OphR y para una posible quinolinato fosforibosil transferasa inducible por ftalato (OphE). En *Burkholderia cepacia* DBO1 OphE está involucrada en la biosíntesis de NAD^+ y confiere a la bacteria una ventaja a la hora de crecer en ftalato en comparación con mutantes deficientes en esta enzima (Chang y Zylstra, 1999). En cuanto a la posición de *ophE*, se encontró siempre separado del resto de genes para la degradación del ftalato mientras que en *B. cepacia* DBO1 se encuentra junto a ellos (Chang y Zylstra, 1999). Finalmente, el gen regulador *ophR* se encontraba en todos los casos corriente arriba de *ophA2* y orientado en el mismo sentido (excepto en *C. indicus* $P73^T$) aunque separado por varios genes codificantes para proteínas de función desconocida o de función aparentemente no relacionada con la degradación del ftalato (figura 1.9A).

Respecto a los genes para la degradación aeróbica del isoftalato (figura 1.7 y tabla S1.2), solo se pudo detectar su presencia en *M. alkaliphilus* $HTCC2654^T$. En concreto, tal y como se observa en la tabla S1.2 y en la figura 1.9B se detetectaron los genes codificantes para las dos subunidades de la isoftalato 3,4-dioxigenasa (IphA1 y IphA2) y de la isoftalato deshidrogenasa (IphB), pero no se pudo detectar la presencia de los genes codificantes para el regulador IphR ni del transportador de isoftalato (IphC) descritos en *Commamonas testosteroni* YZW-D (Wang *et al*., 1995). En cuanto a la organización de los genes se observó que los genes *iphA1*, *iphA2* e *iphB* se encontraban en la misma agrupación génica. Los genes *iphA1* e *iphA2* se encontraban juntos mientras que *iphB* se encontraba separado de *iphA2* por tres genes codificantes para proteínas aparentemente no relacionadas con la degradación del isoftalato (figura 1.9B). Además, se detectaron dos supuestos genes *iphA2* que se orientaban en sentido contrario uno respecto al otro (figura 1.9B).

Finalmente, respecto a la detección de los genes para la degradación aeróbica del tereftalato (figura 1.7), solo se pudo determinar su presencia en *C. baekdonensis* B30 (tabla S1.2). En concreto, se detectó la presencia de los genes codificantes para las tres subunidades de la tereftalato 1,2-dioxigenasa (TphA1, TphA2 y TphA3), así como de la deshidrogenasa (TphB), pero no se pudieron detectar ni los genes codificantes para el transportador de terephthalato (TphC) ni para el regulador transcripcional (TphR) (tabla S1.2) descritos en *Comamonas testosteroni* YZW-D (Wang *et al*., 1995). En cuanto a la organización de los genes los genes *tphA1*, *tphA2*, *tphA3* y *tphB* se encontraban juntos y orientados en el mismo sentido (figura 1.9C).

3. Vía de degradación de homogentisato

La vía de degradación de homogentisato (2,5-dihidroxifenilacetato) ha sido descrita como una vía central para la degradación de tirosina y fenilalanina en bacterias, hongos y mamíferos (Méndez *et al*., 2011). Estos aminoácidos se degradan hasta 4-hidroxifenilpiruvato y 4-hidroxifenilacetato y estos, a su vez, se transforman en homogentisato por la acción de la 4-hidroxifenilpiruvato dioxigenasa y la 4-hidroxifenilacetato 1-monooxigenasa, respectivamente

(Arias-Barrau *et al*., 2004; Méndez *et al*., 2011) (figura 1.10). Otro posible precursor del homogentisato es el 3-hidroxifenilacetato por la acción de la 3-hidroxifenilacetato 6-hidroxilasa (Arias-Barrau *et al*., 2004; Méndez *et al*., 2011).

Como se ha mencionado anteriormente, la vía de degradación de homogentisato es la segunda vía de degradación de compuestos monoaromáticos más detectada en los genomas del linaje Roseobacter (tablas 1.1 y 1.2). La enzima principal para la rotura de homogentisato es la homogentisato 1,2-dioxigenasa (HmgA) que es una enzima de la superfamilia cupina de las extradiol dioxigenasas (Pérez-Pantoja *et al*., 2012).

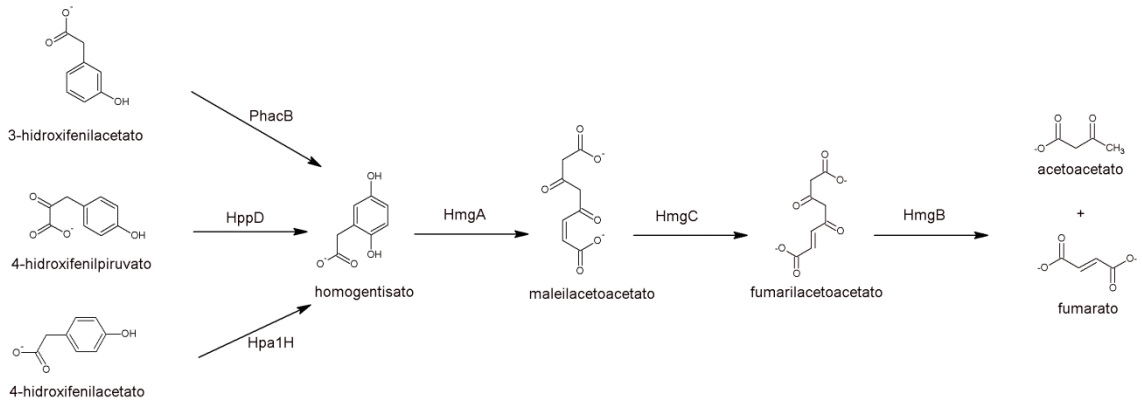

Figura 1.10. Vía de degradación de homogentisato y de compuestos precursores. Abreviatura de las enzimas: HmgA, homogentisato 1,2-dioxigenasa; HmgC, maleilacetoacetato isomerasa; HmgB, fumarilacetoacetato hidrolasa; PhacB, 3-hidroxifenilacetato 6-hidroxilasa; HppD, 4-hidroxifenilpiruvato dioxigenasa; Hpa1H, 4-hidroxifenilacetato 1-hidroxilasa. Figura elaborada a partir información de Méndez *et al*., 2011.

En esta tesis se pudo detectar el gen *hmgA* en 42 genomas del linaje Roseobacter (figura 1.11). Solo 38 de ellos presentarían la vía de degradación del homogentisato completa puesto que no se pudo detectar *hmgC* en *Sulfitobacter* sp. NAS14.1, *Rhodobacteraceae* bacterium HTCC2083 y *Pseudooceanicola batsensis* HTCC2597^T, ni *hmgB* en *Celeribacter indicus* P73^T (figura 1.11). Además, cabe destacar que la presencia de la ruta completa para la degradación del homogentisato era más común en los genomas de los grupos G1, G2 y en la mitad del G3, que en el resto de grupos (figura 1.11).

En 36 de los genomas se detectó la presencia de un gen codificante para un regulador de tipo MarR (*hmgR*) junto a los genes implicados en la degradación del homogentisato (figura 1.12). Sin embargo, solo 33 *hmgR* aparecen en las estructuras génicas alrededor de *hmgA* (figura 1.12), puesto que en *Sulfitobacter* sp. 3SOLIMAR09 y *Sulfitobacter* sp. NAS-14.1 *hmgR* se encontraba junto a *hmgB* y en *S. stellata* E-37^T se encontraba junto a *hmgC* en otras estructuras en las que no se encontraba *hmgA*. Además, en el caso de *C. indicus* P73^T y *Pelagibaca bermudensis* HTCC2601^T se encontró un gen codificante para un regulador de tipo IclR (figura 1.11). En el trabajo de Arias-Barrau y colaboradores (2004) se describió la presencia de un reguladores transcripcionales de tipo IclR junto a los genes implicados en la degradación de homogentisato en varias cepas de *Pseudomonas*. Arias-Barrau y colaboradores (2004) también observaron que en organismos de otros géneros bacterianos, como *Xanthomonas* y *Ruegeria*, junto a los genes *hmg* no siempre se encontraba un regulador *hmgR* de tipo IclR, sino que se podían encontrar diversos tipos de reguladores transcripcionales. Por ejemplo, en el caso de *Ruegeria pomeroyi* DSS-3^T detectaron que el regulador transcripcional *hmgR* situado

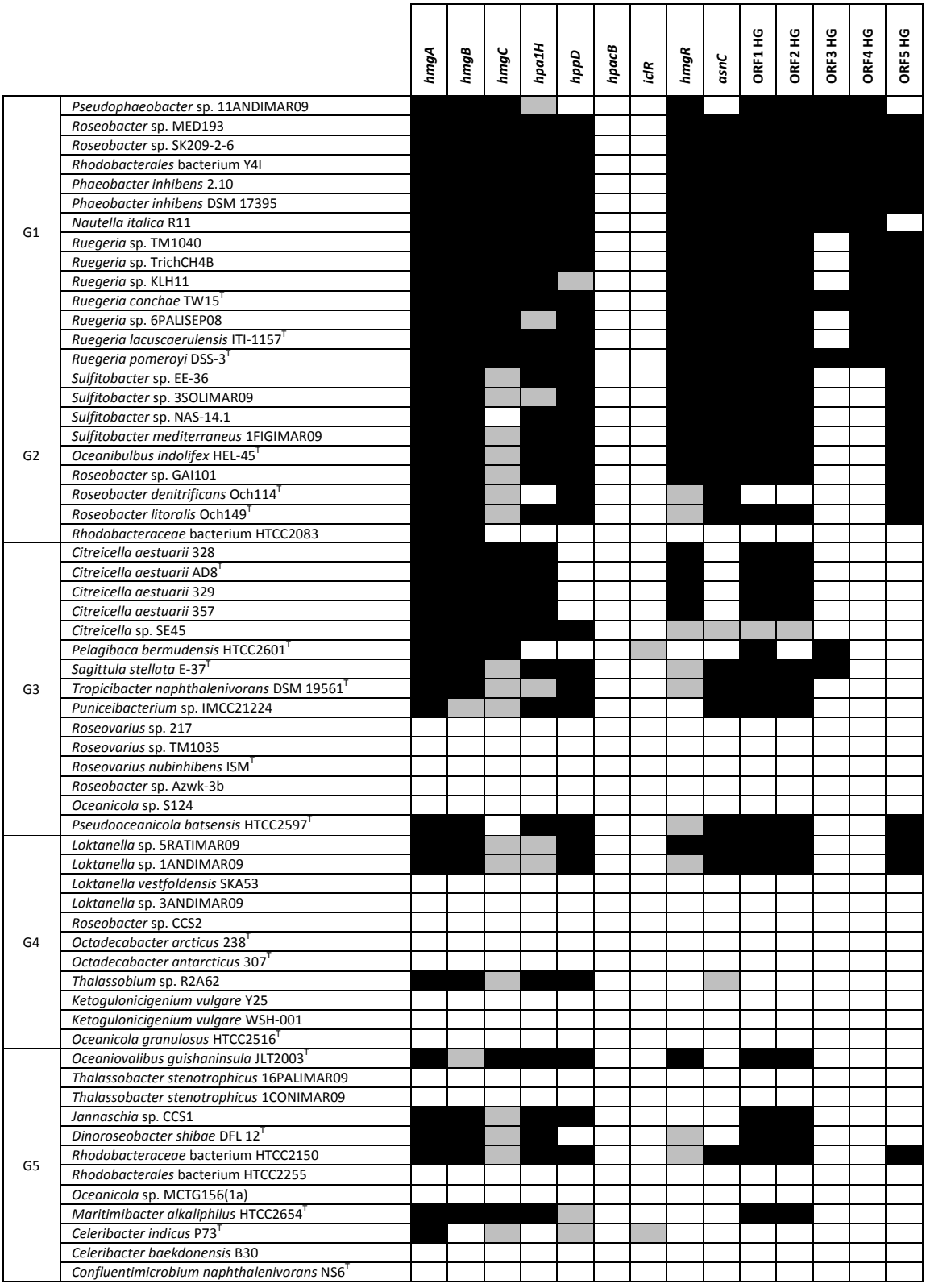

Figura 1.11. Detección de genes codificantes para las proteínas involucradas en la vía del homogentisato en genomas del linaje Roseobacter. Los genomas están ordenados en base a su filogenia (Luo y Moran, 2014; Mas-Lladó, 2017). Código de color: Negro 50 % de identidad de secuencia mínima con la proteína modelo, *E-value* <10^{-5}; gris 30 % de identidad de secuencia mínima, *E-value* <10^{-5}; blanco, no detectado.

junto a los genes *hmg* pertenecia a la familia MarR (Arias-Barrau *et al*., 2004). Por lo tanto, los resultados de esta tesis no solo confirmaban la presencia en *R. pomeroyi* DSS-3^T de un regulador transcripcional de tipo MarR, sino que este es el tipo de regulador transcripcional más extendido junto a los genes *hmg* entre los miembros del linaje Roseobacter.

Con respecto a la organización de los genes para degradación de homogentisato se observó una gran variabilidad ya que se detectaron 24 estructuras génicas distintas alrededor de *hmgA* (figura 1.12). Además, cabe destacar que en *Loktanella* sp. 5RATIMAR09 se detectaron dos copias del gen *hmgA* englobadas en dos organizaciones génicas distintas (figura 1.12). Los genes *hmgA* y *hmgB* se encontraban en la misma agrupación génica en 37 genomas, en 36 de los cuales se encontraban orientados en el mismo sentido (figura 1.12). En el caso de *hmgC,* este podía detectarse en otro lugar del genoma o agrupado junto a *hmgA.* En concreto, se encontraba en la misma agrupación génica que *hmgA* en 30 de los genomas estudiados pero su posición en la agrupación génica era variable (figura 1.12). Tanto la presencia de *hmgC* junto al resto de genes para la degradación de homogentisato como su ausencia al lado de estos ya había sido descrita en otros grupos bacterianos. Ejemplos del primer tipo de ordenación génica serían *Pseudomonas putida* KT2440 y *Sinorhizhobium meliloti* 1021, mientras que el segundo se ha observado en bacterias del orden *Burkholderiales* (Arias-Barrau *et al*., 2004; Pérez-Pantoja *et al*., 2012). Cabe destacar que la mayor parte de *hmgC* identificados en nuestro estudio también presentaron como mínimo un 30 % de identidad de secuencia con el gen *nagL*, implicado en la degradación de gentisato (ver apartado 5 del capítulo 1). Este sería un caso similar al de *pcaF* y *paaJ*, genes de rutas diferentes de degradación de compuestos aromáticos que presentaban ≥ 30 % de identidad de secuencia (ver apartado 2.1.1). Por lo tanto, sería necesario realizar estudios posteriores con los que poder confirmar la posible presencia de enzimas con doble actividad HmgC y NagL, resultado sugerido en base al estudio de similitud de secuencia.

Se observó que la mayoría de los 42 genomas que contenían genes para la degradación de homogentisato presentaban genes codificantes para las enzimas 4-hidroxifenilacetato 1-hidroxilasa (Hpa1H) y/o la 4-hidroxifenilpiruvato dioxigenasa (HppD), implicadas en la degradación de compuestos que se canalizan via homogentisato (figura 1.11). En cambio, solo se pudo detectar la presencia del gen codificante para la 3-hidroxifenilacetato 6-hidroxilasa (PhacB) en *Tropicibacter naphthalenivorans* DSM 19561^T utilizando el criterio de detección más relajado (figura 1.11). Por lo tanto, los resultados sugieren que la mayoría de los miembros del linaje Roseobacter no solo podrían degradar el homogentisato sino también los compuestos 4-hidroxifenilacetato y 4-hidroxifenilpiruvato, aunque no 3-hidroxifenilacetato. Respecto a la localización de los genes codificantes para estas dos enzimas, se observó que la mayoría de genes codificantes para la 4-hidroxifenilacetato 1-hidroxilasa (*hpa1H*) se encontraban junto a los genes para degradación de homogentisato (figura 1.12). En cambio, solo se detectó la presencia del gen codificante para la 4-hidroxifenilpiruvato dioxigenasa (*hppD*) junto a estos genes en *Thalassobium* sp. R2S62 (figura 1.12). En este estudio también se observó la presencia de un gen que codifica para un regulador transcripcional de tipo AsnC en 29 de los 34 genomas en los que se detectó el gen codificante para la enzima HppD (figura 1.11). Esta agrupación ha sido descrita en varios miembros del género *Streptomyces* (Yang *et al*., 2007). En nuestro estudio, la agrupación *hppD-asnC* se encontró siempre separada de la agrupación

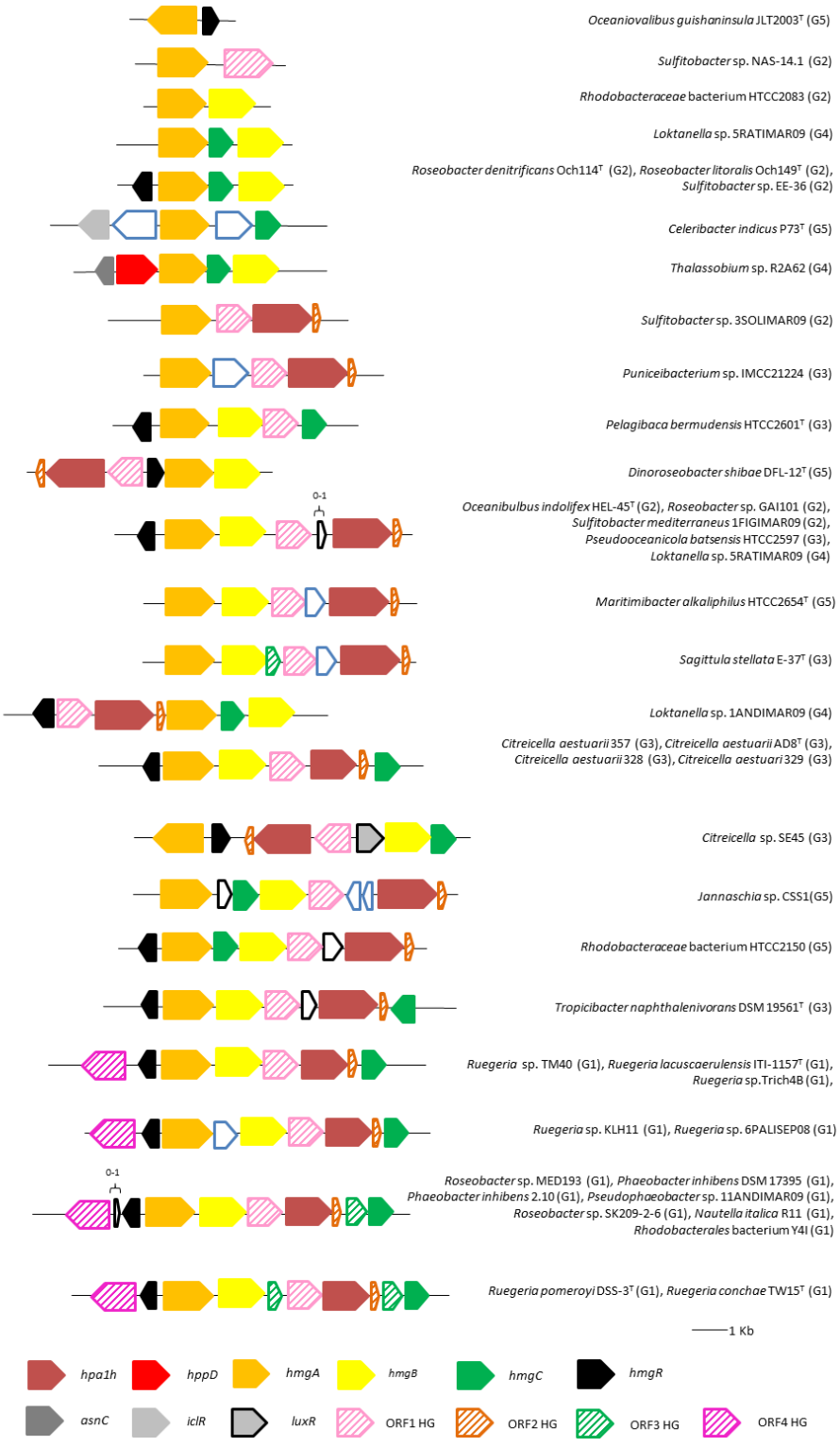

Figura 1.12. Organización de los genes de la vía del homogentisato en los genomas del linaje Roseobacter. Las abreviaturas son las mismas que las de la figura 1.10 junto a: *hmgR,* gen para regulador transcripcional del operón *hmg* de tipo MarR; *iclR,* gen para regulador de tipo IclR; ORF1 HG, gen para metalobetalactamasa; ORF2 HG, gen para proteína hipotética conservada de la familia DUF2783; ORF3 HG, gen para glioxilasa; ORF4 HG, gen para adelinato ciclasa; *asnC,* gen para regulador transcripcional de *hppD*. Los genes de color blanco y contorno negro codifican para proteínas hipotéticas y los genes de color blanco con contorno azul se observaron ocasionalmente, pero no se encuentran conservados, ni codifican para proteínas implicadas en la degradación de compuestos aromáticos. El texto entre paréntesis indica los grupos genómicos.

génica alrededor de *hmgA,* a excepción de en *Thalassobium* sp. R2A62 (figura 1.12), y junto a un gen codificante para una presunta bomba de expulsión (ORF5 HG). Además, el gen codificante para el regulador de tipo AsnC se encontraba siempre orientado en sentido contrario al gen *hppD,* tal y como se observa en *Thalassobium* sp. R2A62 (figura 1.12).

Por lo que respecta a la presencia de genes no relacionados con la degradación de homogentisato, como se puede observar en la figura 1.11 en 38 de los genomas estudiados se encontró conservada la presencia de un gen codificante para una presunta metalobetalactamasa (denominada en este trabajo ORF1 HG) junto a los genes *hmg.* En concreto se detectó ORF1 HG formando parte de la misma estructura génica que *hmgA* en 35 genomas del linaje Roseobacter (figura 1.12). Esta estructura había sido descrita anteriormente en *Jannaschia* sp. CSS1, *S. stellata* E-37^{T} y *C. aestuarii* 357 por Suárez-Suárez (2013). Además, junto a los genes para degradación de homogentisato también se observó la presencia de un gen codificante para una proteína hipotética de la familia pfam DUF 2783 (ORF2 HG) en 37 genomas, de un gen para una presunta glioxilasa (ORF3 HG) en 11 genomas y un gen codificante para una presunta adenilato ciclasa (ORF4 HG) en los 14 genomas del G1 de Roseobacter (figuras 1.12).

4. Vía de degradación de fenilacetil-CoA

El fenilacetato es un compuesto monoaromático que puede ser degradado hacia intermediarios del ciclo de Krebs de dos formas: a través de su paso a 2-hidroxifenilacetato (Cox *et al*., 1996) y a través de un paso previo de activación mediante la unión de un grupo Coenzima A (CoA), dando lugar a fenilacetil-CoA (Pérez-Pantoja *et al*., 2008). La vía de degradación del fenilacetil-CoA es la tercera vía de degradación de compuestos monoaromáticos más detectada en los genomas del linaje Roseobacter (tablas 1.1 y 1.2). La vía está compuesta por reacciones que requieren de la actividad de 8 enzimas (figura 1.13) que actúan en el siguiente orden: la fenilacetil-CoA ligasa (PaaK) se encarga de la activación del fenilacetato y da lugar a fenilacetil-CoA que es hidroxilado por el complejo de la fenilacetil-CoA monooxigenasa (PaaABCDE). El compuesto resultante, que ya ha perdido su carácter aromático, es isomerizado por la 1,2-epoxifenilacetil-CoA isomerasa (PaaG). A continuación, la oxepina-CoA hidrolasa (PaaZ) se encarga de la rotura del compuesto cíclico y finalmente las enzimas enoil-CoA hidratasa (PaaF), 3-oxoadipil-CoA tiolasa (PaaJ) y 3-hidroxiacil-CoA deshidrogenasa (PaaH) se encargan de la degradación del compuesto hasta dar lugar a intermediarios del ciclo de Krebs (Teufel *et al*., 2010; Ferrández *et al*., 1998; Hamlin *et al*., 2009). En cuanto a la enzima acil-CoA tioesterasa (PaaI) no se conoce con exactitud la reacción de la ruta en la que participa, aunque se ha visto que actúa preferentemente sobre substratos de fenilacetil-CoA hidroxilados y poco sobre ésteres de CoA alifáticos, con lo que preferentemente actuaría sobre los primeros intermediarios de la ruta (Teufel *et al*., 2010). Se han descrito otras proteínas relacionadas con la vía de de gradación del fenilacetato: PaaX y PaaY que son reguladores transcripcionales de la ruta; PaaL, que es una permeasa de fenilacetato; y PaaM, que es una porina de la membrana externa implicada en el transporte de fenilacetato (Luengo *et al.* 2001).

En este estudio se detectó la presencia de la ruta completa de los genes para la degradación de fenilacetato en 28 genomas del linaje Roseobacter (figura 1.14). Además, *Roseobacter* sp.

GAI101 podría presentar una vía incompleta puesto que se detectaban todos los genes de la ruta a excepción de *paaA*, *paaB* y *paaC.* En 28 genomas se detectaron dos o tres genes implicados en las últimas reacciones de la ruta (figura 1.14). En concreto, en 18 de ellos se detectaron los genes *paaJ*, *paaF* y *paaH*, en 9 de ellos los genes *paaF* y *paaH,* y en *Octadecabacter antarcticus* 307^{T} *paaJ* y *paaF* (figura 1.14). Cabe recordar que el gen *paaJ* presenta una alta identidad de secuencia con *pcaF* (ver apartado 2.1.1 de este capítulo), lo que explicaría la elevada detección de este gen incluso en los microorganismos que no presentan el resto de genes de la vía del fenilacetato.

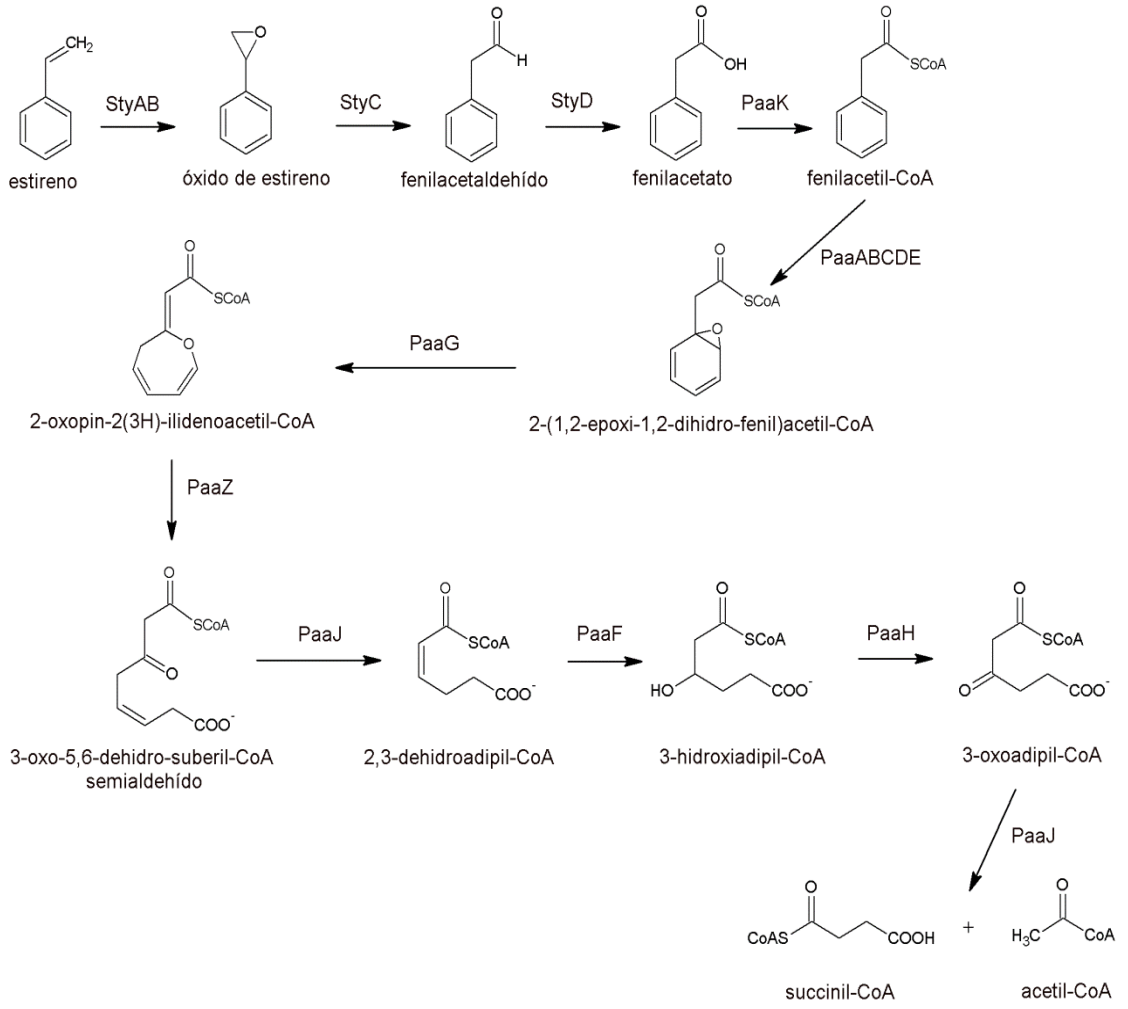

Figura 1.13. Vía de degradación de fenilacetil-CoA y de compuestos cuya degradación procede por esta vía. Abreviatura de las enzimas: PaaK, fenilacetil-CoA ligasa; PaaABCDE, fenilacetil-CoA monooxigenasa; PaaG, 1,2-epoxifenilacetil-CoA isomerasa; PaaZ, oxepina-CoA hidrolasa; PaaJ, 3-oxoadipil-CoA tiolasa; PaaF, enoil-CoA hidratasa; PaaH, 3-hidroxiacil-CoA deshidrogenasa; StyAB, estireno monooxigenasa; StyC, estireno óxido isomerasa; StyD, fenilacetaldehído deshidrogenasa. Figura elaborada a partir de información de Teufel *et al*., 2010.

Además, tal y como se puede observar en la figura 1.14, la detección de la ruta completa para la degradación del fenilacetil-CoA era más alta en genomas de los grupos filogenómicos G1 y G2 que en el resto de grupos filogenómicos, sobre todo en el caso del grupo G1. Por el contrario, no se detectó la ruta completa para la degradación del fenilacetil-CoA en ningún miembro del grupo filogenómico G4 (figura 1.14). La detección desigual de una ruta completa también se había detectado en otras rutas (ver tabla 1.2, apartado 1 del presente capítulo), y

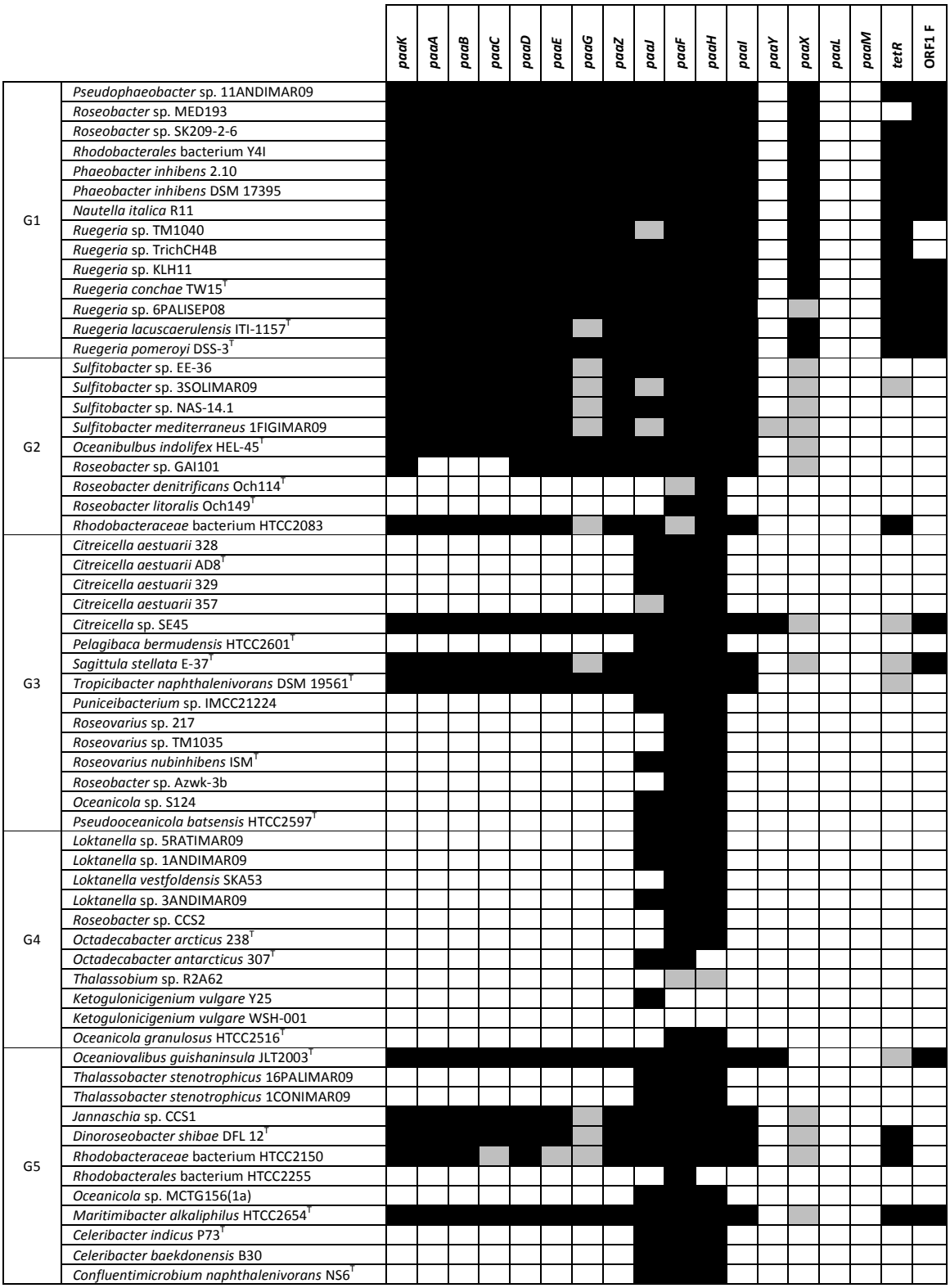

Figura 1.14. Detección de genes codificantes para las proteínas involucradas en la vía del fenilacetil-CoA en genomas del linaje Roseobacter. Los genomas del linaje Roseobacter están ordenados en base a su filogenia (Luo y Moran, 2014; Mas-Lladó, 2017). Código de color: Negro 50 % de identidad mínima de secuencia con la proteína modelo, *E-value* $<10^{-5}$; gris 30 % de identidad mínima de secuencia, *E-value* $<10^{-5}$; blanco, no detectado.

esto sugeria que los miembros de los grupos G1 y G2 serían los que presentarían un mayor potencial para el catabolismo de compuestos aromáticos.

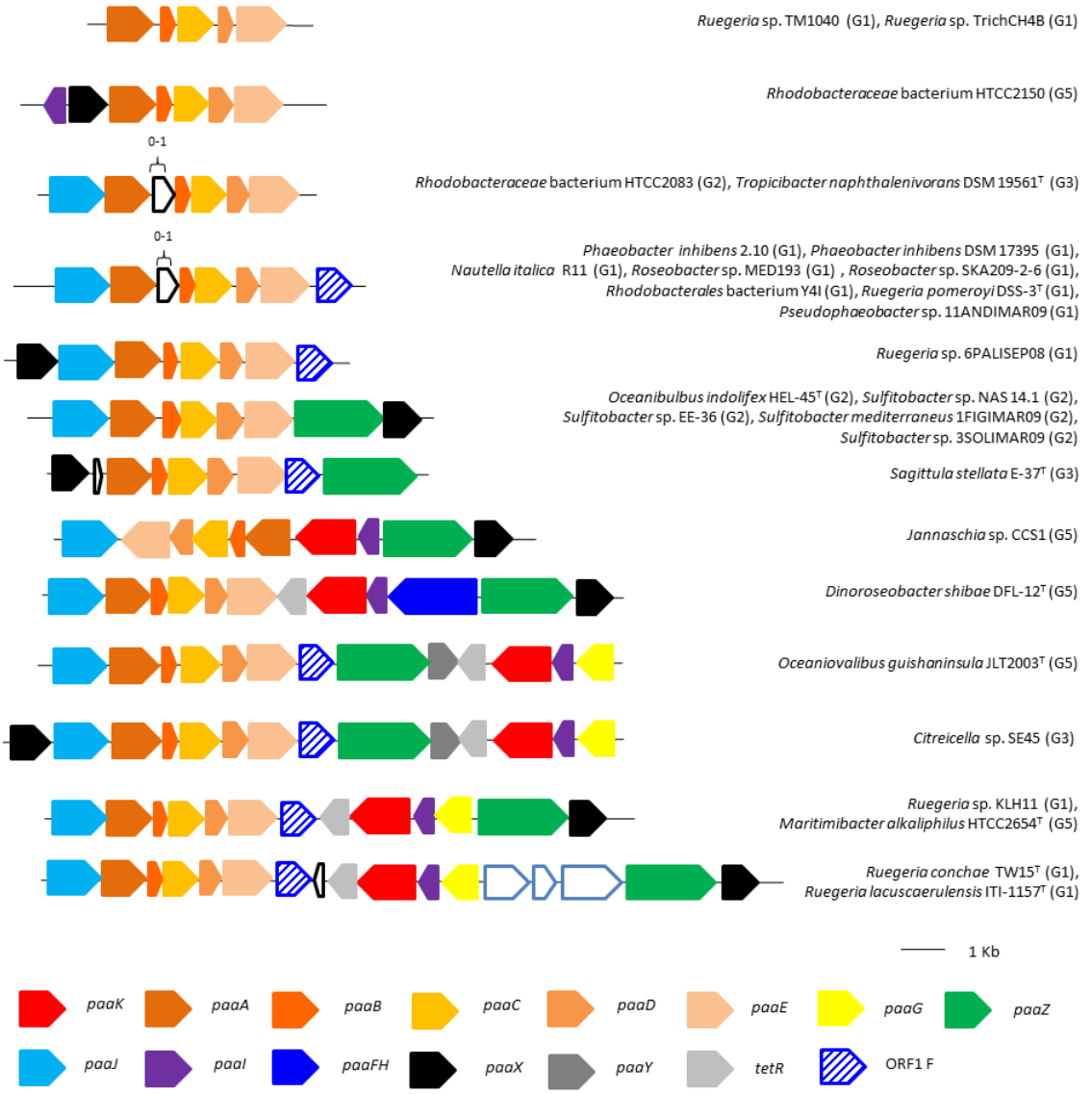

Figura 1.15. Organización de los genes de la vía del fenilacetil-CoA en los genomas del linaje Roseobacter. Las abreviaturas son las mismas que las de la figura 1.13 junto a: *paaI,* gen para acil-CoA tioesterasa; *paaX, paaY* y *tetR,* genes para reguladores transcripcionales; *paaFH,* gen de fusión de entre *paaF* y *paaH* para una enoil-CoA hidratasa y ORF1 F, proteína conservada de función desconocida. Los genes de color blanco y contorno negro codifican para proteínas hipotéticas y los genes de color blanco con contorno azul se observaron ocasionalmente, pero no se encuentran conservados, ni codifican para proteínas implicadas en la degradación de compuestos aromáticos. El texto entre paréntesis indica los grupos genómicos.

En cuanto a la organización de los genes para la degradación del fenilacetato, éstos se agruparon en 13 estructuras distintas (figura 1.15). Por lo tanto, había también variabilidad en cuanto a la organización génica, aunque menor que en los genes para la degradación de protocatecuato (ver figura 1.3). Como se puede observar en la figura 1.15, los genes *paaABCDE* codificantes para la fenilacetil-CoA monooxigenasa siempre se encontraban juntos y orientados en el mismo sentido. En 24 genomas estos genes se encontraban junto a *paaJ* (figura 1.15). Además, en 14 genomas se observó el gen *paaZ* junto al grupo *paa(J)ABCDE* y en la misma orientación, salvo en *Jannaschia* sp. CSS1 donde se encontraba orientado en sentido

contrario (figura 1.15). En 9 de estos 14 genomas, *paaZ* se encontraba separado de *paaABCDE* por varios genes. Estos podían estar implicados en la degradación del fenilacetato o ser un gen para un regulador transcripcional de tipo TetR, un gen codificante para una proteína de función desconocida conservada junto a los genes *paa* (denominado aquí ORF1 F) y que no había sido previamente descrito en otros grupos bacterianos, o genes para proteínas no implicadas en la degradación de compuestos aromáticos (figura 1.15).

También cabe destacar que en la mayoría de organismos estudiados ni *paaF* ni *paaH* se encontraban en la misma agrupación génica que *paaABCDE* puesto que solían encontrarse en otras regiones del genoma. Solo en en el caso de *Dinoroseobacter shibae* DFL 12^T se pudo detectar la presencia de un aparente gen de fusión de *paaF* y *paaH* junto a los genes *paaABCDE* (figura 1.15) que no ha sido descrito previamente en la literatura. En cuanto a la presencia de *paaK*, se pudo observar que solo en 8 de los genomas estudiados se encontraba agrupado junto a *paaABCDE*, y en todos ellos también se encontraba cercano a *paaZ* y orientado en sentido contrario a este (figura 1.15). Además en 6 genomas de roseobacters también se observó la presencia de gen *paaG* cercano a *paaZ* y en sentido contrario a este (figura 1.15).

En cuanto a la regulación de los genes para degradación de fenilacetato, en 15 genomas se detectó el gen regulador *paaX* junto al grupo génico *paaABCDE* aunque su posición era variable (figura 1.15). Solo se pudo determinar la presencia del otro regulador descrito en la literatura, *paaY*, en tres genomas del linaje Roseobacter (figura 1.16). Además, también se pudo determinar la presencia de un gen codificante para un regulador de tipo TetR junto a las agrupaciones de genes alrededor de *paaABCDE* en 7 genomas de Roseobacter (figura 1.15). Estos resultados sugieren que en el linaje Roseobacter podría haber reguladores para los genes de degradación de fenilacetato distintos, o adicionales, a *paaX* y *paaY* que son los únicos descritos hasta el momento (Teufel *et al*., 2010).

Como ocurría para los genes de la ruta del protocatecuato de la vía del beta-cetoadipato, en el caso de los genes de degradación de fenilacetil-CoA también se encontraron 11 agrupaciones génicas que no incluían los genes *paaABCDE*. Nueve de estas estructuras, presentes en 21 genomas de Roseobacter, eran estructuras alrededor del gen *paaK*, cuya presencia junto a los genes *paaABCDE* solo había podido ser detectada en 8 genomas de Roseobacter (figura 1.12). En 5 de estas estructuras, presentes en 15 genomas de Roseobacter *paaK* se encontraba en la misma agrupación génica que *paaZ* (figura 1.16). Además en 3 de estas 5 estructuras también se encontraban agrupados los genes *paaG* y *paaI,* ambos siempre orientados en el mismo sentido que *paaK* (figura 1.16). Estos resultados muestran que *paaK* se solía encontrar separado de los genes *paaABCDE*. Además, tal y como se pudo observar en la figura 1.16, *T. naphthalenivorans* DSM 19561^T presentaba dos copias del gen *paaK*, cada una de ellas localizada en agrupaciones génicas distintas, una de las cuales contenía genes implicados en la degradación de estireno (figura 1.16), compuesto cuya degradación procede vía fenilacetato (figura 1.16). Cabe destacar que *T. naphthalenivorans* DSM 19561^T fue el único de los genomas estudiados en los que se detectaron los genes implicados en la degradación del estireno. También cabe destacar que los genes *paaF* y *paaH* se encontraban en regiones diferentes en los genomas del linaje Roseobacter estudiados, a excepción de *Pseudophaeobacter* sp.

11ANDIMAR09 (figura 1.16) en donde el gen *paaF* en se encontró a 3,4 kb de *paaJ*, separados por tres genes codificantes para proteínas no relacionadas con degradación de fenilacetato.

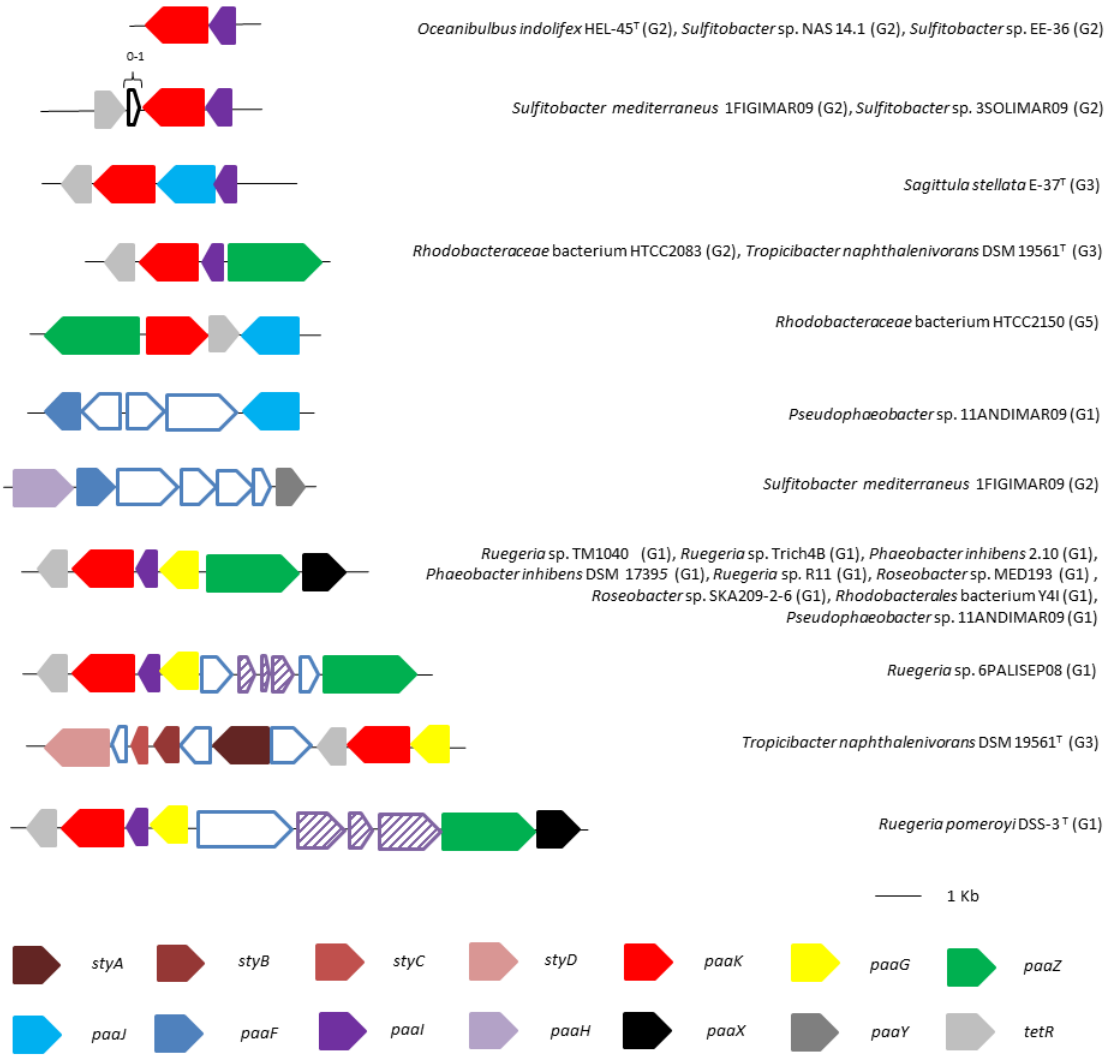

Figura 1.16. Organizaciónde los genes para la degradación de fenilacetil-CoA separadas del *core paaABCDE* en los genomas del linaje Roseobacter. Las abreviaturas son las mismas que las de las figuras 1.11 y 1.13 junto a: *lysR,* gen para regulador transcripcional. Los genes de color blanco y contorno negro codifican para proteínas hipotéticas y los genes de color blanco con contorno azul se observaron ocasionalmente, pero no se encuentran conservados, ni codifican para proteínas implicadas en la degradación de compuestos aromáticos. El texto entre paréntesis indica los grupos genómicos.

En cuanto a la presencia de genes reguladores en estas estructuras génicas separadas del grupo de genes *paaABCDE,* en 10 genomas se observó un regulador de tipo *paaX* junto a *paaZ* (figura 1.16), lo que sugirió su posible regulación por este gen. Por otra parte en 8 de estas estructuras génicas se detectó la presencia de gen codificante para un regulador de tipo TetR junto a *paaK* y orientado en el mismo sentido, a excepción de *Sulfitobacter mediterraneus* 1FIGIMAR09 y *Sulfitobacter* sp. 3SOLIMAR09, en los que *tetR* se encontraba orientado en sentido contrario a *paaK* (figura 1.16). Este tipo de regulador ya se había detectado junto a *paaK* en alguna de las estructuras génicas alrededor de *paaABCDE* (figura 1.15). Por lo tanto, estos resultados sugieren que reguladores de la familia TetR podrían jugar un papel en la expresión de los genes implicados en la degradación del fenilacetato en el linaje Roseobacter, y en particular de *paaK*.

Los resultados obtenidos muestran que en los miembros del linaje Roseobacter los genes implicados en la degradación del fenilacetato se organizaban en varios grupos génicos separados en el genoma independientemente de que los genomas se encontrasen cerrados o en estado de boceto. Esto los diferencia de modelos de degradación como *Escherichia coli* W o *Pseudomonas putida* U en los que los genes implicados en la degradación del fenilacetato se organizan en 3 y 5 operones contiguos, respectivamente (Luengo *et al*., 2001).

5. Vía de degradación de homoprotocatecuato

El homoprotocatecuato (3,4-dihidroxifenilacetato) es un compuesto monoaromático que se puede generar a partir de la degradación de 4-hidroxifenilacetato, compuesto que a su vez puede provenir de la degradación de aminácidos aromáticos. Esta vía catabólica fue la cuarta vía de degradación de compuestos monoaromáticos más detectada en los miembros del linaje Roseobacter (tablas 1.1 y 1.2).

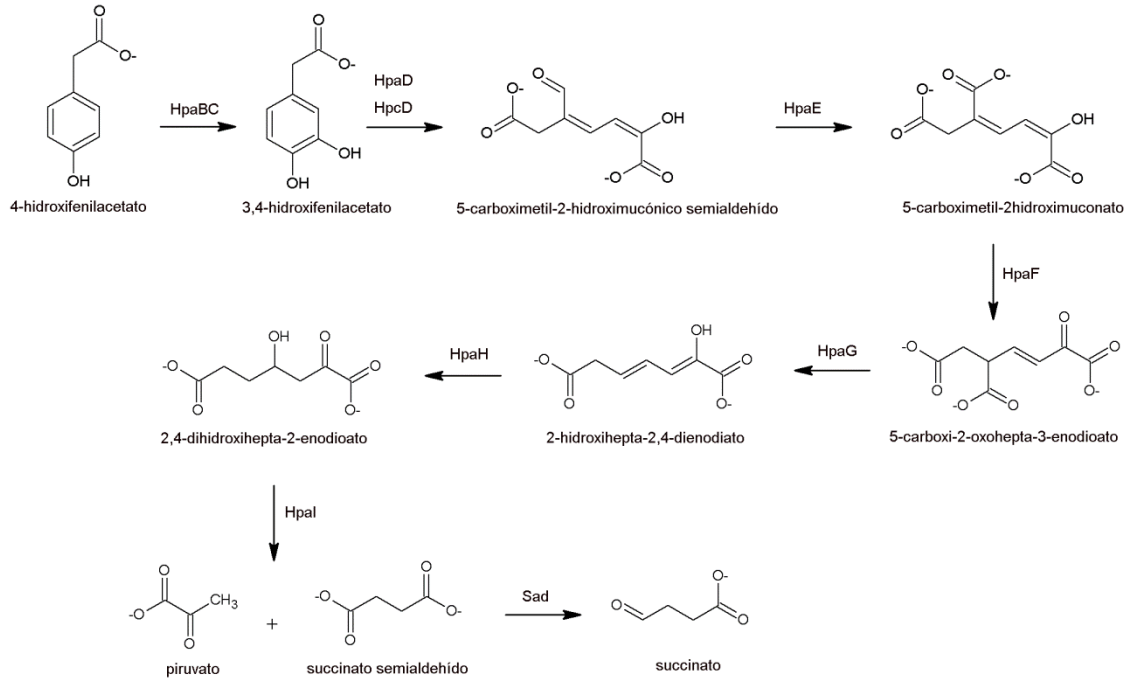

Figura 1.17. Vía de degradación de homoprotocatecuato. Abreviatura de las enzimas: HpaBC, 4-hidroxifenilacetato hidroxilasa; HpaD, homoprotocatecuato 2,3-dioxigenasa; HpaE, 5-carboxi-2-hidroximucónico semialdehído deshidrogenasa; HpaF, 5-carboximetil-2-hidroximucónico delta isomerasa; HpaG, 5-oxopent-3-eno-1,2,5-tricarboxilato descarboxilasa; HpaH, 2-oxo-hept-3-eno-1,7-dioato hidratasa; HpaI; 4-hidroxi-2-oxoheptanodioato aldolasa; Sad, succinato semialdehído deshidrogenasa. Figura elaborada a partir de información de Díaz *et al*., 2001.

