

TESIS DOCTORAL

**DESARROLLO DE SISTEMAS DE EXPRESIÓN
PARA ANÁLISIS METAGENÓMICOS FUNCIONALES E
IDENTIFICACIÓN DE ENZIMAS DE INTERÉS**

Memoria presentada por Laura Terrón González para optar al grado de Doctora
Sevilla, junio de 2014.

El Director de la Tesis

La Doctoranda

Dr. Eduardo Santero Santurino

Laura Terrón González

UNIVERSIDAD PABLO DE OLAVIDE
Dpto. Biología Molecular e Ingeniería Bioquímica

Quiero dar las gracias a mucha gente...

Primero a Eduardo, por acogerme en el grupo, por tu paciencia y enseñarme tanto estos años. Gracias por haberme guiado en esta tesis y por sacar tiempo, gracias por el carrerón final.

A los aspirinos, mi subgrupo de acogida. Carlos, contigo empecé a cacharrear en micro. Gracias también por tu ayuda con el citómetro. A Eva, por tu gran ayuda al principio con el diseño de las clonaciones, entre otras cosas, y seguir resolviendo mis preguntas después. A Amando, gracias también por tu ayuda cuando me ha hecho falta. A Bea, por escuchar mis paranoias no solo científicas y por tu ayuda y tus ánimos sobre todo en la escritura. Van ya unos cuantos años... Espero que nos queden muchos más viajes.

A Guada, mi compi metagenómica en los últimos años, muchas gracias por tu ayuda con los experimentos con el HPLC, ensayos varios y rastreos, por intentar que viera las cosas de otra manera cuando me hacía falta y por los ratitos de desconexión. Seguro que nos quedan más momentos juntas. A Elena, por aguantarme tanto, muchas gracias por las charlas y ayuda con los experimentos cada vez que lo necesitaba, por tus ánimos y por los ratos fuera del laboratorio. Tenemos pendiente una celebración de las gordas, y todo lo que venga después. A Ana Grande y Cris, por haber estado siempre dispuestas a ayudar y a reírnos un rato, y seguir ahí en la distancia. A Ana Chica y Laura Chica, por tantas horas de laboratorio y buenos recuerdos que tengo con vosotras fuera también. A Aroa por tus consejos y creatividad, gracias por la portada, a Laura Grande por tu ayuda cuando tenía alguna pregunta, a Helena por la alegría que transmites, a Inma por tu interés en solucionar nuestras dudas bioinformáticas, a Sofía por los momentos de risa absurda en el laboratorio, a Nuria, Encarni, Ali y todas las demás niñas con las que he coincidido estos años.

A Fernan, por contestar cualquier pregunta y estar siempre tan dispuesto a ayudar. A Inés, por tu ayuda cuando me ha hecho falta, más cerquita estos últimos años. A Paqui, por mostrar interés en cómo van las cosas, solucionar dudas más de una vez y mantener “la tradición”. A Belén por estar también siempre dispuesta a resolver mis preguntas.

¡Cómo ha pasado el tiempo! He estado muy a gusto en Micro, a pesar de los momentos de estrés... Os voy a echar de menos.

Al resto de gente del CABD, a todos con los que me he cruzado estos años. En especial a Genética, donde empecé a cogerle el gustillo a trabajar en un laboratorio, sobre todo a mis

profes del proyecto fin de carrera, a Manolo, que hiciste que me planteara hacer la tesis, y a Ramón.

Al Ministerio por entonces de Educación y Ciencia, por la beca FPU que recibí. Al proyecto Consolider y a toda la gente con la que he coincidido. A Manuel Gómez y Manuel Ferrer por la anotación de los clones metagenómicos.

A mis amigos de fuera del laboratorio, por hacer que desconectara de esto cuando me hacía falta y escuchar mis quejas y agobios.

A toda mi familia, a los que están y a los que ya no, aunque siguen presentes, por su apoyo. A mi hermano Juan por el soporte informático cuando me desesperaba con el ordenador. A mis padres.

GRACIAS A TODOS

ÍNDICE

| | página |
|--|-----------|
| INTRODUCCIÓN..... | 1 |
| 1. DIVERSIDAD MICROBIANA..... | 3 |
| 2. METAGENÓMICA..... | 5 |
| 2.1 Construcción de metagenotecas..... | 6 |
| 2.1.1 Selección y recogida de la muestra..... | 7 |
| 2.1.2 Extracción y tratamiento del ADN..... | 8 |
| 2.1.3 Clonación en el vector adecuado..... | 10 |
| 2.1.4 Introducción de la metagenoteca en bacterias hospedadoras..... | 12 |
| 2.2 Identificación de actividades de interés en metagenotecas..... | 13 |
| 2.2.1 Análisis basados en inducción de la expresión génica..... | 16 |
| 2.2.2 Análisis basados en secuencia..... | 17 |
| 2.2.3 Análisis basados en función..... | 20 |
| 3. MEJORAS EN METAGENÓMICA FUNCIONAL..... | 22 |
| 3.1 Desarrollo de sistemas de rastreo..... | 23 |
| 3.2 Estrategias para el aumento de las probabilidades de expresión del ADN metagenómico..... | 23 |
| 3.2.1 Desarrollo de sistemas de expresión homóloga..... | 24 |
| 3.2.2 Desarrollo de sistemas de expresión heteróloga..... | 27 |
| 4. RESISTENCIAS A ANTIBIÓTICOS BETALACTÁMICOS..... | 30 |
| 5. DEGRADACIÓN DE CONTAMINANTES AROMÁTICOS..... | 31 |
| 5.1 Oxigenasas hidroxilantes de anillos aromáticos..... | 33 |
| 5.2 Dioxigenasas extradiólicas..... | 36 |
| 6. CELULASAS..... | 37 |
| | |
| OBJETIVOS..... | 39 |
| | |
| MATERIALES Y MÉTODOS..... | 43 |
| 1. ESTIRPES, PLÁSMIDOS Y OLIGONUCLEÓTIDOS..... | 45 |

| | |
|---|-----------|
| 2. MEDIOS, CONDICIONES DE CULTIVO Y CONSERVACIÓN..... | 50 |
| 2.1 Medios y condiciones de cultivo..... | 50 |
| 2.2 Conservación del material biológico..... | 51 |
| 3. MANIPULACIÓN DE ADN..... | 51 |
| 3.1 Extracción de ADN plasmídico de <i>Escherichia coli</i> | 51 |
| 3.2 Amplificación de ADN mediante la reacción en cadena de la polimerasa (PCR)..... | 52 |
| 3.3 Asociación de oligonucleótidos complementarios..... | 53 |
| 3.4 PCR recurrente..... | 53 |
| 3.5 Mutagénesis dirigida por PCR (PCR solapante)..... | 54 |
| 3.6 Digestión y modificación de fragmentos de ADN..... | 55 |
| 3.7 Electroforesis de ADN..... | 56 |
| 3.7.1 Electroforesis en geles de agarosa..... | 56 |
| 3.7.2 Electroforesis en geles de poliacrilamida..... | 57 |
| 3.8 Aislamiento y purificación de fragmentos de ADN..... | 57 |
| 3.9 Cuantificación de ADN..... | 58 |
| 3.10 Clonación de fragmentos de ADN en vectores..... | 58 |
| 3.11 Secuenciación de ADN y análisis de secuencias..... | 59 |
| 4. TRANSFERENCIA DE ADN A ESTIRPES BACTERIANAS..... | 59 |
| 4.1 Transformación..... | 59 |
| 4.1.1 Transformación de <i>Escherichia coli</i> mediante choque térmico..... | 59 |
| 4.1.1.1 Preparación de células competentes..... | 59 |
| 4.1.1.2 Transformación..... | 60 |
| 4.1.2 Transformación de <i>Escherichia coli</i> mediante electroporación..... | 61 |
| 4.1.2.1 Preparación de células electrocompetentes..... | 61 |
| 4.1.2.2 Transformación..... | 61 |
| 4.2 Conjugación..... | 61 |
| 4.2.1 Conjugación triparental..... | 61 |
| 5. PLÁSMIDOS CONSTRUIDOS EN ESTA TESIS..... | 63 |
| 5.1 Vectores para metagenotecas..... | 63 |
| 5.2 Subclones con gen(es) metagenómico(s)..... | 66 |

| | |
|--|-----------|
| 5.3 Fusiones traduccionales a <i>lacZ</i> y transcripcionales a <i>gfp</i> | 67 |
| 6. CONSTRUCCIÓN DE ESTIRPES..... | 69 |
| 6.1 Introducción de genes en el cromosoma utilizando productos de PCR..... | 69 |
| 6.1.1 Preparación del ADN para la sustitución..... | 70 |
| 6.1.2 Transformación de los receptores..... | 71 |
| 6.1.3 Comprobación de candidatos..... | 71 |
| 6.1.4 Escisión del gen de resistencia a antibiótico..... | 72 |
| 7. ESTIRPES CONSTRUIDAS EN ESTA TESIS..... | 72 |
| 7.1 Selección de mutantes resistentes espontáneos a antibióticos..... | 75 |
| 8. CONSTRUCCIÓN DE METAGENOTECAS..... | 76 |
| 8.1 Recogida de muestras..... | 76 |
| 8.2 Separación de la fracción bacteriana..... | 76 |
| 8.3 Extracción del ADN..... | 77 |
| 8.4 Preparación del ADN..... | 77 |
| 8.5 Ligación del ADN metagenómico con el vector pMPO579..... | 78 |
| 8.6 Titulación de la metagenoteca..... | 79 |
| 8.7 Empaquetado de las ligaciones..... | 80 |
| 8.8 Infección en masa y almacenamiento de la metagenoteca..... | 80 |
| 9. RASTREO DE ACTIVIDADES ENZIMÁTICAS..... | 81 |
| 9.1 Rastreo de resistencias a antibióticos..... | 82 |
| 9.2 Rastreo de dioxigenasas extradiólicas..... | 83 |
| 9.3 Rastreo de oxigenasas hidroxilantes de anillos aromáticos..... | 84 |
| 9.4 Rastreo de endoglucanasas..... | 84 |
| 9.5 Rastreo de β -glucosidasas..... | 85 |
| 10. SECUENCIACIÓN DE CLONES Y ANÁLISIS DE SECUENCIAS..... | 86 |
| 11. ANÁLISIS DE PROTEÍNAS..... | 87 |
| 11.1 Superproducción de proteínas..... | 87 |
| 11.2 Obtención de extractos totales de cultivos bacterianos..... | 87 |
| 11.3 Electroforesis de proteínas..... | 88 |
| 12. ANÁLISIS DE LA EXPRESIÓN GÉNICA..... | 89 |

| | |
|---|------------|
| 12.1 Ensayos de medida del nivel de fluorescencia..... | 89 |
| 12.2 Ensayos de actividad β -galactosidasa..... | 90 |
| 13. ENSAYOS DE ACTIVIDAD ENZIMÁTICA DIOXIGENASA EXTRADIÓLICA..... | 91 |
| 14. CITOMETRÍA DE FLUJO..... | 93 |
| 15. ENSAYO COLORIMÉTRICO PARA LA DETECCIÓN DE NAFTOLES.. | 94 |
| 16. ENSAYOS MEDIANTE HPLC..... | 95 |
| | |
| RESULTADOS..... | 97 |
| 1. CONSTRUCCIÓN Y VALIDACIÓN DE SISTEMAS DE EXPRESIÓN HETERÓLOGA EN VECTORES PARA METAGENOTECAS..... | 99 |
| 1.1 Construcción de los sistemas de expresión..... | 99 |
| 1.1.1 Construcción de vectores para metagenotecas..... | 99 |
| 1.1.2 Construcción de estirpes hospedadoras..... | 103 |
| 1.2 Validación de los fósmidos de expresión..... | 106 |
| 1.2.1 Análisis de la capacidad de conjugación del fósrido..... | 106 |
| 1.2.2 Medida de niveles de expresión desde el vector..... | 107 |
| 1.2.3 Identificación de distintos niveles de fluorescencia y enriquecimiento de clones mediante la tecnología SIGEX..... | 112 |
| 1.2.4 Construcción de una metagenoteca de suelo de Punta de San García..... | 114 |
| 1.2.5 Rastreo comparativo de resistencias a antibióticos con y sin la utilización de los sistemas de expresión heteróloga..... | 115 |
| 2. CARACTERIZACIÓN DE LOS CLONES CON RESISTENCIAS A CARBENICILINA..... | 121 |
| 2.1 Secuenciación de los clones con resistencias a carbenicilina..... | 121 |
| 2.2 Betalactamasa de clase C y bomba de eflujo del clon ETN1..... | 122 |
| 2.3 Betalactamasa de clase B del clon TN2..... | 123 |
| 2.4 Bomba de eflujo de los clones TN3 y TN4..... | 125 |
| 2.5 Betalactamasa de clase B de los clones N5 y N6..... | 125 |
| 3. IDENTIFICACIÓN DE ACTIVIDADES DE INTERÉS..... | 129 |
| 3.1 Construcción de metagenoteca de suelo con historia de contaminación..... | 129 |
| 3.2 Rastreo de dioxigenasas extradiólicas..... | 130 |

| | |
|--|------------|
| 3.3 Rastreo de oxigenasas hidroxilantes de anillos aromáticos..... | 132 |
| 3.4 Rastreo de endoglucanasas..... | 135 |
| 3.5 Rastreo de β -glucosidasas..... | 137 |
| 4. CARACTERIZACIÓN DE CLONES CON ACTIVIDAD DIOXIGENASA EXTRADIÓLICA..... | 139 |
| 4.1 Secuenciación de clones con actividad dioxigenasa extradiólica..... | 139 |
| 4.1.1 Identificación de genes de dioxigenasas extradiólicas y asignación a clones..... | 140 |
| 4.1.2 Análisis filogenético de las dioxigenasas extradiólicas identificadas..... | 142 |
| 4.2 Ensayos de actividad dioxigenasa extradiólica con distintos sustratos..... | 149 |
| 4.3 Subclonación del gen upoed11..... | 154 |
| 5. CARACTERIZACIÓN DE CLONES CON ACTIVIDAD OXIGENASA HIDROXILANTE DE ANILLOS AROMÁTICOS..... | 157 |
| 5.1 Secuenciación de clones con actividad oxigenasa..... | 157 |
| 5.1.1 Identificación de genes de oxigenasas y asignación a clones..... | 158 |
| 5.1.1.1 Dioxigenasas..... | 161 |
| 5.1.1.2 Monooxigenasas..... | 163 |
| 5.1.2 Análisis filogenético de los citocromos P450..... | 166 |
| 5.2 Detección de monooxigenasas productoras de 1-naftol..... | 169 |
| 5.2.1 Ensayo colorimétrico para la detección de naftoles..... | 169 |
| 5.2.2 Ensayo mediante cromatografía líquida de alta resolución para la detección de 1-naftol..... | 170 |
| 5.3 Ruta de desulfuración de dibenzotiofeno..... | 172 |
| 5.3.1 Subclonación de los genes de la ruta de desulfuración de dibenzotiofeno..... | 175 |
| 5.3.1.1 Activador transcripcional de promotores dependientes de σ^N | 176 |
| 5.3.2 Caracterización de la expresión de los genes de la ruta de desulfuración de dibenzotiofeno..... | 180 |
| DISCUSIÓN..... | 189 |
| 1. CONSTRUCCIÓN Y VALIDACIÓN DE SISTEMAS DE EXPRESIÓN HETERÓLOGA PARA METAGENÓMICA FUNCIONAL..... | 191 |
| 2. OTROS FACTORES IMPORTANTES PARA LA IDENTIFICACIÓN DE ACTIVIDADES DE INTERÉS..... | 198 |

| | |
|---|------------|
| 3. RESISTENCIAS A CARBENICILINA IDENTIFICADAS EN LA METAGENOTECA DE PUNTA DE SAN GARCÍA..... | 201 |
| 4. ENZIMAS IMPLICADAS EN RUTAS DE DEGRADACIÓN DE CONTAMINANTES AROMÁTICOS..... | 204 |
| 5. CARACTERIZACIÓN DEL OPERÓN DE DESULFURACIÓN DE DIBENZOTIOFENO..... | 209 |
| | |
| CONCLUSIONES..... | 215 |
| | |
| ANEXO..... | 219 |
| | |
| BIBLIOGRAFÍA..... | 229 |

ÍNDICE DE TABLAS

| | página |
|--|---------------|
| Tabla 1. Estirpes bacterianas utilizadas en esta tesis | 45 |
| Tabla 2. Plásmidos utilizados en esta tesis | 48 |
| Tabla 3. Oligonucleótidos utilizados en esta tesis | 50 |
| Tabla 4. Porcentaje de bacterias inducidas en una población antes y después de FACS | 113 |
| Tabla 5. Número de clones resistentes a carbenicilina detectados en las diferentes estirpes por cada millón de transconjungantes | 116 |
| Tabla 6. Transferencia por conjugación de los clones resistentes a carbenicilina a las distintas estirpes | 118 |
| Tabla 7. Información asociada a cada tanda de secuenciación por <i>pools</i> | 140 |
| Tabla 8. Relación de clones con actividad dioxigenasa extradiólica y gen o genes posibles responsables de dicha actividad identificados en cada uno de ellos | 141 |
| Tabla 9. Relación de clones con actividad oxigenasa hidroxilante detectada y con posible gen responsable de la actividad asignado, e información sobre su secuenciación | 160 |
| Tabla 10. Monooxigenasas codificadas por genes de clones con actividad oxigenasa y clones en los que se ubican | 163 |
| Tabla 11. Resumen de las condiciones en las que se produce la actividad de desulfuración de dibenzotiofeno en comparación con la de producción de índigo en presencia de los genes <i>dsz</i> | 179 |