En la vía de degradación del homoprotocatecuato intervienen siete enzimas que generan como productos finales succinato y piruvato para su posterior metabolización en el ciclo de Krebs (figura 1.17). La enzima principal que se encarga de la rotura del homoprotocatecuato es la homoprotocatecuato 2,3-dioxigenasa. Se han descrito dos tipos de homoprotocatecuato 2,3-dioxigenasa, una de las cuales pertenece a la superfamilia LigB de las extradiol dioxigenasas y ha sido descrita en *E. coli* (HpaD). En *E. coli* W los genes codificantes para las enzimas implicadas en la degradación del homoprotocatecuato forman el operón *hpaGEDFHI*, que está regulado por un regulador transcripcional codificado por el gen *hpaR* (Díaz *et al*., 2001). En esta cepa, junto a los genes implicados en la degradación del homoprotocatecuato

también se encuentra el operón *hpaBC*, para la degradación de 4-hidroxifenilacetato a homoprotocatecuato, junto a su gen regulador *hpaA* (Díaz *et al*., 2001). Además, en *E. coli* W junto a los genes *hpaABC* se detecta el gen *hpaX* que codifica para una permeasa implicada en el transporte de hidroxifenilacetato (Díaz *et al*., 2001). El segundo tipo de homoprotocatecuato 2,3-dioxigenasa se ha descrito en actinobacterias (HpcD) y es una enzima que pertenece al tipo I de la superfamilia EXDO (Wang y Lipscomb, 1997; Pérez-Pantoja *et al*., 2010a). Puede ser una dioxigenasa dependiente de Fe^{2+}, como la presente en *Brevibacterium fuscum,* o dependiente de Mn^{2+} como la de *Arthrobacter globiformis* (Pérez-Pantoja *et al*., 2010a). Ambos tipos de homoprotocatecuato 2,3-dioxigenasas generan el mismo producto, 5-carboximetil-2-hidroximucónico semialdehído (Pérez-Pantoja *et al*., 2010a).

Como se muestra en la figura 1.18 se detectaron los genes de la ruta completa de la vía del homoprotocatecuato en 22 genomas. Además, en 13 genomas se detectaron al menos dos genes de la ruta, en nueve de los cuales se detectaron los dos últimos genes de la ruta del homoprotocatecuato (*hpaI* y *sad*), sugiriendo la existencia de una ruta incompleta para la degradación del homoprotocatecuato en esos genomas (figura 1.18). En los 22 genomas con ruta completa se observaron los genes *hpaD* y *hpcD* codificantes para los dos tipos de homoprotocatecuato 2,3-dioxigenas descritos. En tres genomas (*R. conchae* TW15^T, *Ruegeria* sp. 6PALISEP08 y *T. naphthalenivorans* DSM 19561^T) se detectó un gen *hpaD*, en 14 genomas se detectó el gen *hpcD,* y en 5 genomas (*Roseobacter* sp. MED193, *Pseudophaeobacter* sp. 11ANDIMAR09, *R. pomeroyi* DSS-3^T, *Roseobacter* sp. SK209-2-6 y *M. alkaliphilus* HTCC2654^T) se detectaron tanto *hpaD* como *hpcD* (figura 1.18). Por lo tanto, los miembros del linaje Roseobacter podían presentar tanto los genes codificantes para la HpaD de la superfamilia LigB de las dioxigenasas extradiólicas, como para la HpcD del tipo I de la superfamilia EXDO, e incluso los dos tipos de genes de forma simultánea. Esta co-ocurrencia de genes codificantes para ambas homoprotocatecuato 2,3-dioxigenasas no se ha descrito hasta el momento en otro grupo bacteriano.

En cuanto a la organización de los genes para degradación de homoprotocatecuato se observaron 13 estructuras génicas distintas (figura 1.19). La estructura más sencilla consistía en el gen *hpcD* junto al gen regulador *hpaR* y un gen codificante para una presunta descarboxilasa dependiente de piridoxal (denominada en este estudio como ORF1 HP). La más compleja estaba formada por todos los genes de la ruta completa de degradación del homoprotocatecuato, incluyendo un gen *hpaD* (codificante para enzima de tipo LigB) y un gen *hpcD* (codificante para enzima de tipo EXDO I), junto al regulador *hpaR* (figura 1.19). En concreto, se observaron 4 tipos de estructuras génicas que incluían el gen *hpaD*, 7 estructuras génicas que incluían el gen *hpcD*, y 2 estructuras, en *Roseobacter* sp. SK209-2-6 y *M. alkaliphilus* HTCC2654^T, que incluían tanto *hpaD* como *hpcD* (figura 1.19). Cabe destacar que en las 13 estructuras génicas detectadas todos los genes se encontraban orientados en el mismo sentido a excepción del regulador *hpaR* que siempre se localizaba en sentido contrario (figura 1.19). Este resultado concordó con la ordenación de los genes descrita en *E. coli* W (Díaz *et al*., 2001). En cuanto a la organización del resto de genes implicados en la degradación del homoprotocatecuato, se pudo ver que la totalidad de genes *hpaE* y *hpaH* detectados en este estudio se encontraban siempre formando una agrupación génica con los genes *hpaD* y/o *hpcD*

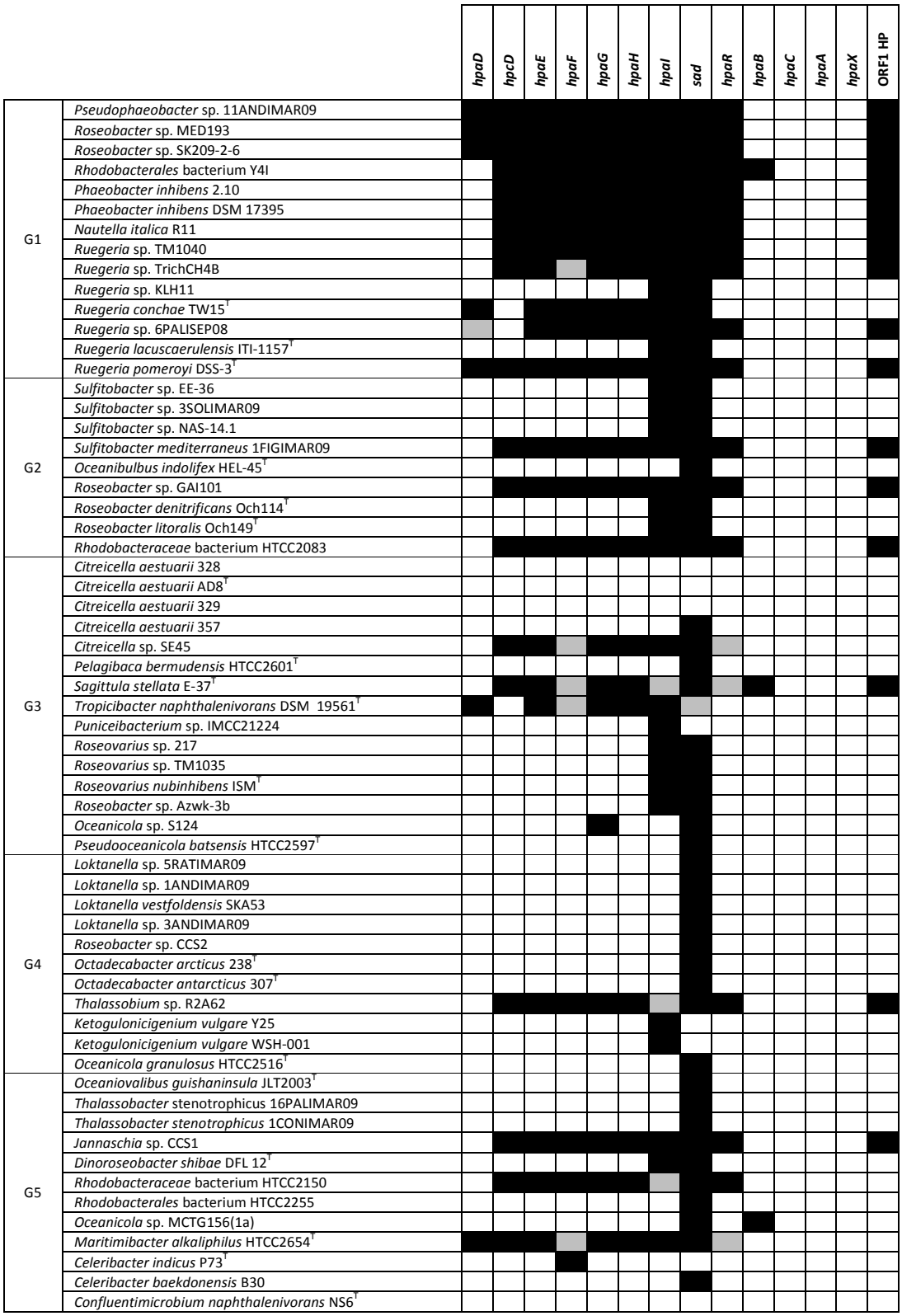

Figura 1.18. Detección de genes codificantes para las proteínas involucradas en la vía del homoprotocatecuato en genomas del linaje Roseobacter. Los genomas del linaje Roseobacter están ordenados en base a su filogenia (Luo y Moran, 2014; Mas-Lladó, 2017). Código de color: Negro 50 % de identidad mínima de secuencia con la proteína modelo, *E-value* <10−5; gris 30 % de identidad mínima de secuencia, *E-value* <10−5; blanco, no detectado.

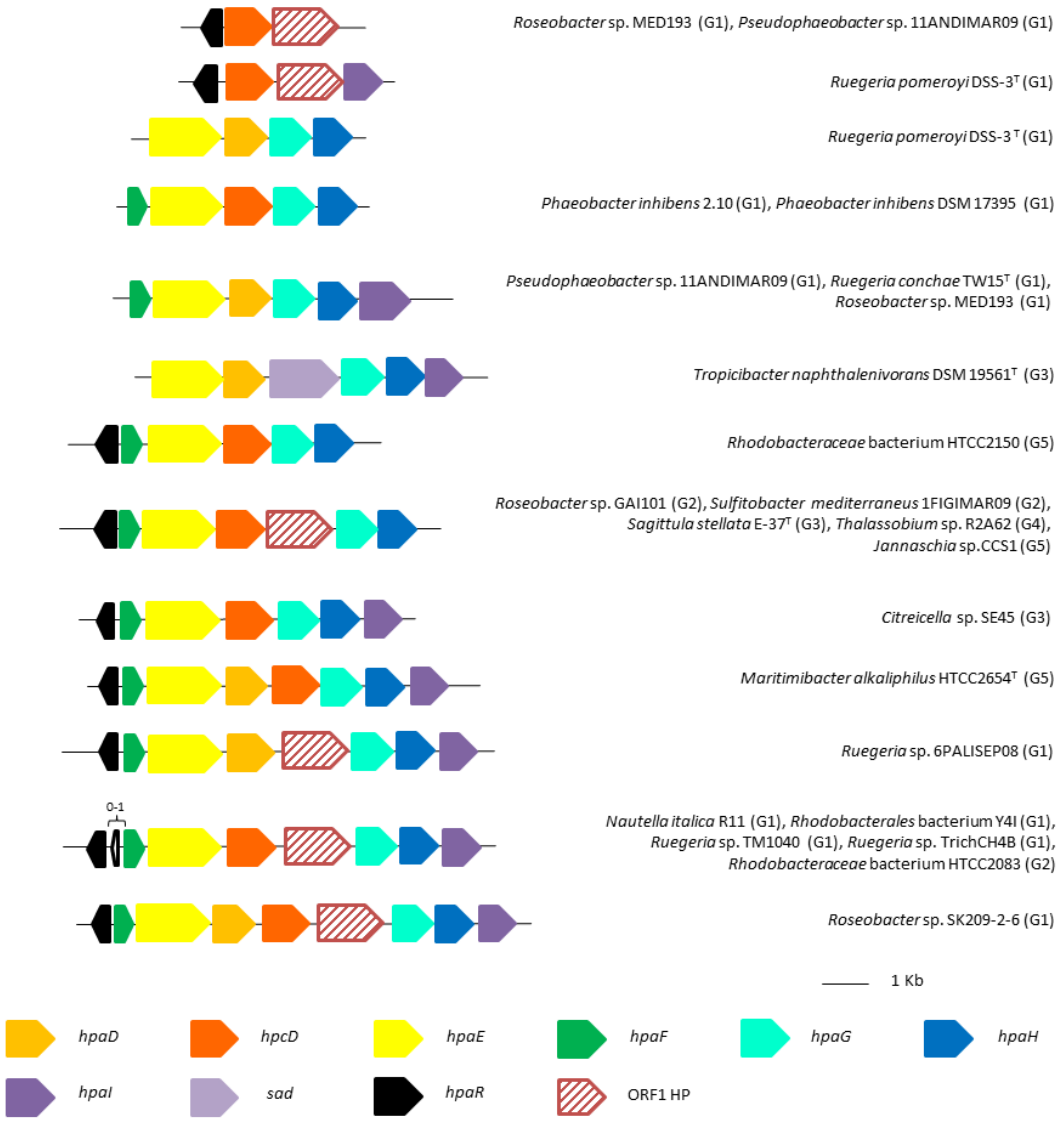

Figura 1.19. Organización de los genes de la vía del homoprotecuato en los genomas del linaje Roseobacter. Las abreviaturas son las mismas que las de la figura 1.17 junto a: *hpcD*, homoprotocatecuato 2,3-dioxigenasa de la superfamilia EXDO; *hpaR,* gen para regulador transcripcional y ORF1 HP, presunta descarboxilasa dependiente de piridoxal. Los genes de color blanco y contorno negro codifican para proteínas hipotéticas. El texto entre paréntesis indica los grupos genómicos.

(figura 1.19). Además, 20 de los 23 *hpaF* detectados y 22 de los 23 *hpaG* detectados también se encontraban en la misma agrupación génica que *hpaD* y/o *hpcD,* a excepción de *hpaF* en *T. naphtahalenivorans* DSM 19561^T, *R. pomeroyi* DSS-3^T y *C. indicus* P73^T, y *hpaG* en *Oceanicola* sp. S124 (figuras 1.18 y 1.19). Respecto a la localización de *hpaI,* este gen se detectó en la misma agrupación que *hpaD* y/o *hpcD* en 14 de los 37 genomas en los que se había detectado (figuras 1.18 y 1.19). En cuanto a *sad*, este gen solía encontrarse en otras regiones del genoma a excepción de en *T. naphthalenivorans* DSM 19561^T (figura 1.19).

Como se puede observar en la figura 1.19 se detectó la presencia de un gen codificante para una presunta descarboxilasa dependiente de piridoxal (ORF1 HP) corriente abajo de *hpaD* y/o *hpcD* en 6 estructuras génicas identificadas en 15 genomas del linaje Roseobacter. Además

también se encontró el gen codificante para esta presunta descarboxilasa separado del gen codificante para la homoprotocatecuato 2,3-dioxigenasa formando dos estructuras adicionales: (1) en *T. naphthalenivorans* DSM 19561^T junto una segunda copia de *hpaH* y *hpaG,* y (2) junto a *hpaI* y *hpaR* en *P. inhibens* 2.10 y *P. inhibens* DSM 17395. La presencia del gen codificante para esta presunta descarboxilasa dependiente de piridoxal junto a los genes de degradación de homoprotocatecuato no ha sido descrita en otros grupos bacterianos. Pero en base a los resultados de esta tesis parece ser una característica conservada en los genomas del linaje Roseobacter, aunque se desconoce su posible rol en la degradación del homoprotocatecuato en este linaje. Para poder entender el papel de esta ORF1 HP se deberían realizar estudios más exhaustivos, tal y como se hizo en el caso de la proteína codificada por el gen *chp* en la ruta del protocatecuato (apartado 2.1.) y tal y como debería hacerse para las proteínas codificadas por las ORF1-3 HG de la ruta del homogentisato (apartado 3) y la ORF1 F de la ruta del fenilacetato (apartado 4). Como ya se ha comentado anteriormente, una posible estrategia para determinar la función de estas proteínas de función desconocia sería realizar estudios de cultivo con mutantes defectivos para estas proteínas, siguiendo el ejemplo del estudio de Price y colaboradores (2018). Por lo tanto, en base a los resultados obtenidos en el linaje Roseobacter parece que es común la presencia de genes codificantes para proteínas de función desconocida hasta la fecha junto a los genes implicados en la degradación de compuestos monoaromáticos.

En cuanto a la presencia de reguladores de los genes implicados en la degradación del homoprotocatecuato solo se detectó la presencia del gen codificante para el regulador HpaR descrito en la literatura en 18 genomas junto a *hpaF* o *hpcD* y orientado en sentido contrario al resto de genes (figura 1.19). Además, en 4 de las agrupaciones génicas alrededor de *hpaD* y/o *hpcD* presentes en 7 genomas no se pudo determinar la presencia de ningún regulador (figura 1.19). Aunque cabe destacar que 3 de estos 7 genomas, *R. pomeroyi* DSS-3^T, *Pseudophaoebacter* sp. 11ANDIMAR09 y *Roseobacter* sp. MED193 presentaban dos agrupaciones génicas, una alrededor de *hpaD* y otra alrededor de *hpcD*, y en esta última sí presentaban un gen *hpaR* (figura 1.19).

Por último, no se pudo detectar la presencia de los genes para degradación de 4-hidroxifenilacetato, *hpaBC* y de su regulador *hpaA*, junto a los genes implicados en la degradación de homoprotocatecuato en ninguno de los genomas de Roseobacter estudiados (figura 1.18). Solo se pudo detectar la presencia de *hpaB* en *Rhodobacterales* bacterium Y4I, *S. stellata* E-37^T y *Oceanicola* sp. MCTG156(1a) mediante el criterio de detección del 50 % de identidad con un *E-value* menor de 10^{-5}. En los tres casos el gen codificante para esta proteína se encontraba en otras regiones del genoma. Por lo tanto estos resultados sugirieron que los miembros del linaje Roseobacter no serían capaces de degradar 4-hidroxifenilacetato a homoprotocatecuato.

6. Vía de degradación del gentisato

El gentisato es un compuesto monoaromático que se puede formar a partir de la degradación de varios compuestos (figura 1.20) como el 3-hidroxibenzoato por acción de la 3-hidroxibenzoato 6-hidroxilasa o el salicilato, un intermediario de la degradación del naftaleno, por acción de la salicilato 1,5-hidroxilasa (Pérez-Pantoja *et al*., 2012). La vía de degradación del

gentisato fue una vía poco detectada en los genomas del linaje Roseobacter puesto que solo se detectó como vía completa en en el 24,6 % de los genomas incluidos en este estudio (tabla 1.2). Tal y como se observa en la figura 1.20 la degradación de gentisato se produce por la acción de tres enzimas, de las cuales la gentisato 1,2-dioxigenasa (NagI) es la enzima principal que se encarga de la rotura oxigenolítica del anillo aromático dando lugar a maleil piruvato. A continuación, la maleilpiruvato isomerasa (NagL) isomeriza el maleilpiruvato a fumarilacetato, el cual es hidrolizado dando lugar a fumarato y piruvato por la acción de la fumarilpiruvato hidrolasa NagK (Pérez-Pantoja *et al*., 2012).

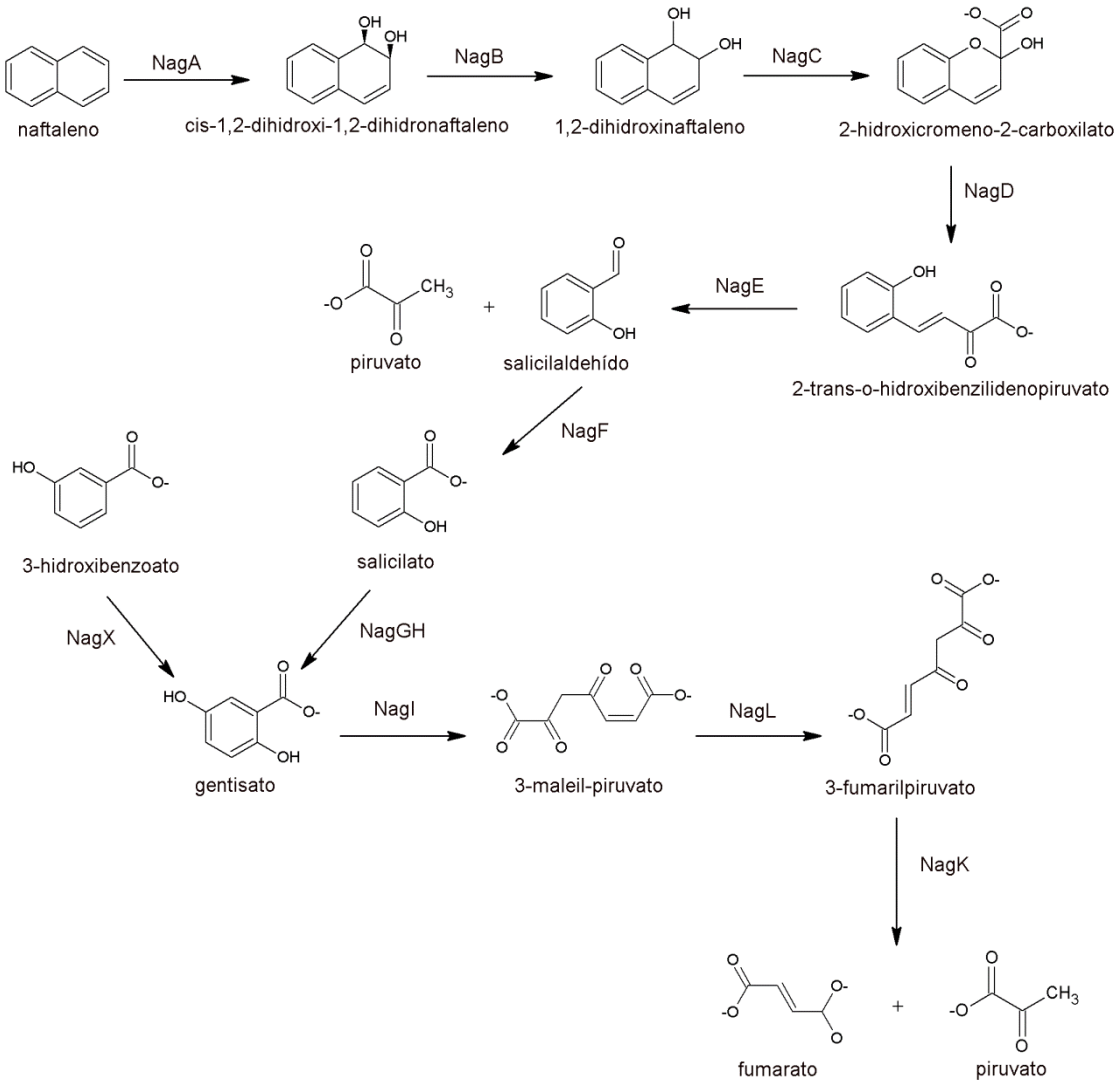

Figura 1.20. Vías de degradación de gentisato, naftaleno, salicilato y 3-hidroxibenzoato. Abreviatura de las enzimas: NagA, naftaleno 1,2-dioxigenasa; NagB, *cis*-1,2-dihidro-1,2-dihidroxinaftaleno deshidrogenasa; NagC, 1,2-dihidroxinaftaleno dioxigenasa; NagD, 2-hidroxicromeno-2-carboxilato isomerasa; NagE, trans-*o*-hidroxibenzilidenopiruvato hidratasa-aldolasa; NagF, salicilaldehído deshidrogenasa; NagGH, salicilato 1,5-hidroxilasa; NagX, 3-hidroxibenzoato 6-hidroxilasa; NagI, gentisato 1,2-dioxigenasa; NagL, maleilpiruvato isomerasa; NagK, fumarilpiruvato hidrolasa. Figura elaborada a partir de Bosch *et al.*, 1999b; Zhou *et al.*, 2001 y Pérez-Pantoja *et al*., 2012).

En este estudio se detectaron los genes *nagI*, *nagL* y *nagK* en 15 genomas del linaje Roseobacter (tabla S1.2). De estos 15 genomas, 7 pertenecían al grupo filogenómico G3 (*T. naphthalenivorans* DSM 19561^T, *Puniceibacterium* sp. IMCC21224, *S. stellata* E-37^T y las cepas

de *C. aestuarii* AD8^T, 328, 329 y 357), 6 al grupo filogenómico G5 (*M. alkaliphilus* HTCC2654^T, *C. indicus* P73^T, *Oceanicola* sp. MCTG156(1a), *Jannaschia* sp. CSS1, *C. baekdonensis* B30 y *Confluentimicrobium naphthalenivorans* NS6^T), 1 al grupo filogenómico G1 (*R. pomeroyi* DSS-3^T) y 1 al grupo filogenómico G2 (*Roseobacter* sp. GAI101). En *C. indicus* P73^T, *Oceanicola* sp. MCTG156(1a) y *Confluentimicrobium naphthalenivorans* NS6^T se encontraron 3, 4 y 2 copias de *nagI*, respectivamente, cada una de las cuales se encontraba en estructuras génicas distintas (figura 1.21).

Los genes para la degradación de gentisato se encontraron formando 14 estructuras génicas diferentes (figura 1.21). Solo en 8 de ellas (presentes en 9 genomas) se observaron los tres genes en la misma estructura (figura 1.21) debido a que la posición del gen *nagL* era variable. El gen *nagI* se encontraba junto a *nagK* y en la misma orientación que este en 6 genomas y junto a *nagL* y en la misma orientación que este en *C. indicus* P73^T y *C. baekdonensis* B30 (figura 1.21). Además, se vió que *nagL* y *nagK* se encontraban juntos y orientados en el mismo sentido en 9 genomas (figura 1.21). El único microorganismo en el que el gen *nagI* no se encontró formando una agrupación con el resto de genes de degradación de gentisato fue *S. stellata* E-37^T. Una posible explicación a la presencia del gen *nagL* alejado del resto de genes *nag* podría ser su alto porcentaje de homología con *hmgC*, tal y como ya se ha comentado en el apartado 3 de este capítulo. El gen *hmgC* codifica para la maleilacetoacetato isomerasa implicada en la degradación del homogentisato y también solía estar separado del resto de genes *hmg* (ver apartado 3). Como las proteínas codificadas por los genes *hmgC* y *nagL* presentaban una elevada identidad de secuencia, no se pudo confirmar que los supuestos genes *hmgC* y *nagL* detectados por predicción genómica en este estudio codificaran para una maleilacetato isomerasa (HmgC) o para una maleilpiruvato isomerasa (NagL). También se observó la presencia del gen *nagX*, codificante para una la 3-hidroxibenzoato-6-hidroxilasa, en los 15 genomas en los que se detectó *nagI* (tabla S1.2). Este gen *nagX* se encontró junto a los genes implicados en la degradación del gentisato en tres de las estructuras génicas presentes en 5 genomas: *Jannaschia* sp. CCS1, *Oceanicola* sp. MCTG156(1a), *Tropicibacter naphthalenivorans* DSM 19561^T, *Ruegeria pomeroyi* DSS-3^T y *C. baekdonensis* B30 (figura 1.21). En estos cinco genomas el gen se encontraba siempre junto a *nagK* y orientado en el mismo sentido (figura 1.21). La enzima 3-hidroxibenzoato 6-hidroxilasa degrada el 3-hidroxibenzoato hasta gentisato (figura 1.20) y la presencia del gen *nagX* codificante para esta enzima junto a los genes implicados en la degradación del gentisato ya había sido descrita en *Polaromonas naphthalenivorans* CJ2 (Park *et al*., 2007). En cuanto a posibles genes reguladores, no se detectó ningún gen codificante para un regulador NagR de tipo LysR (tabla S1.2) como los descritos en *Ralstonia* sp. U2 (Zhou *et al*., 2001). Por lo tanto, no se pudo determinar si estos genes podían estar sometidos a regulación. En 7 de los genomas analizados, los genes codificantes para la degradación de gentisato se localizaban próximos a genes de degradación de naftaleno, formando 7 estructuras génicas diferentes (figura 1.21). Era el caso de *Oceanicola* sp. MCTG156(1a), *Confluentimicrobium naphthalenivorans* NS6^T, *Puniceibacterium* sp. IMCC21224 y las cuatro cepas de *Citreicella aestuarii* analizadas en este estudio (357, 328, 329 y AD8^T) (tabla S1.2). Además, cabe destacar que en *C. indicus* P73^T también se detectaron los genes codificantes para enzimas implicadas en la degradación del naftaleno, aunque en este caso se encontraban alejados de los genes codificantes para la degradación de gentisato

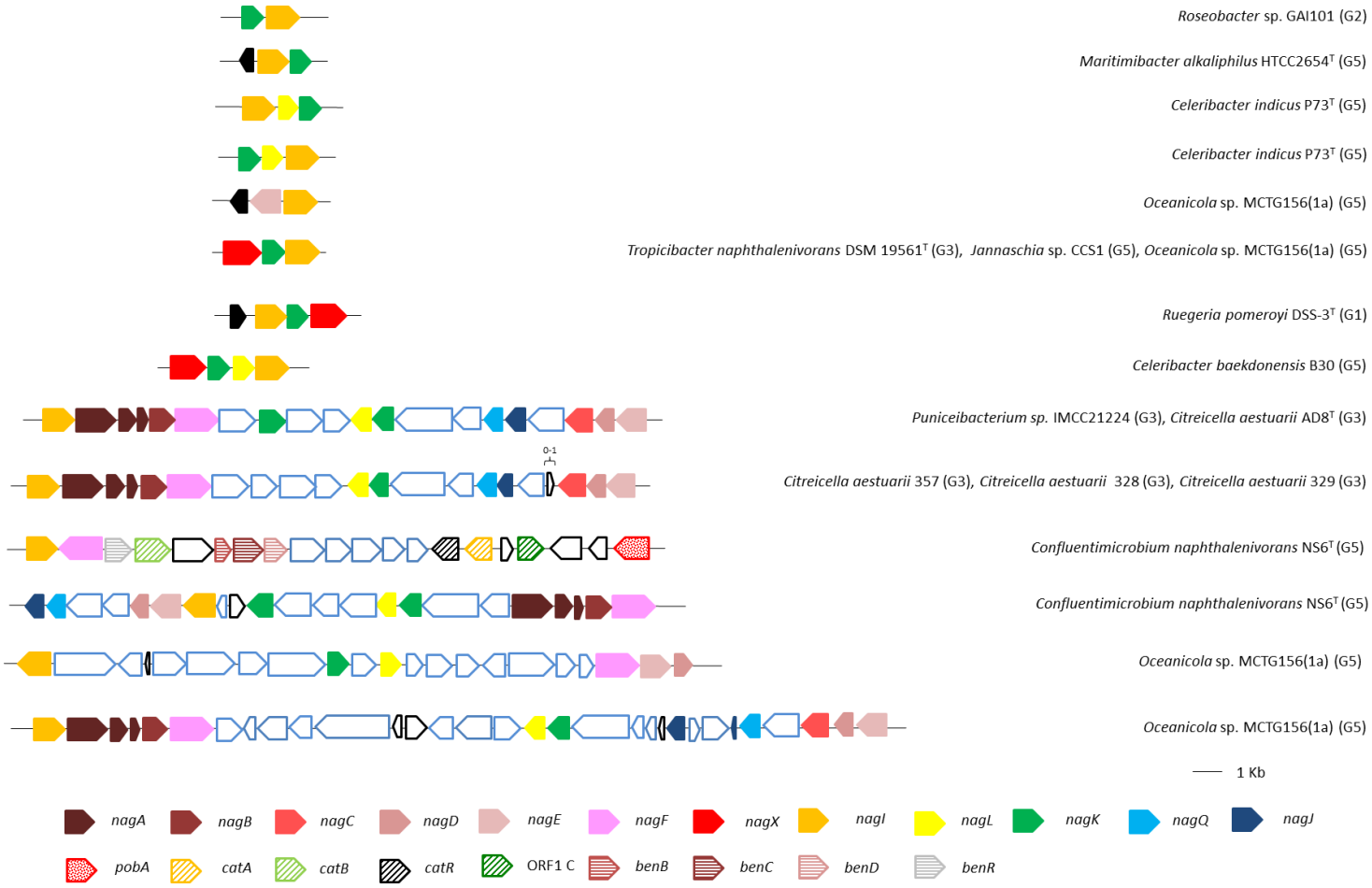

Figura 1.21. Organización de los genes de la vía del gentisato en los genomas del linaje Roseobacter**.** Las abreviaturas son las mismas que las de la figuras 1.5 y 1.20 junto a: *nagQ,* gen para proteína de membrana externa relacionada con el operón *nag* y *nagJ,* gen para glutatión-*S*-transferasa relacionada con el operón *nag*. Los genes de color blanco y contorno negro codifican para proteínas hipotéticas y los genes de color blanco con contorno azul se observaron ocasionalmente, pero no se encuentran conservados, ni codifican para proteínas implicadas en la degradación de compuestos aromáticos. El texto entre paréntesis indica los grupos genómicos.

tabla (S1.2). La presencia de genes para la degradación de naftaleno junto a los genes de degradación del gentisato ya había sido descrita en *C. aestuarii* 357 (Suárez-Suárez, 2013) y en el modelo de degradación *Ralstonia* sp. U2 (Zhou *et al*., 2001). En los genomas de estos roseobacters los genes de degradación del naftaleno y gentisato estaban próximos pero separados por varios genes codificantes para proteínas cuya relación con la degradación de compuestos aromáticos no ha sido demostrada (figura 1.21). Entre estos genes destacaba la presencia de varios genes codificantes para integrasas y transposasas en una de las estructuras génicas de *Oceanicola* sp. MCTG156(1a) (figura 1.21). En cuanto a la ordenación de los genes implicados en la degradación del naftaleno se observó como *nagA*, *nagB* y *nagF* se encontraban juntos y orientados en el mismo sentido junto a *nagI* en seis genomas (figura 1.21). Esta agrupación se encontraba separada de los genes *nagEDC* por varios genes codificantes para proteínas cuya función en la degradación de compuestos aromáticos no ha sido demostrada y por los genes *nagL* y *nagK* (figura 1.21). Por otra parte, en *Confluentimicrobium naphthalenivorans* NS6^{T} junto a una copia del gen *nagI* se detectó el gen *nagF*, varios genes *cat* implicados en la degradación del catecol a beta-cetoadipato, genes *ben* implicados en la degradación de benzoato a catecol y el gen *pobA* implicado en la degradación de 4-hidroxibenzoato a protocatecuato (figura 1.21). Esta agrupación génica de genes *nag*, *cat*, *ben* y *pob* juntos observada en *Confluentimicrobium naphthalenivorans* NS6^{T}, no había sido descrita como tal previamente ni en Roseobacter ni en otros géneros bacterianos, ahora bien no se descarta que pueda darse en más microorganismos.

Por otra parte, en ninguno de los genomas en los que se detectó la presencia de genes para degradación de naftaleno se predijo el gen *nagG* codificante para la salicilato 1,5-hidroxilasa, que cataliza la degradación del salicilato a gentisato (figura 1.20, tabla S1.2). Solo se detectó un posible gen *nagG* en *C. aestuarii* 357 (EIE49957.1) que se encontraba en otra zona del genoma y que presentaba un 39 % de identidad con un *E-value* de 10^{-36} con la subunidad grande (NagG) de la salicilato 1,5-hidroxilasa de *Ralstonia* sp. U2 (AAD12607.1) confirmando el resultado obtenido por la doctora Suárez-Suárez (2013). La función de esta proteína no ha sido demostrada en *C. aestuarii* 357 pero sí se ha demostrado que esta bacteria degrada naftaleno a salicilato y posteriormente a gentisato (Suárez-Suárez, 2013). Por otra parte, el salicilato también se podría degradar a catecol mediante una salicilato 1,2-hidroxilasa (S12H) aunque solo se pudo detectar la presencia de posible gen para esta enzima con el criterio menos restrictivo (30 % de identidad con un *E-value* menor de 10^{-5}) en *Celeribacter indicus* P73^{T} (tabla S1.2). Por lo tanto, la detección de las dos hidroxilasas de salicilato fue baja en los genomas de Roseobacter estudiados con el criterio utilizado.

Finalmente, también se detectó la presencia de los genes codificantes para la proteína de membrana NagQ y una glutatión-*S*-transferasa relacionada con la degradación del naftaleno (NagJ), en 8 genomas de Roseobacter (tabla S1.2). En 7 de estos genomas los genes codificantes para estas proteínas se encontraban en la misma estructura génica que los genes para degradación de gentisato (figura 1.21). En estos 7 genomas los genes *nagQJ* se encontraban juntos o solo separados por varias transposasas y orientados en el mismo sentido (figura 1.21). En cuanto a su posición respecto al resto de genes de degradación de naftaleno, en *Puniceibacterium* sp. IMCC21224, *Oceanicola* sp. MCTG156(1a) y las cuatro cepas de *C. aestuarii* los genes *nagQJ* se encontraban entre los genes *nagLK* y los genes *nagCDE* (figura 1.21). En *Confluentimicrobium naphthalenivorans* NS6^{T} se encontraban corriente arriba de

nagD (figura 1.21) y en *C. indicus* P73^T corriente abajo de *nagD* (datos no mostrados puesto que los genes para la degradación del naftaleno en *C. indicus* P73^T no se encuentran agrupados junto a los genes implicados en la degradación del gentisato). La presencia de *nagQ* y *nagJ* junto al operón de degradación de naftaleno ya había sido descrito en *Ralstonia* sp. U2 (Zhou *et al*., 2001).

7. Vía de degradación de benzoil-CoA

El benzoato además de ser degradado vía catecol hasta dar lugar a intermediarios del ciclo de Krebs (apartado 2.2.) también puede degradarse mediante un paso previo de activación con un grupo Coenzima A dando lugar a benzoil-CoA (Pérez-Pantoja *et al*., 2010a). La vía de degradación del benzoil-CoA era una vía poco detectada en los genomas de Roseobacter puesto que solo se pudo detectar la presencia de la enzima principal en el 11,5 % de los genomas incluidos en este estudio (tabla 1.1). Hasta la fecha de realización de esta tesis se había detectado la presencia de la vía de degradación del benzoil-CoA en el 16 % de los genomas incluidos en el estudio de Newton y colaboradores (2010) en base a la presencia del gen *boxB* pero no se había detectado la presencia de los genes implicados en la degradación del benzoato vía catecol (ver apartado 2.1.2.).

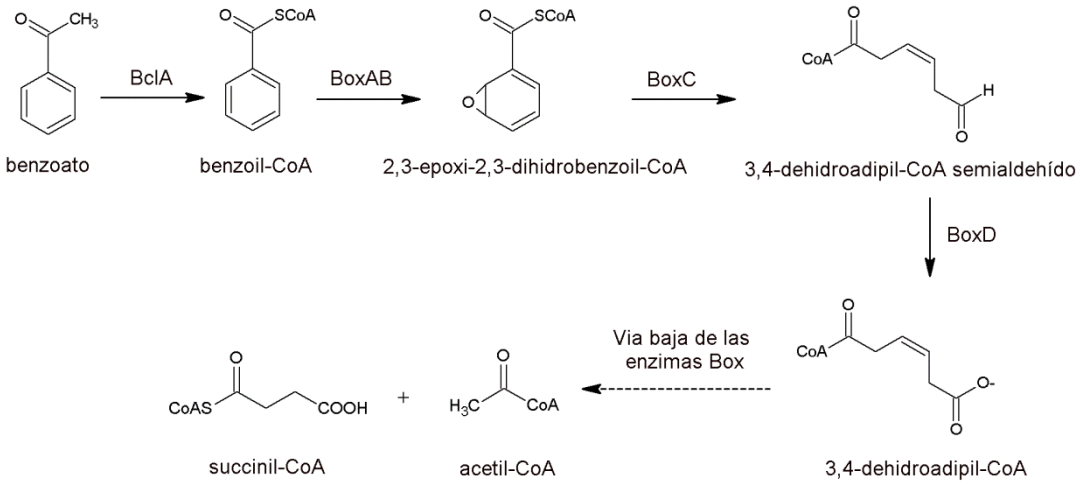

Figura 1.22. Vía de degradación de benzoil-CoA. Abreviatura de las enzimas: BclA, benzoato-CoA ligasa; BoxAB, benzoil-CoA dioxigenasa; BoxC, benzoil-CoA dihidrodiol liasa; BoxD, 3,4-dehidroadipil-CoA semialdehído deshidrogenasa. Figura elaborada a partir de Valderrama *et al*., 2012.

La vía de degradación del benzoil-CoA está compuesta por reacciones en las que intervienen cuatro enzimas (figura 1.22). Estas enzimas actúan en cuatro pasos: una benzoato-CoA ligasa (BclA) se encarga de la activación del benzoato para dar lugar a benzoil-CoA, una dioxigenasa (BoxAB) se encarga de transformar el benzoil-CoA en 2,3-epoxi-2,3-dihidrobenzoil-CoA, perdiendo así el carácter aromático y la liasa BoxC se encarga de la rotura no oxigenolítica del compuesto cíclico 2,3-epoxi-2,3-dihidrobenzoil-CoA. A continuación, una 3,4-dehidroadipil-CoA semialdehído deshidrogenasa (BoxD) da lugar a 3,4-dehidroadipil-CoA, el cual es degradado hacia intermediarios del ciclo de Krebs por varias enzimas implicadas en un metabolismo de tipo beta-oxidación que constituyen la denominada vía baja de las enzimas Box (Valderrama *et al*., 2012; Pérez-Pantoja *et al*., 2010a). En concreto, en esta vía baja de las enzimas Box se cree que se transforma el 3,4-dehidroadipil-CoA a 2,3-dehidroadipil-CoA, el cual se hidrata dando

lugar a beta-hidroxiadipil-CoA. Este compuesto es el sustrato de una deshidrogenasa y da lugar a beta-cetoadipil-CoA que por la acción de una tiolasa da lugar a succinil-CoA y acetil-CoA (Zaar *et al*., 2001; Gescher *et al*., 2005).

En este estudio se detectó la presencia de genes para la enzima principal, benzoil-CoA dioxigenasa (BoxAB), en 7 genomas del linaje Roseobacter: *R. pomeroyi* DSS-3^T, *Ruegeria* sp. KLH11, *Roseobacter* sp. GAI101, *Sulfitobacter mediterraneus* 1FIGIMAR09, *S. stellata* E.37^T, *T. naphthalenivorans* DSM 19561^T y *M. alkaliphilus* HTCC2654^T (tabla S1.2). Además, cabe destacar que en *Jannaschia* sp. CSS1 se detectó el gen para la subunidad BoxB (tabla S1.2). En los 8 genomas en los que se detectó *boxB* también se detectaron los genes *boxC*, *bclA* y el regulador *boxR* (tabla S1.2). No se pudo detectar la presencia del gen *boxD* en ningún genoma de Roseobacter (tabla S1.2) resultado que sugirió que los miembros del linaje Roseobacter no podían transformar el 3,4-dehidroadipil-CoA semialdehído hasta 3,4-dehidroadipil-CoA (figura 1.22). Ahora bien, como ya se ha comentado anteriormente, se conoce que *Sagittula stellata* E-37^T puede catabolizar el benzoato (Gulvik y Buchan 2013) a pesar de que no hayamos podido detectar *boxD* en su genoma (tabla S1.2). Por lo tanto, para poder determinar si los miembros del linaje Roseobacter pueden o no transformar el 3,4-dehidroadipil-CoA semialdehído hasta 3,4-dehidroadipil-CoA deberían realizarse experimentos de cultivo y análisis químicos. Además, para poder confirmar la presencia de este gen, quizás se tendrían que cambiar los criterios de búsqueda, puesto que este enzima podría ser lo suficientemente diferente para que no se hubiera podido detectar debido a los criterios usados.

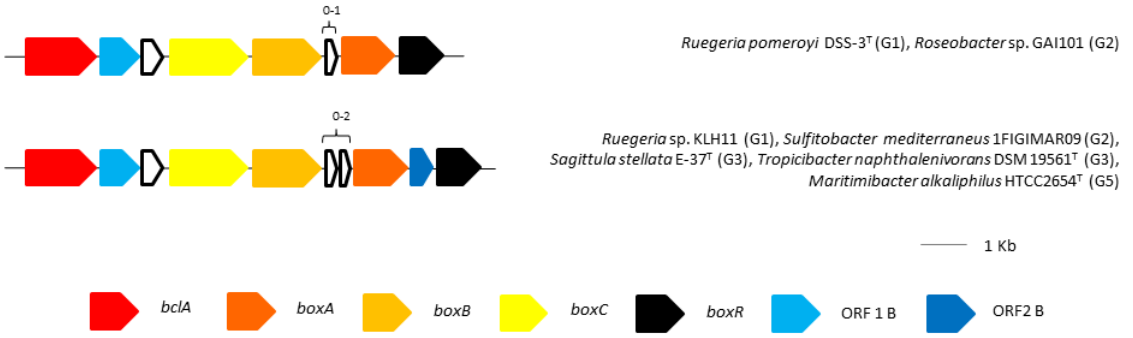

Figura 1.23. Organización de los genes de la vía del benzoil-CoA en los genomas del linaje Roseobacter. Las abreviaturas son las mismas que las de la figura 1.22 junto a: *boxR*, gen para regulador transcripcional; ORF1 B, gen para hidrolasa de la via baja de las enzimas Box y ORF2 B, gen para tioesterasa de via baja de las enzimas Box. Los genes de color blanco y contorno negro codifican para proteínas hipotéticas. El texto entre paréntesis indica los grupos genómicos.

Respecto a la organización de los genes *box* alrededor de los genes *boxAB* se pudieron observar dos estructuras génicas, básicamente debido a la presencia de un gen codificante para una tiosesterasa (figura 1.23) que podría estar involucrada en la vía baja de la degradación del benzoil-CoA en base al trabajo previamente descrito en *R. pomeroyi* DSS-3^T y *Jannaschia* sp. CSS-1 por Valderrama y colaboradores (2012). El gen para esta tioesterasa se encontraba entre los genes *boxA* y *boxR* (figura 1.23). En las dos estructuras se pudo observar que el gen *bclA* se encontraba corriente arriba de *boxC* y orientado en el mismo sentido aunque estaban separados por entre uno y dos genes. Uno de estos genes codificaba para una hidrolasa que podía estar involucrada en la vía baja de las enzimas Box (figuras 1.22 y 1.23). La

presencia de genes implicados en la vía baja de degradación del benzoil-CoA entre *bclA* y *boxC* ya había sido descrita por Valderrama y colaboradores (2012) para *R. pomeroyi* DSS-3^T y *Jannaschia* sp. CSS1. También se observó que los genes *boxA*, *boxB* y *boxC* solían encontrarse juntos y orientados en el mismo sentido aunque en el caso de *R. pomeroyi* DSS-3^T, *S. stellata* E-37^T, *M. alkaliphilus* HTCC2654^T o *T. naphthalenivorans* DSM 19561^T se encontraron genes codificantes para una y dos proteínas hipotéticas separando *boxB* y *boxA* (figura 1.23). En cuanto a la localización del gen regulador *boxR*, este siempre se encontró corriente abajo de *boxCB(A)* y orientado en el mismo sentido (figura 1.23).

Cabe destacar que en 6 de estos 7 microorganismos (excepto en *Ruegeria* sp. KLH11) se encontraron los genes para degradación de benzoil-CoA junto a genes de otras vías de degradación de compuestos aromáticos. En *S. stellata* E-37^T, *M. alkaliphilus* HTCC2654^T, *T. naphthalenivorans* DSM 19561^T y *R. pomeroyi* DSS-3^T se localizaban junto a los genes para la degradación del ferulato (ver apartado 2.3.1). En el caso de *R. pomeroyi* DSS-3^T se encontraron junto a genes para la degradación de gentisato (apartado 6). Además, cabe destacar que *R. pomeroyi* DSS-3^T fue el único de los genomas estudiados en el que el gen codificante para la hidrolasa conservada junto a los genes de degradación de benzoil-CoA también presentaba más de un 30 % de identidad de secuencia con un *E-value* menor de 10^{-5} con la proteína Atu1417 de la vía de degradación del ferulato, sugiriendo que esta proteína podría estar implicada en estas dos vías de degradación. Finalmente en *Roseobacter* sp. GAI101 y *Sulfitobacter mediterraneus* 1FIGIMAR09 los genes para degradación de benzoil-CoA se encontraban junto a los genes para la degradación del homogentisato (apartado 3) y en *Jannaschia* sp. CSS1 se encontraban junto a los genes para la degradación del fenilacetil-CoA (apartado 4). Estos tipos de agrupaciones de genes implicados en vías de degradación de compuestos aromáticos no relacionados no habían sido descritas como tales previamente en Roseobacter. Por otra parte, la mayor parte de genomas en los que se han detectado estas agrupaciones de degradación se encuentran en estado de boceto. Por lo tanto, no se puede saber si estas agrupaciones se encuentran en el cromosoma o en plásmidos. Solo en el caso de *R. pomeroyi* DSS-3^T, cuyo genoma está completado, se puede confirmar que estas agrupaciones se encuentran en el cromosoma.

8. Resumen del potencial degradador de compuestos aromáticos presente en genomas del linaje Roseobacter

Con todos los datos de predicción génica obtenidos se generó un mapa catabólico del potencial degradador de compuestos aromáticos en 61 genomas del linaje Roseobacter (figura 1.24). Cabe destacar que la mayoría de las vías de degradación periféricas de compuestos aromáticos detectadas en los genomas de Roseobacter analizados eran las vías de degradación de compuestos aromáticos precursores del protocatecuato (ferulato, 4-hidroxibenzoato, 3-hidroxibenzoato) y del homogentisato (4-hidroxifenilacetato y 4-hidroxifenilpiruvato), que son los dos compuestos aromáticos centrales para los que los miembros del linaje Roseobacter presentaban un mayor potencial catabólico (figura 1.25). Por lo tanto, era coherente que las vías periféricas más detectadas fueran aquellas que daban lugar a los dos compuestos centrales más detectados: protocatecuato y homogentisato (figura 1.25).

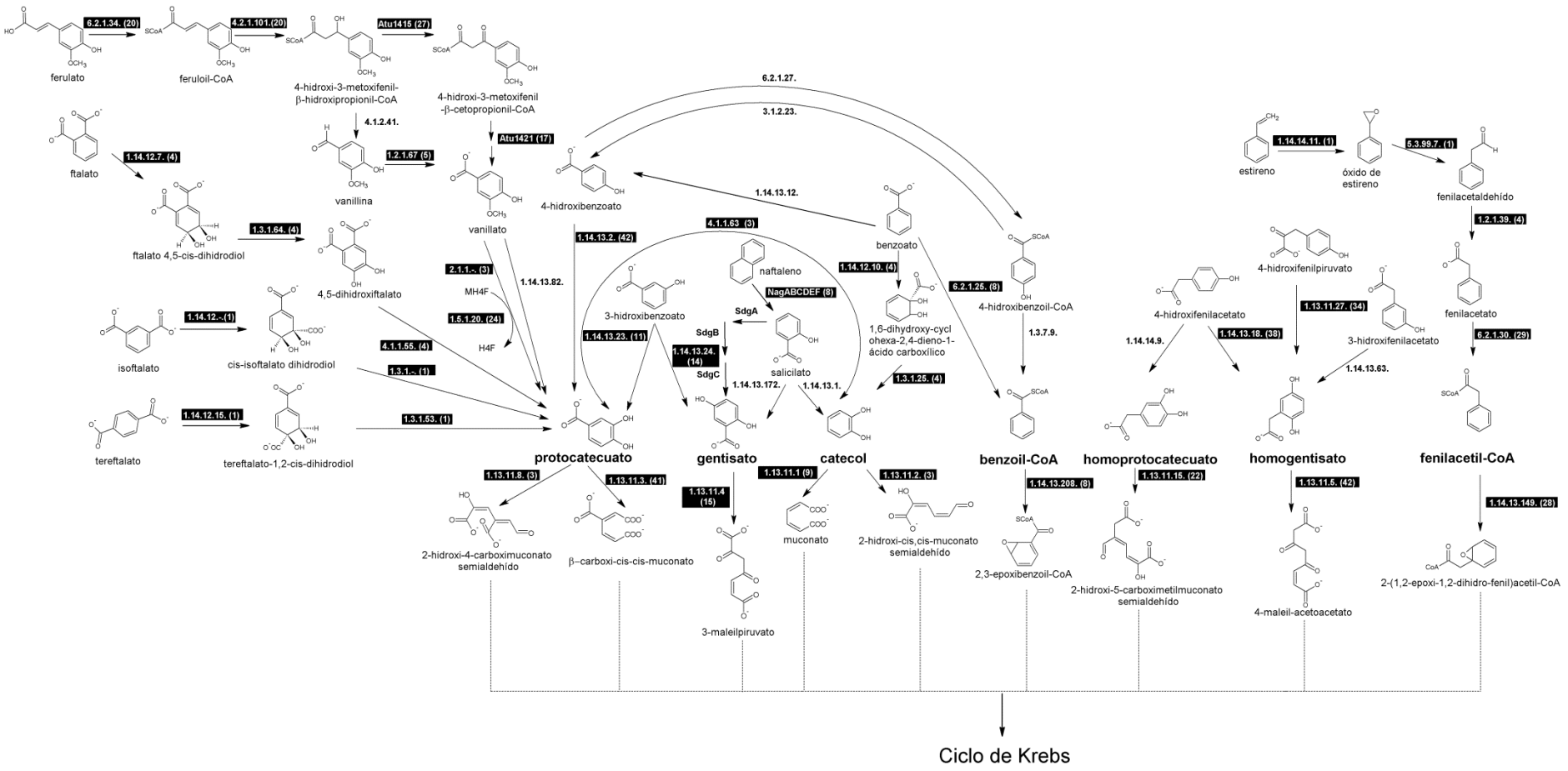

Figura 1.24. Mapa catabólico para la degradación de compuestos aromáticos en el linaje Roseobacter. Todos los números EC y las abreviaturas de esta figura se encuentran junto al correspondiente nombre de la enzima en la tabla S1.2. Los números EC de las enzimas cuyos genes han sido detectados en los miembros del linaje Roseobacter se muestran con fondo negro. Entre paréntesis se indica el número de genomas en el que se han encontrado los genes codificates para cada enzima.