ÍNDICE DE FIGURAS

| | página |
|--|---------------|
| Figura 1. Esquema de la construcción y rastreo de metagenotecas para identificación de enzimas de interés | 15 |
| Figura 2. Esquemmatización de los principales pasos en la degradación aerobia de hidrocarburos aromáticos | 33 |
| Figura 3. Esquema de una PCR recurrente para sintetizar fragmentos de ADN | 54 |
| Figura 4. Esquema de una PCR solapante | 55 |
| Figura 5. Esquema del proceso de introducción de genes en el cromosoma utilizando productos de PCR | 70 |
| Figura 6. Fósmidos derivados de pCC1FOS-CeuI: pMPO571 y pMPO579 | 102 |

| | |
|--|-----|
| Figura 7. Estirpes especializadas derivadas de la estirpe de <i>E. coli</i> EPI300-T1 ^R : MPO553 y MPO554, y esquema del sistema de inducción heteróloga que permite cada una de ellas | 105 |
| Figura 8. Expresión del gen <i>gfp</i> en los fósmidos pMPO578 y pMPO579 en las estirpes EPI300-T1 ^R , MPO553 y MPO554 | 108 |
| Figura 9. Expresión del gen <i>gfp</i> en el fósrido sin terminador y cuando lo precede un terminador de la transcripción, en la estirpe MPO553 | 110 |
| Figura 10. Expresión del gen <i>gfp</i> en la estirpe MPO554 en distintas condiciones de inducción | 111 |
| Figura 11. Expresión del gen <i>gfp</i> en el fósrido sin terminador y cuando lo precede un terminador de la transcripción, en presencia y ausencia de la proteína antitermiadora N | 112 |
| Figura 12. Imágenes de la zona costera de la que se tomó la muestra de suelo para la construcción de la metagenoteca | 114 |
| Figura 13. Patrones de restricción de los 6 clones que confieren resistencia a carbenicilina | 117 |
| Figura 14. Esquema de los clones que confieren resistencia a carbenicilina | 124 |
| Figura 15. Imágenes de la pila de la refinería de la industria petroquímica de la que se tomó la muestra de suelo para la construcción de la metagenoteca | 129 |
| Figura 16. A. Esquema de las reacciones de dioxigenación extradiólica sobre catecol y 2,3-dihidroxibifenilo, generando los correspondientes productos de rotura. B. Colonia capaz de producir 2-hidroximucónico semialdehído a partir de catecol junto a otras sin la actividad en el rastreo de dioxigenasas extradiólicas | 131 |
| Figura 17. Esquema de las reacciones que tienen lugar para la producción de índigo a partir de triptófano | 133 |
| Figura 18. Colonia productora de índigo junto a otras sin la actividad en el rastreo de oxigenasas iniciales | 134 |
| Figura 19. Colonia formadora de un halo en la tinción con rojo Congo por la degradación de la CMC junto a otras sin la actividad en el rastreo de endoglucanasas | 136 |
| Figura 20. Análisis filogenético de las secuencias metagenómicas de dioxigenasas extradiólicas tipo I, subfamilia I.2 | 143 |
| Figura 21. Análisis filogenético de las secuencias metagenómicas de dioxigenasas extradiólicas tipo I, subfamilias I.3, I.4, I.5 y I.6 | 145 |
| Figura 22. Análisis filogenético de las secuencias metagenómicas de dioxigenasas extradiólicas tipo II | 147 |
| Figura 23. Actividad dioxigenasa extradiólica en los clones metagenómicos | 150 |

| | |
|--|-----|
| Figura 24. Representación del número de clones que presenta su máxima actividad dioxigenasa extradiólica con cada sustrato | 153 |
| Figura 25. Visualización en gel de poliacrilamida-SDS de la superproducción de la proteína UPOEDO11 | 155 |
| Figura 26. Esquema de los clones con genes codificantes de dioxigenasas subunidad α | 162 |
| Figura 27. Análisis filogenético de las secuencias metagenómicas de citocromos P450 | 167 |
| Figura 28. Representación de los resultados del ensayo colorimétrico empleando el indicador Fast Blue B para detectar la producción de naftoles | 169 |
| Figura 29. Producción de 1-naftol por células del clon UPO16 cultivadas en presencia de naftaleno | 171 |
| Figura 30. Producción de catecol por células del clon UPO16 cultivadas en presencia de fenol | 172 |
| Figura 31. Ruta de desulfuración de dibenzotiofeno | 174 |
| Figura 32. Esquema del clon con los genes de la ruta de desulfuración de dibenzotiofeno (UPO21) y de los subclones con los genes necesarios para que se produzca la actividad (pMPO587 y pMPO588) | 177 |
| Figura 33. Esquema de las fusiones para el estudio de la expresión de los genes de la ruta de desulfuración de dibenzotiofeno | 181 |
| Figura 34. Expresión de las fusiones de la ruta de desulfuración de dibenzotiofeno | 184 |

ABREVIATURAS

ADN: ácido desoxirribonucleico

ARN: ácido ribonucleico

Ap: ampicilina; **Ap^R:** resistente a ampicilina

ARNi: ARN interferente

ARNm: ARN mensajero

ARNr: ARN ribosómico

BAC: cromosoma artificial bacteriano (*bacterial artificial chromosome*)

Cb: carbenicilina; **Cb^R:** resistente a carbenicilina

Cm: cloranfenicol; **Cm^R:** resistente a cloranfenicol

CMC: carboximetilcelulosa

DBT: dibenzotiofeno

D.O.₆₀₀: densidad óptica medida a 600 nm

EDO: extradiol dioxigenasa / dioxigenasa extradiólica

FACS: clasificación de células activadas por fluorescencia (*fluorescence-activated cell sorting*)

FMN: flavín mononucleótido

GFP: proteína verde fluorescente (*green fluorescent protein*)

Gm: gentamicina; **Gm^R:** resistente a gentamicina

h: hora

HPLC: cromatografía líquida de alta resolución

IPTG: isopropil-β-D-tiogalactósido

kb: kilo pares de bases

kDa: kilodalton

Km: kanamicina; **Km^R:** resistente a kanamicina

LB: medio Luria-Bertani

METREX: *metabolite-regulated expression*

min: minutos

NAD(P)H: nicotin-adenín dinucleótido (fosfato)

Nal: ácido nalidíxico, **Nal^R:** resistente a ácido nalidíxico

nt: nucleótidos

ONPG: o-nitrofenil-β-D-galactósido

orf: pauta abierta de lectura (*open reading frame*)

pb: pares de bases

PBS: tampón fosfato salino (*phosphate buffer saline*)

PCR: reacción en cadena de la polimerasa (*polymerase chain reaction*)

PIGEX: *product induced gene expression*

pT7: promotor del gen *10* del fago T7

p/v: peso/volumen

Rif: rifampicina, **Rif^R:** resistente a rifampicina

RND: resistencia-nodulación-división celular

r.p.m.: revoluciones por minuto

SIGEX: *substrate induced gene expression*

sitio *nut*: sitio de utilización de la proteína antiterminadora N (*N-utilization*); **sitio *nut_L*:** sitio de utilización de N hacia la izquierda (*N-utilization leftward*)

Tc: tetraciclina; **Tc^R:** resistente a tetraciclina

v/v: volumen/volumen

X-gal: 5-bromo-4-cloro-3-indol-β-D-galactopiranosido

INTRODUCCIÓN

1. DIVERSIDAD MICROBIANA

Los microorganismos constituyen un grupo de seres vivos extremadamente heterogéneo. Se han adaptado a vivir en ambientes muy dispares, entre los que se encuentran suelos, superficies oceánicas, profundidades marinas, desiertos, aguas termales, glaciares e incluso organismos superiores; adquiriendo formas de metabolismo muy distintas que les han permitido aprovechar los recursos disponibles en cada uno de esos ambientes. Además de su enorme diversidad, que constituye el mayor reservorio de diversidad genética de la Tierra (Ferrer y col., 2009a; Whitman y col., 1998), hay que destacar su abundancia, habiendo estimaciones que sitúan el número de células procariotas (dejando a un lado los hongos) en $4-6 \times 10^{30}$, incluyendo más de 70 filos (Pace, 2009; Whitman y col., 1998).

Esta enorme diversidad genética da lugar a innumerables tipos de enzimas y productos de interés para el ser humano en diversos campos: farmacología, biomedicina, agricultura, alimentación, eliminación de contaminantes, generación de energía,... Las enzimas son muy interesantes como sustitutas de los catalizadores químicos, pudiendo llevar a cabo por lo general las reacciones en condiciones ambientales normales y permitiendo reducir la generación de residuos. Una ventaja muy importante de estos biocatalizadores se debe a la regio- y enantioselectividad que presentan las enzimas, que las hace muy específicas en cuanto a la posición de la molécula del sustrato en la que actúan y la acción que realizan, permitiendo una mayor selectividad hacia el sustrato en cuestión y reduciendo la producción de moléculas quirales y la generación de subproductos no deseados. Además de las enzimas, también son de interés muchos metabolitos producidos por los microorganismos, como por ejemplo, los metabolitos secundarios biológicamente activos, de gran interés para la industria farmacéutica, representados principalmente por los antibióticos.