El estudio conjunto del potencial catabólico de los siete compuestos aromáticos centrales estudiados y de los correspondientes compuestos aromáticos periféricos en los genomas del linaje Roseobacter permitó observar que este potencial no solo era desigual en los genomas de los cinco grupos filogenómicos, tal y como ya se había descrito en apartados anteriores, sino que también era desigual dentro de los miembros de un mismo grupo filogenómico, sobre todo en el caso de los grupos G3 y G5 (figura 1.25). En concreto se observó que los miembros del grupo G3 *Roseovarius* sp. 217, *Roseovarius* sp. TM1035, *Roseobacter* sp. Azwk-3b, *Roseovarius nubinhibens* ISM^T, *Oceanicola* sp. S124 y *Pseudooceanicola batsensis* HTCC2597^T presentaban menor número de vías de degradación de compuestos aromáticos centrales que el resto de miembros del G3 (figura 1.25). El potencial catabólico de estas seis cepas se debía básicamente a la presencia de los genes implicados en la degradación del protocatecuato y de la vía incompleta para la degradación del fenilacetil-CoA y homoprotocatecuato. Estos resultados eran similares a los de los genomas del G4, grupo filogenómico en el que se había detectado un menor potencial para la degradación de compuestos aromáticos centrales (apartado 1; figura 1.25). En cuanto a los miembros del grupo G5 con un menor potencial para el catabolismo de compuestos aromáticos centrales se observó que *Thalassobacter stenotrophicus* 16PALIMAR09 y 1CONIMAR09 y *Rhodobacterales* bacterium HTCC2255 no presentaban ninguna de las vías de degradación estudiadas de forma completa (figura 1.25). Solo los dos primeros podían presentar las vías de degradación del protocatecuato y del fenilacetil-CoA de forma incompleta (figura 1.25). En cuanto a los genomas en los que se detectaron un mayor número de rutas centrales de degradación de compuestos aromáticos completas se encontraban los miembros del G1 (siendo *Ruegeria* sp. KLH11, *Ruegeria conchae* TW15^T, *Ruegeria* sp. 6PALISEP08 y *Ruegeria lacuscaerulensis* ITI-1157^T los que presentan un menor potencial dentro de este grupo. En otros grupos filogenómicos los genomas donde se encontraron más rutas centrales fueron *Sulfitobacter mediterraneus* 1FIGIMAR09 (G2), *Citreicella* sp. SE45 (G3), *Sagittula stellata* E-37^T (G3), *Tropicibacter naphthalenivoran*s DSM 19561^T (G3), *Jannaschia* sp. CCS1 (G5), *Rhodobacteraceae* bacterium HTCC2150 (G5) y *Maritimibacter alkaliphilus* HTCC2654^T (G5). De entre ellos, *Sagittula stellata* E-37^T era el que presentaba un mayor potencial para el catabolismo de compuestos aromáticos centrales puesto que no solo presentaba los genes codificantes para las vías completas de degradación del protocatecuato, homogentisato, fenilacetato, homoprotocatecuato y gentisato, sino que también presentaba vías presuntamente incompletas para la degradación del catecol a partir de su rotura en posición *orto* y *meta* (figura 1.25).

En cuanto a rutas de degradación periféricas completas, se observó que, en base a la predicción genómica, los microorganismos que presentaban un menor potencial para degradar compuestos aromáticos periféricos (0-2 compuestos) eran: tres miembros del grupo G1 (*Ruegeria conchae* TW15^T, *Ruegeria* sp. 6PALISEP08 y *Ruegeria lacuscaerulensis* ITI-1157^T), dos miembros del grupo G2 (*Roseobacter denitrificans* Och114^T y *Rhodobacteraceae* bacterium HTCC2083), siete miembros del grupo G3 (las cepas de *C. aestuarii* AD8^T, 328 y 329, *Roseovarius* sp. 217, *Roseovarius* sp. TM1035, *Roseovarius nubinhibens* ISM^T y *Roseobacter* sp. Azwk-3b), nueve miembros del grupo G4 (*Loktanella vestfoldensis* SKA53, *Loktanella* sp. 3ANDIMAR09, *Roseobacter* sp. CCS2, *Octadecabacter arcticus* 238^T, *Octadecabacter antarcticus* 307^T, *Thalassobium* sp. R2A62, *Ketogulonicigenium vulgare* Y25,

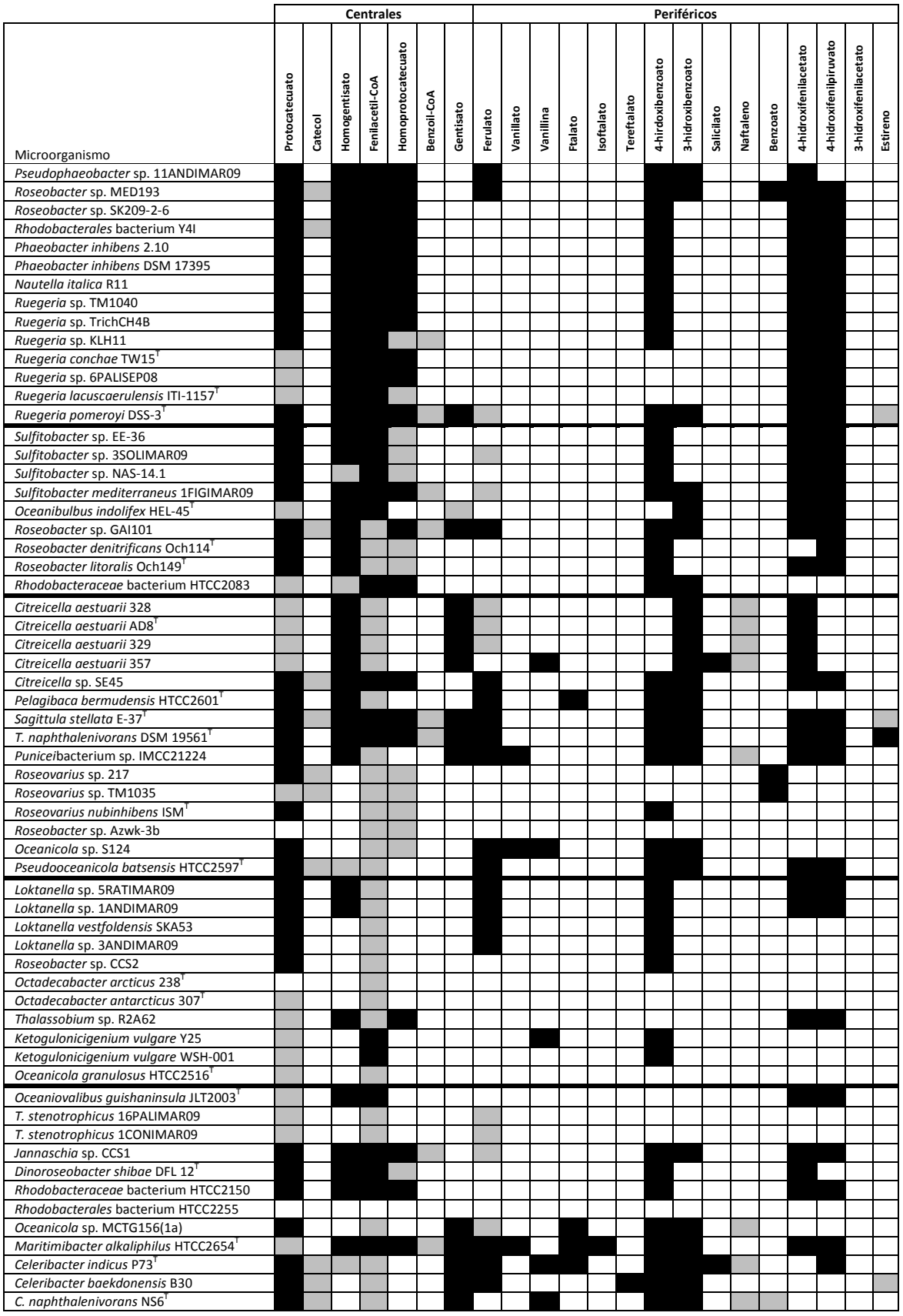

Figura 1.25. Compuestos que potencialmente pueden ser degradados por miembros del linaje Roseobacter en base al estudio de predicción genómica. Los genomas del linaje Roseobacter están ordenados en base a su filogenia (Luo y Moran, 2014; Mas-Lladó, 2017). Código de color: Negro, detección de genes de la vía completa; gris, detección de genes de la vía incompleta y blanco, ausencia de detección de los genes de la vía.

Ketogulonicigenium vulgare WSH-001 y *Oceanicola granulosus* HTCC2516^T) y cinco miembros del grupo G5 (*Oceaniovalibus guishaninsula* JLT2003^T, *Thalassobacter stenotrophicus* 16PALIMAR09 y 1CONIMAR09, *Dinoroseobacter shibae* DFL 12^T y *Rhodobacterales* bacterium HTCC2255). Por lo tanto, entre los microorganismos con un menor potencial para el catabolismo de compuestos aromáticos periféricos se encontraban en general los mismos microorganismos que presentaban un menor potencial para el catabolismo de compuestos aromáticos centrales (figura 1.25). Por otra parte, los microorganismos en los que se detectaron un mayor potencial para degradar compuestos aromáticos periféricos (6 a 8 compuestos) fueron: dos miembros del grupo G3 (*Tropicibacter naphthalenivoran*s DSM 19561^T y *Puniceibacterium* sp. IMCC21224) y dos miembros del grupo G5 (*Maritimibacter alkaliphilus* HTCC2654^T y *Celeribacter indicus* P73^T) (figura 1.25).

9. Crecimiento de miembros del linaje Roseobacter con hidrocarburos como fuente de carbono y energía

Para complementar el estudio del potencial catabólico de los miembros del linaje Roseobacter se determinó la capacidad de varios aislados de este linaje para crecer a partir de protocatecuato, 4-hidroxibenzoato, catecol, gentisato, salicilato, homogentisato, benzoato, fenilacetato y homoprotocatecuato como única fuente de carbono y energía (ver detalles de los experimentos en el apartado 2.1 de material y métodos). Cabe añadir que también se analizó la capacidad de crecimiento de *C. aestuarii* AD8^T en naftaleno y dibenzotiofeno al 0,1 % (p/v). En concreto, se observó que *C. aestuarii* AD8^T podía crecer con naftaleno 0,1% (p/v) y con dibenzotiofeno 0,1 % (p/v), resultado coherente con el descrito en la cepa *C. aestuarii* 357 por Suárez-Suárez (2013).

En general se observó concordancia entre los resultados de predicción genómica y los resultados de los cultivos (tabla 1.3). En concreto, en la degradación del fenilacetato se detectó una correspondencia total entre los resultados de predicción génica y los resultados de cultivo. En el crecimiento con el resto de compuestos estudiados se detectaron algunas discrepancias, siendo las vías de degradación del homogentisato y catecol las vías en las que se detectó una mayor proporción de incongruencias entre los resultados de predicción genómica y los de cultivo. Estas discrepancias fueron de dos tipos. Por un lado se detectó el crecimiento de varios aislados en compuestos cuya vía de degradación no había sido detectada (p.e. crecimiento en gentisato de *O. indolifex* HEL-45^T y crecimiento en homoprotocatecuato de *R. lacuscaerulensis* ITI-1157^T) o se había considerado incompleta (p.e. *O. granulosus* HTCC2516^T en protocatecuato, *Roseobacter* sp. MED193 en catecol y *S. stellata* E-37^T en benzoato) (tabla 1.3). Por otro lado, también se dio el caso de falta de crecimiento de varios aislados en compuestos cuya vía de degradación había sido detectada (p.e. *C. aestuarii* AD8^T, *C. aestuarii* 357 y *S. stellata* E-37^T en homogentisato o de *R. pomeroyi* DSS-3^T en homoprotocatecuato) (tabla 1.3). También cabe destacar que se detectaron crecimientos de aislados en presencia de varios compuestos aromáticos a pesar de presentar vías incompletas para la degradación de esos compuestos. En el caso del crecimiento de *S. stellata* E-37^T en benzoato, este crecimiento fue un resultado coherente con el obtenido por Gulvick y Buchan (2013), a pesar de tener una discrepancia con los resultados de predicción genómica. Por lo tanto, el crecimiento de *S. stellata* E-37^T en benzoato en estos dos experimentos podría sugerir que que *S. stellata* E-37^T presentaría una enzima BoxD o alguna otra enzima que puede reemplazar su función,

diferente al modelo utilizado para su búsqueda, y por ello no pudo ser detectada con los criterios de detección de este estudio.

Tabla 1.3. Crecimiento en medio mineral con diferentes compuestos aromáticos como fuente de carbono y energía. Las casillas marcadas en gris indican discrepancia entre los resultados del cultivo y la predicción genómica para vías completas.

Grupo	Microorganismo	Protocateuato	4-hidroxibenzoato	Catecol	Gentisato	Salicilato	Homogentisato	Benzoato	Fenilacetato	Homoprotocatecuato
G1	*Roseobacter* sp. MED193	++	+	+			+	++	++	++
	Phaeobacter inhibens DSM 17395	++	+	+				–	++	++
	Ruegeria sp. 6PALISEP08	–	–							
	Ruegeria lacuscaerulensis ITI-1157^T	++	–						++	+
	Ruegeria pomeroyi DSS-3^T	+	–	+	++	–		+		–
G2	*Sulfitobacter* sp. EE-36							–		–
	Oceanibulbus indolifex HEL-45^T	++	–		+				++	
	Roseobacter litoralis Och149^T	+	–	++					–	
G3	*Citreicella aestuarii* 328					+				
	Citreicella aestuarii AD8^T	–			++	++	–	–	–	–
	Citreicella aestuarii 329					+				
	Citreicella aestuarii 357	–	–	–	++	++	–	–	–	–
	Pelagibaca bermudensis HTCC2601^T	++	++							
	Sagittula stellata E-37^T	++	+++	++	+		–	++	+++	+
	Roseovarius nubinhibens ISM^T	+	+							
	Pseudooceanicola batsensis HTCC2597^T	+	+	++						
G4	*Oceanicola granulosus* HTCC2516^T	+++	–							
G5	*Thalassobacter stenotrophicus* 16PALIMAR09	–	–							
	Thalassobacter stenotrophicus 1CONIMAR09	–	–							
	Dinoroseobacter shibae DFL 12^T	–	–	–					+	

"+" significa una DO_{600} >0.15; "++"significa 0.6 >DO_{600} <0.2; "+++"significa DO_{600} ≥ 0.6; "–" significa que no se detectó crecimiento y la casilla en blanco significa que no se ha determinado el crecimiento de un determinado microorganismo en esas condiciones.

Por otra parte, no solo se detectaron discrepancias entre los resultados de crecimiento y de predicción genómica de este estudio, sino que también se detectaron algunas discrepancias con resultados de crecimiento de otros estudios. Por ejemplo, siguiendo con *S. stellata* E-37^T, en este estudio este microorganismo creció en catecol y en gentisato, al contrario de lo que habían descrito Gulvick y Buchan y (2013). Otro ejemplo de discrepancia entre los resultados de este estudio y el de Gulvick y Buchan (2013) fue la falta de crecimiento de *R. pomeroyi* DSS-3^T en 4-hidroxibenzoato. Este caso es distinto a los anteriores, puesto que en caso también esperábamos crecimiento de *R. pomeroyi* DSS-3^T en este compuesto al detectar el gen *pobA* y los genes para la degradación de protocatecuato (ver tabla S1.2).

En general, se pueden proponer varias posibles explicaciones para explicar las contradicciones entre los resultados de la predicción genómica y de los cultivos. En los casos en los que no se detectó crecimiento a partir de un compuesto aromático para el cual se habían detectado los genes codificantes para las enzimas implicadas en su degradación, se podría pensar que las condiciones de cultivo ensayadas no fueron adecuadas para el crecimiento con estos compuestos. Por lo tanto, se deberían repetir estos experimentos ensayando varias concentraciones de los compuestos. Otra opción sería realizar cultivos donde también se

añadiese alguno de los compuestos que se han descrito como inductores de la expresión de los genes para degradación de ese compuesto. Esta estrategia fue la que se siguió con *S. stellata* E-37^T para observar crecimiento en protocatecuato, ya que a pesar de presentar la vía completa para la degradación del protocatecuato, solo pudo crecer en este compuesto si el inóculo provenía de un cultivo anterior con 4-hidroxibenzoato. En cuanto a los casos contrarios, en los que se detectó crecimiento a pesar de no haber detectado los genes codificantes para las enzimas implicadas en su degradación, se pueden sugerir varias explicaciones basadas en las limitaciones metodológicas del estudio. En primer lugar, la mayor parte de estos genomas se encuentran en estado de boceto (tabla S1.1) con lo que no puede descartarse la posibilidad de que falte información, aunque esto sea improbable. También podría darse el caso de que los genes no detectados en estos genomas fueran suficientemente diferentes en secuencia a los usados como modelo y por ello no fueran detectados con el criterio de detección usado. Finalmente, podría suceder que estos microorganismos tuvieran vías alternativas para la degradación de estos compuestos que sean desconocidas hasta la fecha. Es más, podría darse una situación similar a la descrita en *Thauera aromatica*, organismo que puede crecer con catecol como fuente de carbono, aunque no se ha detectado ningún gen codificante para enzimas específicas de la degradación del catecol (Ding *et al*., 2008). En concreto, se sabe que *T. aromatica* presenta varias enzimas y reguladores involucrados en la degradación del fenol, 3-hidroxibenzoato, 4-hidroxibenzoato y benzoato cuya expresión se ve inducida cuando este organismo crece en catecol. Por ello, se ha postulado que estas enzimas podrían ser lo suficiente promiscuos como para intervenir también en la degradación del catecol (Ding *et al*., 2008).

CAPÍTULO 2: TOLERANCIA A HIDROCARBUROS EN MIEMBROS DEL LINAJE ROSEOBACTER

El objeto de esta parte de la tesis era determinar la respuesta de aislados pertenecientes al linaje Roseobacter a su exposición a hidrocarburos, tales como el tolueno y el naftaleno, que podrían resultarles tóxicos. En base a los experimentos de Nogales y colaboradores (2007) y Piña-Villalonga (2012) se sabía que los miembros del linaje Roseobacter son componentes mayoritarios en los puertos deportivos de aguas de Mallorca. Otros autores también habían mostrado la proliferación de bacterias de este linaje en presencia de hidrocarburos (ver apartado 3 de la introducción). En el momento de realización de esta tesis se conocía que los miembros del linaje Roseobacter presentaban capacidad degradadora de compuestos aromáticos (ver apartado 3 de la introducción), pero no se había descrito qué mecanismos les permitian tolerar la presencia en el ambiente de estos compuestos potencialmente tóxicos. Solo se disponía de información previa sobre cambios en el proteoma de *R. pomeroyi* DSS-3^T tras exposición a diferentes factores de estrés que incluyeron la incubación con naftaleno y diésel (Christie-Oleza *et al*., 2012), de la cepa *C. aestuarii* 357 creciendo a expensas de naftaleno como fuente de carbono y energía (resultados no publicados del grupo de investigación), y de siete aislados de puertos deportivos de Mallorca (*Ruegeria* sp. 6PALISEP08, *Sulfitobacter mediterraneus* 1FIGIMAR09, *Sulfitobacter* sp. 3SOLIMAR09, *Loktanella* sp. cepas 1ANDIMAR09 y 3ANDIMAR09, y *Thalassobacter stenotrophicus* cepas 16PALIMAR09 y 1CONIMAR09) tras exposición a agua de mar contaminada artificialmente con diésel (Mas-Lladó, 2017). Además, no se había realizado ningún experimento con aislados del linaje Roseobacter expuestos a tolueno, un hidrocarburo monoaromático de alta toxicidad utilizado en estudios de tolerancia a hidrocarburos en otras bacterias (Ramos *et al*., 2002; Segura *et al*., 2007). Por ello en este trabajo se propuso realizar un experimento de exposición a naftaleno y tolueno de aislados del linaje Roseobacter con procedimientos similares a los utilizados en otras bacterias para estudiar tolerancia a hidrocarburos (Segura *et al*., 2007) y determinar si en los miembros del linaje Roseobacter se daba una respuesta similar a la documentada en microorganismos modelo de degradación de hidrocarburos del género *Pseudomonas* (Ramos *et al*., 2015).

1. Aproximación proteómica a la respuesta a naftaleno y tolueno en tres miembros del linaje Roseobacter

Para analizar la respuesta a los hidrocarburos naftaleno y tolueno se realizó un experimento preliminar en el que se incluyeron todos los aislados de puertos deportivos de Mallorca cuyos genomas se habían analizado en esta tesis, y que incluían los que se utilizaron en el experimento con diésel mencionado anteriormente (Mas-Lladó, 2017). Todos estos aislados se obtuvieron en un estudio anterior mediante siembra en placa en un medio de cultivo estándar para bacterias marinas (agar marino) sin hidrocarburos (Piña-Villalonga, 2012). La predicción de genes para el catabolismo de compuestos aromáticos realizada en esta tesis mostró la presencia de las rutas centrales mayoritarias en los genomas de estas bacterias, a excepción de las dos cepas de *T. stenotrophicus* (ver por ejemplo la figura 1.25, capítulo 1). Para estos aislados, y el resto de bacterias del linaje Roseobacter analizados en esta tesis, se hizo también una predicción génica de posibles bombas de tipo RND similares a las descritas en otras bacterias para tolerancia a hidrocarburos (ver apartado 2.1 de este capítulo). En los experimentos preliminares se analizaron los extractos proteicos de todos los aislados por electroforesis SDS-PAGE (ver apartado 3.1. de material y métodos para el detalle

experimental). Este análisis electroforético evidenció cambios en algunos perfiles proteicos tras la exposición a hidrocarburos (datos no mostrados). Ante la imposibilidad de analizar todos los aislados por proteómica masiva se seleccionaron finalmente tres microorganismos: (1) *R. pomeroyi* DSS-3^T, como modelo de un Roseobacter bien estudiado, no degradador de naftaleno ni de tolueno, y con el que ya se habían realizado estudios de respuesta a naftaleno (Christie-Oleza *et al*., 2012); (2) *C. aestuarii* 357, como modelo de un Roseobacter degradador de naftaleno (Suárez-Suárez, 2013), y (3) *Sulfitobacter* sp. 3SOLIMAR09. En los experimentos preliminares realizados en esta tesis se observaron cambios en los perfiles electroforéticos de proteínas de 3SOLIMAR09 tras exposición a naftaleno 0,1 % (p/v) y tolueno 0,1 % y 0,3 % (v/v), en comparación con los observados en cultivos control no tratados (ver figura suplementaria S2.1). En el genoma de esta bacteria se había predicho la presencia de genes catabólicos para degradación de proteocatecuato, homogentisato y fenilacetato (ver figuras 1.2, 1.11 y 1.14 del capítulo 1), así como para una bomba de tipo AcrAB-TolC (ver apartado 2.1 y tabla S2.1).

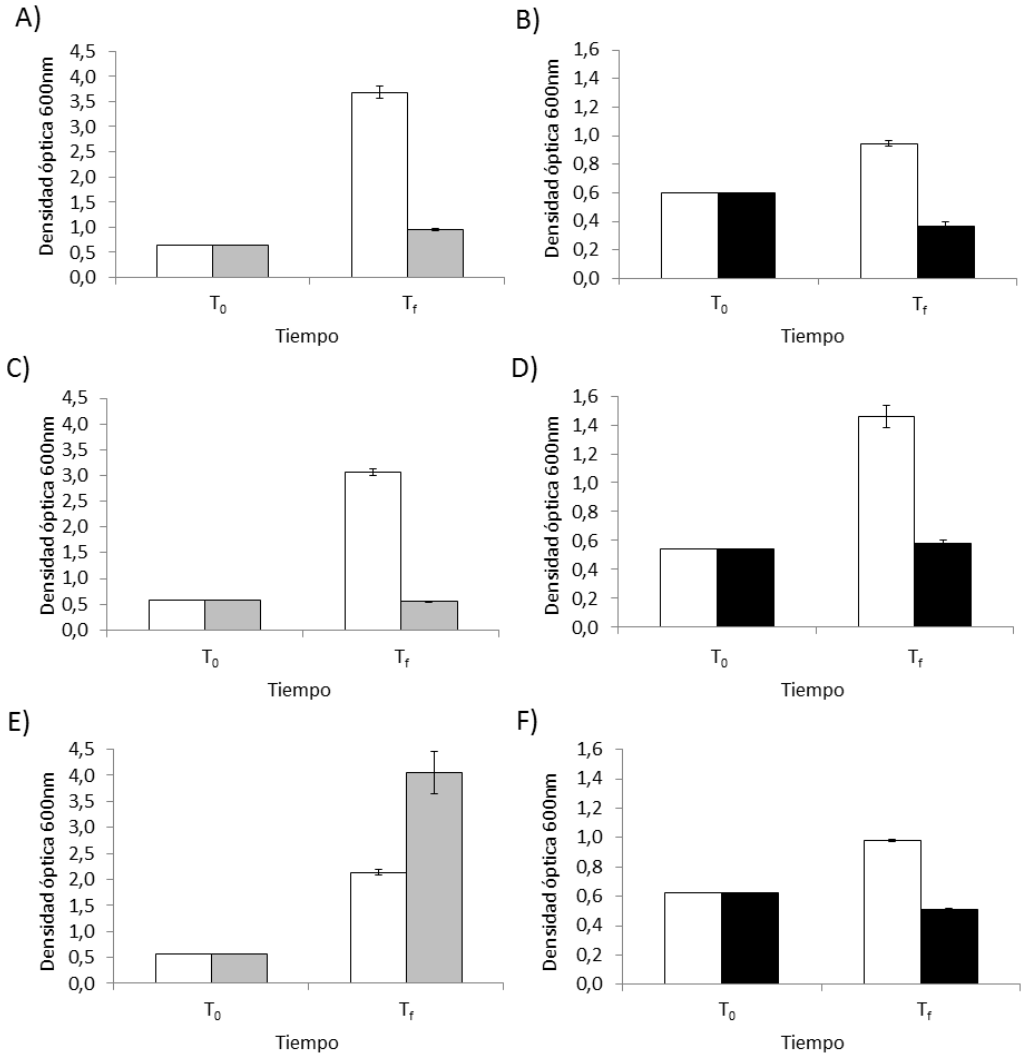

Figura 2.1. Densidad óptica de los cultivos en los tratamientos con hidrocarburos. En color blanco se muestran los cultivos control, en gris los cultivos tratados con naftaleno y en negro los tratados con tolueno. Abreviaturas: T_0, inicio del tratamiento; Tf, final del tratamiento. Paneles A, C y E, exposición a naftaleno 0,1 % (p/v) durante 24 h. Paneles B, D y F, exposición a tolueno 0,3 % (v/v) durante 3 h. Paneles A-B, *R. pomeroyi* DSS-3^T; paneles C-D, *Sulfitobacter* sp. 3SOLIMAR09; paneles E-F, *C. aestuarii* 357. Se representa la media y desviaciones estándar de tres réplicas.

La exposición a hidrocarburos para los análisis por proteómica masiva se realizó como se detalla en el apartado 2.3 de material y métodos. Se realizó un seguimiento de la densidad óptica de los cultivos antes y después de la exposición a los hidrocarburos (figura 2.1). En el caso de la exposición a naftaleno en *R. pomeroyi* DSS-3^T y *Sulfitobacter* sp. 3SOLIMAR09 se observó una menor densidad óptica en los cultivos tratados que en sus respectivos controles (figuras 2.1A y 2.1C). La densidad óptica de los cultivos de estas dos cepas se mantuvo estable tras la exposición a naftaleno, lo que sugiere que en presencia de este compuesto se redujo el crecimiento de estas bacterias. En cambio, los cultivos de *C. aestuarii* 357 expuestos durante 24 h a naftaleno alcanzaron una densidad óptica mayor que sus respectivos controles (figura 2.1E). Estos resultados sugieren que la cepa 357, que es degradadora de naftaleno, habría podido usar este compuesto como fuente de carbono y energía adicional y, por ello, habría podido alcanzar densidades ópticas mayores que en condiciones control, en las que solo están disponibles los nutrientes del caldo marino. En los ensayos de viabilidad realizados se comprobó que los tres microorganismos estudiados seguían siendo viables tras 24 h de exposición a naftaleno 0,1 % (p/v). Por lo que respecta a la exposición a tolueno (figuras 2.1B, 2.1D y 2.1F) en los tres microorganismos se observó una disminución de la densidad óptica respecto a los cultivos control tras 3 h de exposición a tolueno 0,3 % (v/v). Además, no se pudo detectar crecimiento en ninguna de las placas de agar marino sembradas tras 3 h de exposición a tolueno como control de viabilidad. Estos resultados parecían evidenciar una pérdida de viabilidad de las células en todos los casos tras la exposición a tolueno.

El análisis de proteómica masiva se realizó a nivel de proteína total y de una fracción enriquecida en proteínas de membrana para los tratamientos con naftaleno (ver detalle experimental en el apartado 3.2 de material y métodos). En el caso de la exposición a tolueno, debido a la limitación de disponibilidad de tiempo de análisis proteómico, se realizó un análisis menos exhaustivo utilizando solo la fracción enriquecida en proteínas de membrana (proteoma fraccionado), ya que podían verse diferencias entre los perfiles proteicos de los cultivos control y tratados en los geles de proteínas realizados (ver ejemplo de 3SOLIMAR09 en figura S2.1).

Como se puede observar en la tabla 2.1 en los estudios de proteoma fraccionado se detectaron entre un 23 y un 33 % de las proteínas anotadas. En cuanto al proteoma total, la detección fue de entre un 30 y un 39 % de las proteínas anotadas en los genomas. Respecto a las proteínas que resultaron significativas en la comparación entre las condiciones control y los tratamientos con los criterios establecidos en este estudio (ver apartado 4 de material y métodos), en los análisis de proteoma fraccionado se detectaron 390,5 ± 90,5 proteínas de media (entre un 6 y un 11 % de las proteínas anotadas en los genomas) y en los proteomas totales 454 ± 181,9 proteínas de media (entre un 7 y un 15 % de las proteínas anotadas). Todas las proteínas que presentaban diferencias significativas en el análisis se muestran en las tablas suplementarias S2.2 a S2.10. Cabe recordar que los criterios utilizados en esta tesis para considerar una proteína como significativa fueron: detección mínima de dos péptidos no redundantes, ANOVA $p<0,05$ y variación mínima 2X en la abundancia relativa normalizada al comparar condición control y tratamientos. La asignación funcional se hizo con el servidor BlastKOALA de KEGG (Kanehisa *et al.*, 2016b).

Tabla 2.1. Proteínas detectadas en los experimentos con naftaleno y tolueno en los tres microorganismos estudiados.

Muestra	Nº proteínas anotadas	Nº proteínas detectadas[a]	Nº proteínas significativas[a]	Nº proteínas significativas con asignación funcional[b]
Control–Naftaleno (Proteoma total)				
R. pomeroyi DSS-3^T	4279	1305 (30,50)	664 (15,52)	308 (46,38)
Sulfitobacter sp. 3SOLIMAR09	3295	1236 (37,51)	348 (10,56)	141 (40,52)
C. aestuarii 357	4573	1805 (39,47)	350 (7,65)	151 (43,14)
Control–Naftaleno (Proteoma fraccionado)				
R. pomeroyi DSS-3^T	4279	1025 (23,95)	430 (10,05)	204 (47,44)
Sulfitobacter sp. 3SOLIMAR09	3295	908 (27,56)	343 (10,41)	144 (41,98)
C. aestuarii 357	4573	1516 (33,15)	273 (5,97)	117 (42,85)
Control–Tolueno (Proteoma fraccionado)				
R. pomeroyi DSS-3^T	4279	945 (22,08)	462 (10,80)	217 (46,97)
Sulfitobacter sp. 3SOLIMAR09	3295	777 (23,58)	326 (9,89)	144 (44,17)
C. aestuarii 357	4573	1353 (29,59)	509 (11,13)	233 (45,78)

a: Porcentajes respecto a proteínas anotadas en el genoma
b: Porcentajes respecto a proteínas significativas en el análisis

Comparando el porcentaje de proteínas detectadas en los proteomas totales de este estudio con los obtenidos en estudios de proteoma total realizados con anterioridad por miembros de nuestro de grupo de investigación en la Universidad de Warwick se pudo observar que estos eran similares. En el estudio de la doctora Brunet-Galmés (Brunet-Galmés, 2016) se detectaron entre un 29,8 y un 37,3 % de las proteínas anotadas de 5 cepas de *Pseudomonas stutzeri* tras exposición a naftaleno y salicilato. En otro estudio con siete aislados del linaje Roseobacter expuestos a agua de mar suplementada con diésel al 0,1 % (v/v) se detectaron entre un 27,7 y un 39,5 de las proteínas anotadas de cada microorganismo, y en el caso concreto de la cepa 3SOLIMAR09 el porcentaje fue del 38,5 % (Mas-Lladó, 2017), similar al obtenido en este trabajo.

1.1. Cambios proteicos globales asociados a la respuesta a naftaleno y tolueno

Para estudiar de forma global los cambios asociados a los tratamientos aplicados en cada uno de los microorganismos estudiados se realizó un análisis de componentes principales (PCA) para cada bacteria por separado. Para ello se utilizó una matriz con la abundancia relativa normalizada de las proteínas significativas detectadas en cada condición y microorganismo (ver sección 4 de material y métodos). En todos los experimentos y con todos los microorganismos un solo componente principal (PC1) explicaba más del 90 % de la varianza observada en los proteomas, separándose los proteomas de las condiciones control de aquellos correspondientes a los tratamientos (datos no mostrados). Por lo tanto, se producían cambios en los proteomas de las cepas estudiadas tras la adición de los dos tipos de hidrocarburos aromáticos ensayados.

Para poder comparar conjuntamente la respuesta de los tres organismos en cada una de las condiciones estudiadas se definieron grupos de proteínas homólogas (apartado 4 de material y métodos). Para ello se utilizaron las proteínas que el análisis mostraba como significativas con

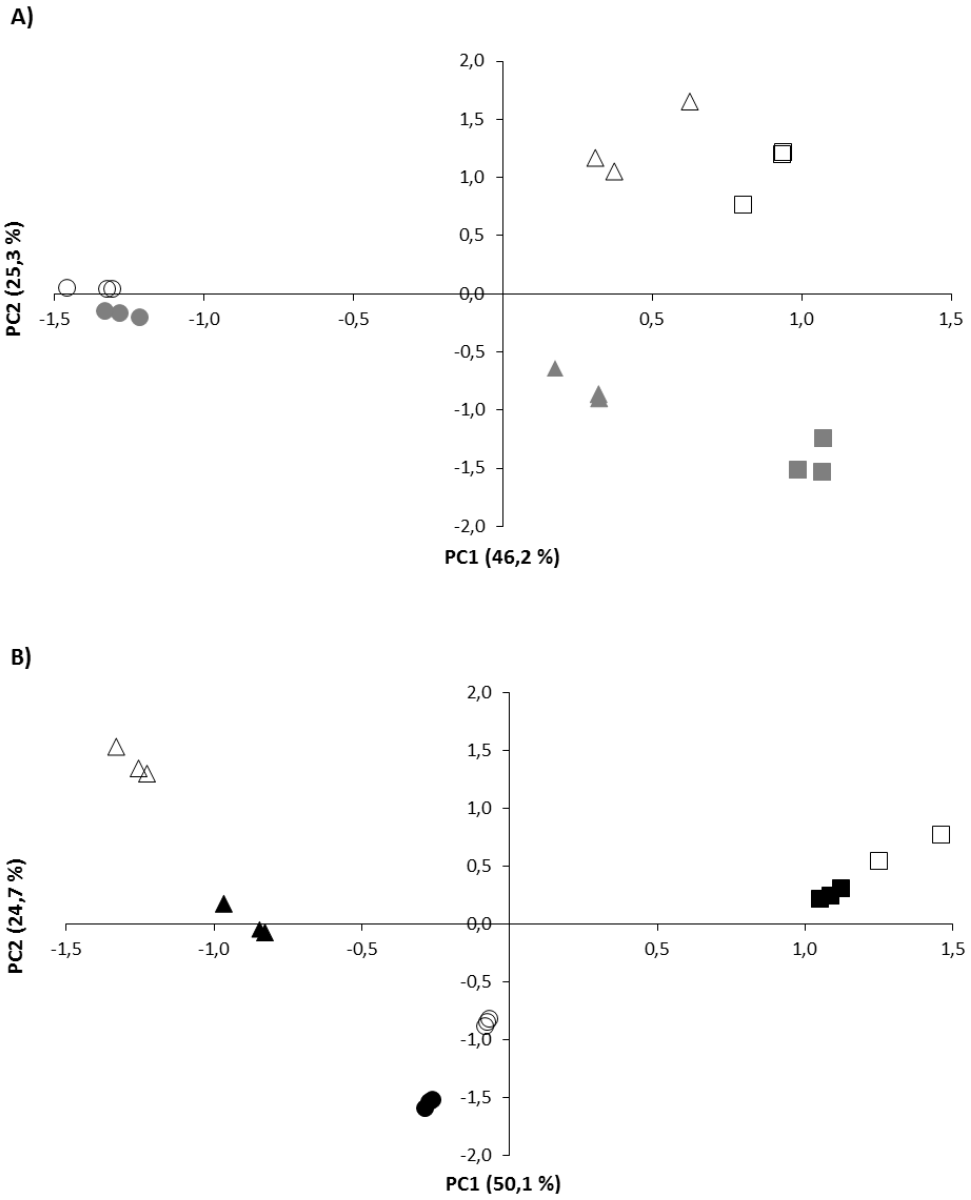

Figura 2.2. Comparación mediante análisis de componentes principales de los proteomas analizados en este estudio. Panel A, respuesta a naftaleno a nivel de proteoma total. Panel B, respuesta a tolueno en el proteoma fraccionado. *R. pomeroyi* DSS-3^T (triángulos), *Sulfitobacter* sp. 3SOLIMAR09 (cuadrados) y *C. aestuarii* 357 (círculos). Los símbolos de color blanco se corresponden a los controles, los de color gris a los tratados con naftaleno y los de color negro a los tratados con tolueno.

los criterios utilizados en este estudio. El análisis por PCA basado en los agrupamientos de proteínas homólogas (figura 2.2) mostró que los triplicados de los proteomas, tanto del control como de los tratados con hidrocarburos, para cada microorganismo aparecían agrupados, aunque a su vez separados entre ellos. Es decir, en todos los casos los proteomas de los cultivos control y tratados de cada microorganismo aparecían diferenciados, pero a su vez, eran más más próximos entre ellos que con los controles o los tratados de los otros microorganismos. Este es un resultado esperable porque los tres microorganismos estudiados son genómicamente, y por tanto proteómicamente, distintos. Por lo que respecta a la respuesta al tratamiento con naftaleno a nivel de proteoma total, dos componentes principales (PC1 y PC2) explicaron el 71,5 % de la varianza de las muestras (figura 2.2A). Además, se observó que en el caso del tratamiento con naftaleno el microorganismo en el que

había menor diferencia entre los proteomas de los cultivos control y tratados era *C. aestuarii* 357 (figura 2.2A). Este resultado concordó con lo esperado puesto que *C. aestuarii* 357 era el único de los tres organismos estudiados que era capaz de degradar el naftaleno y por lo tanto era esperable un menor efecto tóxico del naftaleno para este microorganismo. En el caso del tratamiento con tolueno, los dos componentes principales explicaron el 74,8 % de la varianza de las muestras y en este caso no se observaron diferencias en la respuesta de *C. aestuarii* 357 en comparación con los otros dos microorganismos estudiados (figura 2.2B). En cuanto al tipo de proteínas que tuvieron un mayor peso en el análisis de componentes principales para separar los proteomas de los tres microorganismos en el caso del estudio de tolerancia a naftaleno se observaron: transportadores de tipo ABC, TRAP y RND, factores sigma, ATPasas, chaperonas, proteínas de función desconocida como la proteína SPO2567 de *R. pomeroyii* DSS-3^T (Christie-Oleza *et al*., 2012), proteínas implicadas en el metabolismo de compuestos de reserva (PHA) y de hidrocarburos (ver información más desarrollada en el apartado 2 de este capítulo). En el caso de los tratamientos con tolueno se observaron principalmente proteínas implicadas en el metabolismo general, algunas proteínas de membrana y transportadores de tipo ABC y TRAP (tabla S2.11).

Para analizar la respuesta a la adición de naftaleno y tolueno en los 3 microorganismos estudiados se clasificaron las proteínas significativas de cada experimento en diferentes funciones biológicas utilizando el servidor de anotación BlastKOALA de KEGG (Kanehisa *et al*., 2016b). Como se puede observar en la tabla 2.1 solo se pudo asignar función a un 40–47 % de las proteínas detectadas que habían pasado los criterios de cribado para ser consideradas significativas. La segunda limitación de este análisis fue que una parte importante de las proteínas con función asignada quedó clasificada en categorías generales lo cual no permitía asignarles función sin hacer un análisis detallado proteína a proteína (ver más adelante). Entre estas estaban por ejemplo las proteínas descritas en *C. aestuarii* 357 para la degradación de naftaleno (Suárez-Suárez, 2013). El detalle de la clasificación de las proteínas estudiadas en las categorías biológicas de KEGG se encuentra en las tablas suplementarias S2.2 a S2.10. En la tabla 2.2 se muestra la abundancia relativa normalizada de las proteínas asignadas a funciones celulares en los diferentes microorganismos en el experimento de exposición a naftaleno. Se observaron diferencias significativas en la abundancia relativa de proteínas asignadas a varias categorías funcionales tras la exposición a naftaleno. Como se puede observar en la tabla 2.2, la respuesta a nivel de proteoma de *R. pomeroyi* DSS-3^T y *Sulfitobacter* sp. 3SOLIMAR09 tras exposición a naftaleno era más similar entre ellas que a la respuesta de *C. aestuarii* 357. Este resultado concuerda el análisis de PCA (figura 2.2) y podría estar relacionado con el hecho de que *C. aestuarii* 357 fuera el único de los tres microorganismos estudiados que era capaz de degradar naftaleno.

En primer lugar se observaron cambios a nivel de proteínas de transporte de membrana, como por ejemplo una abundancia relativa de transportadores de tipo ABC era significativamente inferior tras la exposición a naftaleno en *R. pomeroyi* DSS-3^T y *Sulfitobacter* sp. 3SOLIMAR09 pero no en *C. aestuarii* 357 (tabla 2.2). También se observó una mayor abundancia relativa de proteínas implicadas en exportación de proteínas desde el citoplasma al exterior celular y/o periplasma en *R. pomeroyi* DSS-3^T y *Sulfitobacter* sp.3SOLIMAR09 tras exposición a naftaleno (tabla 2.2). Por otra parte, tras la exposición a naftaleno se observó de forma significativa una

Tabla 2.2. Clasificación en categorías funcionales de las proteínas significativas con función asignada (KEGG) del proteoma total tras exposición a naftaleno. Se muestran los promedios y desviación estándar de las abundancias relativas normalizadas (n=3). Se marcan como negativas todas aquellas categorías funcionales a las que KEGG no asignó proteínas.

Categoría funcional KEGG	*R. pomeroyi* DSS-3^T		*Sulfitobacter* sp. 3SOLIMAR09		*C. aestuarii* 357	
	Control	Naftaleno	Control	Naftaleno	Control	Naftaleno
Metabolismo de los aminoácidos						
Metabolismo de glicina, serina y treonina	0,019 ±0,005[a]	0,005±0,001[a]	–	–	–	–
Metabolismo de la fenilalanina	0,146 ± 0,018[a]	0,049 ± 0,013[a]	0,012 ± 0,004	0,003 ± 0,000	–	–
Metabolismo de otros aminoácidos						
Metabolismo del glutatión	0,168 ± 0,017[a]	0,886 ± 0,051[a]	0,126 ± 0,025	0,198 ± 0,017	1,145 ± 0,154[a]	2,071 ± 0,412[a]
Metabolismo de los carbohidratos						
Metabolismo del butanoato	0,430 ± 0,023[a]	0,123 ± 0,012[a]	0,034 ± 0,007[a]	0,007 ± 0,01[a]	0,052 ± 0,012	0,078 ± 0,019
Metabolismo del glioxilato y dicarboxilato	–	–	–	–	0,680 ± 0,112[a]	0,096 ± 0,009[a]
Metabolismo del piruvato	0,014 ± 0,002[a]	0,005 ± 0,001[a]	–	–	0,059 ± 0,005[a]	0,013 ± 0,003[a]
Metabolismo energético						
Metabolismo del nitrógeno	–	–	1,787 ± 0,904	0,324 ± 0,108	0,007 ± 0,001[a]	0,000 ± 0,000[a]
Fosforilación oxidativa	0,330 ± 0,039[a]	0,104 ± 0,010[a]	0,198 ± 0,009[a]	0,377 ± 0,047[a]	0,280 ± 0,008[a]	0,061 ± 0,004[a]
Metabolismo del azufre	0,037 ± 0,005[a]	0,014 ± 0,001[a]	–	–	–	–
Metabolismo de cofactores y vitaminas						
Metabolismo del folato	0,013 ± 0,004[a]	0,071 ± 0,007[a]	–	–	–	–
Metabolismo de glicanos						
Biosíntesis de peptidoglicano	–	–	–	–	0,015 ± 0,001	0,016 ± 0,003
Metabolismo de lípidos						
Metabolismo de glicerofosfolípidos	0,019 ± 0,005[a]	0,005 ± 0,001[a]	1,599 ± 0,526[a]	0,157 ± 0,022[a]	–	–
Transporte de membrana						
Transportadores ABC	20,480 ± 4,796[a]	2,816 ± 0,163[a]	28,952 ± 3,612[a]	9,202 ± 1,230[a]	2,017 ± 0,035	2,917 ± 0,817
Movilidad celular						
Ensamblaje del flagelo	–	–	–	–	0,055 ± 0,015[a]	0,005 ± 0,001[a]
Comunicación celular						
Quorum sensing	5,437 ± 0,221[a]	1,917 ± 0,045[a]	–	–	0,219 ± 0,018[a]	0,043 ± 0,012[a]
Transducción de señales						
Sistema de dos componentes	2,672 ± 0,360	1,593 ± 0,201	–	–	2,814 ± 1,126	0,545 ± 0,154

a: Diferencias significativas ($p < 0,05$) en abundancia relativa normalizada entre control y tratamiento (t-Student).

Tabla 2.2. Clasificación en categorías funcionales de las proteínas significativas con función asignada (KEGG) del proteoma total tras exposición a naftaleno. Se muestran los promedios y desviación estándar de las abundancias relativas normalizadas (n=3). Se marcan como negativas todas aquellas categorías funcionales a las que KEGG no asignó proteínas (continuación).

Categoría funcional KEGG	*R. pomeroyi* DSS-3Ta		*Sulfitobacter* sp. 3SOLIMAR09		*C. aestuarii* 357	
	Control	Naftaleno	Control	Naftaleno	Control	Naftaleno
Crecimiento celular y muerte						
Ciclo celular	0,828 ± 0,054[a]	0,661 ± 0,021[a]	–	–	0,179 ± 0,015	0,205 ± 0,059
Metabolismo de nucleótidos						
Metabolismo de purinas	0,034 ± 0,006[a]	0,122 ± 0,015[a]	–	–	–	–
Metabolismo de pirimidinas	0,187 ± 0,010[a]	1,452 ± 0,102[a]	0,456 ± 0,112[a]	2,154 ± 0,361[a]	–	–
Reparación y replicación						
Reparación por escisión de bases	0,061 ± 0,025	0,019 ± 0,002	0,006 ± 0,001[a]	0,002 ± 0,000[a]	0,035 ± 0,002[a]	0,006 ± 0,001[a]
Reparación de la excisión de nucleótidos	0,002 ± 0,001	0,004 ± 0,002	0,077 ± 0,010[a]	0,299 ± 0,042[a]	–	–
Replicación del ADN	0,022 ± 0,003[a]	0,009 ± 0,001[a]	0,006 ± 0,002	0,002 ± 0,000	0,080 ± 0,011[a]	0,178 ± 0,015[a]
Recombinación homóloga	0,003 ± 0,003	0,001 ± 0,000	0,005 ± 0,000[a]	0,025 ± 0,004[a]	0,001 ± 0,000[a]	0,000 ± 0,000[a]
Traducción						
Biosíntesis de aminoacil ARNt	0,297 ± 0,019[a]	0,470 ± 0,031[a]	–	–	–	–
Ribosoma	15,395 ± 1,099[a]	5,422 ± 0,281[a]	1,740 ± 0,205[a]	3,885 ± 0,339[a]	–	–
Plegamiento, exportación y degradación						
Exportación de proteínas	0,254 ± 0,023[a]	0,693 ± 0,019[a]	0,207 ± 0,011[a]	0,603 ± 0,031[a]	–	–
Degradación de ARN	3,616 ± 0,957[a]	12,748 ± 1,752[a]	9,307 ± 0,764[a]	24,315 ± 0,798[a]	0,068 ± 0,015[a]	0,014 ± 0,002[a]

a: Diferencias significativas ($p < 0,05$) en abundancia relativa normalizada entre control y tratamiento (t-Student).

mayor abundancia relativa de proteínas implicadas en la degradación de ARN en *R. pomeroyi* DSS-3^T y *Sulfitobacter* sp. 3SOLIMAR09 (tabla 2.2). Además, en *R. pomeroyi* DSS-3^T también se observó de forma significativa una menor abundancia relativa de proteínas ribosómicas, de proteínas del ciclo celular y de la replicación del ADN (tabla 2.2) tras la exposición a naftaleno. Este resultado sugirió que esta bacteria podía estar ajustando su metabolismo a unas condiciones de no crecimiento al encontrarse en una situación desfavorable por la exposición a naftaleno. En cambio, en el proteoma de *C. aestuarii* 357 se observó de forma significativa una mayor abundancia relativa de proteínas relacionadas con la replicación de ADN tras exposición a naftaleno, así como una mayor abundancia relativa de proteínas del ciclo celular, aunque en este caso no era un cambio significativo. Estos resultados concuerdan con los valores de densidad óptica de los cultivos (figura 2.1), que indicaban el crecimiento de *C. aestuarii* 357 en el tratamiento con naftaleno en comparación con el control. Se observaron también cambios en la abundancia relativa de proteínas implicadas en fosforilación oxidativa (tabla 2.2). En concretó se observó, de forma significativa, una menor abundancia relativa de proteínas implicadas en la fosforilación oxidativa en *R. pomeroyi* DSS-3^T y *C. aestuarii* 357 y una mayor abundancia de estas en *Sulfitobacter* sp. 3SOLIMAR09 tras la exposición a naftaleno (tabla 2.2). En *Pseudomonas putida* DOT-T1E se ha descrito un aumento de las proteínas implicadas en la fosforilación oxidativa tras la exposición a hidrocarburos para obtener energía para el funcionamiento de las bombas de expulsión de hidrocarburos (Udaondo *et al*., 2012). Por lo tanto, el resultado en *Sulfitobacter* sp. 3SOLIMAR09 se correspondería con lo esperado según lo descrito en la literatura. Otro cambio significativo observado fue una mayor abundancia relativa de proteínas implicadas en el metabolismo del glutatión en *R. pomeroyi* DSS-3^T y *C. aestuarii* 357 tras la exposición a naftaleno (tabla 2.2). El glutatión forma parte de la respuesta a estrés oxidativo de las células. Se ha descrito el aumento de las proteínas implicadas en el metabolismo del glutatión tras la exposición a hidrocarburos debido al incremento en la producción de peróxido de hidrógeno y especies reactivas del oxígeno que se produce por la interferencia de los hidrocarburos en la cadena de transporte electrónico (Udaondo *et al*., 2012; Ramos *et al*., 2015). Por lo tanto, los resultados sugieren que el naftaleno afectaría a los tres microorganismos estudiados, en relación al estrés oxidativo, aunque de forma diferente. En *Sulfitobacter* sp. 3SOLIMAR09 no se observó un cambio significativo a nivel de la abundancia relativa de las proteínas de respuesta a estrés oxidativo. En cambio, en el caso de *R. pomeroyi* DSS-3^T y *C. aestuarii* 357 se observó, de forma significativa, una mayor abundancia relativa de proteínas del metabolismo del glutatión, indicador de que puede ocurrir daño oxidativo. Finalmente, también se observó una menor abundancia relativa de las proteínas implicadas en *quorum sensing* tras la exposición al naftaleno en *R. pomeroyi* DSS-3^T y *C. aestuarii* 357 (tabla 2.2). Por lo tanto, en estas dos bacterias la exposición a naftaleno parecía afectar a la comunicación celular.