Por tanto, el estudio de los microorganismos y de las relaciones entre ellos tiene una gran importancia, no solo por su conocimiento en sí sino también por la enorme fuente de recursos que suponen. Tradicionalmente nuestro conocimiento sobre los microorganismos se ha limitado a aquellos que éramos capaces de cultivar en el laboratorio. Sin embargo, se estima que la inmensa mayoría de los microorganismos (más del 99 %) no son capaces de crecer con las condiciones de cultivo disponibles actualmente (Amann y col., 1995; Rappe y Giovannoni, 2003), lo que dejaría una enorme cantidad de información fuera de nuestro alcance con esas técnicas tradicionales.

A pesar de que la enorme diversidad de microorganismos, así como la variabilidad en sus requerimientos de crecimiento y el desconocimiento de los mismos, hacen que resulte imposible abarcar mediante técnicas dependientes cultivo la totalidad de los microorganismos existentes, ni tan siquiera un alto porcentaje de ellos, las estrategias de cultivo pueden ser mejoradas (Tyson y Banfield, 2005). Hay distintas razones que pueden influir en la imposibilidad de cultivar muchos de los microorganismos de los que aún no sabemos nada (o muy poco). Entre ellas se encuentran la ausencia de otros organismos con los que tienen relaciones de mutualismo o sinergismo, falta de nutrientes o superficies necesarios, presencia de compuestos inhibidores, combinaciones inadecuadas de temperatura, presión o composición de la fase gaseosa, acumulación de productos de desecho tóxicos de su propio metabolismo, una tasa de crecimiento demasiado lenta o la rápida dispersión de las células en lugar de formar colonias (Handelsman 2004; Simu y Hagstrom, 2004).

En ocasiones, simplemente la variación de algunas de las condiciones en los medios de cultivo (adición de nuevos compuestos, incremento del tiempo de incubación, variación de las condiciones de exposición a la luz, etc.) permite la identificación de nuevos organismos, pero en otros casos se requieren estrategias más complejas. Una opción que ha permitido identificar nuevos microorganismos, es reemplazar los medios de cultivo convencionales ricos en nutrientes por medios con baja concentración de nutrientes, evitando el dominio por parte de las especies microbianas de rápido crecimiento de manera que pueden ser detectados nuevos microorganismos oligotrofos (Connon y Giovannoni, 2002; Zengler y col., 2002). Por otro lado, se han desarrollado técnicas más sofisticadas para el aislamiento de bacterias de hábitats complejos, como la construcción de ambientes simulando los naturales (Kaeberlein, Lewis y Epstein, 2002; Wery y col., 2003) y la encapsulación de células en microgotas de gel (Zengler y col., 2002).

En los últimos años ha tomado fuerza el estudio de esta diversidad de microorganismos no cultivados por el extraordinario potencial que representa. Por una parte, se están desarrollando nuevas técnicas de cultivo que permiten incrementar el número de especies microbianas que pueden ser estudiadas en cultivos de laboratorio de forma aislada. Por otra, se ha desarrollado una alternativa para la exploración de toda esta diversidad, que permite su estudio a partir de la extracción directa del material genético de los organismos presentes en una muestra ambiental, sin necesidad de cultivo previo: la metagenómica. La combinación de ambas técnicas permitirá un avance más rápido en el conocimiento sobre el mundo microbiano.

2. METAGENÓMICA

La metagenómica permite el estudio de los microorganismos a partir de su material genético, sin necesidad de cultivo, accediéndose de este modo a la información genética tanto de los organismos susceptibles de ser cultivados en un laboratorio como de los que no. Por ello, presenta un potencial enorme de acceso a la gran fuente de recursos que supone la diversidad microbiana. Aunque se pueden realizar análisis metagenómicos de microorganismos eucarióticos, por su mayor simplicidad la mayoría de estos análisis se centran, como en esta tesis, en el grupo de microorganismos que cuenta con una mayor diversidad, que son las bacterias, aun siendo conscientes de la importancia del estudio del conjunto de microorganismos (bacterias, arqueas, protistas y hongos).

La idea de clonar directamente el ADN ambiental en vectores fue publicada por primera vez por Pace y col. en 1985, que realizaron estudios de diversidad de microorganismos a partir de la secuenciación de los genes codificantes del ARNr 5S y 16S, con lo que se dieron cuenta de que la inmensa mayoría de los microorganismos del medio ambiente detectados por estas técnicas eran distintos a los conocidos por los microbiólogos. Por otro lado, Healy col. en 1995, fueron los primeros en detectar funciones de interés codificadas por ADN ambiental. En 1998, Handelsman y col., emplearon el término metagenoma para referirse al conjunto de genomas de una muestra ambiental, y resaltaron la importancia de la clonación de ese material genético para acceder a la fuente de productos naturales proporcionada por los microorganismos, teniendo en cuenta que una minúscula parte de ellos eran cultivables. A partir de entonces, la metagenómica ha permitido el descubrimiento de muchísimas nuevas enzimas (oxigenasas, lipasas, proteasas, celulasas, betalactamasas,...) y productos de interés generados por estas (como los antibióticos). Esta disciplina abre una ventana de posibilidades enorme, y como muestra de ello está el hecho de que gran parte de las nuevas enzimas con funciones de interés identificadas están poco relacionadas con aquellas previamente conocidas con esa función o son completamente nuevas (Daniel, 2004).

Para realizar un análisis metagenómico, en primer lugar, habría que seleccionar el ambiente objeto de estudio, aislar ADN de ese ambiente en particular y construir una metagenoteca, que engloba al conjunto de clones que albergan fragmentos de ADN de los microorganismos de esa muestra ambiental. A continuación se realizarían los análisis pertinentes según el objetivo en cada caso.

La metagenómica permite abarcar el estudio de los microorganismos desde dos perspectivas. Por un lado, se pueden llevar a cabo estudios filogenéticos que permitan el análisis de la diversidad de una muestra, para lo que principalmente se emplean como marcadores los genes que codifican el ARNr 16S. Y por otro lado, se pueden realizar análisis de los distintos tipos de fisiología y metabolismo presentes en esa muestra, idealmente asociándolos a los distintos tipos de microorganismos, así como la identificación de enzimas y productos metabólicos de interés. En la presente tesis se tratará este último punto.

En ambos sentidos, la metagenómica ha permitido grandes avances en el conocimiento que se posee sobre la comunidad de microorganismos del planeta, a una velocidad muy superior a la que permite el estudio basado en el cultivo de los organismos. Sin embargo, la metagenómica también presenta limitaciones, entre las que destacan la frecuente dificultad para obtener un material genético en buenas condiciones y suficiente cantidad a partir de una muestra ambiental, lo que repercute en una baja representatividad de la metagenoteca, y por otra parte, la dificultad en la identificación funcional de actividades de interés, relacionada con problemas en la expresión de los genes en hospedadores heterólogos y con la disponibilidad de rastreos adecuados.

A lo largo de todo el proceso hay muchos factores a tener en cuenta que pueden influir en el éxito del análisis metagenómico. Cada paso, desde la recogida de la muestra hasta la selección de la estrategia de rastreo, conlleva irremediamente una serie de sesgos, que dependerán de las decisiones tomadas en cada momento, las cuales deben ser adoptadas teniendo en cuenta el objetivo final. En la figura 1 se muestra un esquema de la construcción y análisis de metagenotecas.

2.1 Construcción de metagenotecas

Las bibliotecas metagenómicas o metagenotecas son colecciones de genes construidas a partir del ADN total directamente extraído de una muestra ambiental, que se clona en un vector replicativo.

La construcción de una metagenoteca conlleva la selección del ambiente a estudiar, toma de la muestra, extracción y tratamiento del ADN, clonación en vectores e introducción en la bacteria hospedadora (figura 1). El conjunto de bacterias con un vector que albergue un fragmento aleatorio de ADN ambiental (clones) formará la metagenoteca. Hay que ser conscientes de que una metagenoteca es una representación parcial de la muestra. No es un

proceso sencillo ni existe un protocolo único, debido a la heterogeneidad del material de partida y a las múltiples aplicaciones posibles, en función de lo cual se construirá un tipo de metagenoteca u otro. Algunos de los protocolos descritos se pueden encontrar en las publicaciones de Taupp y col. (2009), Simon y Daniel (2010), Martínez y Osburne (2013) y Sabehi y Béjà (2013).