Respecto a los resultados obtenidos tras la exposición a tolueno (tabla S2.12), se observaron también diferencias en la abundancia relativa de proteínas clasificadas en varias categorías funcionales en cada uno de los microorganismos estudiados. En el análisis de estos resultados debe tenerse en cuenta que el estudio se realizó con el proteoma fraccionado y que, según los resultados de los controles de viabilidad y la densidad óptica de los cultivos tras el tratamiento (figura 2.1.), se habría producido lisis celular tras el tratamiento con tolueno. Centrando el análisis en proteínas con una probable localización en la membrana celular puede verse que

algunas de las diferencias observadas en los tratamientos con tolueno se correspondieron con los resultados obtenidos en naftaleno (tablas 2.2 y S2.12). Por ejemplo, tras exposición a tolueno se observó también, de forma significativa, una menor abundancia relativa de transportadores ABC, en este caso incluso en la cepa *C. aestuarii* 357, y de proteínas implicadas en *quorum sensing* en *R. pomeroyi* DSS-3^T y *C. aestuarii* 357 (tabla S2.12). En cuanto a otros sistemas de transporte de membrana tras la exposición a tolueno, también se observó una menor abundancia relativa de proteínas de sistemas de secreción bacteriano en los tres microorganismos, aunque la diferencia no era significativa en el caso de *Sulfitobacter* sp. 3SOLIMAR09, y en sistemas de transducción de señales de dos componentes en *C. aestuarii* 357 (tabla S2.12).

2. Análisis detallado de las proteínas observadas en la respuesta a naftaleno

Una vez determinados los tipos de categorías funcionales en las que se detectaron cambios en abundancias proteicas relativas tras la exposición a naftaleno se estudiaron en detalle las proteínas individuales que mostraban diferencias de abundancia significativas según los criterios establecidos en este estudio para cada uno de los microorganismos estudiados (ver detalles de las proteínas en las tablas S2.3-S2.8). La mayoría de proteínas detectadas que resultaban significativas aparecían tanto en el proteoma total como en el proteoma fraccionado. Por lo tanto el análisis se basó en el proteoma total y se incluyeron además las proteínas que se detectaban exclusivamente en el proteoma fraccionado. Para poder entender la función de cada una de las proteínas detectadas que resultaban significativas se optimizó de forma manual la clasificación funcional de KEGG para poder asignar una función a todas aquellas proteínas que no habían podido ser asignadas funcionalmente por KEGG o bien habían sido asignadas a categorías generales.

2.1. Proteínas de membrana

En la literatura se ha descrito que las bombas de expulsión de la membrana son el principal mecanismo para la tolerancia a compuestos aromáticos en bacterias Gram negativas (Ramos *et al*., 2015). Además, como se ha mencionado anteriormente se habían observado cambios significativos a nivel de transportadores ABC y sistemas de exportación de proteínas en algunos de los microorganismos estudiados (tabla 2.2). Por ello, en primer lugar, se analizaron en detalle los cambios que se producían a nivel de transporte de membrana tras la exposición a naftaleno.

En la tabla 2.3 se puede observar la clasificación de las proteínas de membrana detectadas en el proteoma total tras la optimización manual de la anotación de KEGG, y de sus correspondientes abundancias relativas. Estas abundancias no se corresponden exactamente con las que se muestran en la tabla 2.2 porque algunas proteínas de transporte se detectaron tras la revisión manual al detalle de los resultados. Se analizaron proteínas de los siguientes sistemas de transporte: porinas, transportadores de tipo TRAP, RND y MFS, y proteínas de la familia RDD (Pfam PF06271). También se incluyeron todas aquellas proteínas de membrana detectadas de función desconocida (tabla 2.3), ya que en esta categoría se incluía la proteína SPO2567 de *Ruegeria pomeroyi* DSS-3^T y sus correspondientes homólogos. Como se ha mencionado anteriormente el análisis por PCA evidenció la importancia de esta proteína en la

Tabla 2.3. Clasificación de las proteínas de membrana del proteoma total tras exposición a naftaleno. Se muestran los promedios y desviación estándar de las abundancias relativas normalizadas (n=3) tras la clasificación manual en categorías funcionales. Se marcan como negativas todas aquellas categorías funcionales a las que ni KEGG ni la anotación manual asignaron proteínas.

Categoría funcional	***R. pomeroyi* DSS-3^T**		***Sulfitobacter* sp. 3SOLIMAR09**		***C. aestuarii* 357**	
	Control	Naftaleno	Control	Naftaleno	Control	Naftaleno
Porinas	0,116 ± 0,025[a]	0,032 ± 0,003[a]	1,264 ± 0,135[a]	0,599 ± 0,137[a]	1,367 ± 0,154[a]	0,154 ± 0,022[a]
Transportadores ABC	27,040 ± 4,697[a]	4,952 ± 0,209[a]	38,115 ± 3,544[a]	11,993 ± 1,190[a]	2,16 ± 0,033	2,954 ± 0,811
Transportadores TRAP	3,394 ± 0,687[a]	0,600 ± 0,100[a]	0,011 ± 0,007	0,002 ± 0,000	1,387 ± 0,166[a]	0,663 ± 0,049[a]
Proteínas familia RDD (Pfam PF06271)	0,010 ± 0,006[a]	0,471 ± 0,032[a]	0,004 ± 0,003[a]	0,041 ± 0,006[a]	PF[b]	PF[b]
Transportadores RND	0,091 ± 0,004[a]	0,921 ± 0,056[a]	PF[b]	PF[b]	PF[b]	PF[b]
Transportadores MFS	–	–	0,218 ± 0,040[a]	0,034 ± 0,003[a]	–	–
Proteínas para la exportación de proteínas	0,396 ± 0,034[a]	0,992 ± 0,034[a]	0,207 ± 0,011[a]	0,603 ± 0,031[a]	PF[b]	PF[b]
Proteínas de membrana de función desconocida	0,553 ± 0,073[a]	8,299 ± 1,200[a]	1,084 ± 0,108[a]	13,574 ± 1,965[a]	0,320 ± 0,034[a]	3,331 ± 0,660[a]

a: Diferencias significativas ($p < 0,05$) en abundancia relativa normalizada entre control y tratamiento (t-Student).

b: PF significa que no se detectaron proteínas pertenecientes a estas categorías funcionales en el estudio del proteoma total pero si se detectaron en el estudio del proteoma fraccionado tras la exposición a naftaleno.

diferenciación de los proteomas de cultivos tratados con naftaleno (ver apartado 1.1). Además la proteína SPO2567 se había observado también previamente en proteomas de cultivos de esta bacteria en presencia de hidrocarburos o altas temperaturas (Christie-Oleza *et al*., 2012).

Como puede verse en la tabla 2.3 se observaron cambios a nivel de proteínas pertenecientes a distintas familias de transportadores de membrana y de porinas. En concreto, se observó de forma significativa una menor abundancia relativa de porinas en los tres microorganismos tras la exposición a naftaleno. Por otra parte, además de la significativa menor abundancia relativa de proteínas identificadas como transportadores de membrana de tipo ABC en *R. pomeroyi* DSS-3^T y *Sulfitobacter* sp. 3SOLIMAR09, se observó una menor abundancia relativa de transportadores de tipo TRAP en *R. pomeroyi* DSS-3^T y *C. aestuarii* 357, y de transportadores de tipo MFS en *Sulfitobacter* sp. 3SOLIMAR09 tras la exposición a naftaleno (tabla 2.3). Las porinas son proteínas presentes en la membrana externa de bacterias Gram negativas, que actúan como canales de transporte de diferentes moléculas hidrofílicas pequeñas y en general presentan una baja especificidad de sustrato, aunque hay diversos tipos de porinas con diversos grados de selectividad y complejidad (Hemamalini y Khare, 2014). Los transportadores de tipo ABC se encargan del transporte a través de membrana de compuestos como lípidos, esteroides, péptidos, fármacos e hidrocarburos aromáticos (Ramos *et al*., 2015). Los transportadores de tipo TRAP se encargan del transporte a nivel de membrana de compuestos como ácidos dicarboxílicos C4 o alfa cetoácidos (Mulligan *et al*., 2011). Finalmente, los transportadores de la superfamilia MFS facilitan el movimiento de pequeños solutos, como cationes monovalentes y divalentes, compuestos lipofílicos y antimicrobianos a través de las membranas celulares en respuesta a gradientes quimiosmóticos (Pao *et al*., 1998; Ramos *et al*., 2015). Por lo tanto la exposición a naftaleno tendría un efecto negativo sobre la difusión pasiva (porinas) y sobre distintos tipos de sistemas de transporte. Este resultado podría ser interpretado como una respuesta de la célula para modular su permeabilidad a los solutos del medio a partir del cambio en sus sistemas de transporte habituales. De esta forma, podría estar evitando que entraran compuestos tóxicos para ella al disminuir la presencia de sistemas de transporte menos selectivos, como son las porinas, e incrementando los sistemas de resistencia como los transportadores de tipo RND. Este resultado es coherente con el obtenido por Roma-Rodrigues y colaboradores (2010) en *Pseudomonas putida* KT2440 cuando se exponía a fenol. En cuanto al resultado de los transportadores ABC, su menor abundancia tras la exposición a naftaleno era una respuesta contraria a la esperada según los resultados de Ramos y colaboradores (2015), ya que habían sido descritos como transportadores implicados en la tolerancia a solventes orgánicos. Ahora bien, podría darse el caso de que los tres microorganismos de este estudio no presentaran bombas de expulsión de tipo ABC, a diferencia de *Pseudomonas putida,* y que los transportadores ABC detectados estuvieran implicados en la captación de nutrientes. Por lo tanto, la menor abundancia de transportadores ABC tras la exposición a naftaleno podría ser explicada como una estrategia para evitar la entrada del naftaleno mediante estos transportadores.

Para otros tipos de transportadores de membrana se observó una mayor abundancia relativa tras el tratamiento con naftaleno (tabla 2.3). Así, se observó de forma significativa una mayor abundancia de proteínas de membrana implicadas en la exportación de proteínas en el proteoma total de *R. pomeroyi* DSS-3^T y *Sulfitobacter* sp. 3SOLIMAR09 (tabla 2.3). En el caso de *C. aestuarii* 357 no se observó en el proteoma total pero sí en el estudio del proteoma

fraccionado, aunque en este caso esta mayor abundancia relativa no fue significativa (tabla 2.3). Las proteínas de membrana para la exportación de proteínas se encargan del transporte activo de proteínas desde el citoplasma hasta el periplasma y/o exterior celular (Kanehisa *et al*., 2016a). Entre las proteínas que forman parte de esta categoría se encuentran las proteínas de la vía Sec dependiente y de los sistemas Tat (*twin arginine translocation*). También se observó una mayor abundancia relativa significativa de proteínas de la familia RDD tras la exposición a naftaleno en *R. pomeroyi* DSS-3^T y *Sulfitobacter* sp. 3SOLIMAR09. En *C. aestuarii* 357 no se pudo detectar este tipo de proteínas en el estudio del proteoma total, pero sí se pudo detectar que una proteína de tipo RDD (EIE51875.1) era 6,2X más abundante tras la exposición a naftaleno respecto al control en el estudio del proteoma fraccionado (tabla S2.7). Las proteínas de la familia RDD se llaman así porque tienen 3 aminoácidos conservados: una arginina (R) y dos aspartatos (D) en una región con dos hélices transmembrana (Pfam PF06271). No se conoce la función de este tipo de proteínas, pero se ha hipotetizado que podrían estar involucradas en el transporte de un conjunto aún desconocido de ligandos. La mayor abundancia relativa de proteínas de la familia RDD detectadas se debió a la mayor detección de la proteína SPO2568 (AAV95816.1) de *R. pomeroyi* DSS-3^T (56,6X respecto al control, tabla S2.2) y de sus correspondientes homólogos en *Sulfitobacter* sp. 3SOLIMAR09 (MM82_00823, 9,6X respecto al control, tabla S2.3) y *C. aestuarii* 357 (EIE51875.1, 6,2X respecto al control en el proteoma fraccionado, tabla S2.7). La mayor abundancia relativa de la proteína SPO2568 de *R. pomeroyi* DSS-3^T tras la exposición a naftaleno ya había sido descrito por Christie-Oleza y colaboradores (2012). Estos autores observaron que la abundancia de esta proteína era hasta 8 veces mayor tras exposición a naftaleno 0,5 % (p/v) respecto al control. Por lo tanto, nuestros resultados respaldarían la hipótesis de que SPO2568 y sus correspondientes homólogos en *Sulfitobacter* sp.3SOLIMAR09 y *C. aestuarii* 357 también podrían participar en la respuesta a naftaleno.

Finalmente, también se observó una mayor abundancia relativa de transportadores de tipo RND (tabla 2.3) tras la exposición a naftaleno. En la literatura se ha descrito que estos transportadores son importantes para la tolerancia a solventes orgánicos (Ramos *et al*., 2015). Por ello se prestó un especial interés a este tipo de transportadores y se hizo un análisis a nivel genómico como complemento al análisis proteómico. Los resultados obtenidos se muestran a continuación.

Con el análisis genómico se pretendía determinar si los tres microorganismos estudiados en los experimentos de tolerancia a naftaleno y tolueno tenían potencialmente bombas de expulsión de tipo RND similares a las descritas en la literatura en relación a la tolerancia a solventes (Ramos *et al*., 2002). Mediante identidad de secuencia se analizó la presencia en estos genomas de genes para 24 proteínas, 18 de las cuales eran subunidades de seis bombas de tipo RND implicadas en tolerancia a solventes y las seis restantes eran proteínas implicadas en la regulación de estas bombas. El análisis se hizo también en el resto de genomas del linaje Roseobacter analizados en esta tesis (ver resultados en la tabla suplementaria S2.1). Los microorganismos modelo empleados para este análisis se muestran en la tabla M3 del apartado 1 de material y métodos. Con esta aproximación tan solo se pudieron detectar genes para los transportadores de membrana interna y para las proteínas de fusión a la membrana de las bombas RND, pero no para las proteínas de membrana externa ni posibles reguladores. En base a esto, se consideró un resultado positivo de predicción de la presencia de una bomba

de expulsión de tipo RND si se detectaba la proteína de membrana interna con el criterio estándar definido en esta tesis (al menos 50 % de identidad en secuencia aminoacídica, *E-value* < 10^{-5}) y la proteína de fusión con al menos un 30 % de identidad con la proteína modelo (*E-value* < 10^{-5}). Cuando ambas proteínas se detectaban con valores de identidad menores del 50 % y mayores o iguales a un 30 %, se consideró un resultado incierto. La detección de este tipo de transportadores en los genomas de Roseobacter analizados no fue homogénea en los 5 grupos filogenómicos del linaje (tabla S2.1). Con los criterios utilizados en este estudio estas bombas se detectaron en los genomas de todos los miembros del grupo filogenómico G3 (excepto en *Tropicibacter naphthalenivorans* DSM 19561^T) mientras que en el resto de grupos filogenómicos su detección era menor (tabla S2.1). Por lo tanto, en base a la mayor detección de proteínas de tipo RND similares a las de los modelos de degradación de hidrocarburos en los genomas del grupo G3, los microorganismos de este grupo tendrían potencialmente más mecanismos de tolerancia a hidrocarburos basados en bombas RND. Ahora bien, al no detectar las proteínas de membrana externa ni los posibles reguladores de las bombas de tipo RND estudiadas no se pudo afirmar con fiabilidad la presencia de genes para este tipo de bombas en los genomas ni que estas pudieran tener una expresión regulada. Entrando en detalle en los resultados de los tres microorganismos con los que se había realizado el estudio proteómico, en *R. pomeroyi* DSS-3^T se detectaron posibles bombas de tipo RND que presentaban al menos un 30 % de identidad con un *E-value* < 10^{-5} con las proteínas modelo AcrAB, de *Escherichia coli* str. K-12 substr. MG1655, TtgAB, TtgDE y TtgGH de *Pseudomonas putida* DOT-T1E y MexAB *Pseudomonas aeruginosa* PAO1 (tabla S2.1). En *Sulfitobacter* sp. 3SOLIMAR09 se detectó la presencia del gen para la proteína de membrana interna con más del 50 % de identidad con un *E-value* < 10^{-5} con la proteína modelo AcrB de *Escherichia coli* str. K-12 substr. MG1655 y la proteína de fusión con al menos un 30 % de identidad con un *E-value* < 10^{-5} con la proteína modelo AcrA de *Escherichia coli* str. K-12 substr. MG1655 (tabla S2.1). Finalmente, en *C. aestuarii* 357 se detectó la presencia del gen para la proteína de membrana interna con más del 50 % de identidad con un *E-value* < 10^{-5} con las proteínas modelo AcrB de *Escherichia coli* str. K-12 substr. MG1655, SrpB de *Pseudomonas putida* S12, TtgG de *Pseudomonas putida* DOT-T1E y MexB de *Pseudomonas aeruginosa* PAO1. Además, se detectó un gen para la proteína de fusión con al menos un 30 % de identidad con un *E-value* < 10^{-5} con las proteínas modelo AcrA de *Escherichia coli* str. K-12 substr. MG1655, SrpA de de *Pseudomonas putida* S12, TtgH de *Pseudomonas putida* DOT-T1E y MexA de *Pseudomonas aeruginosa* PAO1 (tabla S2.1).

A continuación se compararon los resultados de predicción génica de las bombas de tipo RND con los resultados del estudio proteómico. Se obtuvieron resultados dispares puesto que en los proteomas no siempre se pudieron detectar las mismas proteínas que se habían detectado por predicción génica. En *R. pomeroyi* DSS-3^T se pudieron detectar 5 subunidades de bombas de tipo RND en el estudio del proteoma total (AAV94225.1, AAV94226.1, AAV95593.1, AAV95594.1 y AAV94784.1, cuya abundancia relativa era, respectivamente, 21,2X, 39,3X, 2,1X, 5,7X y 12,6X mayor tras la exposición a naftaleno respecto al control, tabla S2.2). Estas proteínas no se habían detectado en el estudio de predicción génica porque todas ellas tenían un porcentaje de identidad de secuencia menor del 30 % con las proteínas de las bombas usadas como modelo. Cabe destacar que la presencia de la proteína AAV94784.1 (SPO1497) también fue 12 veces mayor tras la exposición a naftaleno en el experimento de Christie-Oleza

y colaboradores (2012). Sin embargo, en el experimento realizado en esta tesis no se vieron otras posibles bombas de expulsión detectadas en el estudio anterior (Christie-Oleza *et al*., 2012). En cuanto al proteoma fraccionado de *R. pomeroyi* DSS-3^T, se detectaron 3 subunidades de bombas de tipo RND (AAV94533.1, AAV94685.1 y AAV94812.1, tabla S2.5). La abundancia relativa de la tercera de ellas fue 12,7X superior tras la exposición a naftaleno que respecto al control mientras que AAV94533.1 y AAV94685.1 tenían abundancias relativas 2X y 3X, respectivamente, en el control respecto al tratamiento con naftaleno (tabla S2.5). Solo una de estas proteínas (AAV94685.1) presentaba un porcentaje de identidad de secuencia de al menos un 30 % de identidad con la proteína de fusión a la membrana de varias de las bombas RND modelo, y por lo tanto, era la única de las proteínas detectadas en el proteoma que había sido incluida en el estudio de la presencia de bombas de tipo RND por predicción genómica en *R. pomeroyi* DSS-3^T (tabla S2.1). En concreto AAV94685.1 presentaba una identidad de secuencia de al menos un 30 % y con un *E-value* < 10^{-5} con AcrA de *Escherichia coli* str. K-12 substr. MG1655, SrpA de *Pseudomonas putida* S12, TtgA, TtgD y TtgG de *Pseudomonas putida* DOT-T1E y MexA de *Pseudomonas aeruginosa* PAO1 (tabla S2.1). Además, junto al gen codificante para esta proteína se encontraba el gen codificante para la proteína AAV94684.1, que presentaba un porcentaje de identidad de secuencia mayor del 30 % de identidad y con un *E-value* < 10^{-5} con la subunidad transportadora de la membrana interna de varias de las bombas RND modelo, en concreto de AcrB de *Escherichia coli* str. K-12 substr. MG1655, TtgB, TtgE, TtgH de *Pseudomonas putida* DOT-T1E y MexB de *Pseudomonas aeruginosa* PAO1.

Respecto a los cultivos expuestos a naftaleno de *Sulfitobacter* sp. 3SOLIMAR09 y *C. aestuarii* 357, no se detectaron proteínas de este tipo en el proteoma total pero sí en el estudio del proteoma fraccionado. La única subunidad de una bomba de tipo RND detectada en el proteoma fraccionado de *Sulfitobacter* sp. 3SOLIMAR09 (MM82_02902, tabla S2.6), era 3 veces más abundante en el control respecto al tratamiento con naftaleno. Sin embargo, tenía menos de un 30 % de identidad de secuencia con las subunidades de las bombas RND utilizadas como modelo y, por lo tanto, no se correspondía con la bomba de tipo AcrAB-TolC detectada en la predicción génica. En el caso de *C. aestuarii* 357, las dos subunidades de una bomba de tipo RND detectadas en el estudio del proteoma fraccionado (EIE49510.1 y EIE49511.1), eran 5,6X y 4,8X más abundantes respectivamente tras la exposición a naftaleno respecto al control (tabla S2.7), y sí se habían podido detectar por predicción genómica. Así, EIE49511.1 se correspondía a una subunidad transportadora de la membrana interna con un porcentaje de identidad de secuencia con las proteínas modelo AcrB de *Escherichia coli* str. K-12 substr. MG1655, SrpB de *Pseudomonas putida* S12, MexB de *Pseudomonas aeruginosa* PAO1y TtgH de *Pseudomonas putida* DOT-T1E, mayor del 50 % de identidad con un *E-value* < 10^{-5}. Mientras que EIE49510.1 se corresponde a la proteína de fusión a la membrana con un porcentaje de identidad de secuencia mayor del 30 % de identidad con un *E-value* < 10^{-5} con las proteínas modelo AcrA de *Escherichia coli* str. K-12 substr. MG1655, SrpA de *Pseudomonas putida* S12, TtgA, TtgD y TtgG de *Pseudomonas putida* DOT-T1E y MexB de *Pseudomonas aeruginosa* PAO1. Por lo tanto, tanto en *R. pomeroyi* DSS-3^T como en *C. aestuarii* 357 se detectó una mayor abundancia de varias subunidades de bombas de tipo RND tras la exposición a naftaleno, resultado observado en *Pseudomonas putida* DOT-T1E tras la exposición a otro compuesto aromático como es el tolueno (Udaondo *et al*., 2012). Así pues, parece que en algunos miembros del linaje Roseobacter también se sobreexpresan bombas de

tipo RND en respuesta a la exposición a compuestos aromáticos, aunque estas bombas no siempre son parecidas a las descritas en tolerancia a hidrocarburos en *Pseudomonas* (Ramos *et al.*, 2002; Udaondo *et al.*, 2012), posiblemente por la distancia filogenética que hay entre los dos grupos.

Finalmente también se observó una abundancia relativa de proteínas de membrana de función desconocida significativamente mayor tras la exposición a naftaleno (tabla 2.3). Esto se debió básicamente a la mayor abundancia relativa de la proteína SPO2567 de *R. pomeroyi* DSS-3^T y de sus correspondientes homólogos en *Sulfitobacter* sp. 3SOLIMAR09 (MM82_00822) y *C. aestuarii* 357 (EIE51874.1) y de la proteína SPO3325 de *R. pomeroyi* DSS-3^T. Por lo que respecta a la primera, tras la exposición a naftaleno se observó que la abundancia de SPO2567 de *R. pomeroyi* DSS-3^T (AAV95815.1) era 120X respecto al control (tabla S2.2), y la de sus correspondientes homólogos en *Sulfitobacter* sp. 3SOLIMAR09 y *C. aestuarii* 357 era de 47,3X y 23X, respectivamente (tablas S2.3 y S2.4). La mayor abundancia relativa de la proteína hipotética de membrana SPO2567 fue descrito por Christie-Oleza y colaboradores (2012) en *R. pomeroyi* DSS-3^T tras la exposición a naftaleno, diésel y elevadas temperaturas. En concreto estos autores detectaron que la abundancia de la proteína SPO2567 de *R. pomeroyi* era 13 veces mayor tras la exposición a naftaleno y 11 veces mayor tras la exposición a elevadas temperaturas. Además, en un estudio anterior de nuestro grupo realizado por Mas-Lladó (2017) con el aislado *Sulfitobacter* sp. 3SOLIMAR09 también se observó que la abundancia del homólogo de SPO2567 (MM82_00822) era 11,6X respecto al control tras exposición de este aislado a agua de mar suplementada con diésel 0,1 % (v/v). Por lo tanto, nuestros resultados respaldaron los resultados previos indicando que esta proteína se expresa claramente tras exposición a hidrocarburos. Por lo que respecta a la proteína SPO3325 (AAV96552), la abundancia observada de esta proteína en cultivos tratados con naftaleno de *R. pomeroyi* DSS-3^T en este estudio fue cuatro veces superior a la de los controles (tabla S2.2). Este valor fue inferior al detectado en el estudio de Christie-Oleza y colaboradores (2012) tras exposición a naftaleno (11X). Por otra parte, cabe destacar que en este caso no se pudieron encontrar proteínas homólogas de SPO3325 en *C. aestuarii* 357 y *Sulfitobacter* sp. 3SOLIMAR09 con los criterios utilizados en este estudio (al menos 50 % de identidad en al menos 50 % de la secuencia y con un *E-value* menor de 10^{-5}). Christie-Oleza y colaboradores (2012) sugirieron un posible rol de las proteínas SPO2567, SPO2568 y SPO3325 en la estabilización de la membrana en *Ruegeria pomeroyi* DSS-3^T. A la vista de los resultados obtenidos en este trabajo, y teniendo en cuenta que proteínas homólogas a SPO2567 se observaron en los tres aislados en esta tesis se propuso generar un mutante de *R. pomeroyi* DSS-3^T defectivo para el gen codificante para esta proteína (ver a continuación en el apartado 3 de este capítulo).

2.2. Otras proteínas

Además de los cambios a nivel de proteínas de membrana, también se observaron cambios en la abundancia relativa de proteínas de otras categorías funcionales tras la exposición a naftaleno. En *R. pomeroyi* DSS-3^T se observó una mayor abundancia relativa significativa de ATPasas implicadas en la obtención de energía a partir de la hidrólisis de ATP tras la exposición a naftaleno (figura 2.3). Este resultado sugería que *R. pomeroyi* DSS-3^T debía aumentar su metabolismo energético para poner en marcha mecanismos con los que poder hacer frente a la presencia de naftaleno. En *C. aestuarii* 357 no se detectaron ATPasas en el proteoma total,

pero sí se detectaron en el proteoma fraccionado, aunque al igual que en *Sulfitobacter* sp. 3SOLIMAR09 no se producían cambios significativos en las abundancias relativas de estas proteínas en los controles y los cultivos tratados con naftaleno. A pesar de no detectar cambios significativos a nivel de ATPasas tras la exposición a naftaleno en *Sulfitobacter* sp. 3SOLIMAR09, la diferencia significativa en la abundancia relativa de las proteínas implicadas en la fosforilación oxidativa en este microorganismo (tabla 2.2) también indicaba un aumento de su metabolismo energético.

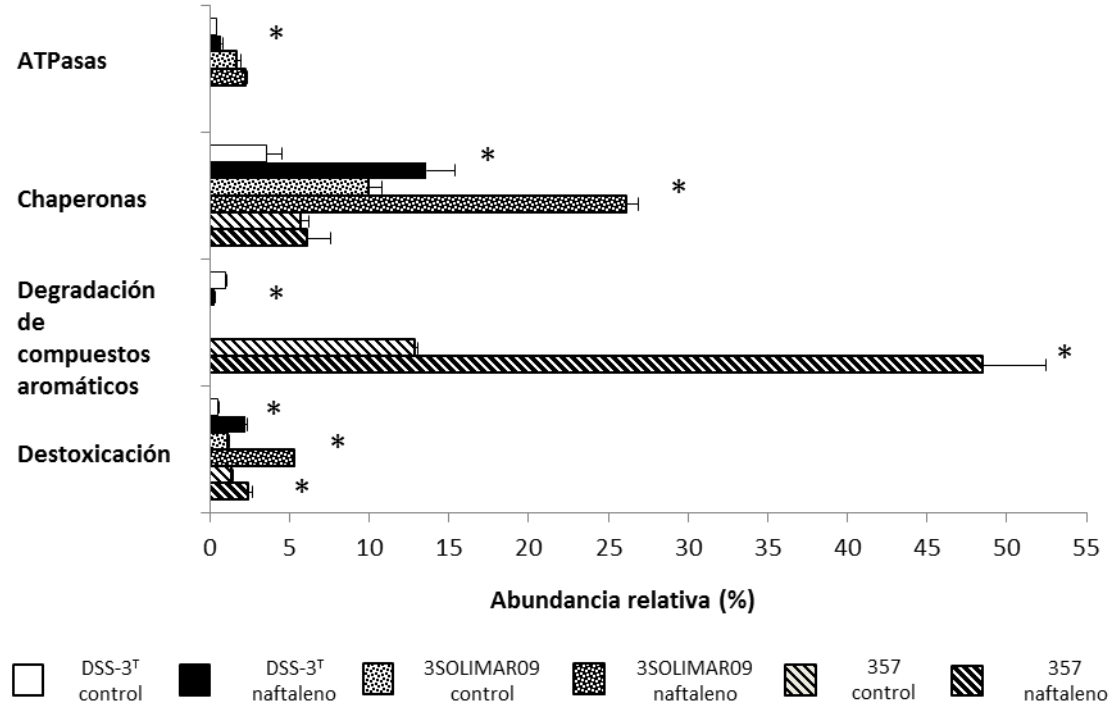

Figura 2.3. Cambios producidos tras la exposición a naftaleno en diferentes categorías de proteínas en los microorganismos estudiados. Se muestran los promedios y desviación estándar de las abundancias relativas normalizadas (n=3) tras la clasificación manual en categorías funcionales. Los cambios de abundancia relativa estadísticamente significativos para cada cepa y condición (t-Student, $p < 0,05$) se encuentran marcados con un asterisco.

Otro tipo de proteínas en las que se observaron cambios tras la exposición a naftaleno fueron las chaperonas (figura 2.3). En concreto se observó una diferencia significativa de su abundancia relativa en *R. pomeroyi* DSS-3^T y *Sulfitobacter* sp. 3SOLIMAR09 tras la exposición a naftaleno (figura 2.3). Entrando en detalle, en ambas bacterias se observó una mayor abundancia relativa de proteínas de las principales chaperonas celulares, la chaperona DnaKJE o Hsp70 (DnaK, DnaJ o GrpE) y el sistema GroESL o Hsp60 (GroEL o GroES). Estas dos chaperonas son importantes para el plegamiento correcto de las proteínas, su expresión aumenta en condiciones de estrés y su función depende de la hidrólisis de ATP (Saibil, 2013). Además en ambas bacterias se observó una mayor abundancia relativa de chaperonas del tipo *small heat shock proteins* (sHsp) como Hsp20 en *Ruegeria pomeroyi* DSS-3^T y IpbA en *Sulfitobacter* sp. 3SOLIMAR09. Estas chaperonas estabilizan proteínas desnaturalizadas para que puedan actuar sobre ellas las chaperonas principales mencionadas anteriormente, impidiendo la formación de agregados irreversibles (Kuczyńska-Wiśnik *et al*., 2002). En *Sulfitobacter* sp. 3SOLIMAR09 se observó también una mayor abundancia relativa de la

chaperona Hsp33 en los tratamientos con naftaleno. Esta chaperona se activa en condiciones de estrés oxidativo (Jakob *et al*., 1999). Se sabe que en *Pseudomonas putida* DOT-T1E, *Pseudomonas putida* S12 y *Pseudomonas putida* KT2440 la presencia de tolueno induce la expresión de varias chaperonas (Udaondo *et al*., 2012). En concreto, en base a análisis informáticos se vio que *Pseudomonas putida* DOT-T1E codificaba al menos para 41 chaperonas, de las cuales solo GroEL, GroES, GrpE, IbpA se expresaban a niveles más altos en presencia de tolueno (Udaondo *et al*., 2012). Además, observaron que *Pseudomonas putida* KT2440, cepa sensible a tolueno, activaba más chaperonas que *P. putida* DOT-T1E y *P. putida* S12, lo que podría estar relacionado con el hecho de que estas dos cepas fueran más eficientes a la hora de eliminar el tolueno (Udaondo *et al*., 2012). Existen otros estudios que también confirman la importancia de las chaperonas en la respuesta a estrés tras la exposición a tolueno, como el de Segura y colaboradores (2005). En este estudio realizaron experimentos con mutantes de *P. putida* DOT-T1E defectivos para las chaperonas CspA, XenA, y Tuf-1, confirmando su papel importante en la respuesta a estrés tras la exposición a tolueno, ya que los mutantes presentaban una menor viabilidad que las cepas silvestres. Por lo tanto, la presencia de naftaleno inducía diferencias significativas en la abundancia relativa de chaperonas, tanto en *Sulfitobacter* sp. 3SOLIMAR09 como en *R. pomeroyi* DSS-3^T, lo que sugería que el naftaleno inducía una mayor situación de estrés en ellas que en *C. aestuarii* 357. Además, se pueden relacionar estos cambios en la abundancia relativa de chaperonas, tanto en *Sulfitobacter* sp. 3SOLIMAR09 como en *R. pomeroyi* DSS-3^T, con el hecho de que hubiera una mayor diferencia entre los proteomas de sus cultivos control y los tratados con naftaleno en el análisis por PCA (figura 2.2).

Otras de las proteínas para las que se observó de forma significativa una mayor abundancia relativa en los proteomas de los tres microorganismos tras la exposición a naftaleno fueron proteínas implicadas en destoxicación celular (figura 2.3). En esta categoría no solo se incluyeron las proteínas implicadas en el metabolismo del glutatión (ver tabla 2.2) sino que también se encontraron una o varias serina proteasas periplasmáticas (tablas S2.2, S2.4 y S2.6). Además en *R. pomeroyi* DSS-3^T y *Sulfitobacter* sp. 3SOLIMAR09 se observaron varias peroxidasas (tablas S2.2 y S2.3). Una de las serina proteasas periplasmáticas detectadas en *R. pomeroyi* DSS-3^T fue la proteína SPO1625 (AAV94912.1, tabla S2.2). En el estudio de Christie-Oleza y colaboradores (2012) habían observado también una abundancia de esta proteína de 19X respecto a su control tras la exposición de *R. pomeroyi* DSS-3^T a naftaleno y habían sugerido su papel en la destoxicación celular. En nuestro estudio se observó que la abundancia de esta proteína AAV94912.1 era 38,5X respecto a su control (tabla S2.2). Además, tal y como se puede observar en la tabla S2.2, en esta bacteria también se observó una mayor abundancia de otras dos serina proteasas periplasmáticas: AAV93831.1 y AAV94622.1 (8,3X y 3X respecto al control, respectivamente). La serina proteasa detectada en *C. aestuarii* 357 (EIE50402.1) es homóloga de SPO1625 (AAV94912.1) de *R. pomeroyi* DSS-3^T y su abundancia era 2,8X respecto al control (tabla S2.4). En el caso de *Sulfitobacter* sp. 3SOLIMAR09, no se detectó ninguna serina proteasa periplasmática en el proteoma total. Pero en el proteoma fraccionado se detectó una proteína (MM82_00912) (tabla S2.6) homóloga de SPO1625 (AAV94912.1) de *R. pomeroyi* DSS-3^T y cuya abundancia era 4,7X respecto a su control (tabla S2.6). Por lo tanto, la mayor abundancia de esta proteína tras la exposición a naftaleno en los tres microorganismos estudiados muestra que es una respuesta común a todos ellos y respaldaría la hipótesis de

Christie-Oleza y colaboradores (2012) sobre el papel de esta proteína en la destoxicación celular en presencia de naftaleno. Esta mayor abundancia de las proteínas implicadas en la respuesta al estrés oxidativo, como son las chaperonas y las proteínas implicadas en la destoxicación, tras la exposición a compuestos aromáticos ya habían sido descritos en *Pseudomonas* como consecuencia del incremento en la producción de especies reactivas del oxígeno y peróxido de hidrógeno que se produce por la interferencia de los compuestos orgánicos con la cadena de transporte de electrones (Ramos *et al*., 2015; Domínguez-Cuevas *et al*., 2006).

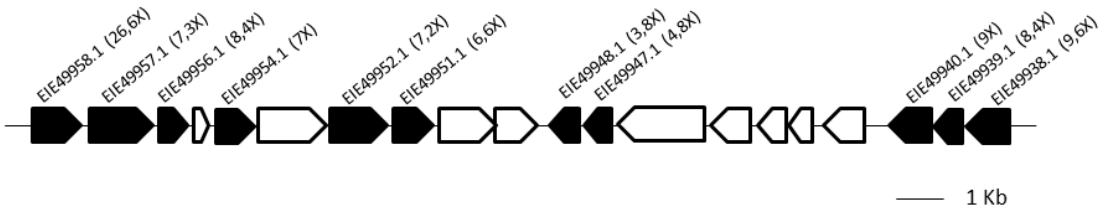

Figura 2.4. Estructura de la región que contiene los genes codificantes para las proteínas implicadas en la degradación de compuestos poliaromáticos en *C. aestuarii* 357. En paréntesis se indica el número de veces que aumenta la detección de estas proteínas tras la exposición a naftaleno. EIE49958.1, gentisato 1,2-dioxigenasa (NagI); EIE49957.1, subunidad Ac de la naftaleno 1,2-dioxigenasa (NagAc); EIE49956.1, subunidad Ad de la naftaleno 1,2-dioxigenasa (NagAd); EIE49954.1, 1,2-dihidroxi-1,2-dihidronaftaleno deshidrogenasa (NagB); EIE49952.1, citocromo P450; EIE49951.1, 5-carboximetil-2-hidroximuconato delta-isomerasa; EIE49948.1, maleilacetoacetato isomerasa (NagL); EIE49947.1, fumarilpiruvato hidrolasa (NagK); EIE49940.1, 1,2-dihidroxinaftaleno dioxigenasa (NagC); EIE49939.1, 2-hidroxicromeno-2-carboxilato isomerasa (NagD); EIE49938.1, trans-*o*-hidroxibenzilidenopiruvato hidratasa-aldolasa (NagE). Imagen modificada de Suárez-Suárez (2013).

Por otra parte, también se pudieron observar cambios significativos a nivel de abundancia de las proteínas involucradas en la degradación de compuestos aromáticos (figura 2.3) en *R. pomeroyi* DSS-3^T y en *C. aestuarii* 357. En *C. aestuarii* 357 tras la exposición a naftaleno se observó una mayor abundancia relativa de las implicadas en la degradación de naftaleno y el gentisato (ver apartado 6 del capítulo 1). Este resultado se corresponde con lo esperado puesto que *C. aestuarii* 357 era el único de los tres microorganismos estudiados que presentaba genes para degradación de naftaleno y que era capaz de degradar este compuesto y sus derivados (Suárez-Suárez, 2013). En la figura 2.4 se puede observar la anotación de la región del genoma de *C. aestuarii* 357 realizada por Suárez-Suárez (2013) en el que se encuentran codificados los genes implicados en la degradación de naftaleno. En esta región se encuentran los genes codificantes para nueve proteínas implicadas en la degradación de naftaleno (EIE49938.1, EIE49939.1, EIE49940.1, EIE49943.1, EIE49944.1, EIE49953.1, EIE49954.1, EIE49957.1 y EIE49956.1) y tres proteínas implicadas en la degradación de gentisato (EIE49958.1, EIE49948.1 y EIE49947.1) (Suárez-Suárez, 2013). En nuestro experimento, en el proteoma total tras la exposición a naftaleno se observó una mayor abundancia respecto a los controles de todas las proteínas de la figura 2.4 marcadas en negro, todas ellas descritas como implicadas en la degradación del naftaleno excepto EIE49951.1 y EIE49952.1. La primera está anotada como una 5-carboximetil-2-hidroximuconato delta-isomerasa implicada en la degradación del homoprotocatecuato (HpaF). En la predicción de genes catabólicos realizada en esta tesis (ver apartado 5 del de capítulo 1) no se detectó ningún gen *hpaF* en el genoma de *C. aestuarii* 357 que cumpliese con los criterios de este estudio (al menos 30 % de homología de secuencia con un *E-value*< 10^{-5} con la proteína

modelo) para poder considerarlo como identificado (ver figura 1.18 del capítulo 1). Respecto a la segunda proteína (EIE49952.1), esta es una posible citocromo P450 aunque su función no ha sido aún confirmada. En la tesis de la doctora Suárez-Suárez (2013) se hipotetizó que podría estar implicada en la hidroxilación del 2-oxohepta-3,5-dienedioiato, compuesto que podría producirse en caso de que el salicilato fuera degradado a partir de la rotura directa del anillo aromático por la gentisato 1,2-dioxigenasa. Por otra parte, también se observó que la abundancia de la fenilacetato-CoA ligasa (EIE49492.1) era 2,5 veces mayor en los controles respecto al tratamiento con naftaleno (tabla S2.4).

En *R. pomeroyi* DSS-3^T se observaron también cambios en la abundancia relativa de proteínas implicadas en degradación de compuestos aromáticos (figura 2.3). En primer lugar, tras la exposición a naftaleno se observó una menor abundancia relativa significativa de proteínas que formaban parte de las vías de degradación del homogentisato, fenilacetil-CoA, benzoil-CoA, y de la rama del protocatecuato de la vía del beta-cetoadipato (tabla S2.2). Tras la exposición a naftaleno solo se observó un aumento respecto a los controles de una proteína anotada en el genoma como salicilato 1,2-hidroxilasa (AAV95760.1) y una citocromo P450 (AAV94909.1) de 3,5X y 2,2X, respectivamente (tabla S2.2). Cabe destacar que esta citocromo P450 de *R. pomeroyi* DSS-3 presentaba un porcentaje de identidad de secuencia del 37 % con un *E-value* < 10^{-5} con la citocromo P450 de *C. aestuarii* 357 mencionada anteriormente. Por ello, sería interesante realizar estudios sobre esta citocromo P450 con los que poder esclarecer su función. Por lo tanto, los resultados sugirieron que, en general, la expresión de los genes implicados en la degradación de compuestos aromáticos en *R. pomeroyi* DSS-3^T era menor en presencia de naftaleno. Sin embargo, llama la atención el hecho de que *R. pomeroyi* DSS-3^T expresase proteínas de degradación de varios compuestos aromáticos en condiciones control, en las que los únicos compuestos aromáticos esperables serían aquellos derivados de aminoácidos presentes en el medio de cultivo.

Finalmente, estudiando en detalle las proteínas individuales que mostraban diferencias de abundancia significativas según los criterios establecidos en este estudio también se pudieron detectar respuestas propias de *Sulfitobacter* sp. 3SOLIMAR09 y *R. pomeroyi* DSS-3^T. En *Sulfitobacter* sp. 3SOLIMAR09 se detectó la presencia de una proteína gamma-glutamil quinasa-GP-reductasa (MM82_02642; tablas S2.3, S2.6 y S2.9) implicada en la síntesis de prolina que es un aminoácido osmoprotector (Zaprasis *et al*., 2013). En concreto se observó una abundancia 3,4X en el proteoma total (tabla S2.3) y 5X en el proteoma fraccionado (tabla S2.6) tras la exposición a naftaleno respecto al control. Además, también era 3X superior tras la exposición a tolueno en el proteoma fraccionado respecto al control (tabla S2.9). En cuanto a *R. pomeroyi* DSS-3^T en el estudio de proteoma fraccionado se detectó la presencia de la proteína MlaC (AAV96947.1, tabla S2.5) cuya función es transferir fosfolípidos de la membrana externa a la membrana interna para mantener la asimetría de la membrana externa (Huang *et al*., 2016). En concreto se vio que la abundacia de esta proteína era 2X tras la exposición a naftaleno respecto al control (tabla S2.5), aunque este resultado no fue estadísticamente significativo.

3. Aproximación al estudio de la función de la proteína SPO2567 de *Ruegeria pomeroyi* DSS-3^T

Como se ha descrito en apartados anteriores algunas proteínas de función desconocida como SPO2567 de *Ruegeria pomeroyi* DSS-3^T, y sus correspondientes homólogas en los otros dos microorganismos estudiados, se sobreexpresaban tras exposición a naftaleno. El hecho de que el resultado observado en esta tesis concordase con los resultados de dos experimentos independentes anteriores (Christie-Oleza *et al*., 2012; Mas-Lladó, 2017) justificaba un análisis de la función de esta proteína. Para ello se realizó un análisis bioinformático de esta proteína y se procedió a la generación de un mutante de *Ruegeria pomeroyi* DSS-3^T deficiente en SPO2567.

3.1. Análisis bioinformático

La proteína SPO2567 (AAV95815.1) es una proteína de 149 aminoácidos codificada entre las posiciones 2719095 y 2719544 del cromosoma de *R. pomeroyi* DSS-3^T y está anotada como proteína hipotética. Pertenece a la superfamilia DUF2852 (Pfam 11014) de función desconocida (Finn *et al*., 2016). Según la información de Pfam en el momento de escribir esta tesis esta superfamilia contenía secuencias de 113 especies bacterianas, prácticamente todas ellas de alfaproteobacterias.

En en análisis bioinformático de la proteína SPO2567 en primer lugar se determinó si era común al resto de miembros del linaje Roseobacter y si estaba presente en otros grupos bacterianos. Para ello se utilizó la herramienta de identificación y anotación de dominios Simple Modular Architecture Research Tool, SMART (Schultz *et al*., 1998; Letunic *et al*., 2015). Con esta herramienta se determinó la presencia de proteínas homólogas a SPO2567 en 232 bacterias, 231 de las cuales pertenecían al filo *Proteobacteria* y una pertenecía a un filo no clasificado. De las 231 proteobacterias que presentaban la proteína SPO2567 o sus correspondientes homólogos 2 pertenecían a la clase *Gammaproteobacteria* [*Neptuniibacter caesariensis* MED92 (Arahal *et al*., 2007) y *Leucothrix mucor* DSM 2157 (Skerman *et al*., 1980)]. Las 229 restantes se encontraban en miembros de la clase *Alphaproteobacteria*. La mayoría de las alfaproteobacterias que presentaban homólogos de esta proteína formaban parte de los órdenes *Rhizobiales* (154) y *Rhodobacterales* (69). Prácticamente la mitad de las bacterias del orden *Rhizobiales* que presentaban esta proteína eran de las familias *Rhizobiaceae* (41) y *Bradyrhizobiaceae* (34), mientras que prácticamente la totalidad de las bacterias del orden *Rhodobacterale*s que la presentaban formaban parte de la familia *Rhodobacteraceae* (68). Por lo tanto, estos resultados indicarían que se trataría de una proteína característica de alfaproteobacterias y conservada entre los miembros de la familia *Rhodobacteraceae*. Se realizó una filogenia de la proteína SPO2567 y las 231 proteínas homólogas detectadas por SMART y se observó que SPO2567 y sus correspondientes homólogas en el linaje Roseobacter se separaban de forma robusta de las proteínas homólogas presentes en los otros grupos bacterianos (figura S2.2). En 56 de los 61 genomas del linaje Roseobacter estudiados en esta tesis se identificaron proteínas homólogas de SPO2567, con una identidad de secuencia mayor del 50 %. Los cinco genomas de Roseobacter en los que no se detectaron homólogos de SPO2567, ni con el criterio de detección del 30 % de identidad, fueron *Roseovarius nubinhibens* ISM^T, *Roseobacter* sp. Azwk-3b, *Ketogulonicigenium vulgare* Y25 y WSH-001 y *Rhodobacterales*

bacterium HTCC2255. La filogenia de las proteínas detectadas en los genomas de Roseobacter estudiados en esta tesis no se correspondía con la filogenia del linaje (figura S2.3).

A continuación, se realizó una predicción de estructura secundaria y terciaria de la proteína SPO2567 con el programa de predicción de estructuras Phyre2 (Kelley *et al*., 2015). Según el resultado de la predicción de estructura secundaria con un nivel de confianza alto la proteína SPO2567 presentaría cuatro regiones de alfa hélice: del residuo 16 al 24, del residuo 31 al 55, del residuo 84 al 115 y del residuo 118 al 129. La segunda de estas regiones correspondería a una estructura de hélice transmembrana. El programa Phyre2 predijo que la región entre los residuos 1–33 estaría orientada hacia el medio extracelular, y que la región entre los residuos 55–149 quedaría orientada hacia el citoplasma celular (figura S2.4). En cuanto al análisis de posibles dominios, el programa tan solo predijo un posible modelo de estructura para los residuos 87–126 con una confianza del 74,4 % y una identidad de secuencia del 33 % del modelo con SPO2567 en esta región. El modelo propuesto correspondía a la estructura cristalina del dominio N-terminal de una metiltransferasa CheR de *Salmonella typhimurium* (dominio d1af7a1, código PDB 1AF7) (Djordjevic y Stock, 1997). No obstante, dado el valor de confianza obtenido, considerado como bajo por el programa Phyre2, y a que solo se encontró un modelo de estructura parcial no fue posible predecir una posible función para la proteína SPO2567 basada en estos resultados. Respecto a la predicción de estructuras de las proteínas homólogas de SPO2567 en *Sulfitobacter* sp. 3SOLIMAR09 y *C. aestuarii* 357 se obtuvo el mismo resultado y en ambos casos se predijo una estructura de hélice transmembrana (entre los residuos 32–54 y entre los residuos 34–56 respectivamente).

Por otra parte, se analizaron las posibles interacciones de la proteína SPO2567 con otras proteínas, utilizando la herramienta STRING v.10 (von Mering *et al*., 2003; Szklarczyk *et al*., 2015). Esta herramienta permite establecer relaciones directas (físicas) e indirectas (funcionales) de una proteína con el resto de proteínas de un genoma, el grado de confianza de esta interacción y si las interacciones con otras proteínas también se encontraba conservada en otros microorganismos. Como se puede observar en la tabla S2.13 STRING detectó que la proteína SPO2567 estaba relacionada con las proteínas SPO2568, SPO2569 y SPO2570, que se encuentran codificadas corriente abajo en el genoma de *R. pomeroyi* DSS-3^T. Este resultado estaba en concordancia con el aumento de la detección de la proteína SPO2568 que se observó acompañando al aumento de SPO2567 en el proteoma de esta bacteria tras la exposición a naftaleno (ver apartado 2.1 de este capítulo). Este resultado podía indicar que estas proteínas se expresaban conjuntamente. En todos los genomas del linaje Roseobacter que se incluyen en la base de datos de STRING se daba esta misma relación entre las proteínas homólogas a SPO2567 con las homólogas de SPO2568, SPO2569 y SPO2570. Por otra parte, también se observó una relación de co-ocurrencia de SPO2567 con 7 proteínas (SPO3462, SPO2758, SPO0125, SPO3896, SPO1310, SPO0934 y SPO2258; tabla S2.13) cuyos genes se encontraban en otras regiones del genoma que SPO2567, y cuyos homólogos pudieron ser detectados en más del 50 % de los genomas de *Rhodobacteraceae* de la base de datos de STRING en los que se había detectado SPO2567. Desafortunadamente, la mayoría de estas proteínas con las que STRING detectó las relaciones de co-ocurrencia, estaban anotadas como proteínas hipotéticas. Por ello, se trató de refinar la identificación de las cinco proteínas hipotéticas que se muestran en la tabla S2.13 mediante la herramienta Conserved Domain Search del NCBI (Marchler-Bauer *et al*., 2015) (tabla S2.14)

Con la herramienta Conserved Domain Search del NCBI (Marchler-Bauer *et al*., 2015), a pesar de poder determinar la presencia de algún tipo de dominio conservado en cada una de estas proteínas hipotéticas, no siempre se les pudo asignar una posible función (tabla S2.14). Por ejemplo, los dominios DUF983 y DUF3168 no tenían una función asignada, aunque se cree que las proteínas con dominio DUF3168 (Pfam 11367) son componentes de bacteriófagos (Finn *et al*., 2016). En base a los dominios encontrados en el resto de proteínas, SPO0125 podría presentar actividad de tipo esterasa, SPO1310 podría estar implicada en la resistencia al telurito y SPO0934 podría ser una posible sulfato permeasa. Además, tanto las proteínas SPO2758 como la SPO0934 presentaban dominios transmembrana por lo que podrían ser proteínas de membrana.