2.1.1 Selección y recogida de la muestra

El primer paso es la recogida de la muestra que constituirá el material de partida. Puede tener muy distinto origen: suelo de un bosque, arena del desierto, agua de la superficie oceánica, sedimentos de las profundidades marinas, aguas termales, etc. Por ejemplo, si se pretende emplear la metagenoteca para la búsqueda de enzimas implicadas en rutas de degradación de contaminantes, una buena elección será tomar la muestra de un suelo contaminado; mientras que si lo que interesan son enzimas que funcionen en condiciones extremas, como altas temperaturas, se podrá seleccionar una muestra de aguas termales. Uno de los tipos de ambientes más estudiados es el suelo, que con frecuencia posee una gran diversidad de microorganismos, en respuesta a su gran heterogeneidad, que resulta en una gran variedad de nichos microbianos. Aunque también existen suelos pobres, un gramo de suelo puede contener del orden de 10^9 - 10^{10} bacterias posiblemente de miles de especies diferentes (Rosselló-Mora y Amann, 2001; Schloss y Handelsman, 2006).

Existe la opción de realizar un enriquecimiento de la muestra antes de proceder a la extracción del material genético, sobre todo intentando favorecer el crecimiento de organismos que pensemos que pueden portar las actividades de interés (por ejemplo mediante crecimiento en medio con algún compuesto determinado como única fuente de carbono o nitrógeno, o con algún tóxico en bajas concentraciones). Puede ser una buena práctica en ambientes con baja densidad de células, como ocurre normalmente con ambientes extremos. Aunque no es recomendable en términos generales, ya que conlleva una disminución de la diversidad de la muestra, en ocasiones ha dado buenos resultados para el descubrimiento de enzimas de interés (Entcheva y col., 2001; Knetsch y col., 2003; Gabor y col., 2004a). Daniel (2004) comparó los clones encontrados para una determinada actividad con dos metagenotecas construidas a partir de una misma muestra, una con la muestra sin enriquecer (Knetsch y col., 2003a) y la otra tras someter la muestra a un proceso de enriquecimiento (Knetsch y col., 2003b). Para encontrar el mismo número de clones positivos, con la

metagenoteca de la muestra enriquecida necesitaron analizar aproximadamente la mitad de clones que con la de la muestra sin enriquecer. Sin embargo, de los clones positivos de la metagenoteca sin enriquecer hubo un 50 % cuyas secuencias de los genes de interés no estaban relacionadas con las de otros conocidos frente a un 10 % en el caso de los clones de la metagenoteca enriquecida.

2.1.2 Extracción y tratamiento del ADN

Se ha publicado una gran variedad de protocolos para la extracción del ADN de una muestra ambiental, por ejemplo los descritos por Steffan y col. (1988), Zhou, Bruns y Tiedje (1996) y Miller y col. (1999). No hay un procedimiento de extracción ideal para cualquier muestra, y cualquier método conlleva sesgos diferentes en la diversidad bacteriana representada (Martin-Laurent y col., 2001). A continuación se comentan los aspectos más críticos. El tamaño de la muestra de partida para la obtención de una cantidad adecuada de ADN dependerá en gran parte del tipo de muestra y de la densidad de bacterias que contenga.

El material genético se puede extraer directamente a partir de la muestra, mediante métodos denominados de extracción directa. Alternativamente se puede llevar a cabo una extracción indirecta, en la que previamente a la extracción del material genético se realiza una separación de los microorganismos del soporte sólido donde están adheridos, mediante mecanismos como la homogeneización con mortero o licuadora o adición de resinas de intercambio iónico, seguido por gradiente de densidad o centrifugación diferencial, como con la resina Nycodenz (Jacobsen y Rasmussen, 1992; Lindahl y Bakken, 1995; Courtois y col. 2001; Gabor y col., 2003; Daniel, 2005), tras lo cual se obtiene una fracción enriquecida en bacterias. Esto reduce la proporción de ADN eucariota clonado, que no será capaz de expresarse en hospedadores bacterianos (Steffan y col., 1988; Gabor y col., 2003), aunque disminuye la cantidad de ADN final recuperado, como se comenta más adelante. En cuanto a la posible pérdida de diversidad asociada a este tratamiento, los estudios realizados al respecto no dan una respuesta clara. Por ejemplo, las comparaciones hechas por Gabor y col. (2003) y Luna y col. (2006) mostraron una aparente mayor diversidad empleando métodos de extracción indirecta y directa, respectivamente; mientras que Courtois y col. (2001) no observaron grandes diferencias.

Mediante métodos de extracción directa se pueden conseguir cantidades de ADN 10-100 veces mayores (Gabor y col, 2003; Courtois y col. 2001). Por otro lado, los métodos

indirectos facilitan la obtención de fragmentos más largos, más limpios y con menos contaminación de ADN eucariota (Lombard y col. 2011). Según el tipo de método de extracción empleado, la lisis celular se producirá sobre la totalidad de la muestra o sobre la fracción bacteriana previamente separada del resto de partículas. En ambos casos, se pueden emplear procedimientos de lisis más o menos agresivos, en función del tipo de organismos que se espere en la muestra o que en principio interesen más. No todos los organismos presentan la misma susceptibilidad a los métodos de lisis celular, por lo que el método elegido puede favorecer la representatividad de unos u otros (Courtois y col., 2001; Stach y col., 2001; Gabor y col., 2003). Por una parte, los tratamientos suaves pueden sesgar en contra de bacterias más robustas pero, por otra, tratamientos de lisis muy agresivos para evitar este sesgo pueden fragmentar demasiado el ADN (Lombard y col., 2011). Dentro de los procedimientos de lisis se encuentran tratamientos físicos, químicos y enzimáticos, que incluso pueden combinarse para lograr una lisis más eficiente (Zhou y col., 1996; Frostegard y col., 1999; Courtois y col., 2001; Lakay y col., 2007).

Otro aspecto clave es el tratamiento de limpieza del ADN para que este quede con una calidad adecuada que permita la obtención de un número suficiente de clones, lo que dependerá de las características de la muestra. En ocasiones, la contaminación con compuestos como los ácidos húmicos requiere tratamientos muy agresivos para que el ADN presente una calidad aceptable, lo que implica que el tamaño final del ADN se vea bastante reducido. Es importante la eliminación de esos compuestos, ya que se unen al ADN e inhiben las posteriores reacciones enzimáticas (Daniel, 2005). La separación de la fracción bacteriana del resto de la muestra y su lavado antes de la lisis contribuyen a preservar la integridad del ADN genómico, al reducir los tratamientos de limpieza posteriores necesarios (Fortin y col., 2004). Hay multitud de opciones para la purificación del ADN, y puede ser necesario combinar varias de ellas en el caso de suelos particularmente ricos en ácidos húmicos (Ogram y col., 1987; Steffan y col., 1988; Miller y col., 1999; Fortin y col., 2004; Desai y Madamwar, 2007; Lakay y col., 2007).

En función del tipo de metagenoteca que se quiera construir, que se comentan más adelante, se necesitan fragmentos de ADN de menor o mayor tamaño, que pueden ir desde unas pocas hasta cientos de kb. En ocasiones, la rotura mecánica que acompaña a la purificación del ADN deja el material genético demasiado fragmentado para la construcción de metagenotecas con un tamaño de inserto grande. Otras veces sí se consiguen fragmentos de gran peso molecular, que pueden requerir una reducción de tamaño mediante fragmentación

mecánica (al azar) o enzimática (digestiones parciales con enzimas de restricción). Cuando se necesitan fragmentos de tamaño reducido normalmente se recurre a la digestión enzimática.

2.1.3 Clonación en el vector adecuado

El ADN metagenómico ya preparado se liga con el vector seleccionado. Para la elección del vector se tendrán en cuenta distintos criterios, como el tamaño de los fragmentos de ADN metagenómico, el tipo de análisis a realizar posteriormente (basado en secuencia o en función, sobre lo que se tratará más adelante) o la estirpe en la que se realizarán los rastreos funcionales. Los tipos de vectores empleados son los plásmidos, vectores de fago lambda, BACs, cósmidos y fósidos.

Para fragmentos de ADN metagenómico pequeños (insertos de hasta 10 kb) se utilizan plásmidos convencionales, que pueden tener orígenes de replicación para distintas especies bacterianas, aunque lo más habitual es que sean replicativos en *Escherichia coli*; o vectores derivados del fago lambda (insertos de hasta 12-20 kb), replicativos exclusivamente en *E. coli*. En el otro extremo se encuentran los cromosomas artificiales bacterianos (BACs), que pueden albergar insertos muy grandes, incluso de hasta 300 kb. Tienen el sistema de replicación del plásmido F, lo que favorece la estabilidad de metagenotecas con vectores con insertos tan grandes, al presentar solo 1-2 copias por célula y un sistema que garantiza el reparto de copias del plásmido entre las bacterias descendientes. Sin embargo, limita su replicación a *E. coli* (Shizuya y col., 1992). Además, la manipulación del ADN durante la construcción de la metagenoteca hasta su introducción en la bacteria hospedadora es mucho más compleja al ser necesario mantener *in vitro* la integridad de fragmentos de ADN de gran tamaño.

Los cósmidos son plásmidos que pueden ser empaquetados en cabezas del fago lambda debido a la presencia de sitios *cos* (Collins y Hohn, 1978), mejorando por tanto la frecuencia de incorporación en la bacteria hospedadora, que se realiza por transfección en vez de por transformación. A diferencia de los vectores anteriores, estos son selectivos de tamaño porque la suma del vector más el inserto no puede superar la capacidad de empaquetamiento en una cabeza del fago lambda, que es de unas 50-52 kb. Los cósmidos, en función de su sistema de replicación, pueden ser de amplio espectro, pudiendo replicarse en bacterias diferentes, o de espectro reducido. Muchos de estos cósmidos están basados en plásmidos de alto número de copias, lo que proporciona gran inestabilidad a las bibliotecas metagenómicas sobre todo

cuando el tamaño del inserto es grande. Este motivo condujo al desarrollo de los fósmidos que, conservando el sitio *cos* para poder ser empaquetados en cabezas de lambda, tienen el sistema de replicación del plásmido F (Kim y col, 1992), lo que implica que estos vectores se mantienen muy establemente en un bajo número de copias (1-2), aunque en principio solo se puedan dividir en *E. coli*. Los fósmidos tienen por tanto propiedades de los cósmidos y de los BACs.

La ventaja que proporcionan las metagenotecas de fragmentos pequeños es que son más fáciles de construir, ya que por lo general es más sencillo obtener cantidad suficiente de fragmentos de ADN de buena calidad de tamaño pequeño. Sin embargo, con fragmentos pequeños se necesitarán más clones para lograr una representatividad similar a la de una metagenoteca construida con fragmentos de mayor tamaño. En el caso de los análisis basados en función en los que se fuerce la expresión de los genes metagenómicos desde promotores heterólogos incorporados al vector (como se verá más adelante), en principio se conseguirá la expresión de una mayor proporción de los genes si los fragmentos son pequeños, ya que en fragmentos grandes la probabilidad de que se transcriban desde el promotor heterólogo por lo general se verá reducida conforme aumente la distancia al promotor heterólogo debido a la presencia de terminadores de la transcripción en el ADN metagenómico. Por otra parte, si no se va a aprovechar la transcripción heteróloga desde el vector para realizar estos rastreos funcionales, sino que se confía en la expresión propia de los genes, cuanto mayor sea el tamaño del inserto menos clones hará falta analizar para encontrar la función de interés. Sin embargo, solo una fracción probablemente minoritaria de los genes podrán expresarse por sí mismos, y en muchas ocasiones serán aquellos genes procedentes de las bacterias más cercanas a la hospedadora. Por otro lado, los fragmentos grandes posibilitan la detección de actividades que requieran varios genes, e incluso de rutas completas. Para los análisis mediante secuenciación, rastreos por homología de secuencia, las metagenotecas de insertos grandes son más adecuadas, ya que la secuenciación (de cada clon) proporcionará más información, además de posibilitar la identificación de secuencias completas de operones o rutas y facilitar el ensamblaje de distintos fragmentos del mismo genoma.

El número de copia de los vectores también es importante. Por un lado, con vectores de alto número de copias la estabilidad de la metagenoteca normalmente será menor, sobre todo en el caso de vectores que alberguen fragmentos grandes, pero por otro, aumentará la dosis génica. Un aumento en la dosis génica puede permitir detectar una determinada actividad cuya expresión sea demasiado baja en la bacteria hospedadora, aunque también podría

provocar la pérdida de una parte de la información genética, debido al perjuicio causado por la expresión en exceso de genes cuyos productos sean tóxicos para el hospedador. Existen vectores que permiten mantener la metagenoteca en baja copia e incrementar el número de copias en condiciones de inducción (p. ej. pCC1FOS, de Epicentre).

A estos vectores se les pueden incorporar distintos elementos en función de los tipos de análisis a realizar con la metagenoteca (como origen de transferencia conjugativa, orígenes de replicación para otras especies o promotores heterólogos), como se comentará en el apartado 3.

2.1.4 Introducción de la metagenoteca en bacterias hospedadoras

Por último, se introducen los vectores con los fragmentos de ADN metagenómico en bacterias hospedadoras mediante transformación o infección. La más empleada es *E. coli*, porque la mayoría de los vectores disponibles se pueden mantener y replicar en esta estirpe y su frecuencia de transformación es muy alta. Por otro lado, en el caso de uso de cósmidos o fósidos, *E. coli* se puede transfectar con la metagenoteca empaquetada *in vitro* en cabezas del fago lambda. Se seleccionan las bacterias que contengan vector (normalmente mediante selección con algún antibiótico al que el vector en cuestión proporcione resistencia), que en su conjunto formarán la metagenoteca, y por lo general se mantienen congeladas a -80 °C ya sea en alícuotas del cultivo que contiene al total de clones (metagenoteca en masa) o en cultivos de los clones individuales (metagenoteca ordenada).

En la representatividad de una biblioteca metagenómica influyen, además de los sesgos introducidos en los distintos pasos del proceso, el tamaño de la misma y la diversidad de la muestra ambiental. Por ello, la cobertura genómica de una metagenoteca no es fácil de calcular. Con los medios actuales, solo se puede tener una cobertura grande de la muestra en el caso de comunidades muy poco diversas, como en el caso descrito por Tyson y col. (2004) con una metagenoteca de un biofilm microbiano crecido en aguas expuestas a drenaje ácido de minas, con muy pocas especies dominantes. Pero en la mayoría de los casos, para que una metagenoteca cubra la mayor parte de la diversidad de una muestra se necesitaría un número enorme de clones, probablemente muchos millones si se construye una metagenoteca con fragmentos de ADN relativamente grandes, de unas 40 kb, como la que se realizan utilizando cósmidos o fósidos. Por ejemplo, si se asume que el tamaño medio del genoma de una

bacteria es de 4 Mb y que 1 g de suelo contiene unos 10^5 genomas diferentes, una metagenoteca compuesta por 40.000 fósmidos que alberguen cada uno fragmentos de 40 kb podría cubrir como máximo unos 400 genomas bacterianos completos si no se repitiese ninguna secuencia, representando por tanto una ínfima parte de los existentes en esa muestra de suelo (Vieites y col., 2009). Por tanto, hay que tener siempre presente que una metagenoteca es una colección de genes de un nicho concreto en un momento dado, pero que no es una representación completa del mismo, sino que solo refleja una parte, probablemente pequeña, de la diversidad existente. Dada esta baja representatividad, se puede decir que los genes obtenidos en una metagenoteca corresponden a los genes de los organismos más abundantes en el nicho del que se tomó la muestra, si dejamos al margen el sesgo que se pueda haber producido al construir la metagenoteca.

2.2 Identificación de actividades de interés en metagenotecas

Como ya se ha comentado en apartados anteriores, la enorme diversidad microbiana existente en el planeta supone una fuente de información y recursos muy valiosa, y la metagenómica está permitiendo el acceso a parte de esos recursos que hasta hace unos años estaba fuera de nuestro alcance. Los estudios metagenómicos realizados reflejan esa enorme diversidad, mostrando que más del 60 % de las secuencias descubiertas están alejadas de otras secuencias conocidas (Ferrer y col., 2009b). En la actualidad se emplean gran variedad de enzimas en procesos industriales, pero la demanda continúa creciendo, y la metagenómica es probablemente la herramienta que mejor puede proporcionar nuevas moléculas de utilidad (Langer y col. 2006).

Una vez que se tengan metagenotecas construidas con las muestras seleccionadas, se pueden realizar análisis en busca de secuencias, actividades enzimáticas o productos metabólicos de interés. Se distinguen dos grandes grupos de estrategias para el análisis de metagenotecas, según estén basadas en las secuencias del material genético contenido en los clones o en la función que los productos génicos de este puedan desarrollar.

Los análisis basados en secuencia incluyen tanto la secuenciación del ADN metagenómico y la búsqueda de secuencias homólogas a las de genes de interés almacenadas en las bases de datos, como la búsqueda de secuencias de genes de un determinado tipo mediante amplificación con oligonucleótidos degenerados o mediante hibridación con sondas para esa clase de genes. Por otro lado, en los análisis basados en función se buscan

actividades responsables de un determinado fenotipo, identificándose posteriormente el gen o genes que la codifican. Cada tipo de estrategia presenta sus ventajas y limitaciones. Los análisis basados en secuencia permiten la identificación de genes que pueden ser de interés aunque no se estén expresando en el hospedador heterólogo seleccionado, pero limitan el descubrimiento de nuevas enzimas a aquellas que muestren cierta similitud con otras descritas previamente. Los análisis basados en función han demostrado que las nuevas enzimas descubiertas no están limitadas a simples variaciones sobre secuencias conocidas, sino que también incluyen secuencias completamente nuevas y funcionalidades novedosas (Ferrer y col., 2009a), pero presentan como limitación principal la necesidad de que los genes responsables sean capaces de expresarse, además de que su producto sea activo, en la estirpe que los alberga, sumado a la disponibilidad de un sistema de rastreo adecuado que permita la identificación de los clones portadores de la actividad de interés. La combinación de ambas aproximaciones puede paliar las limitaciones de cada una de ellas.