3.2. Generación de un mutante de *R. pomeroyi* DSS-3^T defectivo en SPO2567

Con el fin de estudiar la función de la proteína SPO2567 se generó un mutante de *R. pomeroyi* DSS-3^T defectivo para esta proteína llamado de aquí en adelante SPO2567$^-$ (ver apartado 5 de material y métodos).

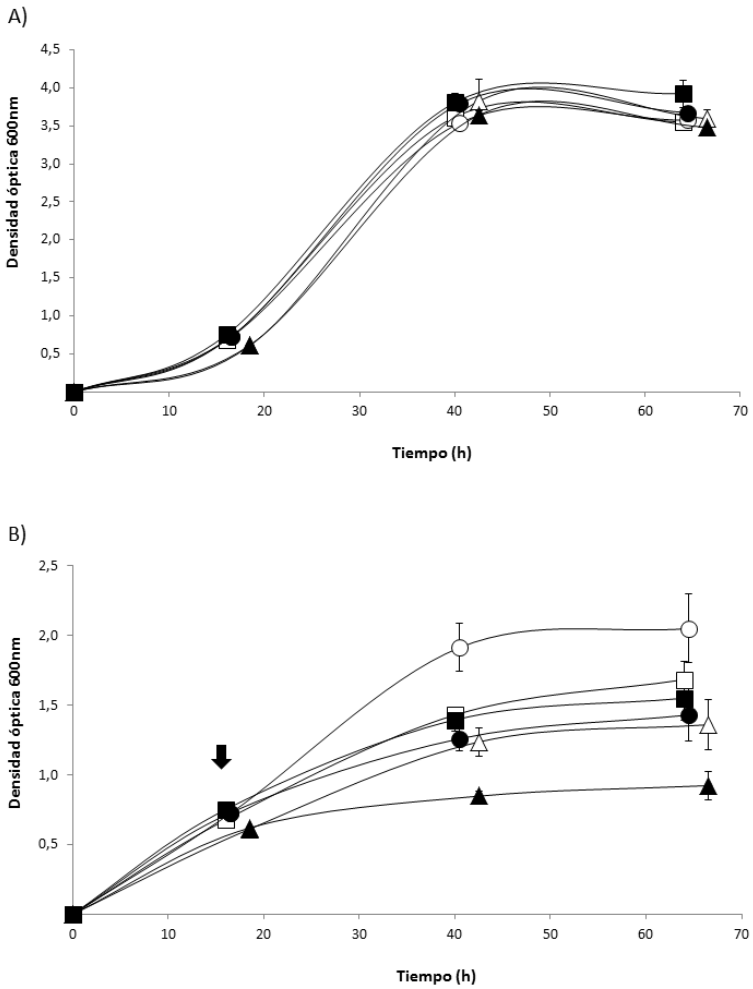

Figura 2.5. Densidad óptica de los cultivos de la cepa silvestre *R. pomeroyi* DSS-3^T (símbolos blancos) y el mutante SPO2567$^-$ (símbolos negros). Panel A, cultivos en caldo marino. Panel B, cultivo en caldo marino con naftaleno 0,1 % (p/v). Se realizaron tres experimentos independientes que se indican con diferentes símbolos: experimento 1 (cuadrado), experimento 2 (círculo) y experimento 3 (triángulo). En el panel B la flecha indica el momento en el que los cultivos se expusieron a naftaleno.

En primer lugar se realizaron experimentos para comprobar si el mutante SPO2567$^-$ presentaba una menor viabilidad que la cepa silvestre en condiciones control con caldo marino y cuando se adicionaba naftaleno. En la figura 2.5 se muestran los resultados de densidad óptica de los cultivos realizados. Como se puede observar en la figura 2.5A no se pudieron detectar diferencias entre la densidad óptica que alcanzaba la cepa silvestre respecto a la mutante cuando crecían en caldo marino. En cuanto a los cultivos que se expusieron a naftaleno, se observó una disminución de la densidad óptica tras la exposición a naftaleno tanto en la cepa silvestre como en el mutante SPO2567$^-$ (figura 2.5B). Las diferencias fueron significativas (t-Student, $p< 0,05$) tras 24 h de exposición a naftaleno en los experimentos 2 y 3 (figura 2.5B). Durante la realización de estos cultivos se recogieron muestras a las 24 y 48 h de incubación desde el inicio de los tratamientos con naftaleno y se sembraron varias diluciones para determinar el número de viables. En todos los experimentos se observó una reducción en el número de viables de entre 4 y 5 órdenes de magnitud (a las 24 y 48 h, respectivamente) entre los cultivos control y los tratados con naftaleno. Sin embargo, aunque esta tendencia se observaba en los tres experimentos la variabilidad de los recuentos no permitió establecer diferencias significativas entre ellos (ver ejemplo en figura suplementaria S2.5).

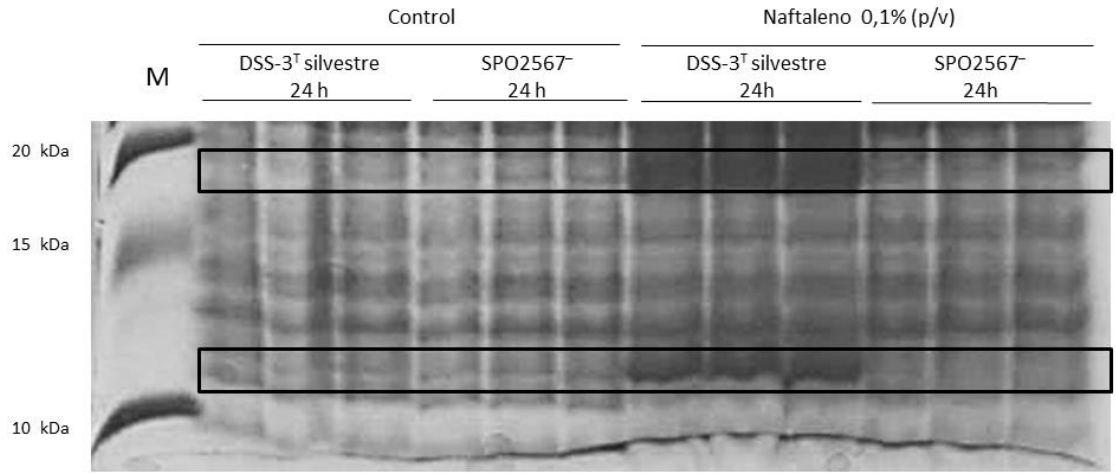

Figura 2.6. Cambios proteicos asociados a la exposición a naftaleno en la cepas silvestre y el mutante SPO2567$^-$ de *R. pomeroyi* DSS-3^T.

Finalmente, se obtuvo el proteoma total de las células de los cultivos control y los expuestos a naftaleno tras 24 y 48 h del inicio de tratamiento de los 3 experimentos realizados y se analizaron los perfiles proteicos en geles de SDS-PAGE. Los perfiles proteicos correspondientes a las 24 y 48 h eran muy similares por lo que se muestran solo los resultados de los primeros. Como se puede observar en la figura 2.6 no se pudieron detectar diferencias remarcables entre los perfiles proteicos de las cepas silvestre y mutante cuando crecían en caldo marino (control) pero sí se observaron diferencias tras 24 h de exposición a naftaleno 0,1 % (p/v). En primer lugar, se observó la expresión de una proteína en el rango de tamaño de la proteína SPO2567 (16,39 kDa) en la cepa silvestre de *R. pomeroyi* DSS-3^T tras 24 h de exposición a naftaleno, que no se observaba en la cepa mutante SPO2567$^-$. Esta proteína tampoco se observaba con la misma intensidad en los cultivos control sin naftaleno, ni de la cepa silvestre ni del mutante. Por otra parte, en la cepa silvestre también se observó la expresión de una proteína con un peso molecular entre 10 y 15 kDa tras la exposición a naftaleno que no aparecía en los perfiles de la cepa mutante ni con la misma intensidad en los controles. En

base a las proteínas que se observaron en el experimento de proteómica masiva de la cepa DSS-3^T silvestre, cuya abundancia relativa era mayor tras 24 h de exposición a naftaleno, y cuyo tamaño estaría en este rango de peso se detectaron dos proteínas: SPO3014 (12,43 kDa) y SPO1886 (11,5 kDa), cuya abundancia era 10,2X y 3X, respectivamente, tras la exposición al naftaleno (tabla S2.2). Ambas proteínas están anotadas como hipotéticas en el genoma de *R. pomeroyi* DSS-3^T. Tras el análisis de estas dos proteínas con la herramienta Conserved Domain Search del NCBI (Marchler-Bauer *et al*., 2015) no se pudo asignar ningún dominio funcional que permitiese inferir una posible función a estas proteínas. En base a estos resultados preliminares con el mutante defectivo para SPO2567 se demostró que presentaba un fenotipo diferente al de la cepa silvestre. Por ello, en un futuro sería interesante realizar experimentos adicionales de viabilidad y estudiar el fenotipo del mutante, no solo en respuesta a naftaleno, sino también a otros tipos de estrés como la temperatura y el diésel, condiciones que según Christie-oleza y colaboradores (2012) también inducían la expresión de esta proteína. Además, también sería interesante estudiar los mecanismos que presenta la célula para suplir la possible función de esta proteína en el mutante SPO2567$^-$ cuando se expone a naftaleno.

CAPÍTULO 3: METAGENÓMICA DE COMUNIDADES MICROBIANAS MARINAS

El objetivo principal de esta parte de la tesis era estudiar las diferencias a nivel taxonómico y funcional en las comunidades microbianas de aguas costeras no contaminadas y aguas contaminadas con hidrocarburos, y si estos cambios podían reproducirse mediante experimentos de microcosmos. En especial y en relación con los capítulos anteriores, se estudió en profundidad la presencia del linaje Roseobacter y el potencial para el catabolismo de compuestos aromáticos de los microorganismos presentes en estas aguas. Para ello se generaron ocho metagenomas, uno de agua control no contaminada, uno de agua de un puerto deportivo y seis de aguas limpias sometidas a varios tratamientos en experimentos de microcosmos (ver apartado 7.2.1 de material y métodos).

Todas las muestras de agua para análisis de metagenomas se tomaron en los meses de marzo-abril de dos años consecutivos. En este periodo, que coincide con el final del invierno, la columna de agua está mezclada, la concentración de clorofila en el agua indica la proliferación de fitoplancton y la abundancia de poblaciones del linaje Roseobacter es relevante, como muestran estudios previos (Nogales *et al*., 2007; Aguiló-Ferretjans *et al*., 2008; Piña-Villalonga, 2012). Los parámetros físico-químicos y biológicos de las muestras de agua utilizadas se muestran en la tabla S3.1. La temperatura del agua estaba en el rango de 15–16 °C, los valores de salinidad en 37,2–37,6 ‰, y la concentración de clorofila total era de 0,55 µg/l en agua del puerto deportivo y estaba entre 0,21 y 0,48 µg/l en aguas control. El número de procariotas totales estaba entre 5,9 x 10^5 ± 2,36 x 10^5 y 1,62 x 10^6 ± 4,26 x 10^5 células/ml en las muestras control y era de 1,4 x 10^6 ± 3,54 x 10^5 células/ml en la muestra del puerto deportivo.

1. Experimentos de microcosmos de laboratorio

Resultados previos del grupo de investigación mostraron la proliferación de bacterias del linaje Roseobacter cuando se adicionaba diésel (con o sin nitrógeno) a aguas costeras de Mallorca (control) en experimentos de laboratorio (Lanfranconi *et al*., 2010, Piña-Villalonga, 2012). En estos experimentos previos se utilizó únicamente el estudio de las secuencias de 16S ARNr como marcador molecular para seguir los cambios que se producían en las comunidades microbianas. Por lo tanto, no se pudieron seguir posibles cambios en la funcionalidad de las comunidades. En esta tesis se planteó realizar un análisis metagenómico de las comunidades microbianas que se desarrollaban en microcosmos contaminados. Con ello se pretendía tener una aproximación a los cambios taxonómicos y funcionales que se producían en los microcosmos y establecer comparaciones con las comunidades ambientales de aguas contaminadas o limpias. Los microcosmos se realizaron en dos años consecutivos (2013 y 2014). El tratamiento realizado fue la adición de bajas concentraciones de diésel (10 mg/l) y $NaNO_3$ (2,63 mg/l) porque se tenían evidencias que esto favorecía la proliferación de poblaciones del linaje Roseobacter (Piña-Villalonga, 2012). En el momento en que se tomaron las muestras de agua se determinó que el bacterioplancton de las aguas control utilizadas para la preparación de los microcosmos no se encontraba limitado por nitrógeno, ya que no se observaron incrementos significativos del número de procariotas totales tras la adición de un pulso de nitrógeno (ver figura S3.1).

Durante la incubación de los microcosmos se siguió la dinámica de las comunidades microbianas mediante la realización de recuentos al microscopio (DAPI y FISH) y de perfiles electroforéticos T-RFLP de 16S ADNr. En el primer caso se determinó la abundancia de

procariotas totales y de poblaciones de interés (p.e. linaje Roseobacter). Con los perfiles de T-RFLP se analizaron cambios en la composición de las comunidades microbianas. También se analizó la abundancia de fragmentos terminales (T-RFs) diagnósticos de la presencia de bacterias del linaje Roseobacter (Piña-Villalonga, 2012). Esta información se utilizó para decidir las muestras que se procesarían posteriormente para el análisis metagenómico.

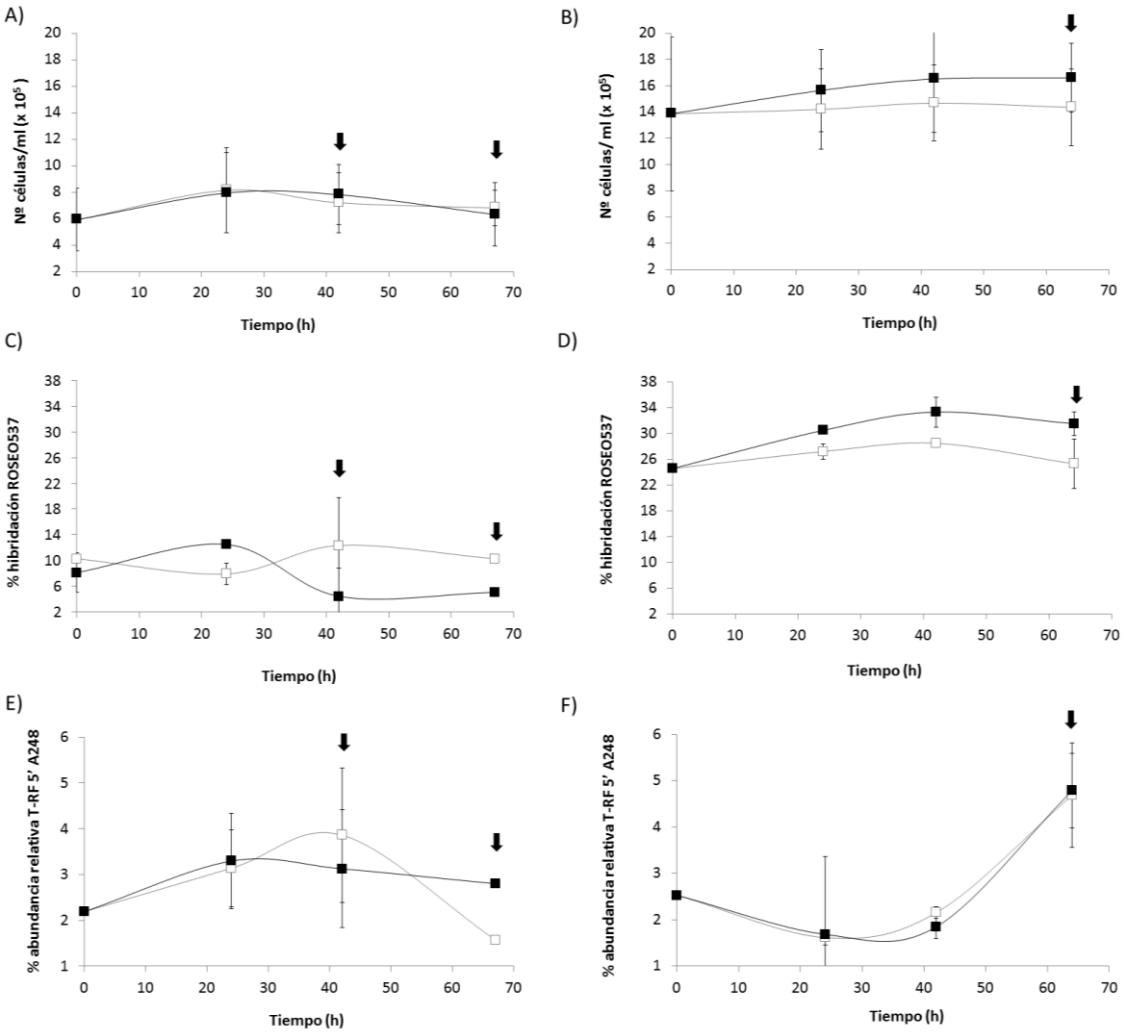

Figura 3.1. Cambios en las comunidades microbianas en experimentos de microcosmos. Paneles A, C y E microcosmos de abril de 2013. Paneles B, D y F, microcosmos M1 de abril de 2014. Paneles A y B, número de procariotas totales. Paneles C y D, porcentaje de hibridación con la sonda de Roseobacter ROSEO537. Paneles E y F, porcentaje de abundancia relativa del T-RF 5' A248. Se muestra la media de las dos réplicas. Los símbolos de color blanco se corresponden a los microcosmos controles mientras que los de color negro a los microcosmos tratados con diésel y $NaNO_3$. Se marcan con flechas los momentos en los que se recogieron muestras para generar los metagenomas.

En ninguno de los experimentos de microcosmos realizados (2013 y 2014 M1 y M2) se detectó un incremento significativo del número de procariotas totales (tinción DAPI) en respuesta a los tratamientos (figuras 3.1A-B). En todos los casos, la mayoría de procariotas detectados formaban parte del dominio *Bacteria*, puesto que entre un 69 y un 92 % de los procariotas totales hibridaron con las sondas universales EUBI-III (datos no mostrados). En los microcosmos de 2013 y en el primer experimento del 2014 (M1) se observó un aumento en el porcentaje de hibridación con la sonda ROSEO537 (Eilers *et al*., 2001) a las 24 h de incubación con diésel y nitrógeno (figura 3.1C) y hasta el final del experimento en el segundo caso (figura

3.1D). Este resultado sugería un aumento en el número de células del linaje Roseobacter en estas condiciones. El resultado fue diferente en los microcosmos M2 (figura S3.2). En ese caso se detectó una menor hibridación de la sonda ROSEO537 durante la incubación que al principio de la misma (figura S3.2B). Además, se vio que el menor porcentaje de hibridación se daba a las 24 h de incubación. Cabe destacar que el porcentaje de hibridación con la sonda ROSEO537 fue mayor en los dos microcosmos de 2014 (alrededor del 20 % de hibridación) que en los microcosmos de 2013 (alrededor del 10 % de hibridación) (figuras 3.1C, 3.1D y S3.2B). Este aumento de detección de Roseobacter por FISH no quedó reflejado en la abundancia relativa del T-RF 5'-A248 identificativo del linaje Roseobacter (Piña-Villalonga, 2012), que no era mayor en los tratamientos en comparación con los controles (figuras 3.1E-F y S3.2C). A pesar de no poder correlacionar la abundancia del T-RF 5'-A248 característico del linaje Roseobacter con la adición de diésel y nitrógeno, en los dos experimentos de microcosmos de 2014 se pudo detectar una mayor abundancia del linaje Roseobacter a las 42 y 64 h de incubación (figuras 3.1F y S3.2C). En cambio, en los microcosmos de 2013 se detectó una mayor abundancia del linaje Roseobacter a las 24 y 48 h de incubación (figura 3.1E).

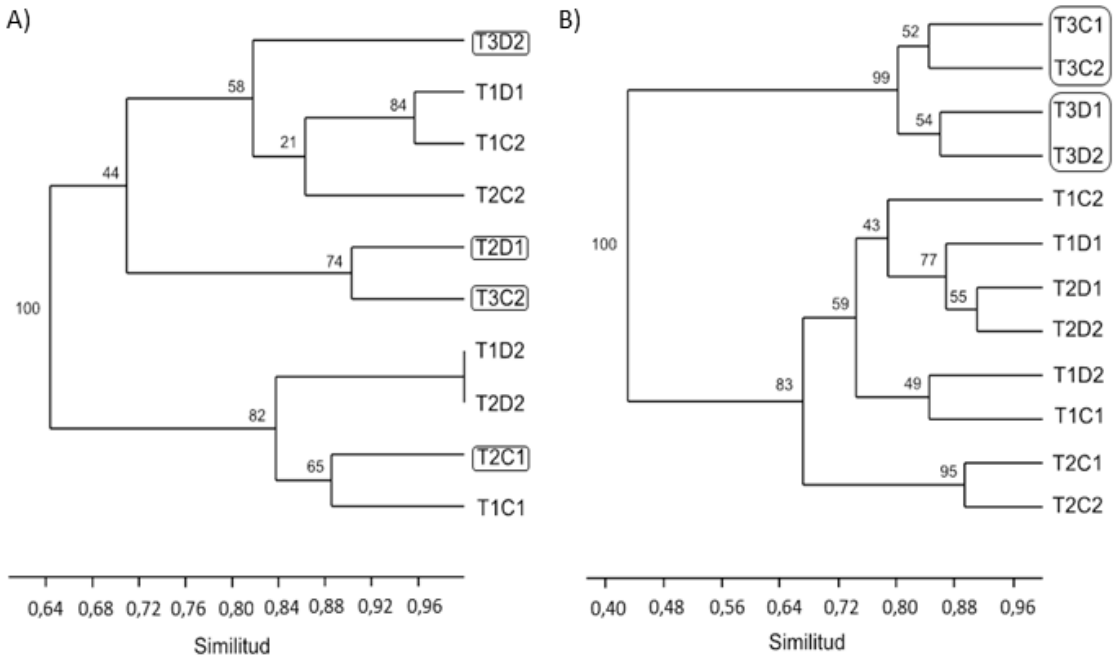

Figura 3.2. Similitud entre las comunidades microbianas en los experimentos de microcosmos. Dendograma calculado con el método UPGMA utilizando el índice de similitud de Bray-Curtis basado en los fragmentos 5' obtenidos tras la digestión con AluI y HhaI. Panel A, experimento de microcosmos de 2013. Panel B, experimento de microcosmos M1 de 2014. Se indican los valores de *bootstrap* (100 réplicas) de cada rama. Los recuadros indican las muestras a partir de las cuales se generaron metagenomas. El nombre de las muestras indica el tiempo de incubación (T1=24 h, T2= 42 h, T3=67 h en los microcosmos de 2013 y T3=64 h en los microcosmos de 2014), el tratamiento (C= control y D= diésel con nitrógeno) y el número de réplica (1 o 2).

En base a la abundancia relativa de los diferentes picos electroforéticos obtenidos tras la digestión con las enzimas AluI y HhaI del 16S ADNr amplificado se determinaron los cambios en la composición de la comunidad microbiana total en cada microcosmos y los distintos tiempos de incubación. Los resultados previos de Lanfranconi (2010) y Piña-Villalonga (2012) en experimentos similares indicaban cambios en la composición de las comunidades microbianas tras el tratamiento con diésel a partir de las 17 h de incubación. Por lo tanto, con este análisis preliminar se pretendía confirmar que se habían producido cambios en la

composición de las comunidades microbianas antes de procesar las muestras para metagenómica. En el experimento de microcosmos de 2013 se observaron diferencias entre las muestras control (C) y las tratadas con diésel y nitrógeno (D) tanto tanto a las 24 h (T1), como posteriormente a las 42 h (T2) y a las 67 h (T3) (ver figura 3.2A). En este experimento una de las réplicas de los microcosmos (réplica 1) se procesó, filtrando la totalidad del agua para recoger la microbiota, a las 42 h y la segunda réplica a las 67 h. Se generaron 4 metagenomas denominados T2C1 2013, T3C2 2013, T2D1 2013 y T3D2 2013, a partir las cuatro muestras correspondientes a estos tiempos (figura 3.2A). De esta manera se obtendría información de los cambios funcionales y en composición que hubieran tenido lugar a tiempos intermedios y finales de incubación. En el experimento de 2014 se decidió procesar únicamente microcosmos incubados a tiempo final (en este caso 64 h). En los experimentos de microcosmos de 2014 se observó también una diferenciación de las comunidades microbianas tras el tratamiento con diésel y nitrógeno (figura 3.2B y figura suplementaria S3.3). En el primer experimento (M1) se pudo observar como todas las muestras recogidas tras 64 h de incubación (T3) se agrupaban juntas con al menos un 80 % de similitud (figura 3.2B), pero separadas en función del tratamiento. Como los perfiles de las dos réplicas de los controles (T3C1 y T3C2) y las dos de los tratados (T3D1 y T3D2) presentaban alrededor de un 84 % y un 86 % de similitud entre ellas, respectivamente, se decidió mezclar ADN total de las dos réplicas (control y tratamiento) y generar dos metagenomas de las muestras compuestas del microcosmos M1 (T3C1 2014 y T3D1 2014).

2. Análisis de los metagenomas

2.1. Características de los metagenomas

Mediante secuenciación Illumina se generaron entre 7 y 11 millones de lecturas crudas para cada uno de los 8 metagenomas del estudio. Tras el cribado de las secuencias que cumplían los requisitos de calidad establecidos (ver apartado 8 de material y métodos) se obtuvieron entre 6 y 10 millones de lecturas de cada metagenoma para el análisis (tabla 3.1). Este número de lecturas era el resultado del sumatorio de los dos conjuntos de lecturas pareadas (R1 y R2) más todas las lecturas desapareadas (R3). Los resultados en el número de *contigs* obtenidos tras realizar ensamblajes fueron dispares en los diferentes metagenomas como puede verse en la tabla.

Tabla 3.1. Datos de los metagenomas generados en este estudio.

Metagenomas	Número de lecturas		*Contigs*	Longitud máxima (pb)[a]
	Crudas	Cribadas		
Ambientales				
Control	9.262.082	8.636.829	1.418.030	6.278
Puerto	7.125.396	6.396.092	157.777	197.625
Microcosmos				
T2C1 2013	11.844.560	10.771.457	1.184.391	30.182
T2D1 2013	9.814.390	8.798.600	661.523	24.163
T3C2 2013	8.506.924	7.649.310	720.101	32.792
T3D2 2013	9.522.134	8.766.188	951.827	21.033
T3C1 2014	7.734.328	7.180.581	1.987.172	9.250
T3D1 2014	7.148.688	6.693.392	1.781.649	12.290

a: longitud del *contig* más largo.

Tabla 3.2. Detalle de las lecturas analizadas con MG-RAST de los metagenomas generados en este estudio.

Metagenomas	Número de lecturas			Longitud media (pb)	% mol G+C	Asignación funcional (%)[d]			
	Inicial[a]	Unión[b]	Final[c]			Proteínas de función conocida	Proteínas de función desconocida	Genes rRNA	No codificantes
Ambientales									
Control	8.199.950	5.758.817	5.718.961	273 ± 92	36 ± 10	75,76	23,84	0,40	0,00
Puerto	5.858.160	4.281.866	4.234.510	251 ± 92	49 ± 13	72,79	26,53	0,44	0,25
Microcosmos									
T2C1 2013	10.090.814	8.170.986	8.074.216	234 ± 83	42 ± 13	57,78	38,26	2,10	1,90
T2D1 2013	7.097.356	7.084.072	6.971.778	225 ± 77	40 ± 12	64,73	31,82	2,51	0,94
T3C2 2013	8.180.400	6.014.640	5.961.754	231 ± 79	43 ± 13	53,51	42,07	2,72	1,70
T3D2 2013	8.276.636	6.539.206	6.449.524	242 ± 83	42 ± 12	68,01	28,87	3,12	0,00
T3C1 2014	6.790.832	4.726.576	4.697.504	268 ± 90	41 ± 12	72,75	26,86	0,39	0,00
T3D1 2014	6.382.070	4.306.478	4.267.075	270 ± 89	40 ± 11	78,38	21,24	0,38	0,00

a: Corresponde únicamente a lecturas pareadas (R1 y R2), descartando las lecturas desapareadas (R3).
b: N° de secuencias tras la unión de las lecturas pareadas con la opción *Join paired ends* de MG-RAST.
c: N° de secuencias tras el control de calidad (QC) de MG-RAST.
d: Asignación funcional en base a la clasificación con las bases de datos subsistemas SEED (Overbeek *et al*., 2005; 2014), COG (Tatusov *et al*., 2000), NOG (Jensen *et al*., 2008) y KO (Kanehisa *et al*., 2016a).

El análisis taxonómico y funcional de los metagenomas se realizó con la herramienta MG-RAST (Meyer *et al*., 2008) aplicando la opción de unión de lecturas pareadas (*Join paired ends*, ver apartado 8 de material y métodos). Tras el solapamiento de las lecturas pareadas MG-RAST realizó un análisis de calidad (QC) a las secuencias resultantes y cómo se puede observar en la tabla 3.2 entre el 98,4–99,4 % de las lecturas pasaron este análisis de calidad. Entre el 95,6–99,6 % de las secuencias que pasaron el análisis de calidad eran secuencias codificantes para proteínas. En concreto entre un 53,5–78,3 % de las secuencias codificaban para proteínas de función conocida. Finalmente, cabe destacar que solo entre 0,38–3,12 % de las secuencias codificaban para rRNA y entre el 0–1,9 % de las secuencias no eran codificantes (tabla 3.2).

2.2. Caracterización taxonómica de los metagenomas

Se determinó la composición de microorganismos en los metagenomas en base a las secuencias que habían pasado el análisis de calidad de MG-RAST (tabla 3.2) a partir de dos aproximaciones distintas: (1) mediante su clasificación en MG-RAST usando la base de datos de REFSeq (Pruitt *et al*., 2007) y (2) mediante su clasificación con SILVAngs (Pruesse *et al*., 2007) (ver apartado 8 de material y métodos). Los resultados que se obtuvieron eran similares pero la resolución era mayor en el primer caso, por lo que solo se muestran los resultados de MG-RAST.

En primer lugar se analizó el grado de cobertura de cada metagenoma mediante el estudio de las curvas de rarefacción que proporcionaba MG-RAST (Meyer *et al*., 2008), calculadas en base al número de especies distintas anotadas en función del número de secuencias analizadas. Como se puede observar en la figura S3.4 en los ocho metagenomas se obtuvieron curvas de rarefacción que tenían tendencia a la saturación, por lo tanto en todos ellos se obtuvo una representación adecuada de la diversidad microbiana. En concreto, se pudieron detectar entre 5819 y 7906 especies distintas en cada uno de los metagenomas analizados (figura S3.4). Todos los metagenomas tenían índices de diversidad de Shannon (H) comprendidos entre 5,5 y 5,8 y valores de equitabilidad (J) comprendidos entre 0,61 y 0,67 (tabla S3.2). En concreto se observó que el metagenoma con un mayor índice de diversidad y de equitabilidad fue el obtenido con la muestra del puerto, sugiriendo no solo que este era el metagenoma con un mayor número de especies distintas, sino que también las especies de este metagenoma eran las que presentaban entre ellas un mayor grado de uniformidad en abundancia (tabla S3.2).

En cuanto a la clasificación taxonómica, entre el 92 y el 97 % de las secuencias de todos los metagenomas pertenecían al dominio *Bacteria*, excepto en el caso del metagenoma de puerto en el que solo el 83 % de las secuencias pertenecían a este dominio (tabla 3.3). En la muestra del puerto el menor porcentaje de secuencias asignadas al dominio *Bacteria* fue acompañado de un mayor porcentaje de secuencias asignadas al dominio *Eukarya* (tabla 3.3). Este resultado era esperable ya que en el agua de puerto la concentración de clorofila era mayor (Tabla S3.1). Además ya se ha descrito previamente que en los puertos, por su mayor concentración de nutrientes, hay una mayor concentración de algas eucariotas que en las aguas oligotróficas (Nogales, 2010; Davidson *et al*., 2014). También cabe destacar que hubo una menor asignación de secuencias al dominio *Eukarya* en los tres metagenomas de los microcosmos tratados con diésel y nitrógeno respecto a sus respectivos controles (tabla 3.3). La menor presencia de eucariotas en experimentos de microcosmos tras la adición de diésel ya había sido descrita

previamente por Lanfranconi (2010). Por otra parte, se pudo ver que tanto el porcentaje de secuencias del dominio *Archaea* como de virus era similar en todos los metagenomas, con abundancias relativas de entre 0,2 y 1,3 % (tabla 3.3). Además, en el caso de las secuencias asignadas al dominio *Archaea* también se observó una menor asignación a este dominio en los tres metagenomas de microcosmos tratados con diésel y nitrógeno respecto a sus respectivos controles (tabla 3.3). Este resultado sugeriría un efecto negativo del diésel sobre la presencia de miembros del dominio *Archaea*, tal y como también ocurría con los miembros del dominio *Eukarya*, pero esto debe confirmarse con estudios adicionales.

Tabla 3.3. Asignación de las secuencias de los metagenomas a los diferentes dominios.

Metagenomas	Porcentaje de lecturas (%)				
	Bacteria	***Archaea***	***Eukarya***	**Virus**	**No clasificadas**
Ambientales					
Control	96,60	0,45	1,78	0,57	0,60
Puerto	83,82	0,23	14,68	1,00	0,27
Microcosmos					
T2C1 2013	94,84	0,40	3,20	1,14	0,42
T2D1 2013	95,61	0,31	2,52	1,15	0,41
T3C2 2013	92,49	0,39	5,31	1,31	0,50
T3D2 2013	96,17	0,28	2,13	1,10	0,32
T3C1 2014	95,93	0,32	2,55	0,69	0,51
T3D1 2014	97,10	0,32	1,24	0,98	0,36

A continuación se analizó la composición a nivel de familias en los ocho metagenomas, centrando el estudio en las familias del dominio *Bacteria* a las que se habían asignado al menos un 1 % de las lecturas en algún metagenoma (figura 3.3, tabla S3.3). Tal y como se puede observar en la figura 3.3, la mayoría de familias detectadas en los metagenomas estudiados pertenecían a las clases *Alphaproteobacteria* y *Gammaproteobacteria* del filo *Proteobacteria* y al filo *Bacteroidetes*. Este resultado era similar al obtenido por Kisand y colaboradores (2012) en metagenomas generados a partir de agua control y del puerto de Génova. En concreto, ellos observaron que en estas muestras predominaban las clases *Alphaproteobacteria*, *Gammaproteobacteria* y *Bacteroidetes*, detectando también en su caso la clase *Actinobacteria* como una de las clases dominantes en las aguas controles (Kisand *et al*., 2012). Por otra parte, no todas las lecturas de nuestro estudio pudieron ser asignadas a familias descritas según el sistema de clasificación de MG-RAST, y en todos metagenomas se observó que al menos un 1 % de las lecturas se asignaron a clases pero no se clasificaban en familias. Tal y como se puede observar en la figura 3.3 todas las cianobacterias no clasificadas a nivel de familia pertenecían al orden *Chroococcales* y a su vez, la mayoría pertenecía al género *Synechococcu*s. Entre las alfaproteobacterias no clasificadas se encontraban secuencias de los grupos SAR11 y SAR116 (Giovannoni *et al*., 1990). Finalmente, entre las gammaproteobacterias no clasificadas se encontraban secuencias de los grupos SAR86 (Britschgi y Giovannoni, 1991) y OM60, SAR92, y BD1-7, todos ellos grupos pertenecientes a las

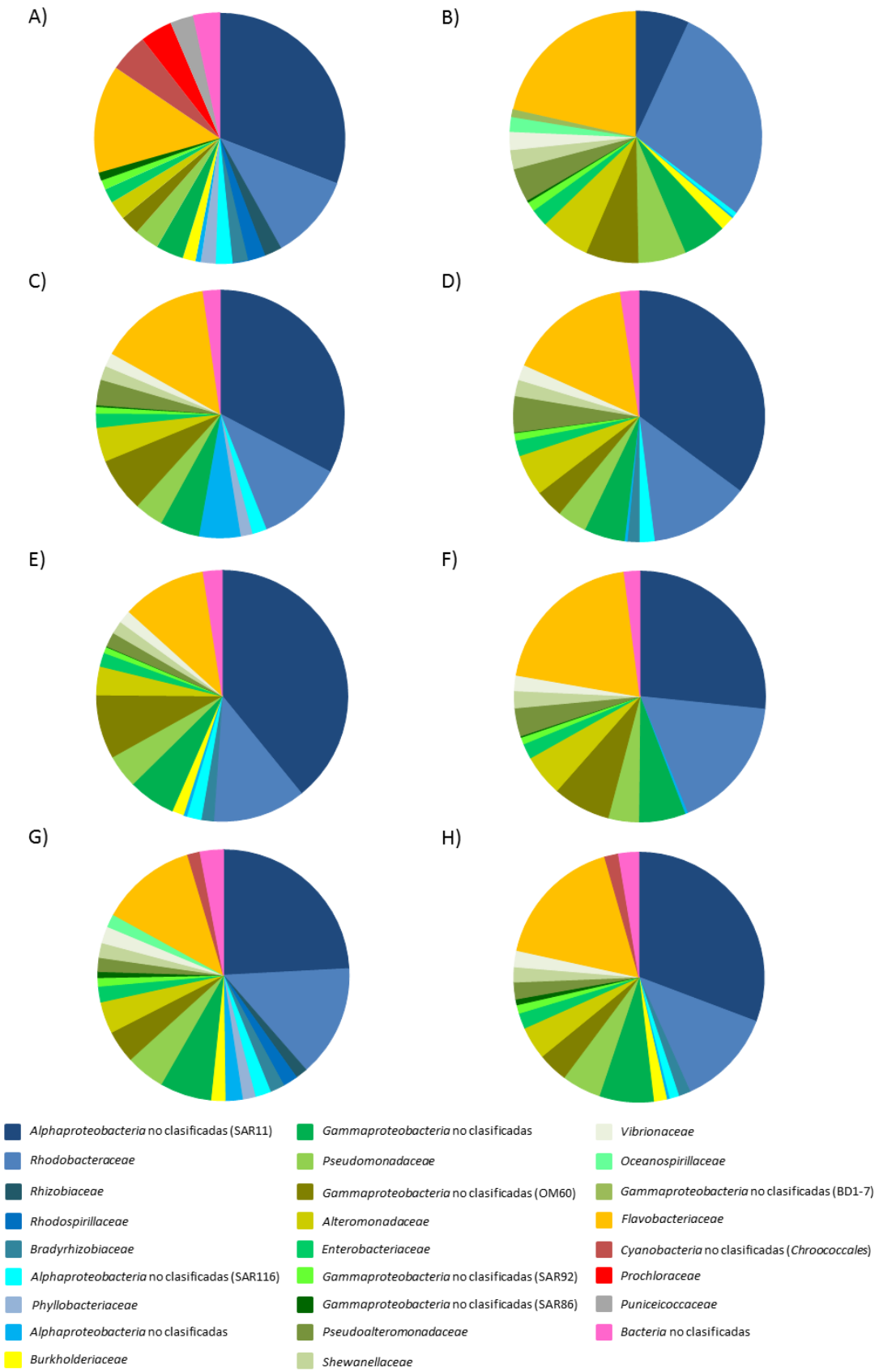

Figura 3.3. Abundancia relativa de las secuencias asignadas a familias del dominio *Bacteria* en los metagenomas analizados. Solo se muestran las familias que representan al menos un 1 % de la abundancia relativa en alguna de las genotecas. Panel A, Control. Panel B, Puerto. Panel C, T2C1 2013. Panel D, T2D1 2013. Panel E, T3C2 2013. Panel F, T3D2 2013. Panel G, T3C1 2014. Panel H, T3D1 2014.

gammaproteobacterias marinas oligotróficas (OMG) (Cho y Giovannoni, 2004). Todos estos grupos son importantes en medio marino y en particular en los ambientes analizados en este estudio (Nogales *et al.* 2007; Aguiló-Ferretjans *et al*., 2018). Por esta razón se revisaron manualmente los resultados de la clasificación taxonómica para especificar los grupos a los que pertenecían las secuencias no clasificadas siempre que fue posible.

Comparando los resultados del metagenoma de agua control con el de puerto (figura 3.3A-B y tabla S3.3) se observó que en el metagenoma del puerto había un menor porcentaje de secuencias asignadas a alfaproteobacterias y cianobacterias, y una mayor proporción de secuencias de gammaproteobacterias, aunque con una menor proporción de SAR86 que en la muestra control, y bacteroidetes. La menor proporción de secuencias asignadas a alfaproteobacterias se debió básicamente a la disminución en el porcentaje de secuencias pertenecientes a los grupos SAR11 y SAR116 y de las familias *Rhizobiaceae*, *Rhodospirillaceae*, *Bradyrhizobiaceae* y *Phyllobacteriaceae*. Además, cabe destacar que la única familia de alfaproteobacterias cuya proporción era mayor en el metagenoma de puerto respecto al metagenoma control fue la familia *Rhodobacteraceae* (figura 3.3A-B y tabla S3.3). La mayor proporción de secuencias asignadas a gammaproteobacterias en el metagenoma de puerto se debió, básicamente, al mayor porcentaje de secuencias asignadas a los grupos OM60 y BD1-7 y a las familias *Alteromonadaceae*, *Oceanospirillaceae*, *Pseudoalteromonadaceae*, *Pseudomonadaceae*, *Shewanellaceae* y *Vibrionaceae* (figura 3.3A-B, tabla S3.3). Respecto al mayor porcentaje de lecturas asignadas al filo *Bacteroidetes* en el metagenoma de puerto, este se debía al mayor porcentaje de lecturas asignadas a la familia *Flavobacteriaceae* (figura 3.3A-B, tabla S3.3). Estos resultados están de acuerdo con los obtenidos por Nogales y colaboradores (2007) y Aguiló-Ferretjans y colaboradores (2008) en muestras tomadas anteriormente en estas localizaciones.

En cuanto a los experimentos de microcosmos, se observaron cambios en la composición de las comunidades bacterianas tras la exposición a diésel y nitrógeno. Sin embargo, los cambios no eran tan evidentes como las diferencias observadas entre el metagenoma control y el metagenoma de puerto (figura 3.3). En el experimento realizado en 2013, tras 42 h de incubación en presencia de diésel y nitrógeno se observó una mayor proporción de lecturas asignadas a SAR11 y a las familias *Rhodobacteraceae*, *Alteromonadaceae*, *Pseudoalteromonadaceae* y *Flavobacteriaceae* y una menor proporción de lecturas asignadas a alfaproteobacterias no clasificadas y al grupo OM60 respecto a T2C1 2013 (figura 3.3C-D). En cuanto a los experimentos de microcosmos tras 64 h y 67 h de incubación en presencia de diésel y nitrógeno, no se obtuvo el mismo resultado en 2013 que en 2014 (figura 3.3E-H). En el experimento de 2013 tras 67 h de incubación con diésel y nitrógeno se observó una mayor proporción de secuencias de las familias *Rhodobacteraceae*, *Alteromonadaceae*, *Pseudoalteromonadaceae* y *Flavobacteriaceae* y una menor proporción de lecturas asignadas a los grupos SAR11, SAR116 y OM60 y a las familias, *Burkholderiaceae* y *Pseudomonadaceae* respecto a T3C2 (figura 3.3E-F). En el experimento de 2014, tras 64 h de incubación con diésel y nitrógeno se observó un mayor porcentaje de lecturas asignadas a la familia *Flavobacteriaceae* y un menor porcentaje de lecturas asignadas a las familias de la clase *Alphaproteobacteria* respecto a T3C1 2014 (figura 3.3G-H, tabla S3.3). El menor porcentaje de secuencias asignadas a alfaproteobacterias se debió a una disminución de las lecturas asignadas a todas las familias y grupos detectados en el metagenoma T3C1 2014 a excepción

del grupo SAR11 (figura 3.3G-H, tabla S3.3). En cuanto a la clase *Gammaproteobacteria*, se mantuvo el porcentaje de secuencias asignadas a esta clase aunque se observaron diferencias en cuanto a la abundancia de varias familias (figura 3.3G-H, tabla S3.3). En concreto, se observó una menor proporción de lecturas asignadas al grupo OM60 y a la familia *Oceanospirillaceae* y una mayor proporción de las lecturas asignadas a las familias *Alteromonadaceae* y *Pseudoalteromonadaceae* (figura 3.3G-H, tabla S3.3). No obstante, comparando los resultados de los dos experimentos se observa una respuesta común en microcosmos tratados: una mayor proporción de las lecturas asignadas a las familias *Flavobaceriaceae, Alteromonadaceae* y *Pseudoalteromonadaceae* y una menor proporción de las lecturas asignadas a SAR116. Esta era una respuesta esperable en base a los resultados de los experimentos previos similares a los realizados en esta tesis con microcosmos tratados con diésel de Lanfranconi (2010) y diésel y nitrógeno de Piña-Villalonga (2012). En cuanto a los resultados de los grupos SAR11 y de la familia *Rhodobacteraceae*, solo los resultados del experimento de 2013 están de acuerdo con los resultados previos (Lanfranconi 2010; Piña-Villalonga, 2012). No se esperaba detectar un mayor número de lecturas de SAR11 tras 64 h de incubación con diésel y nitrógeno respecto a su control (figura 3.3G-H), puesto que, SAR11 es un grupo de alfaproteobacterias oligotróficas, que disminuye en la zona de influencia de puerto (Nogales *et al*., 2007). Por lo tanto, es común que este grupo sea minoritario en aguas de puerto y en las aguas que simulen sus condiciones. Así como tampoco se esperaba una menor proporción de lecturas de la familia *Rhodobacteraceae* tras 64 h de incubación con diésel y nitrógeno respecto a su control (figura 3.3G-H), puesto que, los miembros de la familia *Rhodobacteraceae*, junto a las gammaproteobacterias y bacteroidetes, son los organismos predominantes en aguas de puerto (Nogales *et al*., 2007). Por ello la mayor detección de la familia *Rhodobacteraceae* en las aguas contaminadas (puerto y microcosmos contaminados de 2013) estaban de acuerdo con los resultados esperados. Analizando en detalle el resto de resultados de los metagenomas de 2013 y 2014, se detectaron varias discrepancias, tanto entre los resultados de los diferentes microcosmos, como entre los resultados de los microcosmos y la bibliografía. En primer lugar, se observó que la abundancia relativa de SAR86 era muy baja y prácticamente igual entre los microcosmos control y contaminados (figura 3.3C-H, tabla S3.3). El grupo SAR86 es un grupo de bacterias oligotróficas importante para diferenciar aguas controles de aguas contaminadas con hidrocarburos, puesto que se había mostrado que su abundancia relativa disminuía en presencia de hidrocarburos (Nogales *et al*., 2007; Lanfranconi *et al*., 2010), tal y como se observa en el metagenoma de puerto (figura 3.3A-B, tabla S3.3). También se observó una menor abundancia relativa del grupo OM60 en los tres metagenomas de microcosmos con diésel y nitrógeno respecto a sus controles (figura 3.3C-3.3H, tabla S3.3). Anteriormente se había observado que el grupo OM60 tenía una mayor abundancia relativa en aguas contaminadas con hidrocarburos (Nogales *et al*., 2007; Lanfranconi *et al*., 2010). Respecto a la menor abundancia de lecturas asignadas a la familia *Oceanospirillaceae* en T3D1 2014 respecto a T3C1 2014, este resultado no se había detectado en ninguno de los otros experimentos de microcosmos y entra en contradicción con los resultados observados en el metagenoma de puerto (figura 3.3A-B y G-H, tabla S3.3). En estudios de metagenómica de aguas contaminadas con vertidos de petróleo se observó una mayor proporción de lecturas asignadas a miembros de la familia *Oceanospirillaceae* en presencia de hidrocarburos (Dubinsky *et al*., 2013), tal y como se observa en el metagenoma de puerto (figura 3.3A-B, tabla S3.3). Sin embargo en los experimentos de microcosmos de

Lanfranconi y colaboradores (2010), se observó que no todas las poblaciones de *Oceanospirillaceae* aumentaban tras la exposición a hidrocarburos en experimentos de microcosmos suplementados con diésel. En concreto, se observó una mayor abundancia de *Oceanospirillaceae* en los microcosmos suplementados con diésel generados a partir de agua de mar muestreada en verano, mientras que no se observó este aumento en los microcosmos que se generaron a partir de agua recogida a final de invierno (Lanfranconi *et al*., 2010). El agua con la que se generaron los microcosmos T3C1 2014 y T3D1 2014 se recogió en abril de 2014. Por lo tanto, se puede hipotetizar que esta sería más parecida al agua de invierno que no a la de verano, y por ello sería esperable que no se observara un aumento de *Oceanospirillaceae* en presencia de diésel y nutrientes. Finalmente, cabe destacar que solo se detectó la presencia de cianobacterias derivadas del orden *Chroococcales*, básicamente del género *Synechococcus*, en el experimento de microcosmos de 2014 (figura 3.3G-H, tabla S3.3) y en el metagenoma control (figura 3.3A, tabla S3.3). Por lo tanto, T3D1 2014 era el único metagenoma contaminado con hidrocarburos en el que se detectaban estos microorganismos, y además, en un número de lecturas similar a T3C1 2014. En el estudio de Lanfranconi y colaboradores (2010) se observó que en los microcosmos generados a partir de aguas de verano, las cianobacterias de tipo *Synechococcus* desaparecían tras la exposición a diésel, mientras que en los microcosmos generados a partir de agua de invierno, no se podían observar cambios claros tras la exposición a diésel y nitrógeno (Lanfranconi *et al.*, 2010). Por lo tanto, al igual que ocurría en el caso del estudio de las *Oceanospirillaceae* se puede hipotetizar que el agua recogida en abril de 2014 sería más parecida al agua de invierno del experimento previo que no a la de verano, y por ello sería posible que no se observara una menor abundancia de *Synechococcus* en presencia de diésel y nutrientes.

En vista de los cambios observados en la abundancia de lecturas asignadas a la familia *Rhodobacteraceae* tras la exposición a diésel y los objetivos particulares de esta tesis doctoral, se estudió con detalle si se producían cambios a nivel de géneros o especies pertenecientes a esta familia en las condiciones estudiadas. En concreto se analizó la abundancia de secuencias pertenecientes al linaje Roseobacter, teniendo en cuenta aquellas especies o grupos a las que al menos se hubieran asignado un 0,1 % de las lecturas totales de algún metagenoma. El metagenoma en el que se observó una mayor abundancia de lecturas asignadas a especies del linaje Roseobacter fue el metagenoma de puerto (tabla 3.4), como era esperable por los datos previos sobre composición de comunidades bacterianas en estos ambientes (Nogales *et al*., 2007; Aguiló-Ferretjans *et al*., 2008, Piña-Villalonga, 2012). En el metagenoma de agua de puerto se observó que las secuencias de este linaje se asignaban a un mayor número de géneros que en el resto de metagenomas (tabla 3.4). Este resultado se corresponde con lo esperado puesto que en la tesis de la doctora Piña-Villalonga (2012) se había observado una mayor diversidad de filotipos de Roseobacter en aguas de puerto respecto a aguas controles. En el metagenoma control las abundancias relativas superiores correspondieron a secuencias afiliadas con *Ruegeria* sp. TM1040, *R. pomeroyi* DSS-3^T, *Roseobacter denitrificans* Och114^T, *Dinoroseobacter shibae* DFL-12^T y la cepa HTCC2255. En ningún caso las lecturas correspondientes a estas bacterias alcanzaban abundancias del 1 % respecto al total de secuencias (porcentajes entre 0,37 y 0,69). Las secuencias afiliadas con todas estas poblaciones, excepto la cepa HTCC2255, tenían también abundancias relativas altas en el

Tabla 3.4. Abundancia relativa de las lecturas asignadas a bacterias del linaje Roseobacter en los metagenomas analizados en este estudio. Se muestran valores de porcentaje sobre el total de secuencias afiliadas taxonómicamente de cada metagenoma. Los valores negativos indican que no se asignaron lecturas a bacterias del linaje Roseobacter en el metagenoma.

Grupo genómico	Genoma de referencia	Ambientales		Microcosmos 2013				Microcosmos 2014	
		Control	Puerto	T2C1	T2D1	T3C2	T3D2	T3C1	T3D1
G1	*Roseobacter* sp. MED193	0,17	0,33	0,19	–	0,18	–	0,21	–
	Roseobacter sp. SK209-2-6	0,08	0,11	–	0,07	0,06	0,07	0,08	0,06
	Rhodobacterales bacterium Y4I	–	0,12	0,06	0,06	0,06	0,07	0,07	0,06
	Phaeobacter inhibens 2.10	0,08	0,15	0,07	0,08	0,08	0,09	0,09	0,07
	Nautella italica R11	0,08	0,10	0,06	0,05	0,05	0,06	0,06	0,05
	Ruegeria sp. TM1040	0,58	1,08	0,50	0,55	0,50	0,76	0,48	0,36
	Ruegeria sp. KLH11	0,09	0,14	–	–	0,07	0,08	0,08	0,06
	Ruegeria lacuscaerulensis ITI-1157^T	0,10	0,17	0,08	0,08	0,08	–	0,10	0,08
	Ruegeria pomeroyi DSS-3^T	0,63	1,69	0,78	0,79	0,65	1,05	0,84	0,79
G2	*Sulfitobacter* sp. EE-36	0,11	0,24	0,09	0,10	0,09	0,17	–	–
	Sulfitobacter sp. NAS-14.1	0,07	0,48	–	0,11	–	–	0,11	0,15
	Oceanibulbus indolifex HEL-45^T	0,06	0,33	0,08	0,09	0,08	0,11	0,09	0,07
	Roseobacter sp. GAI101	0,15	0,51	0,19	0,19	0,18	0,24	0,20	0,16
	Roseobacter denitrificans Och114^T	0,69	1,84	0,87	0,77	0,72	1,28	1,07	0,84
	Roseobacter litoralis Och149^T	0,08	0,31	0,10	0,10	0,10	0,15	0,11	0,09
	Rhodobacteraceae bacterium HTCC2083	0,15	0,24	0,15	0,14	0,14	0,16	0,17	0,13
G3	*Pelagibaca bermudensis* HTCC2601^T	0,08	0,13	0,08	0,07	0,07	0,09	0,09	0,07
	Sagittula stellata E-37^T	0,11	0,11	0,09	0,09	0,08	0,09	0,11	0,08
	Roseovarius sp. 217	–	0,43	–	0,27	0,18	0,23	0,22	0,27
	Roseovarius sp. TM1035	0,06	0,10	–	0,06	0,06	0,07	0,07	0,06
	Roseovarius nubinhibens ISM^T	0,24	0,30	0,19	0,22	0,21	0,34	0,26	0,19
	Roseobacter sp. Azwk-3b	0,11	0,19	0,11	0,11	0,11	0,13	0,13	0,10
	Pseudooceanicola batsensis HTCC2597^T	0,16	0,24	0,20	0,20	0,14	0,24	0,18	0,19
G4	*Loktanella vestfoldensis* SKA53	0,23	1,65	0,52	0,69	0,55	1,07	0,67	0,93
	Roseobacter sp. CCS2	0,09	1,64	0,58	0,65	0,66	1,35	0,85	0,95
	Octadecabacter antarcticus 307^T	0,19	0,36	0,28	0,33	0,33	0,58	0,27	0,23
	Thalassobium sp. R2A62	0,10	0,24	–	0,12	0,12	0,16	0,14	0,12
	Ketogulonicigenium vulgare WSH-001	0,08	0,18	0,10	0,07	0,11	0,17	0,14	0,12
G5	*Dinoroseobacter shibae* DFL-12^T	0,37	0,66	0,40	0,45	0,37	0,53	0,55	0,37
	Rhodobacteraceae bacterium HTCC2150	0,15	0,18	0,12	0,12	0,11	0,12	0,15	0,10
	Rhodobacterales bacterium HTCC2255	0,40	0,25	3,12	–	0,22	0,33	1,00	0,80
	Maritimibacter alkaliphilus HTCC2654^T	0,21	0,19	0,16	0,20	0,14	0,18	0,24	0,13

metagenoma del puerto (la mayoría de ellas por encima del 1 % respecto al total, y por tanto superiores a las observadas en el metagenoma control). Adicionalmente, en el metagenoma del puerto se observaron secuencias afiliadas a *Roseobacter* sp. CCS2 y *Loktanella vestfoldensis* SKA53 con abundancias relativas superiores (1,65 %). En el metagenoma del puerto se observaron también algunas poblaciones con abundancias relativas menores (entre 0,33 y 0,51 %) pero que o bien no estaban representadas en el metagenoma control o su abundancia relativa era considerablemente inferior. Era el caso de secuencias afiliadas con *Roseobacter* sp. MED193, *Sulfitobacter* sp. NAS-14.1, *Roseobacter* GAI101, *Oceanibulvus indolifex* HEL-45^T, y *Roseovarius* sp. 217. Si se relacionan estos resultados con los resultados de rutas catabólicas predichas en los genomas de estas bacterias (ver capítulo 1) se puede postular el potencial catabólico de roseobacters en las muestras control y de puerto. Esto no deja de ser una aproximación porque microorganismos relacionados filogenéticamente no tienen por qué compartir el mismo tipo de genes catabólicos (ver por ejemplo figura 1.25). Si se comparan los compuestos aromáticos que en teoría podrían ser catabolizados por estas bacterias (ver figura 1.25) se observa un amplio potencial para la degradación de compuestos monoaromáticos en ambas muestras. La diferencia entre ambas muestras estaría en la mayor proporción en el metagenoma del puerto de secuencias de bacterias potencialmente degradadoras de benzoato, catecol y ferulato.

Por lo que respecta a los resultados de los metagenomas de los experimentos de microcosmos se observaron cambios en la abundancia relativa de secuencias afiliadas con distintas especies de Roseobacter tras el tratamiento. Las diferencias fueron más evidentes en los microcosmos preparados en 2013. En estos microcosmos se observó también una cierta dinámica en las poblaciones al comparar los resultados de T2 (42 h) y T3 (67 h). En los microcosmos tratados con diésel y nitrógeno se observó un incremento en abundancia relativa (en torno a o superando el 1 % de abundancia al final de la incubación) de secuencias afiliadas con *Roseobacter* sp. CCS2 y *Loktanella vestfoldensis* SKA53 (tabla 3.4). En los experimentos de microcosmos de 2013 se observó también un incremento en secuencias afiliadas con *Ruegeria pomeroyi* DSS-3^T y *Roseobacter denitrificans* Och114^T (tabla 3.4) e incrementos menores de secuencias afiliadas con *Ruegeria* sp. TM1040, *Roseovarius* sp. 217, *Roseovarius nubinhibens* ISM^T, y *Octadecabacter antarcticus* 307^T (tabla 3.4). Un resultado común en los experimentos de microcosmos de 2013 y 2014 tratados con diésel y nitrógeno fue la disminución de la abundancia relativa de secuencias afiliadas con *Roseobacter* sp. MED193 y la cepa HTCC2255 (tabla 3.4). Por lo tanto, en general en los microcosmos tratados con diésel y nitrógeno se observaba un aumento en la abundancia relativa de secuencias afiliadas a los cinco roseobacters que presentaban una mayor abundancia relativa en el metagenoma de puerto (tabla 3.4). Como se puede observar en la tabla 3.4 las únicas especies del linaje Roseobacter con una mayor abundancia de lecturas asignadas siempre a aguas contaminadas respecto a sus controles fueron *Roseovarius* sp. 217, *Roseobacter* sp. CCS2 y *Loktanella vestfoldensis* SKA53.

2.3. Caracterización funcional de los metagenomas

Para la caracterización funcional de los metagenomas se siguieron varias estrategias (ver apartado 8 de material y métodos para el detalle). En primer lugar se utilizó la herramienta MG-RAST y diferentes métodos de análisis. Los mejores resultados de asignación funcional (mayor porcentaje de secuencias asignadas a funciones en todos los metagenomas) se

obtuvieron con la base de datos de Subsistemas SEED (Overbeek *et al*., 2005; 2014) y por lo tanto son los que se muestran en este capítulo. Como se puede observar en la tabla 3.5, las categorías funcionales a las que se asignaron un mayor porcentaje de lecturas en los ocho metagenomas fueron: metabolismo de proteínas (8,9–9 %), de carbohidratos (8,5–9 %), de aminoácidos y derivados (8,3–8,8 %) y de cofactores, vitaminas, grupos prostéticos y pigmentos (7–7,2 %). Entre un 24,6 y un 25,7 % de las lecturas se asignaron a agrupaciones funcionales pero de función desconocida (categoría subsistemas basados en agrupaciones, 15,8–16,3 %) o a proteínas que pueden catalogarse en más de una función y por ello estarían en más de una categoría (categoría diversa, 8,8–9,4 %), resultado que impedía realizar una asignación a su función concreta. Estas seis categorías funcionales contienen la mayor parte de las proteínas celulares, puesto que engloban las proteínas del metabolismo celular. Por ello, es coherente que se encuentren entre las categorías funcionales dominantes, tal y como ya se ha documentado en otros estudios de metagenómica de comunidades de aguas marinas (Dinsdale *et al*., 2008; Gilbert *et al*., 2008).

Como se puede observar en la tabla 3.5 los valores de abundancias relativas asignados a las diferentes categorías funcionales fueron similares en todos los metagenomas. Por lo que respecta a los metagenomas de muestras ambientales se detectaron diferencias de al menos un 0,1 % de abundancia relativa entre el metagenoma control y el metagenoma de puerto en lecturas asignadas a algunas de las categorías funcionales. En concreto en el metagenoma ambiental control se observó una mayor abundancia relativa de lecturas asignadas a las categorías: metabolismo de los ácidos nucleicos (nucleósidos y nucleótidos); pared celular y cápsula; división y ciclo celular; respuesta a estrés, y metabolismo del carbono, nitrógeno, fósforo, cofactores, vitaminas, grupos prostéticos y pigmentos. En cambio, en el metagenoma de puerto se observaba una mayor abundancia relativa de lecturas asignadas a las categorías funcionales de metabolismo lipídico (ácidos grasos, lípidos e isoprenoides); captación y metabolismo del hierro; metabolismo de compuestos aromáticos; respiración; metabolismo del ARN; transporte de membrana; movilidad y quimiotaxis; fagos, profagos, elementos transponibles y plásmidos, y en la categoría de virulencia, enfermedad y defensa. Este resultado es coherente con el obtenido por Kisand y colaboradores (2012) al comparar metagenomas generados a partir de las aguas del puerto de Génova con metagenomas de agua control. Kisand y colaboradores (2012) observaron que en las aguas del puerto había un enriquecimiento de lecturas asignadas a funciones tradicionalmente ligadas a los organismos copiotróficos, que son los organismos que normalmente se encuentran adaptados a las condiciones de puerto. Entre los organismos copiotróficos se encuentran los miembros del linaje Roseobacter, gammaproteobacterias como las de los géneros *Alteromonadaceae*, *Pseudoalteromonadaceae* y bacteroidetes, todos ellos detectados en el metagenoma de puerto de esta tesis (figura 3.3, tabla S3.3). En concreto, las funciones que estos autores detectaron en mayor abundancia en el puerto estaban ligadas a la transducción de señales, para dar respuesta celular a los estímulos del medio, a bombas de expulsión y sistemas de transporte de cationes y metales pesados, como respuesta a la necesidad de protección a compuestos tóxicos presentes en las aguas de puerto (Kisand *et al*., 2012). Además, en el agua de puerto de Génova también se detectó una mayor abundancia de lecturas asignadas a motilidad celular, tráfico intracelular, secreción, integrasas y transposasas y a varias

Tabla 3.5. Abundancia relativa de las lecturas asignadas a diferentes categorías funcionales. Se muestran valores de porcentaje sobre el total de secuencias afiliadas funcionalmente de cada metagenoma.

	Ambientales		Microcosmos 2013				Microcosmos 2014	
Categoría (Subsistemas SEED)	**Control**	**Puerto**	**T2C1**	**T2D1**	**T3C2**	**T3D2**	**T3C1**	**T3D1**
Aminoácidos y derivados	8,81	8,32	8,78	8,88	8,88	8,71	8,75	8,57
Metabolismo de proteínas	9,05	9,00	9,01	8,92	8,92	8,96	8,93	8,99
Carbohidratos	8,87	8,53	8,98	8,98	8,98	8,92	8,90	8,73
Nucleósidos y nucleótidos	3,49	3,20	3,32	3,35	3,35	3,21	3,30	3,31
Ácidos grasos, lípidos e isoprenoides	2,94	3,04	3,13	3,12	3,12	3,15	3,13	3,13
Metabolismo del nitrógeno	1,14	1,03	1,08	1,13	1,13	1,06	1,13	1,09
Metabolismo del fósforo	1,11	0,85	0,97	0,95	0,95	0,93	1,02	0,94
Metabolismo del potasio	0,23	0,31	0,26	0,26	0,26	0,29	0,25	0,27
Metabolismo del azufre	0,99	0,99	0,96	0,96	0,96	0,96	0,99	0,94
Cofactores, vitaminas, grupos prostéticos y pigmentos	7,17	7,05	7,18	7,20	7,20	7,13	7,12	7,11
Metabolismo de compuestos aromáticos	1,04	1,54	1,42	1,45	1,45	1,48	1,38	1,44
Captación y metabolismo del hierro	0,58	0,82	0,75	0,72	0,72	0,84	0,74	0,82
Transporte de membrana	2,75	2,97	2,87	2,85	2,85	3,00	2,90	2,93
Metabolismo del ADN	4,11	4,06	4,04	3,98	3,98	4,06	4,02	4,12
Metabolismo del ARN	4,46	4,91	4,54	4,44	4,44	4,65	4,50	4,59
Regulación y señalización celular	1,17	1,25	1,23	1,20	1,20	1,25	1,18	1,19
División y ciclo celular	1,76	1,51	1,60	1,62	1,62	1,56	1,66	1,64
Pared celular y cápsula	4,01	3,78	3,82	3,82	3,82	3,78	3,92	3,92
Latencia y esporulación	0,15	0,20	0,17	0,15	0,15	0,18	0,15	0,17
Movilidad y quimiotaxis	0,45	0,68	0,56	0,57	0,57	0,58	0,59	0,64
Fotosíntesis	0,39	0,43	0,31	0,33	0,33	0,31	0,31	0,29
Respiración	3,22	3,40	3,25	3,31	3,31	3,20	3,31	3,16
Metabolismo secundario	0,48	0,43	0,49	0,51	0,51	0,48	0,52	0,49
Respuesta a estrés	2,77	2,63	2,72	2,78	2,78	2,69	2,80	2,69
Virulencia, enfermedad y defensa	1,81	2,75	2,20	2,18	2,18	2,37	2,13	2,27
Fagos, profagos, elementos transponibles y plásmidos	1,33	1,73	1,45	1,49	1,49	1,49	1,34	1,50
Subsistemas basados en agrupaciones	16,33	15,87	15,97	15,85	15,85	15,95	15,96	16,12
Diversa	9,38	8,74	8,92	9,00	9,00	8,80	9,07	8,94

vías de degradación de compuestos aromáticos, todas ellas, funciones relacionadas con la estrategia de de los organismos copiotróficos, que les ayudan a mejorar el acceso a los nutrientes y facilitan su adaptación a los estímulos antropogénicos presentes en los puertos (Kisand *et al*., 2012).

Respecto a los cambios en la abundancia de lecturas asignadas a las diferentes categorías funcionales en los experimentos de microcosmos no se observaron cambios notables tras los tratamientos. En el experimento de microcosmos de 2013 tras 42 h de incubación no se pudieron apreciar diferencias mayores al 0,1 % en la abundancia relativa de las lecturas asignadas a ninguna de las categorías funcionales entre el microcosmos control (T2C1 2013) y el microcosmos tratado con diésel y nitrógeno (T2D1 2013) (tabla 3.5). La similitud entre la asignación funcional de T2C1 2013 y T2D1 2013 estaría de acuerdo con los escasos cambios a nivel taxonómico observados entre estos dos metagenomas (figura 3.3C-D). Estos resultados sugerían que 42 h de incubación con diésel y nitrógeno no eran suficientes para provocar cambios evidentes a nivel taxonómico ni a nivel funcional respecto al control. Tras 67 h de incubación de estos microcosmos se observó una abundancia relativa superior (diferencia de al menos 0,1 %) de las lecturas asignadas a las categorías de aminoácidos y derivados y nucleósidos y nucleótidos en el microcosmos control (T3C2 2013) respecto al microcosmos tratado con diésel y nitrógeno (T3D2 2013). En el microcosmos T3D2 2013 se observaba una mayor abundancia relativa de lecturas asignadas a captación y metabolismo del hierro; metabolismo del ARN, transporte de membrana; y virulencia, enfermedad y defensa respecto a T3C2 2013 (tabla 3.5). Algunas de estas categorías también tenían abundancias relativas superiores en el metagenoma de puerto en comparación con el metagenoma ambiental control (tabla 3.5). Respecto al experimento de microcosmos de 2014 tras 64 h de incubación, también se observaron cambios a nivel funcional entre los metagenomas generados a partir del microcosmos control (T3C1 2014) y del microcosmos tratado con diésel y nitrógeno (T3D1 2014), aunque estos cambios se daban en un menor número de categorías funcionales y eran cambios menores (tabla 3.5). En concreto, tras 64 h de incubación se observó una abundancia relativa superior (diferencia de al menos un 0,1 %) de lecturas asignadas a las categorías de aminoácidos y derivados; carbohidratos; respiración, y respuesta a estrés en T3C1 2014 respecto a T3D1 2014. En el metagenoma T3D1 2014 se observó una abundancia relativa superior de lecturas asignadas a las categorías metabolismo del ADN; fagos, profagos, elementos transponibles y plásmidos, y virulencia, enfermedad y defensa respecto a T3C1 2014. En el metagenoma de los microcosmos tratados de 2014 se observaron también abundancias ligeramente superiores en las categorías: metabolismo de compuestos aromáticos; captación y metabolismo de hierro, y transporte de membrana (tabla 3.5).

En resumen, a pesar de haber detectado diferencias en la composición taxonómica de los metagenomas, sobre todo a nivel de los microcosmos ambientales, estas diferencias taxonómicas no iban acompañados de grandes cambios a nivel de las categorías funcionales globales, especialmente en los microcosmos. De forma que, en principio, el potencial, a nivel de grandes categorías funcionales, de las comunidades estudiadas sería similar, tal y como ya se había visto en otros estudios como el de Sunagawa y colaboradores (2015). En concreto, estos autores al comparar la variabilidad taxonómica y funcional de aguas oceánicas, veían que, a pesar de observar una alta variabilidad taxonómica, esta iba acompañada de una distribución relativamente estable de las categorías funcionales (Sunagawa *et al*., 2015). Por lo

tanto, la redundancia funcional observada en las aguas oceánicas podía ser un indicativo de que la variabilidad funcional es independiente de la variabilidad taxonómica (Sunagawa *et al*., 2015). Además, Bell y colaboradores (2005) sugirieron que la redundancia funcional en las comunidades microbianas podía conferir una capacidad de tamponamiento para los ecosistemas en caso de pérdida de biodiversidad. Volviendo a los resultados de este trabajo, la aproximación metagenómica para comparar comunidades ambientales de aguas limpias y agua de puerto fue útil porque se detectaron diferencias en varias categorías funcionales (tabla 3.5). En cambio esta aproximación no resultó tan informativa para estudiar el efecto del diésel y el nitrógeno en la funcionalidad de las comunidades microbianas en microcosmos con cortos tiempos de incubación. En este caso, se podría sugerir una aproximación metatranscriptómica basada en el estudio del ARN (Moran *et al*., 2013), una aproximación metaproteómica basada en el estudio de las proteínas (Zarraonaindia *et al*., 2013) o una aproximación metabolómica (Bargiela *et al*., 2015) para completar el estudio de los cambios taxonómicos y funcionales que se producen en respuesta a cambios en las condiciones ambientales.

Para obtener una visión más detallada de las diferencias entre los metagenomas se analizaron las proteínas de las categorías funcionales mostradas en la tabla 3.5. En particular, se revisaron las proteínas para las que se observó una mayor abundancia relativa de lecturas en el metagenoma del puerto respecto al metagenoma control y en los metagenomas de microcosmos tratados con diésel y nitrógeno respecto a sus respectivos controles. En el sistema de clasificación de lecturas en categorías funcionales que realiza MG-RAST una misma proteína de referencia puede ser clasificada en diferentes categorías funcionales (p.e. EC 2.3.1.16 tabla S3.5 y S3.9) por lo que en algunos casos tan solo pudieron proponerse funciones generales. Esto ocurría principalmente con proteínas del metabolismo celular central. Por lo que respecta a funciones metabólicas, en primer lugar se observaron diferencias entre los metagenomas control y los contaminados (puerto y microcosmos tratados con diésel) en la abundancia relativa de lecturas asociadas al metabolismo de ácidos grasos, al metabolismo de polihidroxialcanoatos, a la síntesis de biotina y a la síntesis síntesis de cofactores y vitaminas (tetrapirroles, riboflavina) (tablas S3.4 y S3.5). Las funciones enzimáticas asociadas fueron las de acil transferasas (EC 2.3.1.16, EC 2.3.1.41), isomerasas (EC 5.1.2.3, EC 5.3.3.8) y oxidoreductasas (EC 1.3.1.34, EC 1.3.99.2). Otras funciones del metabolismo central para las que la abundancia relativa era mayor en los metagenomas del puerto y microcosmos contaminados respecto a sus controles eran el ciclo del glioxilato y rutas anapleróticas (fosfoenolpiruvato carboxiquinasa) (tabla S3.6). Se detectaron también abundancias relativas superiores de lecturas asignadas a aminoacil-tRNA sintetasas asociadas a la síntesis de algunos aminoácidos como la lisina (tabla S3.7). Además, también se detectaron cambios en el metabolismo de la cisteína, glicina, serina, isoleucina y metionina en los metagenomas del puerto y microcosmos contaminados respecto a sus controles, pero no en todos los casos el porcentaje de abundancia relativa de lecturas asignadas a estas categorías era superior en las muestras contaminadas (tabla S3.8).

En cuanto a la replicación del ADN y la división celular (tablas S3.9 y S3.10) en los metagenomas contaminados respecto a sus controles se observaron abundancias relativas superiores de lecturas asignadas a DNA polimerasas, topoisomerasas, proteínas implicadas en la partición del cromosoma y de plásmidos, sistemas de reparación de ADN y proteínas de

formación del septo. En el estudio de Kisand y colaboradores (2012) también se había descito una mayor abundancia relativa de de las proteínas implicadas en la replicación del ADN en el puerto de Génova. Por lo tanto, estos resultados sugerían que en las muestras contaminadas con hidrocarburos (puerto y microcosmos contaminados con diésel y nitrógeno) podría darse un aumento de la división celular y por ello, también se observaría un aumento de las lecturas asignadas a la replicación del ADN. Este resultado estaría de acuerdo con el aumento, aunque no significativo, del número de procariotas totales observado en los microcosmos contaminados respecto a los controles (ver figura 3.1). En cuanto a la mayor abundancia relativa de lecturas asignadas a proteínas implicadas en la reparación del ADN (tabla S3.9), este podría relacionarse con el estrés causado por la exposición a hidrocarburos (Ramos *et al*., 2015).

Por otra parte, se observaron algunas diferencias entre los metagenomas en cuanto a sistemas de transporte. En los cuatro metagenomas de muestras contaminadas con hidrocarburos (puerto y microcosmos con diésel y nitrógeno) se observó una mayor abundancia relativa de lecturas asignadas a un receptor dependiente de TonB que formaba parte de un transportador de hierro (tabla S3.11) y de otro transportador de membrana dependiente de TonB (tabla S3.12). Los transportadores dependientes de TonB (TBDTs) son transportadores de la membrana externa que unen y transportan sideróforos, vitamina B12, complejos de níquel y carbohidratos (Noinaj *et al*., 2010). Según Tang y colaboradores (2012) hay una prevalencia alta de secuencias asignadas a TBDTs en bacterias de los órdenes *Sphingomonadales*, *Alteromonadales*, en *Bacteroidetes* y en el grupo OMG. En nuestro estudio se detectó un mayor número de lecturas asignadas a las familias *Alteromonadaceae* y *Flavobacteriaceae* en los metagenomas contaminados con hidrocarburos (puerto y microcosmos con diésel y nitrógeno), así como un mayor número de lecturas asignadas a los grupos BD1-7 y OM60, ambos OMG, en el metagenoma de puerto (figura 3.3, tabla S3.3). Por lo tanto, se podría relacionar la mayor abundancia relativa de lecturas asignadas a TBDTs en puerto y microcosmos con diésel y nitrógeno, con el mayor número de lecturas asignadas a estas familias y grupos taxonómicos, siendo nuestro resultado coherente con el Tang y colaboradores (2012).

Además, se pudieron observar cambios entre los metagenomas contaminados (puerto y microcosmos con diésel y nitrógeno) y sus respectivos controles a nivel de lecturas asignadas otros tipos de tranportadores (tablas S3.11 y S3.12). En concreto, se observó una menor abundancia relativa de los transportadores de tipo ABC, de los transportadores de tipo Tct (*tricarboxylate transport protein*) y, en menor medida, de las proteínas de la vía de secreción Sec en los metagenomas de muestras contaminadas (puerto y microcosmos con diésel y nitrógeno) (tablas S3.11 y S3.12). Ahora bien, también se detectaron algunas excepciones a estos resultados, como el caso puntual de un transportador ABC de oligopéptidos (TC 3.A.1.5.1) en T3D1 2014 (tabla S3.12) o de la proteína SecF en los metagenomas contaminados (puerto y microcosmos con diésel y nitrógeno) (tabla S3.12). Por lo tanto, en principio se observaba una menor abundancia relativa de las lecturas asignadas a los transportadores de membrana de tipo ABC y Tct y a los sistemas de secreción tras la exposición a hidrocarburos. Una menor detección de los transportadores de tipo ABC también había sido descrita por Kisand y colaboradores (2012) en metagenomas del puerto de Génova. En concreto, en las aguas del puerto observaron una mayor abundancia de transportadores de tipo RND y una

menor abundancia de los transportadores de tipo ABC respecto a los metagenomas de agua control (Kisand *et al*., 2012). La presencia de los transportadores de tipo ABC en una abundancia relativa superior en los controles que en los ambientes contaminados con hidrocarburos podría ser debida a la escasez de nutrientes en las aguas controles oligotróficas, lo que provoca que la captación de nutrientes deba optimizarse con la presencia de transportadores de tipo ABC (Kisand *et al*., 2012). Según Tang y colaboradores (2012), los transportadores ABC son abundantes en miembros del linaje Roseobacter, SAR11 y cianobacterias. Por lo tanto, la mayor presencia de transportadores ABC en los metagenomas control (control y microcosmos controles) se podría relacionar con la mayor presencia de SAR11 en estos metagenomas, y que en el caso del metagenoma control, también con la presencia de lecturas asociadas a cianobacterias (ver figura 3.3, tabla S3.3).

Finalmente, en los metagenomas de muestras contaminadas también se observaron proporciones más altas de lecturas asignadas a funciones de respuesta a estrés, como chaperonas (ClpA, ClpB, HtpG), o proteínas del metabolismo del glutatión (tablas S3.7 y S3.13) y a proteínas de fagos (integrasas, terminasas, recombinasas, y proteínas estructurales (tabla S3.14). La mayor abundancia relativa de las lecturas asignadas a elementos móviles en los metagenomas contaminados con hidrocarburos, sobre todo en el caso del metagenoma de puerto, sugería que las comunidades de ambientes contaminados presentarían un mayor potencial de intercambio genético que las comunidades de aguas control. Esto concuerda con los resultados de afiliación taxonómica de las lecturas, puesto que en las comunidades de agua control predominaba SAR11 (figura 3.3), microorganismo que presenta un genoma mínimo con un bajo número de genes parálogos, sin pseudogenes ni profagos (Grote *et al*., 2012). Además, estos resultados también son coherentes con los del estudio de Kisand y colaboradores (2012), en el que se observó una mayor abundancia relativa de lecturas asignadas a integrasas y transposasas en los metagenomas de aguas del puerto de Génova. Kisand y colaboradores (2012), relacionaron esta mayor abundancia de elementos móviles en aguas de puerto con una mayor facilidad de los organismos que viven en ellas para adaptarse a este medio, en el que se dan estímulos heterogéneos y variables derivados de la acción antropogénica en estos ambientes.

2.4. Caracterización del potencial catabólico de compuestos aromáticos en los metagenomas

Un interés particular en esta tesis era estudiar el potencial catabólico para compuestos aromáticos en los diferentes metagenomas generados y establecer relaciones con la presencia de microorganismos del linaje Roseobacter y/o el grado de contaminación del agua. Se pretendía no solo tener información sobre el tipo y abundancia relativa de lecturas asignadas a genes para degradación de compuestos aromáticos en los diferentes metagenomas, sino también utilizar ensamblajes de secuencias para conocer cómo se organizaban los genes. Por ejemplo, se quería ver si se organizaban o no en estructuras supragénicas como las descritas para los modelos de degradación *Acinetobacter calcoaceticus* ADP-1 (Harwood y Parales, 1996) y *Agrobacterium tumefaciens* (Parke, 1997) o en estructuras parecidas a las observadas en el linaje Roseobacter que se han mostrado en esta tesis (ver capítulo 1).

Con la asignación funcional realizada con MG-RAST (tablas 3.5 y S3.15) se pudieron evidenciar diferencias relativas en la abundancia de genes para degradación de compuestos aromáticos

entre los metagenomas de aguas control y los de aguas contaminadas (puerto y microcosmos con diésel). En concreto, se asociaron lecturas a las rutas del homogentisato, rama del catecol de la vía del beta-cetoadipato, naftaleno, gentisato, salicilato, quinato, benzoato y al metabolismo de fenilacanoatos y compuestos heterocíclicos (tabla S3.15). Pero no en todos los casos el porcentaje de abundancia relativa de lecturas asignadas a estas categorías era superior en las muestras contaminadas que en los controles (tabla S3.15). Sin embargo, los resultados parecían indicar que en los metagenomas de muestras contaminadas había una mayor proporción de lecturas asignadas a compuestos aromáticos complejos (tabla S3.15). En concreto, se detectó una mayor abundancia relativa de las lecturas asignadas a enzimas implicadas en la degradación de fenilalcanoatos (EC 5.1.2.3 y EC 5.3.3.8) en los cuatro metagenomas de muestras contaminadas, y de compuestos aromáticos heterocíclicos (EC 1.3.99.16) en tres de ellos (tabla S3.15). En el estudio de Kisand y colaboradores (2012) también observaron la prevalencia de varias vías de degradación de compuestos aromáticos en las muestras de agua del puerto de Génova respecto al control de aguas limpias, siendo las vías de degradación del estireno y del benzoil-CoA las más importantes en ese caso. Además, se han comparado los resultados de esta tesis con los descritos sobre el vertido de la plataforma DWH observando, también en ellos, que la abundancia relativa de lecturas asignadas a las vías de degradación de compuestos aromáticos no siempre era superior en las muestras contaminadas que en los controles (Lu *et al*., 2012; Mason *et al*., 2012; Bargiela *et al*., 2015). En general, se observaba una mayor proporción de lecturas asignadas a la degradación de compuestos aromáticos en las muestras de la estela del vertido de petróleo respecto a las muestras de agua lejanas a ella, y por tanto, no contaminadas con el vertido (Lu *et al*., 2012; Bargiela *et al*., 2015). Sin embargo, en el estudio de Mason y colaboradores (2012) se observó una menor abundancia relativa de los genes involucrados en la degradación de compuestos aromáticos, como el etilbenceno, el tolueno y los hidrocarburos aromáticos policíclicos, en las muestras de la estela de contaminante respecto a las muestras no contaminadas (Mason *et al*., 2012). En cuanto al tipo de compuestos aromáticos cuyas vías de degradación fueron más prevalentes en los metagenomas de la estela del vertido de la DWH, se detectaron tanto compuestos monoaromáticos, por ejemplo catecol, como compuestos poliaromáticos tales como el naftaleno (Lu *et al*., 2012; Bargiela *et al*., 2015).

Para poder realizar un análisis más exhaustivo sobre el potencial catabólico de compuestos aromáticos, tanto de las aguas crónicamente contaminadas con hidrocarburos, como de las aguas de los microcosmos contaminados con diésel y nitrógeno se decidió realizar un análisis utilizando la base de datos AromaDeg (Duarte *et al*., 2014). Esta base de datos está formada por 3605 proteínas implicadas en la degradación aeróbica de compuestos aromáticos, ha sido manualmente curada y se ha construido en base a un enfoque filogenómico (Duarte *et al*., 2014). La base de datos AromaDeg utiliza un análisis filogenómico para definir diferentes familias de enzimas dentro de cada superfamilia y diferentes grupos filogenéticos de enzimas dentro de cada familia. En general la predicción funcional según el grado de identidad de secuencia no tiene en cuenta que la evolución no solo conserva la función, sino que también genera nuevas funciones en las que el mecanismo bioquímico básico puede ser conservado, mientras que el sustrato o la especificidad del ligando puede cambiar. Al incluir un análisis filogenómico de las proteínas implicadas en la degradación aeróbica de compuestos aromáticos, Aromadeg permite disminuir los errores en la predicción de funciones. Además,

mejora la clasificación funcional de las proteínas de la degradación de compuestos aromáticos. Las mayores limitaciones de Aromadeg son, por un lado, la falta de representación de las enzimas implicadas en la rotura intradiol del catecol, de benzoil-CoA, fenilacetil-CoA y homogentisato, cuya presencia se ha estudiado en esta tesis (capítulo 1). La falta de representación de estas enzimas es una limitación importante para poder correlacionar el potencial catabólico de los metagenomas generados con la presencia de miembros del linaje Roseobacter, cuyo potencial catabólico está formado en gran parte por este tipo de enzimas (ver figura 1.25). Por otra parte, AromaDeg presenta una baja representación de extradiol dioxigenasas de la superfamilia cupina (EXDO cupina) respecto al resto de familias incluidas en la base de datos (ver apartado 3 de la introducción). En concreto, AromaDeg solo se centra en el estudio de la familia de la gentisato 1,2-dioxigenasa obviando las otras familias que forman esta superfamilia (Duarte *et al*., 2014).

Tabla 3.6. Asignación de lecturas de los metagenomas según la base de datos AromaDeg.

Metagenoma	Nº lecturas R1		Abundancia relativa (%)
	Total	Clasificadas con AromaDeg	
Ambientales			
Control	4.099.975	2345	0,06
Puerto	2.929.080	1090	0,04
Microcosmos			
T2C1 2013	5.045.407	2141	0,04
T2D1 2013	4.090.200	1773	0,04
T3C2 2013	3.548.678	1357	0,04
T3D2 2013	4.138.318	1885	0,05
T3C1 2014	3.395.416	1925	0,06
T3D1 2014	3.191.035	1751	0,05

La asignación funcional con AromaDeg se realizó con los dos conjuntos de lecturas pareadas (R1 y R2) y con las lecturas ensambladas (*contigs*). Dado que los resultados con las lecturas pareadas eran similares se muestran solo los correspondientes a las lecturas R1. Tal y como se puede observar en la tabla 3.6, el número de lecturas de cada metagenoma asignado a la secuencias de proteínas de degradación de compuestos aromáticos por AromaDeg varió entre las 1000 y las 2000 lecturas, representando entre el 0,03—0,05 % de las lecturas totales de los metagenomas. Como era esperable, estos porcentajes de asignación fueron inferiores a los obtenidos con la base de datos SEED Subsystems en el análisis de MG-RAST (ver tabla 3.5) ya que la base de datos de AromaDeg es más restringida. La asignación funcional de los ensamblajes no mejoró la detección de enzimas para la degradación de compuestos aromáticos, ni permitió detectar posibles estructuras génicas. En concreto, el estudio de los *contigs* con los criterios de detección usados, no permitió detectar más de un enzima para la degradación de compuestos aromáticos en cada *contig*, ya que AromaDeg solo asignaba una función a cada *contig*. Como resultado, la detección de genes catabólicos con el estudio de los *contigs* fue menor que con el estudio de las lecturas. Por lo tanto, para poder tener una información más detallada sobre el potencial catabólico de compuestos aromáticos de los metagenomas se decidió analizar en detalle solo los resultados de la asignación funcional a partir de las lecturas, resultados que se exponen a continuación.

En todos los metagenomas se detectaron representantes de todas las superfamilias recogidas en la base de datos de AromaDeg: superfamilia de las oxigenasas Rieske con hierro no hemo

(Rieske-NH), superfamilia de las extradiol dioxigenasas del oxígeno quelado adyacente (EXDO I), extradiol dioxigenasas de la superfamilia cupina (EXDO cupina) y extradiol dioxigenasas de la superfamilia LigB (EXDO LigB) (ver apartado 2 de la introducción). Como se puede observar en la tabla 3.7 el 60 % de las lecturas anotadas codificaban para dioxigenasas Rieske-NH mientras que la categoría detectada en menor proporción fue la de la familia del gentisato de la superfamilia EXDO cupina.

A nivel de superfamilias se observó que en el metagenoma de puerto había una menor proporción de lecturas asignadas a Rieske-NH y EXDO I y una mayor proporción de lecturas asignadas a EXDO cupina y sobre todo a EXDO LigB respecto al metagenoma control (tabla 3.7). Respecto a los cambios observados en los microcosmos, estos eran menos evidentes que las diferencias entre el metagenoma de puerto y el agua control, tal y como ya se había observado en la asignación funcional con la base de datos de subsistemas SEED (tabla 3.5). En general se observó también una mayor proporción de lecturas asignadas a EXDO LigB en los metagenomas de los microcosmos tratados con diésel y nitrógeno respecto a sus controles. Además, se pudieron observar diferencias entre los metagenomas generados a las 42 h de incubación (T2) y los generados tras más de 60 h de incubación (T3) e incluso entre los metagenomas generados a T3 en 2013 y 2014 (tabla 3.7).

Tabla 3.7. Abundancia relativa de lecturas asignadas a las superfamilias de enzimas para degradación de compuestos aromáticos de la base de datos AromaDeg en los metagenomas. Se muestran valores de porcentaje sobre el total de secuencias asignadas funcionalmente por AromaDeg de cada metagenoma.

Metagenomas	Superfamilia			
	Rieske-NH	EXDO I	EXDO LigB	EXDO Cupina
Ambientales				
Control	60,98	30,96	5,88	2,17
Puerto	57,71	27,52	12,57	2,20
Microcosmos				
T2C1 2013	59,46	28,35	10,32	1,87
T2D1 2013	60,41	27,81	10,55	1,24
T3C2 2013	61,97	28,00	8,99	1,03
T3D2 2013	59,84	27,59	11,25	1,33
T3C1 2014	62,18	28,26	7,64	1,92
T3D1 2014	59,05	29,64	10,05	1,26

A continuación, se analizó con detalle a qué familias de proteínas dentro de cada subfamilia se asignaban las secuencias de los metagenomas. En los metagenomas ambientales se observaron diferencias a nivel de familias de forma más evidente que en los metagenomas de microcosmos tratados y controles (figura 3.4). En concreto, se observó una mayor abundancia relativa de la familia del bifenilo en el metagenoma del puerto en comparación con el metagenoma control de aguas limpias (figuras 3.4A-B). En el metagenoma de puerto también se observó una mayor abundancia relativa respecto al control de la familia I.1 de la superfamilia EXDO I, que incluye enzimas implicadas en la degradación de diversos compuestos (figuras 3.4A-B). Además, se observó que la mayor abundancia relativa de las lecturas asignadas a la superfamilia LigB en el metagenoma de puerto (tabla 3.7) se debía tanto al mayor número de lecturas asignadas a las familias del protocatecuato como del homoprotocatecuato (figuras 3.4A-B).

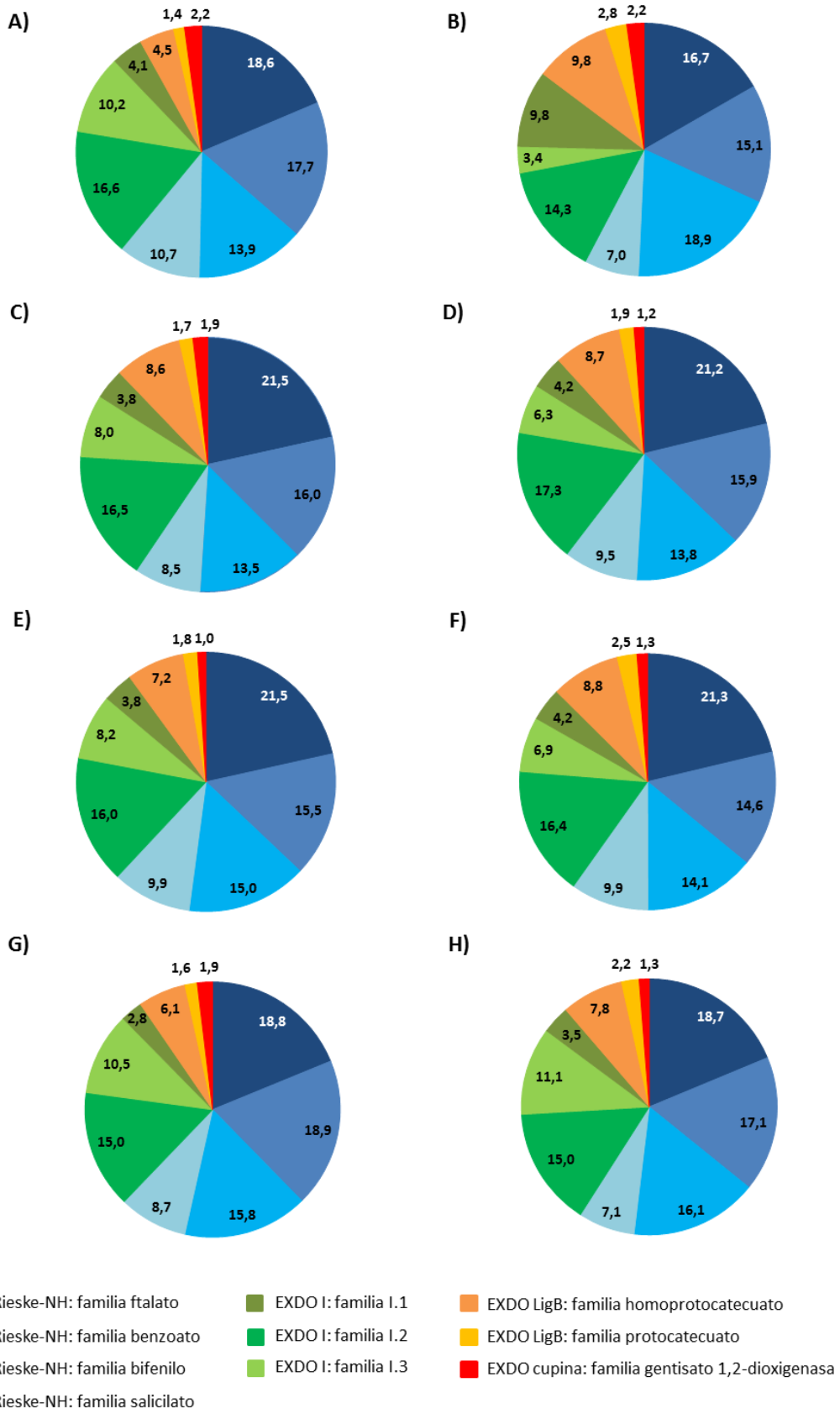

Figura 3.4. Asignación funcional de las lecturas de los metagenomas a las familias de enzimas para degradación de compuestos aromáticos de AromaDeg. Los números indican el porcentaje de abundancia relativa de lecturas asignadas a cada familia. Panel A, Control. Panel B, Puerto. Panel C, T2C1 2013. Panel D, T2D1 2013. Panel E, T3C2 2013. Panel F, T3D2 2013. Panel G, T3C1 2014. Panel H, T3D1 2014.

En cuanto a los metagenomas generados a partir de microcosmos, las diferencias observadas en las abundancias relativas de las diferentes familias eran poco evidentes tanto a T2 como a T3 (figuras 3.4C-3.4H). En el caso de los metagenomas a T3, se observó una mayor abundancia relativa de las lecturas asignadas a las familias del protocatecuato y homoprotocatecuato de la superfamilia LigB tras la exposición a diésel y nitrógeno (figuras 3.4E-3.4H), tal y como también se observó en el metagenoma de puerto (figuras 3.4A y 3.4B).

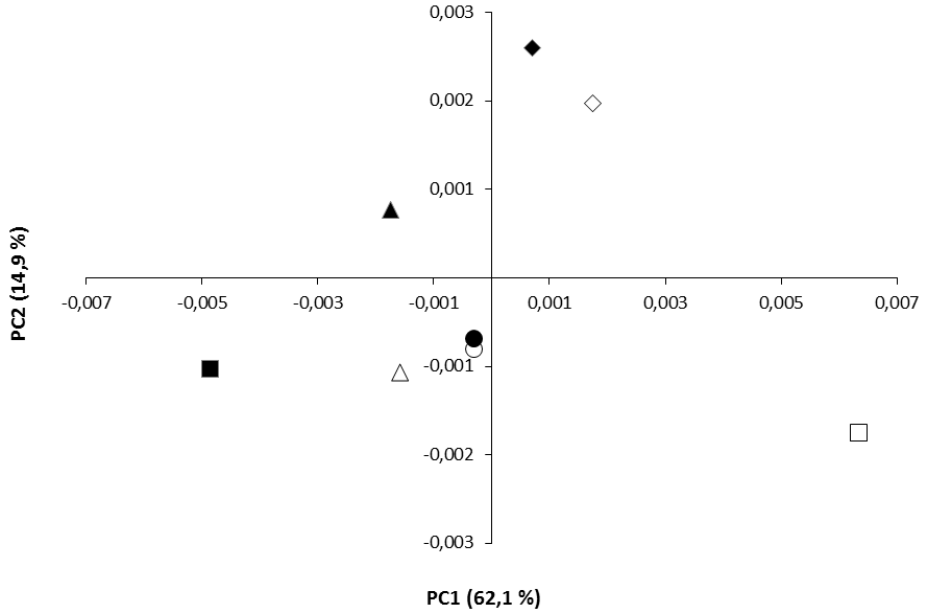

Figura 3.5. Comparación mediante PCA de los metagenomas en base a la asignación funcional con AromaDeg. Metagenomas control y puerto (cuadrados), metagenomas T2C1 2013 y T2D1 2013 (círculos), metagenomas T3C2 2013 y T3D2 2013 (triángulos), metagenomas T3C1 2014 y T3D1 2014 (rombos). Los símbolos de color blancos se corresponden a los metagenomas de muestras control y los de color negro a las muestras contaminadas con hidrocarburos (puerto y tratamientos con diésel).

Con el fin de incrementar la resolución del análisis se analizaron los porcentajes de abundancia relativa de lecturas afiliadas a cada uno de los grupos filogenéticos que componen las diferentes familias de AromaDeg (tabla S3.18) y se compararon las muestras control con las muestras contaminadas (puerto y microcosmos con diésel). Los grupos filogenéticos de cada familia de AromaDeg son grupos respaldados por valores de *bootstrap* elevados, a los que se les ha asignado función a partir de la presencia de proteínas de función validada dentro de cada grupo (Duarte *et al*., 2014). Para comparar de forma global la composición a nivel de grupos filogenéticos de AromaDeg de los ocho metagenomas se realizó un análisis de componentes principales (figura 3.5). A este nivel de resolución se observaron diferencias entre los metagenomas de la muestra control y del puerto, así como en los metagenomas de los microcosmos control y tratados con diésel. En concreto, las mayores diferencias se dieron entre el metagenoma control y el metagenoma de puerto (figura 3.5). En el caso de los metagenomas de los microcosmos se observaron mayores diferencias entre el tratamiento y el control en el tiempo de incubación final (T3), sobre todo entre los microcosmos T3C2 y T3D2 (figura 3.5). Por lo tanto, este resultado confirma las diferencias observadas entre los metagenomas control y los contaminados con hidrocarburos a nivel de familias de enzimas de AromaDeg, sobretodo entre el metagenoma control y el de puerto (figura 3.4). Además, indica que las diferencias en cuanto a enzimas de degradación entre los metagenomas generados a

partir de microcosmos control y microcosmos tratados con diésel y nitrógeno eran mayores a medida que aumentaba el tiempo de exposición a estas condiciones (figura 3.5). La poca diferencia observada en los metagenomas tras la exposición a diésel y nitrógeno fue un resultado inesperado en base al planteamiento inicial del experimento, y debido a la información previa disponible (Lanfranconi *et al*., 2010; Piña-Villalonga, 2012), pero no del todo inesperado en base a la falta de grandes cambios en la asignación funcional y taxonómica de las lecturas de los metagenomas que se había obtenido con MG-RAST (figura 3.3; tabla 3.5). Además debe tenerse en cuenta que los compuestos aromáticos son componentes minoritarios del diésel por lo que no sería esperable que la proporción de genes de degradación de estos compuestos fuese alta en una comunidad microbiana marina, ni siquiera tras exposición a diésel. Por otro lado, es posible que los tiempos de incubación ensayados quizás no fueran suficientes para poder observar cambios en el catabolismo de aromáticos asociados a la exposición a diésel y nitrógeno. Sin embargo, la extensión de los tiempos de incubación quedó descartada porque experimentos previos demostraron que se producía una completa alteración de la comunidad microbiana en los microcosmos debido al confinamiento del agua (Lanfranconi, 2010). Se analizó cuáles eran los grupos filogenéticos de proteínas de degradación que diferenciaban los metagenomas de las muestras control y contaminadas (puerto y tratamientos con diésel). En primer lugar, se determinó que las diferencias entre los metagenomas control y puerto se debían principalmente a siete grupos de enzimas que se muestran en la figura 3.6.

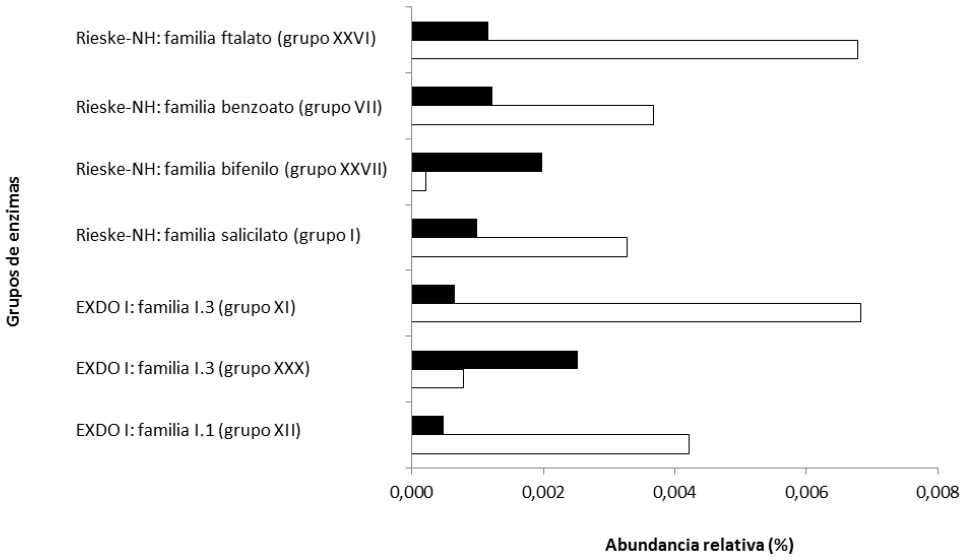

Figura 3.6. Grupos de enzimas de AromaDeg que discriminan los metagenomas ambientales. Se indica la abundancia relativa (porcentaje) de las lecturas asignadas a estos grupos. Las barras de color blanco se corresponden al metagenoma control y las de color negro al de puerto.