Las dos estrategias han permitido el descubrimiento de nuevas enzimas de interés, aunque, por los motivos comentados, aquellas completamente diferentes a lo conocido hasta el momento solo se han podido identificar gracias al análisis basado en función. Como ejemplos se encuentran, entre otras, la identificación de nuevas dioxigenasas (Morimoto y Fujii, 2009), nitrito reductasas (Bartossek y col., 2010) y glicerol deshidratasas (Knietsch y col., 2003c), mediante rastreos basados en secuencia; y de nuevas dioxigenasas (Suenaga y col., 2007), esterases (Rhee y col., 2004), xilanasas (Gong y col., 2013), betalactamasas (Allen y col., 2009) y actividades antimicrobianas (Biver y col., 2013), mediante rastreos basados en función.

Frecuentemente, en una metagenoteca determinada hay pocos clones con la secuencia o función de interés, por lo que para identificarlos normalmente es necesario analizar un gran número de clones, y aun así puede que no se encuentre ninguno. La frecuencia de organismos que contengan los genes deseados en la muestra ambiental empleada para la construcción de la metagenoteca influirá en el número de clones, y en consecuencia en la cantidad de ADN metagenómico, que se necesite analizar (Daniel 2005). La utilización de métodos de rastreo adecuados también es un factor muy importante, que puede determinar el éxito o fracaso de la búsqueda. Puede haber genes que se identifiquen por función y no por secuencia porque no presenten una similitud suficiente con ninguno de los descritos hasta el momento; y otros que se identifiquen por secuencia y no por función porque no se estén expresando, porque se estén expresando pero las enzimas producidas no sean activas en ese hospedador heterólogo, o

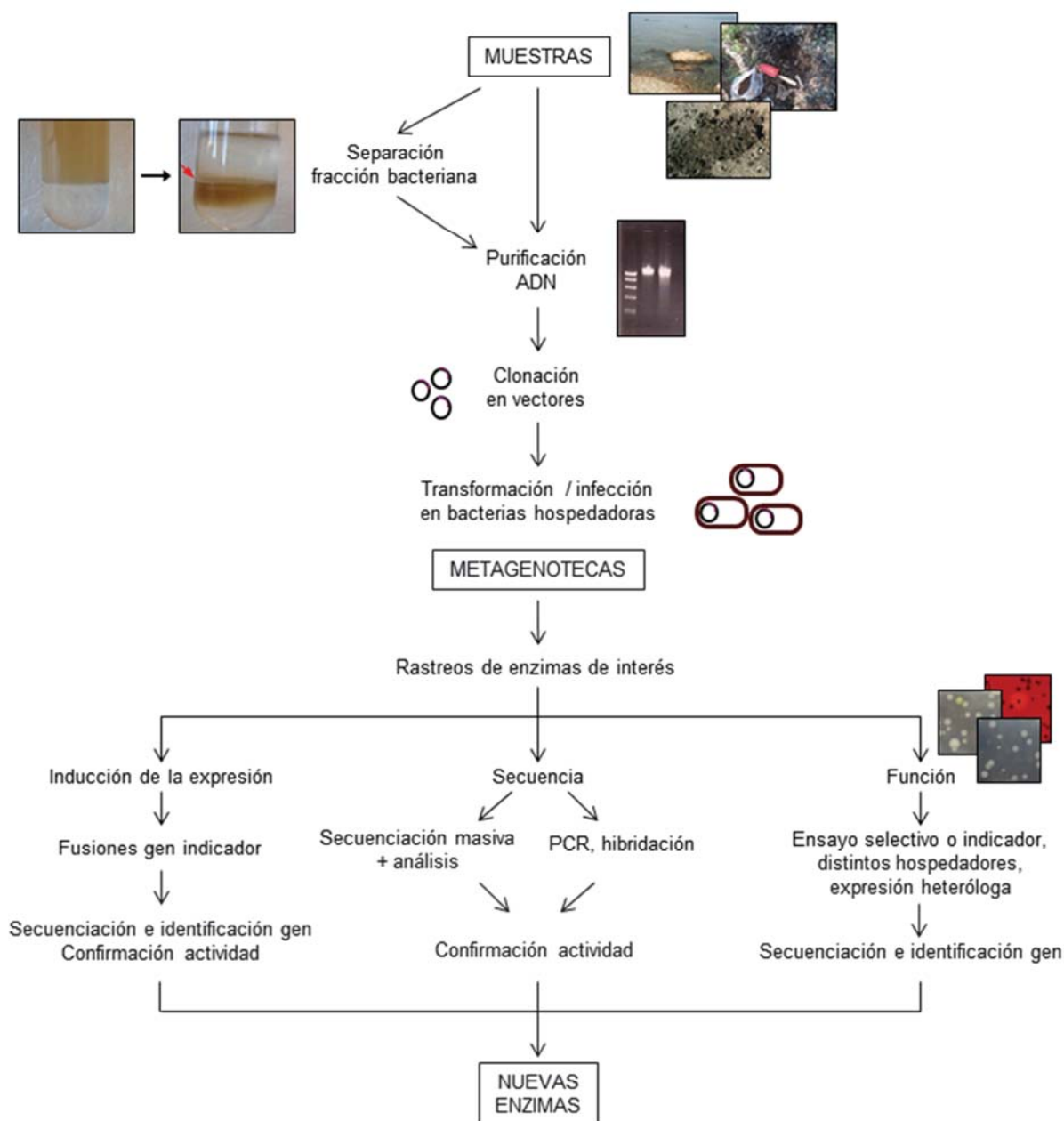


Figura 1. Esquema de la construcción y rastreo de metagenotecas para identificación de enzimas de interés. Se indican los pasos principales en la construcción de la metagenoteca y en los análisis de búsqueda, ya sean basados en inducción de la expresión génica, secuencia o función.

porque el sistema de rastreo empleado no sea apropiado. De este modo, genes interesantes presentes en una metagenoteca pueden ser identificados o pasar desapercibidos, en función del rastreo seleccionado.

Además de las estrategias basadas en secuencia y en función, con posterioridad se desarrolló una tercera estrategia basada en la inducción de la expresión génica, con la que se puede detectar la expresión de un gen de interés sin fenotipo detectable mediante su

asociación con un gen indicador. Por ahora ha sido mucho menos utilizada que las otras dos, pero ofrece una alternativa que podría permitir la identificación de algunos genes interesantes no identificados mediante los otros dos tipos de análisis.

El conocimiento sobre la dinámica funcional de una comunidad microbiana y el descubrimiento de nuevos genes de interés se puede acelerar si se combina la metagenómica con otras disciplinas, como la metatranscriptómica, que permite conocer aquellos genes presentes en una metagenoteca que se están transcribiendo en unas condiciones determinadas; y la metaproteómica, que ofrece una caracterización a gran escala del conjunto de proteínas de una comunidad en un momento dado (Simon y Daniel, 2011). Además, toda la información obtenida resulta de gran utilidad para el diseño exitoso de nuevas enzimas mediante ingeniería molecular, que permitiría modificar determinadas características de las enzimas conocidas con objetivos concretos (Ferrer y col., 2008).

A continuación se desarrollan las distintas estrategias comentadas, que se esquematizan en la figura 1.

2.2.1 Análisis basados en inducción de la expresión génica

Los análisis basados en la inducción de la expresión génica incluyen métodos de rastreo que permiten detectar indirectamente la expresión de los genes de interés (cuya actividad no da lugar por sí misma a fenotipos detectables), al provocar la expresión de un gen indicador, como *gfp*. Son análisis sensibles a la orientación de los genes de interés.

Dentro de estos tipos de rastreos, se encuentra *substrate-induced gene expression* (SIGEX) (Uchiyama y col., 2005). Se basa en que con frecuencia la expresión de genes reguladores de rutas catabólicas está inducida por un determinado sustrato y en que en ocasiones está controlada por un regulador codificado por un gen cercano. Para llevar a cabo esta técnica, se necesita un vector al que se le ha introducido un gen *gfp* sin promotor justo aguas abajo del sitio de clonación del ADN metagenómico, de manera que se estaría poniendo la expresión de este gen bajo el control de promotores que estuvieran presentes en el ADN metagenómico. En presencia del inductor se esperaría un aumento en los niveles de expresión del operón, y, si la transcripción alcanza al gen indicador, también en los de este, lo que se reflejaría en un aumento del nivel de fluorescencia emitida. Esos clones que emiten niveles de fluorescencia más elevados que el resto se pueden detectar mediante microscopía de fluorescencia o clasificación de células activadas por fluorescencia (FACS, *Fluorescence-*

Activated Cell Sorting). No se podrán detectar los genes catabólicos expresados constitutivamente ni aquellos cuyos reguladores transcripcionales estén codificados por genes alejados de los catabólicos, de manera que no se encuentren en el mismo clon (Uchiyama y col., 2005). Otro inconveniente de esta técnica es que la molécula proporcionada podría inducir otros genes (es relativamente frecuente la promiscuidad en los sistemas de inducción), dando lugar a falsos positivos, aunque estos se podrían descartar posteriormente.