La mayoría de estos grupos tenían una abundancia relativa mayor en el metagenoma control, a excepción de las secuencias afiliadas al grupo XXVII de la familia del bifenilo (Rieske-NH) y el grupo XXX de la familia I.3 (EXDO I) cuya abundancia relativa era mayor en la muestra de puerto (figura 3.6). En base a los sustratos descritos para estos grupos de enzimas (Duarte *et al*., 2014) puede postularse una posible función en algunos de ellos. Por ejemplo, el grupo XXVII de la familia del bifenilo está formado por enzimas que tienen como sustrato la tetralina,

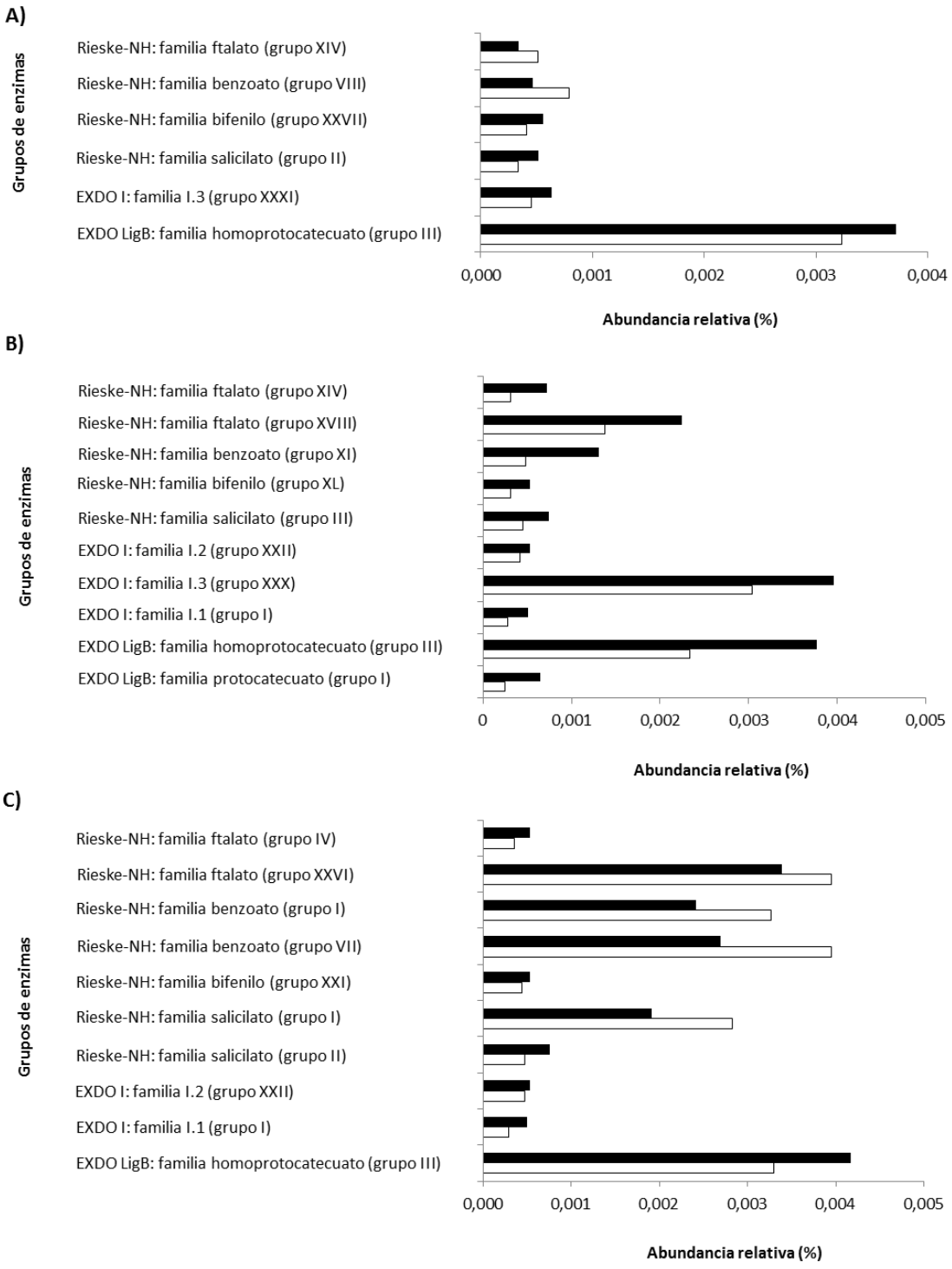

Figura 3.7. Grupos de enzimas de AromaDeg que discriminan los metagenomas de los microcosmos. Se indica la abundancia relativa (porcentaje) de las lecturas asignadas a estos grupos. Panel A, T2C1-T2D1 2013. Panel B, T3C2-T3D2 2013. Panel C, T3C1-T3D1 2014. Las barras de color blancos se corresponden a los microcosmos control y las de color negro a los tratados con diésel y nitrógeno.

el bifenilo y el etilbenceno y las enzimas del grupo XXX de la familia I.3 actúan sobre dihidroxibifenilos (tabla S3.18). Finalmente, las enzimas del grupo I de la familia del salicilato tienen actividad salicilato 1-hidroxilasa (tabla S3.18) y al resto de grupos de enzimas de la figura 3.6 no se les puede asignar una función concreta (tabla S3.18). La tetralina es un derivado del naftaleno (Krichko *et al*., 1969), el etilbenceno, el bifenilo y los dihidroxibifenilos

son compuestos presentes en el petróleo (Sarna *et al*., 1994; Adams y Richardson, 1953). Por lo tanto, la mayor abundancia relativa de secuencias relacionadas con degradación de contaminantes en el metagenoma del puerto podría explicarse por el mayor grado de contaminación de estas aguas en comparación con aguas control. Por otra parte, el salicilato es un intermediario de la degradación de naftaleno, y por tanto también está relacionado con contaminación, pero también es una fitohormona producida por plantas y algas (Mikami *et al*., 2016; Mori *et al*., 2017). Esto podría explicar la presencia de secuencias relacionadas con el grupo I de la familia del salicilato en ambos metagenomas.

Tabla 3.8. Sustratos de los grupos de enzimas importantes para diferenciar los metagenomas de microcosmos. Se indica el tipo de microcosmos en el que predominan estos grupos de enzimas: D+N, microcosmos tratados con diésel y nitrógeno y C, microcosmos control.

Superfamilia	Familia	Grupo	Compuesto	Microcosmos
Rieske-NH	Ftalato	IV	Desconocido	D+N
Rieske-NH	Ftalato	XIV	Desconocido	C
Rieske-NH	Ftalato	XVIII	4-aminobencenosulfonato	D+N
Rieske-NH	Ftalato	XXVI	Desconocido	C
Rieske-NH	Benzoato	I	Desconocido	C
Rieske-NH	Benzoato	VII	Desconocido	C
Rieske-NH	Benzoato	VIII	*p*-cumato	C
Rieske-NH	Benzoato	XI	Benzoato y 2-clorobenzoato	D+N
Rieske-NH	Bifenilo	XXI	Desconocido	D+N
Rieske-NH	Bifenilo	XXVII	Tetralina, bifenilo y etilbencenos	D+N
Rieske-NH	Bifenilo	XL	Bifenilo, bencenos y sustratos relacionados (clorobenceno, triclorobenceno, tetraclorobenceno, isopropilbenceno, tolueno y dibenzotiofeno)	D+N
Rieske-NH	Salicilato	I	Salicilato	C
Rieske-NH	Salicilato	II	Salicilato	D+N
Rieske-NH	Salicilato	III	Desconocido	D+N
EXDO I	I.1	I	2,3-dihidroxi-*p*-cumato y 2,3-dihidroxibenzoato	D+N
EXDO I	I.2	XXII	Catecol	D+N
EXDO I	I.3	XXX	2,3-dihidroxibifenilo y 3,4-dihidroxi-9,10-secoandrosta-1,3,5(10)-trien-9,17-diona propionato.	D+N
EXDO I	I.3	XXXI	2,3-dihidroxibifenilo	D+N
EXDO LigB	Homoprotocatecuato	III	2-aminofenol	D+N
EXDO LigB	Protocatecuato	I	Protocatecuato	D+N

Se analizaron también los grupos filogenéticos de cada una de las familias de AromaDeg que tuvieron un mayor peso a la hora de diferencias los metagenomas de los tres microcosmos tratados con diésel y nitrógeno de sus respectivos controles (figura 3.7). Las familias concretas a las que se asignaron las secuencias variaron en los dos experimentos realizados (2013 y 2014) y entre los dos tiempos de incubación en el experimento de 2013. La única excepción fue el grupo III de la familia del homoprotocatecuato, que se observó en los tres metagenomas de microcosmos generados. La mayoría de los grupos que diferenciaban los metagenomas correspondientes a las condiciones control y tratamiento tenían abundancias relativas superiores en los microcosmos tratados con diésel y nitrógeno. Los tipos de sustratos sobre los

que actuarían estos grupos de enzimas en base a la información disponible (Duarte *et al*., 2014) se muestran en la tabla 3.8. Como se ha mencionado anteriormente con el salicilato, algunos de los compuestos potencialmente metabolizables podrían provenir de fuentes naturales. Sería, por ejemplo, el caso del *p*-cumato y el 2,3-dihidroxi-*p*-cumato que provienen de la degradación de terpenos (Agulló *et al*., 2017). También del catecol, protocatecuato, benzoato y el 2,3-dihidroxibenzoato que son intermediarios comunes en el metabolismo de compuestos aromáticos derivados de la lignina (Bugg *et al*., 2011; Marín *et al*., 2012). Otros compuestos, en cambio, pueden ser relacionados con contaminación. Sería el caso del 2-clorobenzoato, que es un intermediario del catabolismo de los bifenilos policlorados y herbicidas (Dennison *et al*., 2010). La tetralina, el bifenilo, el 2,3-dihidroxibifenilo, el etilbenceno, el benceno y sustratos relacionados (clorobenceno, triclorobenceno, tetraclorobenceno, isopropilbenceno, tolueno y dibenzotiofeno) son compuestos presentes en el diésel (Liang *et al*., 2005), que deriva de combustibles fósiles (Sarna *et al*., 1994; Adams y Richardson, 1953). Finalmente, el 2-aminofenol y 4-aminobencenosulfonato son compuestos derivados de la anilina, compuesto derivado del benceno (Hu *et al*., 2007). Por lo tanto, que detectásemos una mayor abundancia relativa de genes para enzimas implicadas en la degradación de componentes del diésel en los metagenomas suplementados con diésel era un resultado esperado.

Finalmente para cada grupo de enzimas que resultaban diferenciales entre muestras control y contaminadas (puerto y microcosmos contaminados con diésel), se analizó la afiliación taxonómica de las secuencias de referencia de la base de datos AromaDeg a las que se habían asignado las lecturas. Con ello se pretendía relacionar a qué grupos taxonómicos se corresponderían los genes catabólicos detectados. Para ello, se calcularon los porcentajes de abundancia relativa de las lecturas asignadas a cada microorganismo dentro de cada grupo de AromaDeg. En concreto, el estudio se centró en aquellos microorganismos a los que al menos se había asignado un 1 % de las lecturas de un grupo de AromaDeg en uno de los dos metagenomas comparados (control y contaminado). Los resultados para los metagenomas ambientales control y del puerto se muestran en la tabla 3.9. El número de especies afiliadas a las lecturas era mayor en aquellas muestras en las que la abundancia relativa de lecturas también era mayor (ver figura 3.6). Así, para todos los grupos, excepto el XXVII de la familia del bifenilo y XXX de la familia EXDO I I.3 (los dos más abundantes en el meteganoma del puerto), se observó un mayor número de especies afiliadas a las lecturas en el metagenoma de la muestra control (tabla 3.9).

Por lo que respecta a las dioxigenasas tipo Rieske-NH, se observó que las secuencias afiliadas al grupo XXVI de la familia del ftalato se relacionaba sobre todo con bacterias del grupo SAR11 (tabla 3.9), grupo más detectado en el metagenoma control (figura 3.3A-B, tabla S3.3). Por lo tanto, las diferencias en abundancias relativas para esta categoría entre los metagenomas del agua control y del puerto estaban de acuerdo con la composición taxonómica de los metagenomas determinada por MG-RAST (ver figura 3.3). En cuanto al grupo VII de la familia del benzoato se observó una variedad de bacterias que incluían representantes de alfa (incluyendo representantes del linaje Roseobacter), beta y gamma proteobacterias (tabla 3.9). No obstante, había diferencias en la proporción de lecturas afiliadas a cada microorganismo (ver por ejemplo las secuencias afiliadas a *Curvibacter lanceolatus*, tabla 3.9), lo que sería

Tabla 3.9. Asignación taxonómica de las lecturas asignadas a grupos de enzimas de AromaDeg que diferencian los metagenomas ambientales. Se indica el porcentaje de las lecturas asignadas a cada especie. Se marcan como negativos todos aquellos grupos de enzimas de un determinado microorganismo en los que no se asignaron lecturas.

Superfamilia	Familia	Grupo	Microorganismo	Familia[a]	Nº de acceso	Porcentaje (%)	
						Control	Puerto
Rieske-NH	Ftalato	XXVI	*Candidatus* Pelagibacter ubique HTCC7211	*Alphaproteobacteria* (SAR11)	ZP_05069914.1	82,37	79,41
			HF4000 007D16	*Alphaproteobacteria* (*Rickettsiales*)	ABZ06218.1	7,91	17,65
			Marinimicrobia bacterium SCGC AAA076-D02	*Candidatus* Marinimicrobia (SAR406)	WP_020023984.1	6,47	–
			Candidatus Pelagibacter ubique SCGC AB-629-G21	*Alphaproteobacteria* (SAR11)	WP_018038917.1	3,24	2,94
	Benzoato	VII	*Verminephrobacter eiseniae* EF01-2^{T}	*Comamonadaceae*	YP_997431.1	15,89	8,33
			Thiomonas sp. FB-6	*Betaproteobacteria (Burkholderiales)*	WP_018913328.1	15,23	36,11
			Gammaproteobacteria marina HTCC2080	*Gammaproteobacteria* (OM60)	ZP_01626570.1	9,93	8,33
			Pusillimonas noertemannii	*Alcaligenaceae*	WP_017523152.1	8,61	5,56
			Sphingomonas wittichii RW1^{T}	*Sphingomonadaceae*	YP_001262032.1	8,61	2,78
			Curvibacter lanceolatus	*Comamonadaceae*	WP_019578166.1	7,95	22,22
			Variovorax paradoxus S110	*Comamonadaceae*	YP_002942564.1	6,62	–
			Jannaschia sp. CCS1	*Rhodobacteraceae*	YP_509285.1	5,96	–
			Alcaligenes faecalis	*Alcaligenaceae*	WP_003799263.1	5,30	5,56
			Ruegeria pomeroyi DSS-3^{T}	*Rhodobacteraceae*	YP_164964.1	3,31	5,56
			Hoeflea phototrophica	*Phyllobacteriaceae*	WP_007197248.1	2,65	–
			Pseudomonas chlororaphis	*Pseudomonadaceae*	WP_009042646.1	2,65	2,78
			Azoarcus toluclasticus	*Zoogloeaceae*	WP_018992065.1	1,99	–
			Azoarcus sp. BH72	*Zoogloeaceae*	YP_932627.1	1,32	–
			Bordetella sp. FB-8	*Alcaligenaceae*	WP_019936573.1	1,32	–
			Comamonas testosteroni S44	*Comamonadaceae*	ZP_07044600.1	0,66	2,78
			Achromobacter piechaudii ATCC 43553	*Burkholderiaceae*	ZP_06686030.1	0,66	–
			Burkholderia sp. 383	*Burkholderiaceae*	YP_366762.1	0,66	–
			NOR51-B	*Gammaproteobacteria* (OM60)	ZP_04959249.1	0,66	–
	Bifenilo	XXVII	*Conexibacter woesei* DSM 14684^{T}	*Conexibacteraceae*	YP_003392246.1	33,33	1,72
			Cycloclasticus sp. P1	*Piscirickettsiaceae*	YP_003392246.1	22,22	1,72
			Azotobacter vinelandii DJ	*Pseudomonadaceae*	YP_002801321.1	11,11	–
			Sphingobium chlorophenolicum L-1	*Sphingomonadaceae*	YP_004555535.1	11,11	–
			Sphingomonas sp. P2	*Sphingomonadaceae*	BAC65446.1	11,11	50,00
			Pseudomonas extremaustralis	*Pseudomonadaceae*	WP_010567587.1	11,11	1,72
			Novosphingobium aromaticivorans	*Sphingomonadaceae*	NP_049186.1	–	24,14
			Rhodococcus jostii RHA1	*Nocardiaceae*	YP_708484.1	–	6,90
			Dechloromonas aromatica RCB	*Azonexaceae*	YP_286446.1	–	5,18

a: En caso de microorganismos que no pertecen a una familia clasificada, se indicará el filo y entre paréntesis el grupo taxonómico más concreto en el que estén descritos.

Tabla 3.9. Asignación taxonómica de las lecturas asignadas a grupos de enzimas de AromaDeg que diferencian los metagenomas ambientales. Se indica el porcentaje de las lecturas asignadas a cada especie. Se marcan como negativos todos aquellos grupos de enzimas de un determinado microorganismo en los que no se asignaron lecturas (continuación).

Superfamilia	Familia	Grupo	Microorganismo	Familia[a]	Nº de acceso	Porcentaje (%)	
						Control	Puerto
Rieske-NH	Bifenilo	XXVII	*Rhodococcus* sp. TFB	*Nocardiaceae*	ACL31215.1	–	5,18
			Kordiimonas gwangyangensis	*Kordiimonadaceae*	WP_020398650.1	–	1,72
			Sphingopyxis macrogoltabida	*Sphingomonadaceae*	AAN26443.1	–	1,72
	Salicilato	I	*Sphingomonas* sp. P2	*Sphingomonadaceae*	BAC65453.1	58,21	48,27
			Pseudoxanthomonas spadix BD-a59	*Xanthomonadaceae*	YP_004931580.1	19,40	37,93
			Cycloclasticus sp. PY97M	*Piscirickettsiaceae*	WP_016390224.1	8,95	6,90
			Burkholderia sp Ch1-1	*Burkholderiaceae*	ZP_06846381.1	8,21	–
			Mycobacterium sp. S65	*Mycobacteriaceae*	AAQ12027.1	2,24	3,45
			Rhodococcus sp. JVH1	*Nocardiaceae*	WP_009478601.1	1,49	3,45
			Mycobacterium sp. MCS	*Mycobacteriaceae*	YP_638812.1	0,75	–
			Novosphingobium aromaticivorans	*Sphingomonadaceae*	NP_049179.1	0,75	–
EXDO I	I.3	XI	*Marinimicrobia* bacterium SCGC AAA076-D02	*Candidatus* Marinimicrobia (SAR406)	WP_020023533.1	84,28	89,47
			Actinobacterium SCGC AAA015-N04	*Actinobacteria*	WP_020053106.1	11,78	–
			Puniceispirillum marinum IMCC1322	*Alphaproteobacteria* (SAR116)	YP_003550915.1	1,07	–
			Brucella inopinata	*Brucellaceae*	WP_008511539.1	0,71	–
			Xanthobacter autotrophicus Py2	*Xanthobacteraceae*	YP_001415837.1	0,71	–
			Azorhizobium caulinodans ORS 571^{T}	*Xanthobacteraceae*	YP_001525796.1	0,36	–
			Martelella mediterranea	*Aurantimonadaceae*	WP_018063960.1	0,36	–
			Octadecabacter acticus 238^{T}	*Rhodobacteraceae*	ZP_05066473.1	0,36	–
			Stappia aggregata IAM 12614^{T}	*Rhodobacteraceae*	ZP_01550172.1	0,36	–
			Bradyrhizobium japonicum USDA 6^{T}	*Bradyrhizobiaceae*	YP_005609355.1	–	5,26
			Celeribacter baekdonensis	*Rhodobacteraceae*	WP_009570204.1	–	5,26
	I.3	XXX	Gammaproteobacteria HIMB55	*Gammaproteobacteria* (OM60)	WP_009470530.1	96,87	72,97
			Gammaproteobacteria marina HTCC2080	*Gammaproteobacteria* (OM60)	ZP_01626269.1	3,12	4,05
			Gammaproteobacteria marina IMCC3088	*Gammaproteobacteria* (OM60)	WP_009576803.1	–	22,97
	I.I	XII	Gammaproteobacteria marina SAR86D	*Gammaproteobacteria* (SAR86)	WP_010159958.1	72,86	50,00
			Psychroflexus torquis ATCC 700755^{T}	*Flavobacteriaceae*	ZP_01256066.1	12,14	21,43
			Hirschia baltica ATCC 49814^{T}	*Hyphomonadaceae*	YP_003060960.1	10,71	14,28
			Parvibaculum lavamentivorans DS-1^{T}	*Rhodobiaceae*	YP_001412815.1	3,57	14,28
			Porphyrobacter sp. AAP82	*Sphingomonadaceae*	WP_017665458.1	0,71	–

a: En caso de microorganismos que no pertecen a una familia clasificada, se indicará el filo y entre paréntesis el grupo taxonómico más concreto en el que estén descritos.

indicativo de que estos genes catabólicos estarían presentes en bacterias diferentes en las muestras control y del puerto. Esto mismo pero más acusado se observó en el grupo XXVII de la familia del bifenilo (tabla 3.9). Las lecturas del metagenoma del puerto pertenecientes a este grupo se afiliaban principalmente con secuencias de *Sphingomonadales* (*Sphingomonas* sp. P2 y *Novosphingobium aromaticivorans*), mientras que las encontradas en el metagenoma control se afiliaban a *Conexibacter woesei* DSM14684 y *Cycloclasticus* sp. P1 (ver tabla 3.9). Para los grupos de familias de extradiol dioxigenasas EXDO I se obtuvieron resultados similares (tabla 3.9). Es interesante destacar que la mayoría de las lecturas del grupo XII de la familia I.1 (EXDO I) se afiliaban con gammaproteobacterias del grupo SAR86 (tabla 3.9), grupo más detectado en el metagenoma control (figura 3.3A-B, tabla S3.3). También era interesante que las secuencias del grupo XI de la familia I.3 se afiliasen a genomas de bacterias marinas no cultivables obtenidos por SCG y a *Puneicispirillum marinum* IMCC1322 (grupo SAR116) y las del grupo XXX de esta misma familia se afiliasen a gammaproteobacterias del grupo OM60 (tabla 3.9). En el metagenoma control se detectó una mayor abundancia de lecturas asignadas al grupo SAR116, mientras que en el metagenoma de puerto se detectó un mayor número de lecturas asignadas al grupo OM60 (figuras 3.3A-B, tabla S3.3). Por lo tanto, estos resultados (ver figura 3.6 y Tabla 3.9) se correspondían con la asignación taxómica obtenida por MG-RAST (ver figura 3.3A-B).

Teniendo en cuenta el interés particular de esta tesis en las bacterias del linaje Roseobacter, se estudiaron con detalle los grupos en los que se encontraban secuencias afiliadas a estas bacterias (aunque existía la limitación de la baja representación en AromaDeg de grupos de enzimas interesantes en el contexto de este linaje). En el metagenoma control se observaron secuencias asignadas a *Jannaschia* sp. CSS1 dentro del grupo VII del benzoato (Rieske-NH) y a *Octadecabacter arcticus* 238^{T} dentro del grupo XI de la familia I.3 (EXDO) que no se observaban en el metagenoma de puerto (tabla 3.9). En el metagenoma de puerto se observaron secuencias asignadas a *Celeribacter baekdonensis* dentro del grupo XI de la familia I.3 (EXDO) que no se observaban en el metagenoma control (tabla 3.9). Además, se observaron secuencias asignadas a *Ruegeria pomeroyi* DSS-3^{T} dentro del grupo VII del benzoato (Rieske-NH) tanto en el metagenoma control como en el metagenoma de puerto, aunque el porcentaje era mayor en el metagenoma de puerto (tabla 3.9). Al desconocer el substrato sobre el que actúan las enzimas de estos dos grupos de las familias del benzoato (Rieske-NH) y de la familia I.3 (EXDO) (tabla S3.18), no se pudo relacionar la posible presencia de estos cinco roseobacters con la degradación de un determinado compuesto.

El mismo ejercicio de afiliación taxonómica de las lecturas correspondientes a los diferentes grupos de enzimas se realizó con los metagenomas de los experimentos de microcosmos. Los resultados se muestran en las tablas S3.19-S3.21. El número de especies afiliadas a las lecturas era mayor en aquellas muestras en las que la abundancia relativa de lecturas asignadas a un grupo enzimático también era mayor (figura 3.7; tablas S3.19-S3.21). Al igual que para los metagenomas ambientales se observaba también variabilidad en el porcentaje de lecturas asignado a cada microorganismo de referencia, tanto por lo que respectaba a la comparación control *vs* tratamiento como entre diferentes tiempos de tratamiento (microcosmos de 2013). Nuevamente esto sería indicativo de la posible presencia de estos genes catabólicos en bacterias diferentes en las muestras de microcosmos control y tratadas con diésel y nitrógeno. En los metagenomas de los microcosmos también se estudiaron con detalle los grupos en los que se encontraban secuencias afiliadas a roseobacters. En el caso de los microcosmos a T2,

no se pudo relacionar la posible presencia de miembros del linaje Roseobacter con secuencias codificantes para degradación de compuestos aromáticos, ni en el metagenoma del microcosmos control ni en el tratado con diésel y nitrógeno (tabla S3.19). En el tratamiento con diésel a T3 del experimento 2013 se observaron secuencias afiliadas a *Jannaschia* sp. CCS1, *Roseobacter* sp. MED193 y *Roseobacter* sp. SK209-2-6 dentro del grupo XVIII de la familia del ftalato (Rieske-NH), a *Octadecabacter arcticus* 238^T dentro del grupo III de la familia del salicilato (Rieske-NH) y a *Roseobacter* sp. MED193 dentro del grupo XI de la familia del benzoato (Rieske-NH), que no se observaban en el microcosmos control T3C2 2013 (tabla S3.20). Las enzimas del grupo XVIII de la familia del ftalato (Rieske NH) tienen como sustrato el 4-aminobencenosulfonato (ver tabla 3.8), derivado de la anilina, que a su vez deriva del benceno (Hu *et al*., 2007). Las enzimas del grupo XI de la familia del benzoato (Rieske-NH), tienen como sustratos el benzoato y el 2-clorobenzoato (ver tabla 3.8). En cuanto a las enzimas del grupo III de la familia del salicilato (Rieske-NH), se desconoce el sustrato sobre el que actúan (ver tabla 3.8). Por lo tanto, parece que las secuencias relacionadas con roseobacters en el metagenoma T3D2 2013, que no se detectaban en el metagenoma T3C2 2013 podrían relacionarse con un mayor potencial para degradar 4-aminobencenosulfonato, benzoato y el 2-clorobenzoato. Por otra parte, tanto en T3C2 2013 como en T3D2 2013 se detectaron secuencias asignadas a *Maritimibacter alkaliphilus* HTCC2654^T dentro del grupo XVIII de la familia del ftalato (Rieske-NH), a *Citreicella aestuarii* 357 dentro del grupo XL de la familia del bifenilo (Rieske-NH) y a *Rhodobacteraceae* bacterium HTCC2083 y *Maritimibacter alkaliphilus* HTCC2654^T dentro del grupo I de la familia del protocatecuato (LigB) (tabla S3.20). Todas estas secuencias tenían abundancias relativas superiores en el metagenoma T3C2 2013, excepto las de *Maritimibacter alkaliphilus* HTCC2654^T dentro del grupos XVIII de la familia del ftalato (Rieske-NH). Además, solo en T3D2 2013 se detectaron secuencias asignadas a *Jannaschia* sp. CCS1, *Roseobacter* sp. MED193 y *Roseobacter* sp. SK209-2-6 dentro del grupo XVIII de la familia del ftalato, a *Roseobacter* sp. MED193 dentro de la familia XI del benzoato y a *Octadecabacter arcticus* 238^T dentro de la familia III del salicilato (tabla S3.20). Las enzimas del grupo XL de la familia del bifenilo (Rieske-NH) tienen como sustratos bifenilo, benceno y sustratos relacionados (clorobenceno, triclorobenceno, tetraclorobenceno, isopropilbenceno, tolueno, dibenzotiofeno) (ver tabla 3.8). En cuanto a las enzimas del grupo I de la familia del protocatecuato (LigB), estas tienen como sustrato al protocatecuato (ver tabla 3.8). Por lo tanto, parece que las secuencias relacionadas con roseobacters en el metagenoma T3C2 2013, se vinculaban con un mayor potencial para degradar sustratos como el 4-aminobencenosulfonato, protocatecuato, bifenilo, benceno y sustratos relacionados. En cuanto a las secuencias relacionadas con roseobacters en el metagenoma T3D2 2013, estas también se vinculaban con un mayor potencial para degradar 4-aminobencenosulfonato, además de otros sustratos como son el benzoato y 2-clorobenzoato. Finalmente, respecto al tratamiento con diésel a T3 del experimento 2014, se observaron secuencias afiliadas a *Ruegeria pomeroyi* DSS-3^T y *Jannaschia* sp. CCS1, dentro del grupo VII de la familia del benzoato (Rieske-NH), siendo las de *Jannaschia* sp. CCS1 solo detectadas en el metagenoma T3C1 2014 (tabla S3.21). Además, se observaron secuencias afiliadas a *Sagittula stellata* E-37^T, dentro del grupo I de la familia del benzoato (Rieske-NH) tanto en T3C1 2014 como en T3D1 2014 (Tabla S3.21). Sin embargo, no se pudo relacionar la posible presencia de estas lecturas relacionadas con enzimas de roseobacters, con la degradación de un determinado compuesto, al desconocer el substrato sobre el que actúan las enzimas de estos dos grupos (tabla S3.18).

Finalmente, se comprobó si el tipo de enzimas asignadas a las secuencias de Roseobacter mostradas en las tablas 3.9, S3.20 y S3.21 eran coherentes con los tipos de enzimas detectados en estos microorganismos que se han descrito en el capítulo 1. Sin embargo la mayoría de los grupos de enzimas asignados por AromaDeg a secuencias de Roseobacter en los metagenomas no habían sido estudiados en esta tesis, en concreto, solo se habían estudiado enzimas del grupo I de la familia del protocatecuato (LigB) (tablas S1.2 y S3.18). La observación de secuencias asignadas al grupo I de la familia del protocatecuato (LigB) en *Maritimibacter alkaliphilus* HTCC2654^{T} y *Rhodobacteraceae* bacterium HTCC2083 era coherente con los resultados de predicción génica, ya que estos dos microorganismos presentan genes codificantes para protocatecuato 4,5-dioxigenasa (tablas S1.2 y S3.20).

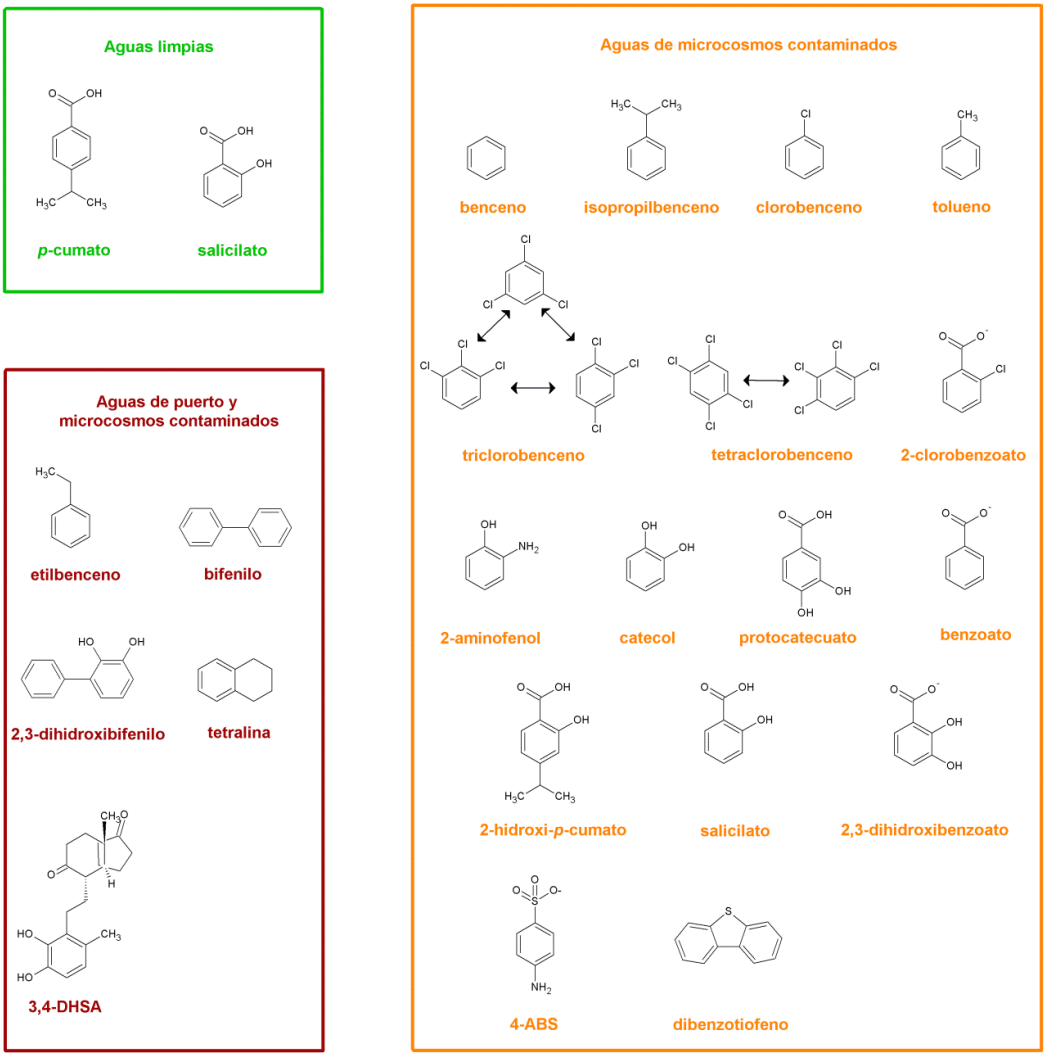

Figura 3.8. Tipo de compuestos sobre los que actúan los enzimas correspondientes a las secuencias génicas que discriminan entre los metagenomas control y los contaminados con hidrocarburos. Abreviaturas de los compuestos: 4-ABS, 4-aminobencenosulfonato y 3,4-DHSA, 3,4-dihidroxi-9,10-secoandrost-1,3,5(10)-trieno-9,17-diona.

Como conclusión de los análisis de genes catabólicos en los metagenomas y a modo de resumen se muestra una representación del tipo de compuestos que serían sustratos potenciales de los grupos de enzimas de AromaDeg a las que se habían asignado lecturas y que tenían un mayor peso a la hora de separar los metagenomas de las aguas control de los de aguas contaminadas. Tal y como se puede observar en la figura 3.8 se detectaron diferencias

sobre los tipos de compuestos aromáticos que potencialmente podrían ser degradados en aguas control respecto a las aguas contaminadas. En concreto se observó que en los metagenomas de muestras control había mayor potencial para degradar *p*-cumato y salicilato a catecol. Además, cabe recordar que la mayoría de los grupos de enzimas más abundantes en los metagenomas control actuaban sobre compuestos desconocidos por lo que no se puede inferir los posibles sustratos (figuras 3.6 y 3.7 y tabla S3.18). Por el contrario, en los metagenomas contaminados con hidrocarburos (puerto y microcosmos con diésel y nitrógeno), se había detectado una mayor abundancia relativa de secuencias relacionadas con degradación de compuestos aromáticos, entre ellos compuestos poliaromáticos (figura 3.8). El mayor potencial para el catabolismo de compuestos aromáticos en el metagenoma de puerto respecto al control era coherente con el obtenido en el estudio de Bargiela y colaboradores (2015) en sedimentos marinos crónicamente contaminados con hidrocarburos. En concreto, estos autores observaron que las poblaciones bacterianas presentes en los sedimentos de zonas crónicamente contaminadas con hidrocarburos presentaban una mayor diversificación de su potencial catabólico. Por ello, sugirieron que las poblaciones que habitan lugares crónicamente contaminados con hidrocarburos estarían adaptadas a este hábitat, ya que poseerían una gran diversidad de enzimas implicadas en la degradación rápida y eficaz de los compuestos derivados del petróleo y por ello, responderían más rápidamente a la degradación de hidrocarburos tras un derrame de petróleo (Bargiela *et al*., 2015).

CONCLUSIONES

1. La predicción génica realizada en esta tesis confirmó la predominancia de cuatro de las seis rutas centrales de degradación de compuestos aromáticos propuestas anteriormente en el linaje Roseobacter: protocatecuato (rotura *orto*), fenilacetil-CoA, homogentisato y homoprotocatecuato. En cambio, otras rutas centrales como la de degradación del gentisato, benzoato, catecol (rotura *orto* y *meta*) y protocatecuato (rotura *meta*) eran poco frecuentes. Estos resultados, junto con la predicción de genes para rutas periféricas, indicarían un buen potencial catabólico de bacterias del linaje Roseobacter para compuestos aromáticos que convergen en protocatecuato y fenilacetato. Ambos son compuestos centrales en la degradación de la lignina, presente habitualmente en ambientes marinos donde prolifera Roseobacter.

2. La predicción de genes para el catabolismo de compuestos aromáticos centrales y periféricos fue desigual entre los cinco grupos filogenómicos del linaje Roseobacter e incluso dentro de los miembros de un mismo grupo. De forma característica en genomas del grupo genómico 4 se predijo un número menor de genes y vías catabólicas. Estos resultados no serían coherentes con una herencia vertical de genes para el catabolismo de compuestos aromáticos en el linaje Roseobacter sino más bien con procesos evolutivos de ganancia o pérdida de genes.

3. Se observaron múltiples agrupaciones para los genes de las diferentes rutas catabólicas con pocos ejemplos de estructuras que incluían todos los genes de una ruta. Esto diferencia las bacterias del linaje Roseobacter estudiadas de otros modelos de degradación de compuestos aromáticos donde se observan estructuras supraoperónicas. También fue infrecuente la detección de agrupaciones que incluyesen genes para diferentes rutas catabólicas. Estos resultados indicarían que no existe una presión para la selección de *clusters* catabólicos en genomas de bacterias del linaje Roseobacter. Por otra parte, no se observó una relación clara entre tipo de ordenación génica y la filogenia, lo que reforzaría la hipótesis de que la presencia de estos genes en los genomas es debida a procesos de ganancia o pérdida de genes.

4. La tolerancia a hidrocarburos aromáticos de las tres bacterias del linaje Roseobacter analizadas en esta tesis (*Ruegeria pomeroyi* DSS-3^T, *Sulfitobacter* sp. 3SOLIMAR09 y *Citreicella aestuarii* 357) fue mayor en el caso del naftaleno que en el caso del tolueno, probablemente debido a la solubilidad y toxicidad de este último. Las condiciones ensayadas en este trabajo fueron adecuadas para analizar mediante proteómica la respuesta a naftaleno. En cambio, para analizar la respuesta a tolueno deberían modificarse los tiempos de exposición y/o la concentración de este compuesto. Por otra parte, aunque el análisis de proteoma fraccionado puede aportar información adicional puede decirse que el análisis del proteoma total proporciona información suficiente y adecuada sobre la respuesta celular a la exposición a hidrocarburos aromáticos en estas bacterias.

5. Tras exposición a naftaleno se observó una respuesta común en los tres aislados estudiados que incluía de forma genérica: (1) cambios a nivel de transporte de membrana que implicaron la disminución de proteínas de transporte de solutos, el incremento de proteínas de sistemas de secreción y el aumento de bombas de expulsión de tipo RND; (2) inducción de mecanismos de respuesta a estrés oxidativo, y (3) inducción de la expresión de chaperonas. Estos cambios se correspondían con los descritos en la literatura para tolerancia a hidrocarburos en bacterias

del género *Pseudomonas*. En el caso de *C. aestuarii* 357 la detección de proteínas de degradación de naftaleno puede considerarse un mecanismo de destoxicación adicional.

6. La sobreproducción de proteínas de función desconocida tras la exposición a naftaleno (por ejemplo las proteínas SPO2567 o SPO2568 de *R. pomeroyi* DSS-3^T y sus correspondientes homólogos en los otros dos aislados analizados en esta tesis) demuestra que tenemos un conocimiento incompleto de la respuesta de estas bacterias a este contaminante. La disminución de la capacidad de crecimiento en presencia de naftaleno de un mutante de *R. pomeroyi* DSS-3^T defectivo en la proteína SPO2567 confirmaría la implicación de esta proteína en la respuesta celular a naftaleno. El hecho de que esta proteína sea característica de dos órdenes de *Alphaproteobacteria* (*Rhizobiales* y *Rhodobacterales*), que esté conservada en bacterias de la familia *Rhodobacteraceae*, y que también se exprese en *R. pomeroyi* DSS-3^T en otras condiciones restrictivas (p.e. alta temperatura) indicaría la existencia de un mecanismo adicional de respuesta a estrés en estas bacterias. En el contexto del bacterioplancton marino esto diferenciaría a bacterias del linaje Roseobacter de otras bacterias con las que comparte hábitat.

7. El análisis metagenómico ha evidenciado diferencias entre la composición taxonómica y en categorías funcionales de una muestra de agua contaminada crónicamente (puerto) y una muestra control. Las diferencias funcionales se podrían atribuir en parte a la diferente composición taxonómica de estas comunidades (p.e. la presencia de genes para transportadores TonB con ciertas gammaproteobacterias y bacteroidetes) pero también a la contaminación que se da en puertos (p.e. genes para mecanismos de respuesta a estrés, degradación de compuestos aromáticos o transportadores de membrana de tipo RND). En cambio el análisis metagenómico ha mostrado limitaciones para evidenciar diferencias funcionales tras contaminación artificial en microcosmos de laboratorio. Para ese tipo de experimentos sería recomendable utilizar aproximaciones ómicas más próximas a la expresión génica como la metatranscriptómica o la metaproteómica.

8. El análisis realizado en esta tesis ha mostrado que existe potencial para la degradación de compuestos aromáticos tanto en aguas no contaminadas como en aguas contaminadas por hidrocarburos. Sin embargo, se han evidenciado diferencias en la proporción de lecturas asignadas a diferentes familias de enzimas, lo que indica un potencial catabólico diferente de estas comunidades. A pesar de las limitaciones inherentes a las bases de datos, y al hecho de que se desconocen los sustratos sobre los que actuarían muchas de las familias génicas detectadas, se observó claramente un mayor potencial catabólico para compuestos poliaromáticos que podían proceder de contaminantes (p.e bifenilos y derivados) en agua del puerto y para compuestos monoaromáticos de posible origen vegetal en agua control (*p*-cumato y salicilato). Estos resultados están de acuerdo con las características de ambos ambientes.

9. Si se incrementa el nivel de detalle del análisis con la inferencia taxonómica de las secuencias para genes de degradación de compuestos aromáticos de los metagenomas se observa que incluso para una misma categoría (familia y grupo de enzimas) los microorganismos que contendrían esos genes son diferentes en muestras contaminadas y no contaminadas. Esto permite hipotetizar que estos microorganismos podrían tener diferente

actividad degradadora en el ambiente. Por otra parte, aunque la baja representación de genomas de Roseobacter en la base de datos utilizada para el análisis (AromaDeg) ha limitado la posibilidad de relacionar genes catabólicos con bacterias del linaje Roseobacter en los metagenomas analizados, los resultados de esta tesis evidencian que varios grupos relevantes del bacterioplancton marino (SAR86, SAR116, OM60) serían también degradadores potenciales de compuestos aromáticos en estos ambientes.

BIBLIOGRAFÍA

Adams NG, Richardson DM. 1953. Isolation and identification of biphenyls from West Edmond crude oil. Analytical Chemistry; 25: 1073–1074.

Aemprapa S, Williams PA. 1998. Implications of the *xylQ* gene of TOL plasmid pWW102 for the evolution of aromatic catabolic pathways. Microbiology; 144: 1387–1396.

Aguiló-Ferretjans MM, Bosch R, Martín-Cardona C, Lalucat J, Nogales B. 2008. Phylogenetic analysis of the composition of bacterial communities in human exploited coastal environments from Mallorca Island (Spain). Systematic and Applied Microbiology; 31: 231–240.

Agulló L, Romero-Silva MJ, Domenech M, Seeger M. 2017. *p*-Cymene promotes its catabolism through the *p*-cymene and the *p*-cumate pathways, activates a stress response and reduces the biofilm formation in *Burkholderia xenovorans* LB400. PLoS One; 12: e0169544.

Alonso-Vega P, Normand P, Bacigalupe R, Pujic P, Lajus A, Vallenet D, Carro L, Coll P, Trujillo ME. 2012. Genome sequence of *Micromonospora lupini* Lupac 08, isolated from root nodules of *Lupinus angustifolius*. Journal of Bacteriology; 194: 4135.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research; 25: 3389–3402.

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Applied and Environmental Microbiology; 56: 1919–1925.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Disponible en: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aono R, Tsukagoshi N, Yamamoto M. 1998. Involvement of outer membrane protein TolC, a possible member of the *mar-sox* regulon, in maintenance and improvement of organic solvent tolerance of *Escherichia coli* K-12. Journal of Bacteriology; 180: 938–944.

Arahal DR, Lekunberri I, González JM, Pascual J, Pujalte MJ, Pedrós-Alió C, Pinhassi J. 2007. *Neptuniibacter caesariensis* gen. nov., sp. nov., a novel marine genome-sequenced gammaproteobacterium. International Journal of Systematic and Evolutionary Microbiology; 57: 1000–1006.

Arias-Barrau E, Olivera ER, Luengo JM, Fernández C, Galán B, García JL, Díaz E, Miñambres B. 2004. The homogentisate pathway: a central catabolic pathway involved in the degradation of L-phenylalanine, L-tyrosine, and 3-hydroxyphenylacetate in *Pseudomonas putida*. Journal of Bacteriology; 186: 5062–5077.

Bargiela R, Mapelli F, Rojo D, Chouaia B, Tornés J, Borin S, Richter M, Del Pozo MV, Cappello S, Gertler C, Genovese M, Denaro R, Martínez-Martínez M, Fodelianakis S, Amer RA, Bigazzi D, Han X, Chen J, Chernikova TN, Golyshina OV, Mahjoubi M, Jaouanil A, Benzha F, Magagnini M, Hussein E, Al-Horani F, Cherif A, Blaghen M, Abdel-Fattah YR, Kalogerakis N, Barbas C, Malkawi HI, Golyshin PN, Yakimov MM, Daffonchio D, Ferrer M. 2015. Bacterial population and

biodegradation potential in chronically crude oil-contaminated marine sediments are strongly linked to temperature. Scientific Reports; 29: 11651.

Bastard K, Smith AAT, Vergne-Vaxelaire C, Perret A, Zaparucha A, De Melo-Minardi R, Mariage A, Boutard M, Debard A, Lechaplais C, Pelle C, Pellouin V, Perchat N, Petit JL, Kreimeyer A, Medigue C, Weissenbach J, Artiguenave F, De Berardinis V, Vallenet D, Salanoubat M. 2014. Revealing the hidden functional diversity of an enzyme family. Nature Chemical Biology; 10: 42–49.

Belda E, van Heck RG, José Lopez-Sanchez M, Cruveiller S, Barbe V, Fraser C, Klenk HP, Petersen J, Morgat A, Nikel PI, Vallenet D, Rouy Z, Sekowska A, Martins Dos Santos VA, de Lorenzo V, Danchin A, Médigue C. 2016. The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic chassis. Environmental Microbiology; 18: 3403–3424.

Bell T, Newman JA, Silverman BW, Turner SL, Lilley AK. 2005. The contribution of species richness and composition to bacterial services. Nature; 436: 1157–1160.

Bernal P, Muñoz-Rojas J, Hurtado A, Ramos JL, Segura A. 2007a. *Pseudomonas putida* cardiolipin synthesis mutant exhibits increased sensitivity to drugs related to transport functionality. Environmental Microbiology; 9: 1135–1145.

Bernal P, Segura A, Ramos JL. 2007b. Compensatory role of the *cis-trans*-isomerase and cardiolipin synthase in the membrane fluidity of *Pseudomonas putida* DOT-T1E. Environmental Microbiology; 9: 1658–1664.

Berthe-Corti L, Nachtkamp M. 2010. Bacterial communities in hydrocarbon-contaminated marine coastal environments. *En* Handbook of Hydrocarbon and Lipid Microbiology. (ed. Timmis, K.N.) Springer-Verlag Berlin Heidelberg pp. 2349–2359.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics; 30: 2114–2120.

Bosch R, Moore ERB, García-Valdés E, Pieper DH. 1999a. NahW, a novel, inducible salicylate hydroxylase involved in mineralization of naphthalene by *Pseudomonas stutzeri* AN10. Journal of Bacteriology; 181: 2315–2322.

Bosch R, García-Valdés E, Moore ER. 1999b. Genetic characterization and evolutionary implications of a chromosomally encoded naphthalene-degradation upper pathway from *Pseudomonas stutzeri* AN10. Gene; 236: 149–157.

Brakstad OG, Lødeng AGG. 2005. Microbial diversity during biodegradation of crude oil in seawater from the North Sea. Microbial Ecology; 49: 94–103.

Breese K, Boll M, Alt-Mörbe J, Schägger H, Fuchs G. 1998. Genes coding for the benzoyl-CoA pathway of anaerobic aromatic metabolism in the bacterium *Thauera aromatica*. European Journal of Biochemistry; 256: 148–154.

Brinkhoff T, Bach G, Heidorn T, Liang L, Schlingloff A, Simon M. 2004. Antibiotic production by a *Roseobacter* clade-affiliated species from the German Wadden Sea and its antagonistic effects on indigenous isolates. Applied and Environmental Microbiology; 70: 2560–2565.

Britschgi TB, Giovannoni SJ. 1991. Phylogenetic analysis of a natural bacterioplankton population by rRNA gene cloning and sequencing. Applied and Environmental Microbiology; 57: 1707–1713.

Brito EMS, Guyoneaud R, Goñi-Urriza M, Ranchou-Peyruse A, Verbaere A, Crapez MAC, Wasserman JCA, Duran R. 2006. Characterization of hydrocarbonoclastic bacterial communities from mangrove sediments in Guanabara Bay, Brazil. Research in Microbiology; 157: 752–762.

Brunet-Galmés I. 2016. Understanding the aromatic hydrocarbon degradation potential of *Pseudomonas stutzeri*: a proteo-genomic approach. Tesis doctoral. Universitat de les Illes Balears. 169 p.

Buchan A, Collier LS, Neidle EL, Moran MA. 2000. Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine *Roseobacter* lineage. Applied and Environmental Microbiology; 66: 4662–4672.

Buchan A, Neidle EL, Moran MA. 2001. Diversity of the ring-cleaving dioxygenase gene *pcaH* in a salt marsh bacterial community. Applied and Environmental Microbiology; 67: 5801–5809.

Buchan A, Neidle EL, Moran MA. 2004. Diverse organization of genes of the β-ketoadipate pathway in members of the marine *Roseobacter* lineage. Applied and Environmental Microbiology; 70: 1658–1668.

Buchan A, González JM, Moran MA. 2005. Overview of the marine *Roseobacter* lineage. Applied and Environmental Microbiology; 71: 5665–5677.

Buchan A, González JM. 2010. Roseobacter. *En* Handbook of hydrocarbon and lipid microbiology. (ed. Timmis, K.N.) Springer-Verlag Berlin Heidelberg. pp. 1335–1343.

Buchan A, LeCleir GR, Gulvik CA, González JM. 2014. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nature Reviews Microbiology; 12: 686–698.

Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R. 2011. Pathways for degration of lignin in bacteria and fungi. Natural Product Reports; 28: 1883–1896.

Cámara B, Bielecki P, Kaminski F, dos Santos VM, Plumeier I, Nikodem P, Pieper DH. 2007. A gene cluster involved in degradation of substituted salicylates via *ortho* cleavage in *Pseudomonas* sp. strain MT1 encodes enzymes specifically adapted for transformation of 4-methylcatechol and 3-methylmuconate. Journal of Bacteriology; 189: 1664–1674.

Campillo T, Renoud S, Kerzaon I, Vial L, Baude J, Gaillard V, Bellvert F, Chamignon C, Comte G, Nesme X, Lavire C, Hommais F. 2014. Analysis of hydroxycinnamic acid degradation in *Agrobacterium fabrum* reveals a coenzyme A-dependent, beta-oxidative deacetylation pathway. Applied and Environmental Microbiology; 80: 3341–3349.

Cappello S, Caruso G, Zampino D, Monticelli LS, Maimone G, Denaro R, Tripodo B, Troussellier M, Yakimov M, Giuliano L. 2007. Microbial community dynamics during assays of harbour oil spill bioremediation: a microscale simulation study. Journal of Applied Microbiology; 102: 184–194.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution; 17: 540–552.

Chang HK, Zylstra GJ. 1998. Novel organization of the genes for phthalate degradation from *Burkholderia cepacia* DBO1. Journal of Bacteriology; 180: 6529–6537.

Chang HK, Zylstra GJ. 1999. Role of quinolinate phosphoribosyl transferase in degradation of phthalate by *Burkholderia cepacia* DBO1. Journal of Bacteriology; 181: 3069–3075.

Charlson RJ, Lovelock JE, Andreae MO, Warren SG. 1987. Oceanic phytoplankton, atmospheric sulfur, cloud albedo and climate. Nature; 326: 655–661.

Cho JC, Giovannoni SJ. 2004. Cultivation and growth characteristics of a diverse group of oligotrophic marine *Gammaproteobacteria*. Applied and Environmental Microbiology; 70: 432–440.

Christie-Oleza JA, Fernandez B, Nogales B, Bosch R, Armengaud J. 2012. Proteomic insights into the lifestyle of an environmentally relevant marine bacterium. ISME Journal; 6: 124–135.

Collier LS, Gaines GL 3rd, Neidle EL. 1998. Regulation of benzoate degradation in *Acinetobacter* sp. strain ADP1 by BenM, a LysR-type transcriptional activator. Journal of Bacteriology; 180: 2493–2501.

Cosper NJ, Collier LS, Clark TJ, Scott RA, Neidle EL. 2000. Mutations in *catB*, the gene encoding muconate cycloisomerase, activate transcription of the distal *ben* genes and contribute to a complex regulatory circuit in *Acinetobacter* sp. strain ADP1. Journal of Bacteriology; 182: 7044–7052.

Cox HH, Faber BW, Van Heiningen WN, Radhoe H, Doddema HJ, Harder W. 1996. Styrene metabolism in *Exophiala jeanselmei* and involvement of a cytochrome P-450-dependent styrene monooxygenase. Applied and Environmental Microbiology; 62: 1471–1474.

Cunliffe M. 2011. Correlating carbon monoxide oxidation with *cox* genes in the abundant Marine Roseobacter Clade. ISME Journal; 5: 685–691.

Daims H, Bruhl A, Amann R, Schleifer KH,Wagner M. 1999. The domain-specific probe EUB338 is insufficient for the detection of all *Bacteria*: development and evaluation of a more comprehensive probe set. Systematic and Applied Microbiology; 22: 434–444.

Das N, Chandran P. 2011. Microbial degradation of petroleum hydrocarbon contaminants: an overview. Biotechnology Research International; 2011: 941810.

Davidson K, Gowen R.J, Harrison PJ, Fleming LE, Hoagland P, Moschonas K. 2014. Anthropogenic nutrients and harmful algae in coastal waters. Journal of Environmental Management; 146: 206–216.

Dennis JJ, Zylstra GJ. 1998. Plasposons: modular self-cloning minitransposon derivatives for rapid genetic analysis of gram-negative bacterial genomes. Applied and Environmental Microbiology; 64: 2710–2715.

Dennison SG, O'Brien P, Gopalkrishnan S, Stark BC. 2010. Enhancement of aerobic degradation of benzoate and 2-chlorobenzoate by adapted activated sludge. Microbiological Research; 165: 687–694.

Díaz E, Ferrández A, Prieto MA, García JL. 2001. Biodegradation of aromatic compounds by *Escherichia coli*. Microbiology and Molecular Biology Reviews; 65: 523–569.

Díaz E. 2004. Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. International Microbiology; 7: 173–180.

Ding B, Schmeling S, Fuchs G. 2008. Anaerobic metabolism of catechol by the denitrifying bacterium *Thauera aromatica*–a result of promiscuous enzymes and regulators? Journal of Bacteriology; 190: 1620–1630.

Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues C, Haynes M, Li L, McDaniel L, Moran MA, Nelson KE, Nilsson C, Olson R, Paul J, Brito BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White BA, Rohwer F. 2008. Functional metagenomic profiling of nine biomes. Nature; 452: 629–632.

Djordjevic S, Stock AM. 1997. Crystal structure of the chemotaxis receptor methyltransferase CheR suggests a conserved structural motif for binding *S*-adenosylmethionine. Structure; 5: 545–558.

Domínguez-Cuevas P, González-Pastor JE, Marqués S, Ramos JL, de Lorenzo V. 2006. Transcriptional tradeoff between metabolic and stress-response programs in *Pseudomonas putida* KT2440 cells exposed to toluene. Journal of Biological Chemistry; 281: 11981–11991.

Dreier J, Ruggerone P. 2015. Interaction of antibacterial compounds with RND efflux pumps in *Pseudomonas aeruginosa*. Frontiers in Microbiology; 6: 660.

Duarte M, Jauregui R, Vilchez-Vargas R, Junca H, Pieper D.H.2014. AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics. Database; 2014: bau118.

Dubinsky EA, Conrad ME, Chakraborty R, Bill M, Borglin SE, Hollibaugh JT, Mason OU, M Piceno Y, Reid FC, Stringfellow WT, Tom LM, Hazen TC, Andersen GL. 2013. Succession of hydrocarbon-degrading bacteria in the aftermath of the Deepwater Horizon oil spill in the Gulf of Mexico. Environmental Science and Technology; 47: 10860–10867.

Eck RV, Dayhoff MO. 1966. Atlas of protein sequence and structure 1996. National Biomedical Research Foundation, Silver Springs, Maryland.

Eilers H, Pernthaler J, Peplies J, Glöckner FO, Gerdts G, Amann R. 2001. Isolation of novel pelagic bacteria from the German Bight and their seasonal contributions to surface picoplankton. Applied and Environmental Microbiology; 67: 5134–5142.

Eulberg D, Lakner S, Golovleva LA, Schlömann M. 1998. Characterization of a protocatechuate catabolic gene cluster from *Rhodococcus opacus* 1CP: evidence for a merged enzyme with 4-carboxymuconolactone-decarboxylating and 3-oxoadipate enol-lactone-hydrolyzing activity. Journal of Bacteriology; 180: 1072–1081.

Felsestein J. 1989. PHYLYP- Phylogeny Inference Package (version 3.2). Cladistics; 5: 164–166.

Ferrández A, Miñambres B, García B, Olivera ER, Luengo JM, García JL, Díaz E. 1998. Catabolism of phenylacetic acid in *Escherichia coli*. Characterization of a new aerobic hybrid pathway. Journal of Biological Chemistry; 273: 25974–25986.

Ferrer-Sevillano F, Fernández-Cañón JM. 2007. Novel *phacB*-encoded cytochrome P450 monooxygenase from *Aspergillus nidulans* with 3-hydroxyphenylacetate 6-hydroxylase and 3,4-dihydroxyphenylacetate 6-hydroxylase activities. Eukaryotic Cell; 6: 514–520.

Fetzner S. 2012. Ring-cleaving dioxygenases with a cupin fold. Applied and Environmental Microbiology; 78: 2505–2514.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research; 44: D279–D285.

Geng H, Belas R. 2010. Molecular mechanisms underlying roseobacter-phytoplankton symbioses. Current Opinion in Biotechnology; 21: 332–338.

GESAMP. 2007. Joint group of experts on the scientific aspects of marine environmental protection. Estimates of oil entering the marine environment from sea-based activities. Reports and studies, GESAMP; 75. UN, UNEP, FAO, UNESCO-IOC, WMO, UNIDO, IMO, IAEA, London. pp. 83.

Gescher J, Eisenreich W, Wörth J, Bacher A, Fuchs G. 2005. Aerobic benzoyl-CoA catabolic pathway in *Azoarcus evansii*: studies on the non-oxygenolytic ring cleavage enzyme. Molecular Microbiology; 56: 1586–1600.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA III, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods; 6: 343–345.

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P, Joint I. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS One; 3: e3042.

Gilbert JA, Dupont CL. 2011. Microbial metagenomics: beyond the genome. Annual Review of Marine Science; 3: 347–371.

Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature; 345: 60–63.

Glöckner FO, Amann R, Alfreider A, Pernthaler J, Psenner R, Trebesius K, Schleifer KH. 1996. An in situ hybridization protocol for detection and identification of planktonic bacteria. Systematic and Applied Microbiology; 19: 403–406.

Göbel M, Kassel-Cati K, Schmidt E, Reineke W. 2002. Degradation of aromatics and chloroaromatics by *Pseudomonas* sp. strain B13: cloning, characterization, and analysis of sequences encoding 3-oxoadipate:succinyl-coenzyme A (CoA) transferase and 3-oxoadipyl-CoA thiolase. Journal of Bacteriology; 184: 216–223.

González JM, Mayer F, Moran MA, Hodson RE, Whitman WB. 1997. *Sagittula stellata* gen. nov. sp. nov., a lignin-transforming bacterium from a coastal environment. International Journal of Systematic Bacteriology; 47: 773–780.

González JM, Simó R, Massana R, Covert JS, Casamayor EO, Pedrós-Alió C, Moran MA. 2000. Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. Applied and Environmental Microbiology; 66: 4237–4246.

Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, Rappé MS. 2012. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. mBio; 3: e00252-12.

Guibert LM, Loviso CL, Marcos MS, Commendatore MG, Dionisi HM, Lozada M. 2012. Alkane biodegradation genes from chronically polluted subantarctic coastal sediments and their shifts in response to oil exposure. Microbial Ecology; 64: 605–616.

Gulvik CA, Buchan A. 2013. Simultaneous catabolism of plant-derived aromatic compounds results in enhanced growth for members of the *Roseobacter* lineage. Applied and Environmental Microbiology; 79: 3716–3723.

Gutierrez T, Singleton DR, Berry D, Yang T, Aitken MD, Teske A. 2013. Hydrocarbon-degrading bacteria enriched by the Deepwater Horizon oil spill identified by cultivation and DNA-SIP. ISME Journal; 7: 2091–2104.

Hamlin JN, Bloodworth RA, Cardona ST. 2009. Regulation of phenylacetic acid degradation genes of *Burkholderia cenocepacia* K56-2. BMC Microbiology; 9:222.

Hammer Ø, Harper DAT, Ryan PD. 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica; 4: 9pp.

Hara H, Masai E, Miyauchi K, Katayama Y, Fukuda M. 2003. Characterization of the 4-carboxy-4-hydroxy-2-oxoadipate aldolase gene and operon structure of the protocatechuate 4,5-cleavage pathway genes in *Sphingomonas paucimobilis* SYK-6. Journal of Bacteriology; 185: 41–50.

Harwati TU, Kasai Y, Kodama Y, Susilaningsih D, Watanabe K. 2007. Characterization of diverse hydrocarbon-degrading bacteria isolated from Indonesian seawater. Microbes and Environments; 22: 412–415.

Harwati TU, Kasai Y, Kodama Y, Susilaningsih D, Watanabe K. 2008. *Tranquillimonas alkanivorans* gen. nov., sp. nov., an alkane-degrading bacterium isolated from Semarang Port in Indonesia. International Journal of Systematic and Evolutionary Microbiology; 58: 2118–2121.

Harwati TU, Kasai Y, Kodama Y, Susilaningsih D, Watanabe K. 2009. *Tropicibacter naphthalenivorans* gen. nov., sp. nov., a polycyclic aromatic hydrocarbon-degrading bacterium isolated from Semarang Port in Indonesia. International Journal of Systematic and Evolutionary Microbiology; 59: 392–396.

Harwood CS, Parales RE. 1996. The β-ketoadipate pathway and the biology of self-identity. Annual Review of Microbiology; 50: 553–590.

Hawley ER, Piao H, Scott NM, Malfatti S, Pagani I, Huntemann M, Chen A, Glavina Del Rio T, Foster B, Copeland A, Jansson J, Pati A, Tringe S, Gilbert JA, Lorenson TD, Hess M. 2014. Metagenomic analysis of microbial consortium from natural crude oil that seeps into the marine ecosystem offshore Southern California. Standards in Genomic Sciences; 9: 1259–1274.

Hemamalini R, Khare S. 2014. A proteomic approach to understand the role of the outer membrane porins in the organic solvent-tolerance of *Pseudomonas aeruginosa* PseA. PLoS One; 9: e103788.

Hiromoto T, Matsue H, Yoshida M, Tanaka T, Higashibata H, Hosokawa K, Yamaguchi H, Fujiwara S. 2006. Characterization of MobR, the 3-hydroxybenzoate-responsive transcriptional regulator for the 3-hydroxybenzoate hydroxylase gene of *Comamonas testosteroni* KH122-3s. Journal of Molecular Biology; 364: 863–877.

Hu C, Zhu L, Xia Y. 2007. Direct amination of benzene to aniline by aqueous ammonia and hydrogen peroxide over V−Ni/Al_2O_3 catalyst with catalytic distillation. Industrial and Engineeering Chemistry Research; 46: 3443–3445.

Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics; 26: 680–682.

Huang YM, Miao Y, Munguia J, Lin L, Nizet V, McCammon JA. 2016. Molecular dynamic study of MlaC protein in Gram-negative bacteria: conformational flexibility, solvent effect and protein-phospholipid binding. Protein Science; 25: 1430–1437.

Ishiyama D, Vujaklija D, Davies J. 2004. Novel pathway of salicylate degradation by *Streptomyces* sp. strain WA46. Applied and Environmental Microbiology; 70: 1297–1306.

Iwai S, Chai B, Sul WJ, Cole JR, Hashsham SA, Tiedje JM. 2010. Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. ISME Journal; 4: 279–285.

Iwaki H, Nishimura A, Hasegawa Y. 2012a. Isolation and characterization of marine bacteria capable of utilizing phthalate. World Journal of Microbiology and Biotechnology; 28: 1321–1325.

Iwaki H, Nishimura A, Hasegawa Y. 2012b. *Tropicibacter phthalicus* sp. nov., a phthalate-degrading bacterium from seawater. Current Microbiology; 64: 392–396.

Jakob U, Muse W, Eser M, Bardwell JC. 1999. Chaperone activity with a redox switch. Cell; 96: 341–352.

Jami M, Lai Q, Ghanbari M, Moghadam MS, Kneifel W, Domig KJ. 2016. *Celeribacter persicus* sp. nov., a polycyclic-aromatic-hydrocarbon-degrading bacterium isolated from mangrove soil. International Journal of Systematic and Evolutionary Microbiology; 66: 1875–1880.

Jeffery SW, Humphrey GF. 1975. New spectrophotometric equations for determining chlorophylls *a*, *b*, c_1 and c_2 in higher plants, algae and natural phytoplankton. Biochemie und Physiologie der Pflanzen 167; 191–194.

Jensen LJ, Julien P, Kuhn M, von Mering C, Muller J, Doerks T, Bork P. 2008. eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Research; 36: D250–D254.

Jeong HI, Jin HM, Jeon CO. 2015. *Confluentimicrobium naphthalenivorans* sp. nov., a naphthalene-degrading bacterium isolated from sea-tidal-flat sediment, and emended description of the genus *Confluentimicrobium* Park *et al.* 2015. International Journal of Systematic and Evolutionary Microbiology; 65:4191–4195.

Jiménez JI, Miñambres B, García JL, Díaz E, 2002. Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. Environmental Microbiology; 4: 824–841.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016a. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research; 44: D457–D62.

Kanehisa M, Sato Y, Morishima K. 2016b. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. Journal of Molecular Biology; 428: 726–731.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols; 10: 845–858.

Kieboom J, Dennis JJ, de Bont JA, Zylstra GJ. 1998. Identification and molecular characterization of an efflux pump involved in *Pseudomonas putida* S12 solvent tolerance. Journal of Biological Chemistry; 273: 85–91.

Kim SJ, Park SJ, Cha IT, Min D, Kim JS, Chung WH, Chae JC, Jeon CO, Rhee SK. 2014. Metabolic versatility of toluene-degrading, iron-reducing bacteria in tidal flat sediment, characterized by stable isotope probing-based metagenomic analysis. Environmental Microbiology 16: 189–204.

Kimes NE, Callaghan AV, Aktas DF, Smith WL, Sunner J, Golding B, Drozdowska M, Hazen TC, Suflita JM, Morris PJ. 2013. Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. Frontiers in Microbiology 4: 50.

Kirk TK, Farrell RL .1987. Enzymatic ‘‘combustion’’: the microbial degradation of lignin. Annual Reviews of Microbiology; 41: 465–505.

Kisand V, Valente A, Lahm A, Tanet G, Lettieri T. 2012. Phylogenetic and functional metagenomic profiling for assessing microbial biodiversity in environmental monitoring. PLoS One; 7: e43630.

Kontur WS, Schackwitz WS, Ivanova N, Martin J, Labutti K, Deshpande S, Tice HN, Pennacchio C, Sodergren E, Weinstock GM, Noguera DR, Donohue TJ. 2012. Revised sequence and annotation of the *Rhodobacter sphaeroides* 2.4.1 genome. Journal of Bacteriology; 194: 7016–7017.

Kostka JE, Prakash O, Overholt WA, Green SJ, Freyer G, Canion A, Delgardio J, Norton N, Hazen TC, Huettel M. 2011. Hydrocarbon-degrading bacteria and the bacterial community response in Gulf of Mexico beach sands impacted by the Deepwater Horizon oil spill. Applied and Environmental Microbiology; 77: 7962–7974.

Krichko AA, Skvortsov DV, Titova TA, Filippo BS, Dogadkina NE. 1969. Production of tetralin by the hydrogenation of naphthalene-containing fractions. Chemistry and Technology of Fuels and Oils; 5: 18–22.

Kuczyńska-Wiśnik D1, Kedzierska S, Matuszewska E, Lund P, Taylor A, Lipińska B, Laskowska E. 2002. The *Escherichia coli* small heat-shock proteins IbpA and IbpB prevent the aggregation of endogenous proteins denatured *in vivo* during extreme heat shock. Microbiology; 148: 1757 – 1765.

Lai Q, Cao J, Yuan J, Li F, Shao Z. 2014. *Celeribacter indicus* sp. nov., a polycyclic aromatic hydrocarbon-degrading bacterium from deep-sea sediment and reclassification of *Huaishuia halophila* as *Celeribacter halophilus* comb. nov. International Journal of Systematic and Evolutionary Microbiology; 64: 4160–4167.

Lane DJ. 1991. 16S/23S rRNA Sequencing. *En* Nucleic acid techniques in bacterial systematics. (eds. Stackebrandt, E, Goodfellow, M). John Wiley and Sons Chichester United Kingdom. pp 115–175.

Lanfranconi MP, Bosch R, Nogales B. 2010. Short-term changes in composition of active marine bacterial assemblages in response to diesel oil pollution. Microbial Biotechnology; 3: 607–621.

Lanfranconi MP. 2010. Functional response of marine microbial communities to diesel pollution. Tesis doctoral. Universitat de les Illes Balears. 156 p.

Lang AS, Beatty JT. 2000. Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of *Rhodobacter capsulatus*. Proceedings of the National Academy of Sciences of the United States of America; 97: 859–864.

Leahy JG, Batchelor PJ, Morcomb SM. 2003. Evolution of the soluble diiron monooxygenases. FEMS Microbiology Reviews; 27: 449–479.

Letunic I, Doerks T, Bork P. 2015. SMART: recent updates, new developments and status in 2015. Nucleic Acids Research; 43: D257–D260.

Liang F, Lu M, Keener TC, Liu Z, Khang SJ. 2005. The organic composition of diesel particulate matter, diesel fuel and engine oil of a non-road diesel generator. Journal of Environmental Monitoring; 7: 983–988.

Lu Z, Deng Y, Van Nostrand JD, He Z, Voordeckers J, Zhou A, Lee YJ, Mason OU, Dubinsky EA, Chavarria KL, Tom LM, Fortney JL, Lamendella R, Jansson JK, D'haeseleer P, Hazen TC, Zhou J. 2012. Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. ISME Journal; 6: 451–60.

Luengo JM, García JL, Olivera ER. 2001. The phenylacetyl-CoA catabolon: a complex catabolic unit with broad biotechnological applications. Molecular Microbiology; 39: 1434–1442.

Luo H, Löytynoja A, Moran MA. 2012. Genome content of uncultivated marine *Roseobacters* in the surface ocean. Environmental Microbiology; 14: 41–51.

Luo H, Csuros M, Hughes AL, Moran MA. 2013. Evolution of divergent life history strategies in marine Alphaproteobacteria. mBio; 4: e00373-13.

Luo H, Swan BK, Stepanauskas R, Hughes AL, Moran MA. 2014. Evolutionary analysis of a streamlined lineage of surface ocean Roseobacters. ISME Journal; 8: 1428–1439.

Luo H, Moran MA. 2014. Evolutionary ecology of the marine *Roseobacter* clade. Microbiology and Molecular Biology Reviews; 78: 573–587.

Luo H, Moran MA. 2015. How do divergent ecological strategies emerge among marine bacterioplankton lineages? Trends in Microbiology; 23: 577–584.

Ma Y, Xiong H, Tang S, Yang Q, Li M. 2009. Comparison of the community structure of planktonic bacteria in ballast water from entryships and local sea water in Xiamen Port. Progress in Natural Science; 19: 947–953.

MacLean AM, MacPherson G, Aneja P, Finan TM. 2006. Characterization of the β-ketoadipate pathway in *Sinorhizobium meliloti*. Applied and Environmental Microbiology; 72: 5403–5413.

Manz W, Amann R, Ludwig W, Wagner M, Schelifer KH. 1992. Phylogenetic oligodeoxynucleotide probes for the major subclasses of *Proteobacteria*: problems and solutions. Systematic and Applied Microbiology; 15: 593–600.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Research; 43: D222–D226.

Marcos MS, Lozada M, Di Marzio WD, Dionisi HM. 2012. Abundance, dynamics, and biogeographic distribution of seven polycyclic aromatic hydrocarbon dioxygenase gene variants in coastal sediments of Patagonia. Applied and Environmental Microbiology; 78: 1589–1592.

Marín M, Plumeier I, Pieper DH. 2012. Degradation of 2,3-dihydroxybenzoate by a novel *meta*-cleavage pathway. Journal of Bacteriology; 194: 3851–3860.

Maruyama K, Shibayama T, Ichikawa A, Sakou Y, Yamada S, Sugisaki H. 2004. Cloning and characterization of the genes encoding enzymes for the protocatechuate *meta*-degradation

pathway of *Pseudomonas ochraceae* NGJ1. Bioscience, Biotechnology and Biochemistry; 68: 1434–1441.

Masai E, Shinohara S, Hara H, Nishikawa S, Katayama Y, Fukuda M. 1999. Genetic and biochemical characterization of a 2-pyrone-4,6-dicarboxylic acid hydrolase involved in the protocatechuate 4,5-cleavage pathway of *Sphingomonas paucimobilis* SYK-6. Journal of Bacteriology; 181: 55–62.

Mas-Lladó M. 2017. Proteogenomic and physiologic characterization of new members of Roseobacter lineage from Mallorca harbors. Tesis doctoral. Universitat de les Illes Balears. 156 p.

Mason OU, Hazen TC, Borglin S, Chain PS, Dubinsky EA, Fortney JL, Han J, Holman HY, Hultman J, Lamendella R, Mackelprang R, Malfatti S, Tom LM, Tringe SG, Woyke T, Zhou J, Rubin EM, Jansson JK. 2012. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. ISME Journal; 6: 1715–1727.

Matilla MA, Pizarro-Tobias P, Roca A, Fernández M, Duque E, Molina L, Wu X, van der Lelie D, Gómez MJ, Segura A, Ramos JL. 2011. Complete genome of the plant growth-promoting rhizobacterium *Pseudomonas putida* BIRD-1. Journal of Bacteriology; 193: 1290.

McKew BA, Coulon F, Osborn AM, Timmis KN, McGenity TJ. 2007. Determining the identity and roles of oil-metabolizing marine bacteria from the Thames estuary, UK. Environmental Microbiology; 9: 165–176.

McNutt MK, Camilli R, Crone TJ, Guthrie GD, Hsieh PA, Ryerson TB, Savas O, Shaffer F. 2012. Review of flow rate estimates of the Deepwater Horizon oil spill. Proceedings of the National Academy of Sciences of the United States of America; 109: 20260–20267.

Méndez V, Agulló L, González M, Seeger M. 2011. The homogentisate and homoprotocatechuate central pathways are involved in 3- and 4-hydroxyphenylacetate degradation by *Burkholderia xenovorans* LB400. PLoS One; 6: e17583.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA. 2008. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics; 9: 386.

Mikami K, Mori IC, Matsuura T, Ikeda Y, Kojima M, Sakakibara H, Hirayama T. 2016. Comprehensive quantification and genome survey reveal the presence of novel phytohormone action modes in red seaweeds. Journal of Applied Phycology; 28: 2539–2548.

Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. Genomics; 95:315–327.

Moran MA, Hodson RE. 1994. Dissolved humic substances of vascular plant origin in a coastal marine environment. Limnology and Oceanography; 39: 762–771.

Moran MA, González JM, Kiene RP. 2003. Linking a bacterial taxon to sulfur cycling in the sea: studies of the marine *Roseobacter* group. Geomicrobiology; 20: 375–388.

Moran MA, Buchan A, González JM, Heidelberg JF, Whitman WB, Kiene RP, Henriksen JR, King GM, Belas R, Fuqua C, Brinkac L, Lewis M, Johri S, Weaver B, Pai G, Eisen JA, Rahe E, Sheldon WM, Ye W, Miller TR, Carlton J, Rasko DA, Paulsen IT, Ren Q, Daugherty SC, Deboy RT, Dodson RJ, Durkin AS, Madupu R, Nelson WC, Sullivan SA, Rosovitz MJ, Haft DH, Selengut J, Ward N. 2004. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. Nature; 432: 910–913.

Moran MA, Belas R, Schell MA, González JM, Sun F, Sun S, Binder BJ, Edmonds J, Ye W, Orcutt B, Howard EC, Meile C, Palefsky W, Goesmann A, Ren Q, Paulsen I, Ulrich LE, Thompson LS, Saunders E, Buchan A. 2007. Ecological genomics of marine roseobacters. Applied and Environmental Microbiology; 73: 4559–4569.

Moran MA, Miller WL. 2007. Resourceful heterotrophs make the most of light in the coastal ocean. Nature Reviews Microbiology; 5: 792–800.

Moran MA, Satinsky B, Gifford SM, Luo H, Rivers A, Chan LK, Meng J, Durham BP, Shen C, Varaljay VA, Smith CB, Yager PL, Hopkinson BM. 2013. Sizing up metatranscriptomics. ISME Journal; 7: 237–243.

Mori IC, Ikeda Y, Matsuura T, Hirayama T, Mikami K. 2017. Phytohormones in red seaweeds: A technical review of methods for analysis and a consideration of genomic data. Botanica Marina; 60: 153–170.

Mosqueda G, Ramos JL. 2000. A set of genes encoding a second toluene efflux system in *Pseudomonas putida* DOT-T1E is linked to the *tod* genes for toluene metabolism. Journal of Bacteriology; 182: 937–943.

Mulligan C, Fischer M, Thomas GH. 2011. Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. FEMS Microbiol Reviews; 35: 68–86.

Namiki T, Hachiya T, Tanaka H, Sakakibara Y. 2012. MetaVelvet: an extension of Velvet assembler to *de novo* metagenome assembly from short sequence reads. Nucleic Acids Research; 40: e155.

NCBI Resource Coordinators. 2017. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Research; 45: D12–D17.

Nelson KE, Weinel C, Paulsen IT, Dodson RJ, Hilbert H, Martins dos Santos VA, Fouts DE, Gill SR, Pop M, Holmes M, Brinkac L, Beanan M, DeBoy RT, Daugherty S, Kolonay J, Madupu R, Nelson W, White O, Peterson J, Khouri H, Hance I, Chris Lee P, Holtzapple E, Scanlan D, Tran K, Moazzez A, Utterback T, Rizzo M, Lee K, Kosack D, Moestl D, Wedler H, Lauber J, Stjepandic D, Hoheisel J, Straetz M, Heim S, Kiewitz C, Eisen JA, Timmis KN, Düsterhöft A, Tümmler B, Fraser CM. 2002. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. Environmental Microbiology; 4: 799–808.

Nesemann K, Braus-Stromeyer SA, Thuermer A, Daniel R, Mavrodi DV, Thomashow LS, Weller DM, Braus GH. 2015. Draft genome sequence of the phenazine-producing *Pseudomonas fluorescens* strain 2-79. Genome Announcements; 3: e00130-15

Newton RJ, Griffin LE, Bowles KM, Meile C, Gifford S, Givens CE, Howard EC, King E, Oakley CA, Reisch CR, Rinta-Kanto JM, Sharma S, Sun S, Varaljay V, Vila-Costa M, Westrich JR, Moran MA. 2010. Genome characteristics of a generalist marine bacterial lineage. ISME Journal; 4: 784–798.

Nikaido H, Takatsuka Y. 2009. Mechanisms of RND multidrug efflux pumps. Biochimica et Biophysica Acta; 1794: 769–781.

Nogales B, Aguiló-Ferretjans MM, Martín-Cardona C, Lalucat J, Bosch R. 2007. Bacterial diversity, composition and dynamics in and around recreational coastal areas. Environmental Microbiology; 9: 1913–1929.

Nogales J, Macchi R, Franchi F, Barzaghi D, Fernández C, García JL, Bertoni G, Díaz E. 2007. Characterization of the last step of the aerobic phenylacetic acid degradation pathway. Microbiology; 153: 357–365.

Nogales B. 2010. Harbors and marinas. *En* Handbook of Hydrocarbon and Lipid Microbiology. (ed. Timmis, K.N.) Springer-Verlag Berlin Heidelberg. pp. 2361–2367.

Noinaj N, Guillier M, Barnard TJ, Buchanan SK. 2010. TonB-dependent transporters: regulation, structure, and function. Annual Review of Microbiology; 64: 43–60.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research; 27: 29–34.

Ohji S, Yamazoe A, Hosoyama A, Tsuchikane K, Ezaki T, Fujita N. 2014. The complete genome sequence of *Pseudomonas putida* NBRC 14164^{T} confirms high intraspecies variation. Genome Announcements; 2: e00029-14.

Olivera ER, Miñambres B, García B, Muñiz C, Moreno MA, Ferrández A, Díaz E, García JL, Luengo JM. 1998. Molecular characterization of the phenylacetic acid catabolic pathway in *Pseudomonas putida* U: the phenylacetyl-CoA catabolon. Proceedings of the National Academy of Sciences of the United States of America; 95: 6419–6424.

Osborn AM, Moore RB, Timmis KN. 2000. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environmental Microbiology; 2: 39–50.

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Research; 33: 5691–5702.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Research; 42: D206–D214.

Panke S, Witholt B, Schmid A, Wubbolts MG. 1998. Towards a biocatalyst for (*S*)-styrene oxide production: characterization of the styrene degradation pathway of *Pseudomonas* sp. strain VLB120. Applied and Environmental Microbiology; 64: 2032–2043.

Pao SS, Paulsen IT, Saier MH Jr. 1998. Major Facilitator Superfamily. Microbiology and Molecular Biology Reviews; 62: 1–34.

Park M, Jeon Y, Jang HH, Ro HS, Park W, Madsen EL, Jeon CO. 2007. Molecular and biochemical characterization of 3-hydroxybenzoate 6-hydroxylase from *Polaromonas naphthalenivorans* CJ2. Applied and Environmental Microbiology; 73: 5146–5152.

Parke D. 1996. Conservation of PcaQ, a transcriptional activator of *pca* genes for catabolism of phenolic compounds, in *Agrobacterium tumefaciens* and *Rhizobium* species. Journal of Bacteriology; 178: 3671–3675.

Parke D. 1997. Acquisition, reorganization, and merger of genes: novel management ofthe β-ketoadipate pathway in *Agrobacterium tumefaciens*. FEMS Microbiology Letters; 146: 3–12.

Parke D, D'Argenio DA, Ornston LN. 2000. Bacteria are not what they eat: that is why they are so diverse. Journal of Bacteriology; 182: 257–263.

Parke D, Ornston LN. 2003. Hydroxycinnamate (*hca*) catabolic genes from *Acinetobacter* sp. strain ADP1 are repressed by HcaR and are induced by hydroxycinnamoyl-coenzyme A thioesters. Applied and Environmental Microbiology; 69: 5398-5409.

Peng RH, Xiong AS, Xue Y, Fu XY, Gao F, Zhao W, Tian YS, Yao QH. 2008. Microbial biodegradation of polyaromatic hydrocarbons. FEMS Microbiology Reviews; 32: 927–955.

Pérez-Pantoja D, De la Iglesia R, Pieper DH, González B. 2008. Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium *Cupriavidus necator* JMP134. FEMS Microbiology Reviews; 32: 736–794.

Pérez-Pantoja D, González B, Pieper DH. 2010a. Aerobic degradation of aromatic hydrocarbons. *En* Handbook of Hydrocarbon and Lipid Microbiology (ed. Timmis, K.N.) Springer-Verlag Berlin Heidelberg. pp. 799–837.

Pérez-Pantoja D, Donoso R, Junca H, González B, Pieper DH. 2010b. Phylogenomics of aerobic bacterial degradation of aromatics. *En* Handbook of Hydrocarbon and Lipid Microbiology (ed. Timmis, K.N.) Springer-Verlag Berlin Heidelberg. pp. 1355–1397.

Pérez-Pantoja D, Donoso R, Agulló L, Córdova M, Seeger M, Pieper DH, González B. 2012. Genomic analysis of the potential for aromatic compounds biodegradation in *Burkholderiales*. Environmental Microbiology 14: 1091–1117.

Piña-Villalonga JM. 2012. Diversidad e importancia ecológica del grupo *Roseobacter* en aguas costeras sometidas a impacto antropogénico. Tesis doctoral. Universitat de les Illes Balears. 271 p.

Plaggenborg R, Overhage J, Steinbüchel A, Priefert H. 2003. Functional analyses of genes involved in the metabolism of ferulic acid in *Pseudomonas putida* KT2440. Applied Microbiology and Biotechnology; 61: 528–535.

Prabagaran SR, Manorama R, Delille D, Shivaji S. 2007. Predominance of *Roseobacter*, *Sulfitobacter*, *Glaciecola* and *Psychrobacter* in seawater collected off Ushuaia, Argentina, Sub-Antarctica. FEMS Microbiology Ecology; 59: 342–355.

Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, Kuehl JV, Melnyk RA, Lamson JS, Suh Y, Carlson HK, Esquivel Z, Sadeeshkumar H, Chakraborty R, Zane GM, Rubin BE, Wall JD, Visel A, Bristow J, Blow MJ, Arkin AP, Deutschbauer AM. 2018. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature; 557: 503–509.

Pruesse E, Quast C, Knittel K, Fuchs B.M, Ludwig W.G, Peplies J, Glöckner F.O. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research; 35: 7188–7196.

Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research; 35: D61–D65.

Psarros G, Skjong R, Vanem E. 2011. Risk acceptance criterion for tanker oil spill risk reduction measures. Marine Pollution Bulletin; 62:116–127.

Ramos JL, Marqués S, Timmis KN. 1997. Transcriptional control of the *Pseudomonas* TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators. Annual Reviews of Microbiology; 51: 341–373.

Ramos JL, Duque E, Godoy P, Segura A. 1998. Efflux pumps involved in toluene tolerance in *Pseudomonas putida* DOT-T1E. Journal of Bacteriology; 180: 3323–3329.

Ramos JL, Duque E, Gallegos MT, Godoy P, Ramos-González MI, Rojas A, Terán W, Segura A. 2002. Mechanisms of solvent tolerance in gram-negative bacteria. Annual Reviews of Microbiology; 56: 743–768.

Ramos JL, Sol Cuenca M, Molina-Santiago C, Segura A, Duque E, Gómez-García MR, Udaondo Z, Roca A. 2015. Mechanisms of solvent resistance mediated by interplay of cellular factors in *Pseudomonas putida*. FEMS Microbiology Reviews; 39: 555–566.

Redondo-Nieto M, Barret M, Morrisey JP, Germaine K, Martínez-Granero F, Barahona E, Navazo A, Sánchez-Contreras M, Moynihan JA, Giddens SR, Coppoolse ER, Muriel C, Stiekema WJ, Rainey PB, Dowling D, O'Gara F, Martín M, Rivilla R. 2012. Genome sequence of the biocontrol strain *Pseudomonas fluorescens* F113. Journal of Bacteriology; 194: 1273–1274.

Reisch CR, Moran MA, Whitman WB. 2011. Bacterial catabolism of dimethylsulfoniopropionate (DMSP). Frontiers in Microbiology; 2: 172.

Rojas A, Duque E, Mosqueda G, Golden G, Hurtado A, Ramos JL, Segura A. 2001. Three efflux pumps are required to provide efficient tolerance to toluene in *Pseudomonas putida* DOT-T1E. Journal of Bacteriology; 183: 3967–3973.

Roma-Rodrigues C, Santos PM, Benndorf D, Rapp E, Sá-Correia I. 2010. Response of *Pseudomonas putida* KT2440 to phenol at the level of membrane proteome. Journal of Proteomics; 73: 1461–1478.

Saibil H. 2013. Chaperone machines for protein folding, unfolding and disaggregation. Nature Reviews Molecular Cell Biology; 14: 630–642

Sambrook J, Russel DW. 2001. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory Press, New York. pp. 1659.

Sarna LP, Barrie Webster GR, Friesen-Fischer MR, Sri Ranjan R. 1994. Analysis of the petroleum components benzene, toluene, ethyl benzene and the xylenes in water by commercially available solid-phase microextraction and carbon-layer open tubular capillary column gas chromatography. Journal of Chromatography A; 677: 201–205.

Schäfer A, Tauch A, Jager W, Kalinowski J, Thierbach G, Pühler A. 1994. Small mobilizable muti-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: Selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene; 145: 69–73.

Schettler T. 2006. Human exposure to phthalates via consumer products. International Journal of Andrology; 29: 134–139.

Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America; 95: 5857–5864.

Segura A, Godoy P, van Dillewijn P, Hurtado A, Arroyo N, Santacruz S, Ramos JL. 2005. Proteomic analysis reveals the participation of energy- and stress-related proteins in the response of *Pseudomonas putida* DOT-T1E to toluene. Journal of Bacteriology; 187: 5937–5945.

Segura A, Hurtado A, Rivera B Lazaroaie MM. 2007. Isolation of new toluene-tolerant marine strains of bacteria and characterization of their solvent-tolerance properties. Journal of Applied Microbiology; 104: 1408–1416.

Segura A, Molina L, Fillet S, Krell T, Bernal P, Muñoz-Rojas J, Ramos JL. 2012. Solvent tolerance in Gram-negative bacteria. Current Opinion in Biotechnology; 23: 415–421.

Shintani M, Urata M, Inoue K, Eto K, Habe H, Omori T, Yamane H, Nojiri H. 2007. The *Sphingomonas* plasmid pCAR3 is involved in complete mineralization of carbazole. Journal of Bacteriology; 189: 2007–2020.

Simon R, Priefer U, Pühler A. 1983. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram negative bacteria. Bio/Technology; 1: 784–791.

Simon C, Daniel R. 2011. Metagenomic analyses: past and future trends. Applied and Environmental Microbiology; 77: 1153–1161.

Skerman VBD, McGowan V, Sneath PHA (editors). 1980. Approved Lists of Bacterial Names. International Journal of Systematic Bacteriology; 30: 225–420.

Sleighter RL, Hatcher PG. 2008. Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Marine Chemistry; 110: 140–152.

Suárez-Suárez LY, Brunet-Galmes I, Piña-Villalonga JM, Christie-Oleza JA, Peña A, Bennasar A, Armengaud J, Nogales B, Bosch R. 2012. Draft genome sequence of *Citreicella aestuarii* strain 357, a member of the *Roseobacter* clade isolated without xenobiotic pressure from a petroleum-polluted beach. Journal of Bacteriology; 194: 5464–5465.

Suárez-Suárez LY. 2013. Caracterización genómica y fisiológica de la capacidad degradadora de hidrocarburos aromáticos de *Citreicella aestuarii* 357. Tesis doctoral. Universitat de les Illes Balears. 139 p.

Sugimoto K, Senda T, Aoshima H, Masai E, Fukuda M, Mitsui Y. 1999. Crystal structure of an aromatic ring opening dioxygenase LigAB, a protocatechuate 4,5-dioxygenase, under aerobic conditions. Structure; 7: 953–965.

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S; Tara Oceans coordinators, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P. 2015. Ocean plankton. Structure and function of the global ocean microbiome. Science; 348: 1261359.

Swan BK, Tupper B, Sczyrba A, Lauro FM, Martínez-García M, González JM, Luo H, Wright JJ, Landry ZC, Hanson HW, Thompson BP, Poulton NJ, Schwientek P, Acinas SG, Giovannoni SJ, Moran MA, Hallam SJ, Cavicchioli R, Woyke T, Stepanauskas R. 2013. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proceedings of the National Academy of Sciences of the United States of America; 110: 11463–11468.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Research; 43: D447–D452.

Tang K, Feng X, Tang Y, Blankenship RE. 2009. Carbohydrate metabolism and carbon fixation in *Roseobacter denitrificans* OCh114. PLoS One; 4: e7233.

Tang K, Jiao N, Liu K, Zhang Y, Li S. 2012. Distribution and functions of TonB-dependent transporters in marine bacteria and environments: implications for dissolved organic matter utilization. PLoS One; 7: e41204.

Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Research; 28: 33–36.

Teufel R, Mascaraque V, Ismail W, Voss M, Perera J, Eisenreich W, Haehnel W, Fuchs G. 2010. Bacterial phenylalanine and phenylacetate catabolic pathway revealed. Proceedings of the National Academy of Science of the United States of America; 107: 14390–14395

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acid Research; 25: 4876–4882.

Tropel D, van der Meer JR. 2004. Bacterial transcriptional regulators for degradation pathways of aromatic compounds. Microbiology and Molecular Biology Reviews; 68: 474–500.

Udaondo Z, Duque E, Fernández M, Molina L, de la Torre J, Bernal P, Niqui JL, Pini C, Roca A, Matilla MA, Molina-Henares MA, Silva-Jiménez H, Navarro-Avilés G, Busch A, Lacal J, Krell T, Segura A, Ramos JL. 2012. Analysis of solvent tolerance in *Pseudomonas putida* DOT-T1E based on its genome sequence and a collection of mutants. FEBS Letters; 586: 2932–2938.

Valderrama JA, Durante-Rodríguez G, Blázquez B, García JL, Carmona M, Díaz E. 2012. Bacterial degradation of benzoate: cross-regulation between aerobic and anaerobic pathways. Journal of Biological Chemistry; 287: 10494–10508.

van Berkel WJ, Kamerbeek NM, Fraaije MW. 2006. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. Journal of Biotechnology; 124: 670–689.

van Gorcom RF, Boschloo JG, Kuijvenhoven A, Lange J, van Vark AJ, Bos CJ, van Balken JA, Pouwels PH, van den Hondel CA. 1990. Isolation and molecular characterisation of the benzoate-para-hydroxylase gene (*bphA*) of *Aspergillus niger*: a member of a new gene family of the cytochrome P450 superfamily. Molecular and General Genetics; 223: 192–197.

Velázquez F, Parro V, de Lorenzo V. 2005. Inferring the genetic network of *m*-xylene metabolism through expression profiling of the *xyl* genes of *Pseudomonas putida* mt-2. Molecular Microbiology; 57: 1557–1569.

von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. 2003. STRING: a database of predicted functional associations between proteins. Nucleic Acids Research; 31: 258–261.

Wagner-Döbler I, Biebl H. 2006. Environmental biology of the marine *Roseobacter* lineage. Annual Reviews of Microbiology; 60: 255–280.

Wang YZ, Zhou Y, Zylstra GJ. 1995. Molecular analysis of isophthalate and terephthalate degradation by *Comamonas testosteroni* YZW-D. Environmental Health Perspectives; 103: 9–12.

Wang YZ, Lipscomb JD. 1997. Cloning, overexpression, and mutagenesis of the gene for homoprotocatechuate 2,3-dioxygenase from *Brevibacterium fuscum*. Protein Expresion and Purification; 10: 1–9.

Yakimov MM, Timmis KN, Golyshin PN. 2007. Obligate oil-degrading marine bacteria. Current Opinion in Biotechnology; 18: 257–266.

Yang H, Wang L, Xie Z, Tian Y, Liu G, Tan H. 2007. The tyrosine degradation gene *hppD* is transcriptionally activated by HpdA and repressed by HpdR in *Streptomyces coelicolor*, while *hpdA* is negatively autoregulated and repressed by HpdR. Molecular Microbiology; 65: 1064–1077.

Yoch DC. 2002. Dimethylsulfoniopropionate: its sources, role in the marine food web, and biological degradation to dimethylsulfide. Applied and Environmental Microbiology; 68: 5804–5815.

Zaar A, Eisenreich W, Bacher A, Fuchs G. 2001. A novel pathway of aerobic benzoate catabolism in the bacteria *Azoarcus evansii* and *Bacillus stearothermophilus*. Journal of Biological Chemistry; 276: 24997–25004.

Zaprasis A, Brill J, Thüring M, Wünsche G, Heun M, Barzantny H, Hoffmann T, Bremer E. 2013. Osmoprotection of *Bacillus subtilis* through import and proteolysis of proline-containing peptides. Applied and Environmental Microbiology; 79: 576–87.

Zarraonaindia I, Smith DP, Gilbert JA. 2013. Beyond the genome: community-level analysis of the microbial world. Biology and Phylosophy; 28: 261–282.

Zhang R, Liu B, Lau SCK, Ki JS, Qian PY. 2007. Particle-attached and free-living bacterial communities in a contrasting marine environment: Victoria Harbor, Hong Kong. FEMS Microbiology Ecology; 61: 496–508.

Zhang Y, Sun Y, Jiao N, Stepanauskas R, Luo H. 2016. Ecological genomics of the uncultivated marine *Roseobacter* lineage CHAB-I-5. Applied and Environmental Microbiology; 82: 2100–2111.

Zhou NY, Fuenmayor SL, Williams PA. 2001. *nag* genes of *Ralstonia* (formerly *Pseudomonas*) sp. strain U2 encoding enzymes for gentisate catabolism. Journal of Bacteriology; 183: 700–708.

Zhuang Z, Song F, Zhang W, Taylor K, Archambault A, Dunaway-Mariano D, Dong J, Carey PR. 2002. Kinetic, Raman, NMR, and site-directed mutagenesis studies of the *Pseudomonas* sp. strain CBS3 4-hydroxybenzoyl-CoA thioesterase active site. Biochemistry; 41(37): 11152–11160.

ANEXOS

ANEXO 1: FIGURAS SUPLEMENTARIAS

LISTADO DE FIGURAS SUPLEMENTARIAS

Todas las figuras suplementarias de esta tesis se encuentran disponibles en el siguiente enlace: https://www.dropbox.com/sh/iud7rp0yo8tqzaj/AADpc71tbFRYWJrwAnumVuwJa?dl=0

Figura S2.1. Cambios en el perfil proteico de *Sulfitobacter* sp. 3SOLIMAR09 tras la exposición a hidrocarburos.

Figura 2.2. Filogenia de las proteínas homólogas a SPO2567 de *Ruegeria pomeroyi* DSS-3^T.

Figura S2.3. Filogenia de las proteínas homólogas a SPO2567 de *Ruegeria pomeroyi* DSS-3^T en el linaje Roseobacter.

Figura S2.4. Predicción de la estructura de la proteína SPO2567 de *R. pomeroyi* DSS-3^T.

Figura S2.5. Variación del número de células de *R. pomeroyi* DSS-3^T silvestre y del mutante SPO2567$^-$ tras 48h del inicio de los tratamientos.

Figura S3.1. Análisis de limitación de nitrógeno tras 24 h de incubación en condiciones control y con adición de $NaNO_3$ 16µM.

Figura S3.2. Características del experimento de microcosmos M2 de abril de 2014.

Figura S3.3. Similitud entre las comunidades microbianas del experimento de microcosmos M2 de 2014.

Figura S3.4. Curvas de rarefacción de los ocho metagenomas generados calculadas por MG-RAST (Meyer *et al*., 2008).

ANEXO 2: TABLAS SUPLEMENTARIAS

LISTADO DE TABLAS SUPLEMENTARIAS

Todas las tablas suplementarias de esta tesis se encuentran disponibles en el siguiente enlace: https://www.dropbox.com/sh/iud7rp0yo8tqzaj/AADpc71tbFRYWJrwAnumVuwJa?dl=0

Tabla S1.1. Características de los genomas estudiados.

Tabla S1.2. Detección de genes codificantes para enzimas de las rutas centrales de compuestos monoaromáticos estudiadas, sus reguladores, transportadores y proteínas conservadas asociadas a estas rutas.

Tabla S2.1. Detección de los transportadores de tipo RND descritos en la literatura implicados en tolerancia a solventes (Ramos *et al.*, 2002) en los miembros del linaje Roseobacter.

Tabla S2.2. Proteínas detectadas en el proteoma total de *Ruegeria pomeroyi* DSS-3^T tras 24h de exposición a naftaleno 0,1 % (p/v).

Tabla S2.3. Proteínas detectadas en el proteoma total de *Sulfitobacter* sp. 3SOLIMAR09 tras 24 h de exposición a naftaleno 0,1 % (p/v).

Tabla S2.4. Proteínas detectadas en el proteoma total de *Citreicella aestuarii* 357 tras 24 h de exposición a naftaleno 0,1 % (p/v).

Tabla S2.5. Proteínas detectadas en el proteoma fraccionado de *Ruegeria pomeroyi* DSS-3^T tras 24 h de exposición a naftaleno 0,1 % (p/v).

Tabla S2.6. Proteínas detectadas en el proteoma fraccionado de *Sulfitobacter* sp. 3SOLIMAR09 tras 24 h de exposición a naftaleno 0,1 % (p/v).

Tabla S2.7. Proteínas detectadas en el proteoma fraccionado de *Citreicella aestuarii* 357 tras 24 h de exposición a naftaleno 0,1 % (p/v).

Tabla S2.8. Proteínas detectadas en el proteoma fraccionado de *Ruegeria pomeroyi* DSS-3^T tras 3 h de exposición a tolueno 0,3 % (v/v).

Tabla S2.9. Proteínas detectadas en el proteoma fraccionado de *Sulfitobacter* sp. 3SOLIMAR09 tras 3 h de exposición a tolueno 0,3 % (v/v).

Tabla S2.10. Proteínas detectadas en el proteoma fraccionado de *Citreicella aestuarii* 357 tras 3 h de exposición a tolueno 0,3 % (v/v).

Tabla S2.11. Grupos de proteínas de *Ruegeria pomeroyi* DSS-3^T, *Sulfitobacter* sp. 3SOLIMAR09 y *Citreicella aestuarii* 357 con un mayor peso en el análisis de componentes principales (PCA) en el experimento de tolerancia a tolueno.

Tabla S2.12. Clasificación en categorías funcionales de las proteínas significativas con función asignada del proteoma fraccionado tras exposición a tolueno.

Tabla S2.13. Proteínas relacionadas con SPO2567 en *R. pomeroyi* DSS-3^T según el análisis realizado con la herramienta STRING (von Mering *et al*., 2003; Szklarczyk D *et al*., 2015).

Tabla S2.14. Clasificación de las proteínas hipotéticas de *R. pomeroyi* DSS-3^T relacionadas con SPO2567 mediante el Conserved Domain Search del NCBI (Marchler-Bauer *et al*., 2015)

Tabla S3.1. Caracterización de las muestras de agua.

Tabla S3.2. Índices de diversidad de secuencias en los metagenomas.

Tabla S3.3. Porcentaje de abundancia relativa de las secuencias asignadas a familias del dominio *Bacteria*

Tabla S3.4. Abundancia relativa de las lecturas asignadas a la categoría de metabolismo de ácidos grasos, lípidos e isoprenoides de Subsistemas SEED (Overbeek *et al.,* 2005; 2014).

Tabla S3.5. Abundancia relativa de las lecturas asignadas a la categoría de cofactores, vitaminas, grupos prostéticos y pigmentos de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.6. Abundancia relativa de las lecturas asignadas a la categoría de metabolismo de carbohidratos de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.7. Abundancia relativa de las lecturas asignadas a la categoría de metabolismo de proteínas de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.8. Abundancia relativa de las lecturas asignadas a la categoría de aminoácidos y derivados de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.9. Abundancia relativa de las lecturas asignadas a la categoría de metabolismo de ADN de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.10. Abundancia relativa de las lecturas asignadas a la categoría de división y ciclo celular de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.11. Abundancia relativa de las lecturas asignadas a la categoría de captación y metabolismo del hierro de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.12. Abundancia relativa de las lecturas asignadas a la categoría de transporte de membrana de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.13. Abundancia relativa de las lecturas asignadas a la categoría de respuesta a estrés de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.14. Abundancia relativa de las lecturas asignadas a la categoría de fagos, profagos, elementos transponibles y plásmidos de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.15. Abundancia relativa de las lecturas asignadas a la categoría de metabolismo de compuestos aromáticos de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.16. Abundancia relativa de las lecturas asignadas a la categoría de subsistemas basados en agrupaciones de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.17. Abundancia relativa de las lecturas asignadas a la categoría diversa de Subsistemas SEED (Overbeek *et al*., 2005; 2014).

Tabla S3.18. Abundancia relativa de las lecturas asignadas a los grupos de enzimas de AromaDeg en los ocho metagenomas generados.

Tabla S3.19. Asignación taxonómica de las lecturas asignadas a grupos de enzimas de AromaDeg que diferencian los metagenomas de microcosmos T2C1 2013 y T2D1 2013.

Tabla S3.20. Asignación taxonómica de las lecturas asignadas a grupos de enzimas de AromaDeg que diferencian los metagenomas de microcosmos T3C2 2013 y T3D2 2013.

Tabla S3.21. Asignación taxonómica de las lecturas asignadas a grupos de enzimas de AromaDeg que diferencian los metagenomas de microcosmos T3C1 2014 y T3D1 2014.