Una técnica similar, pero solo aplicable a casos concretos, es la denominada *metabolite-regulated expression* (METREX) (Williamson y col. 2005). Se trata de un biosensor que detecta pequeñas moléculas que inducen *quorum-sensing* que se encuentra en la misma célula que el vector con el ADN metagenómico. El biosensor se compone de un promotor de *quorum-sensing* que controla el gen reportero *gfp*. Si el clon produce un inductor de *quorum-sensing*, la célula produce GFP, y el clon puede ser identificado por microscopía de fluorescencia o FACS.

También se encontraría incluida dentro de este tipo de estrategias la *product induced gene expression* (PIGEX). Uchiyama y Miyazaki (2010) desarrollaron un sistema para la detección de amidasas, en el que colocaron un activador transcripcional sensible a benzoato aguas arriba del gen indicador *gfp*. El clon de la metagenoteca emitirá fluorescencia, y por tanto podrá ser identificado, si produce amidasas capaces de transformar el sustrato aportado (benzamida) en benzoato.

2.2.2 Análisis basados en secuencia

Dentro de los análisis basados en secuencia están incluidas técnicas como la amplificación mediante PCR con oligonucleótidos degenerados (Courtois y col., 2003; Eschenfeldt y col., 2001; Owen y col., 2013; Precigou y col., 2001) y la hibridación con sondas específicas para una determinada clase de gen (Demanèche y col., 2009a; Demanèche y col., 2009b; Knietsch y col., 2003c). En ambos casos se requiere el diseño de oligonucleótidos o sondas a partir de regiones conservadas de genes conocidos. De este modo, solo podrán descubrirse variantes de proteínas ya conocidas. Cuando la actividad de interés requiere varios genes, en muchos casos también es posible la captura del grupo completo de genes mediante este tipo de análisis, siempre que se encuentren en el mismo clon. Se pueden utilizar oligonucleótidos degenerados o sondas para un gen diana, secuenciar el clon que resulte positivo, o al menos el entorno del gen diana, y comprobar que se hallen el resto de

genes. La identificación de genes conservados de muestras ambientales mediante PCR puede hacerse amplificando directamente el ADN aislado de las muestras, sin necesidad de construir metagenotecas, pero esta técnica por lo general da como resultado la amplificación de genes parciales, y la posterior recuperación de los genes completos de ADN aislado de muestras ambientales es complicado (Daniel, 2005).

También se encuentra dentro de este tipo de análisis la secuenciación aleatoria de ADN metagenómico, una estrategia que puede proporcionar gran cantidad de información muy diversa. Para ello, se procede a la secuenciación de clones de la metagenoteca, o incluso directamente del ADN metagenómico tras su purificación. A partir de las secuencias resultantes, se identifican secuencias homólogas al gen o conjunto de genes de interés existentes en las bases de datos mediante análisis bioinformáticos. Un ejemplo de este tipo de análisis es el realizado por Jadeja y col. (2014) para la búsqueda de oxigenasas. Los genes seleccionados se pueden amplificar mediante PCR, si se dispone de la muestra, o pueden ser sintetizados químicamente. Además, puede resultar de interés el estudio del entorno genómico de esos genes. Una vez que se tengan las secuencias, estas se pueden aprovechar para búsquedas de distinta clase.

Este tipo de análisis puede abarcar objetivos más ambiciosos que la identificación de genes de un determinado tipo, como es la obtención de genomas completos o parciales de los organismos que forman la comunidad microbiana de un ambiente y descifrar las relaciones metabólicas entre esos organismos. Dos de los primeros trabajos destacados dentro de este grupo son los realizados por Tyson y col. (2004) y por Venter y col. (2004). En el primer caso, secuenciaron los clones de una metagenoteca del biofilm microbiano del drenaje ácido de minas, dominado por un pequeño número de especies, y obtuvieron los genomas casi completos de las 2 especies dominantes y parciales de otras 3. Ese mismo año, Venter y col. llevaron a cabo el primer proyecto de secuenciación de gran envergadura, consistente en la secuenciación del metagenoma del Mar de los Sargazos. Las secuencias generadas incluían un total de más de mil millones de pares de bases, que se estimó que procedían de al menos 1.800 especies diferentes, identificándose más de 1,2 millones de genes desconocidos hasta el momento (Venter y col., 2004). La efectividad de esta aproximación estará limitada por la alta riqueza de especies, la heterogeneidad en la abundancia de los miembros de la comunidad y por cambios de posición en el genoma (Tyson y col., 2004).

En la actualidad, el rápido desarrollo de las técnicas de secuenciación masiva de nueva generación (NGS, *Next Generation Sequencing*) ha supuesto una mayor velocidad de

secuenciación y una importante reducción en los costes, a la vez que ha permitido prescindir de la clonación previa del ADN a secuenciar en vectores (Shokrallay col., 2012). Esto ha posibilitado la ejecución de una gran cantidad de proyectos de secuenciación de ADN ambiental (p. ej. Qin y col., 2010), que han supuesto una enorme aportación de nuevas secuencias a las bases de datos. Uno de los problemas de estas técnicas de secuenciación consiste en la producción de lecturas de poca longitud, lo que dificulta el ensamblaje para obtener secuencias más largas (contigs). Por otro lado, la generación de tal cantidad de secuencias hace que en muchas ocasiones resulte imposible realizar correcciones manualmente, y limita las anotaciones a métodos automáticos (Ferrer y col. 2005), por lo que la calidad de estas por lo general no es muy alta. Para tratar de asignar función a las secuencias (anotación funcional), se realizan alineamientos de esas secuencias mediante la herramienta BLAST (Altschul y col., 1990) con las secuencias contenidas en las bases de datos. Una restricción importante es que aproximadamente a la mitad de los genes de un metagenoma no se le puede asignar una función (Teeling y Glockner, 2012). Aunque esto no resulta extraño, teniendo en cuenta que incluso los genomas de las bacterias más estudiadas tienen un porcentaje muy elevado de genes sin función asignada (por ejemplo en 2009 un tercio de los genes codificantes de proteínas de *E. coli* K12 carecían de anotación funcional (Hu y col., 2009)). Previsiblemente, conforme se amplíen las bases de datos con secuencias de calidad y con función asignada, esta limitación se irá reduciendo.

Los análisis basados en secuencia presentan la importante ventaja de no requerir la expresión de los genes clonados en hospedadores distintos al de origen. Por otro lado, entre los problemas que presentan se encuentra que es posible identificar secuencias relacionadas con un determinado tipo de gen pero que resulten no estar completas, o que incluso estando completas no den lugar a una proteína funcional. Pero el principal inconveniente es que el descubrimiento de nuevas enzimas está limitado a aquellas que presentan cierto grado de similitud con otras conocidas, de manera que se pueda establecer una relación entre la nueva secuencia y alguna contenida en las bases de datos. Esto no implica que no se puedan descubrir enzimas de gran interés, ya que pequeñas variaciones en la secuencia pueden suponer cambios interesantes en la actividad.

2.2.3 Análisis basados en función

Los análisis basados en función permiten la identificación de enzimas con una determinada actividad completamente diferentes a las descritas con anterioridad, al ser detectadas por su función, reflejada en la aparición de un determinado fenotipo.

Dentro de este tipo de análisis existen distintas estrategias, y es necesario el diseño de un rastreo específico para cada actividad. La estrategia más sencilla es aquella que permite seleccionar los clones portadores de la actividad de interés mediante crecimiento en unas determinadas condiciones, a diferencia del resto de clones. Estos sistemas selectivos permiten el análisis de un gran número de clones en una misma placa. Es el caso de búsqueda de resistencias a antibióticos, metales pesados u otros compuestos tóxicos; y enzimas que permitan el aprovechamiento de un compuesto como fuente de carbono, nitrógeno o azufre, por ejemplo. En esos casos, los rastreos consistirán por lo general en el cultivo de los clones de la metagenoteca en medio sólido suplementado con el compuesto tóxico para la bacteria hospedadora, o con el compuesto que la bacteria ha de utilizar como nutriente para crecer. Algunos ejemplos de enzimas descubiertas empleando este tipo de rastreos son el caso de betalactamasas (Allen y col., 2009), resistencias a níquel (González-Pastor y Mirete, 2010) y nitrilasas (Bayer y col., 2011). Una variación de este tipo de ensayos consiste en el empleo de hospedadores mutantes que requieran la complementación con determinados genes para el crecimiento en las condiciones seleccionadas. Un ejemplo es el uso de un mutante sensible a bajas temperaturas del gen *polA* de *E. coli* para la identificación de ADN polimerasas (Simon y col., 2009).

Otro grupo de rastreos basados en función es el de aquellos en los que se permite el crecimiento de todos los clones a analizar, al cultivarlos en un medio no restrictivo, y se proporcionan unas condiciones en las que se puedan distinguir los clones productores de enzimas de interés, ya sea durante el crecimiento de los clones o posteriormente. Esta distinción frecuentemente consiste en producción de color o formación de halos alrededor de colonias. En muchos casos, se emplean tintes químicos o sustratos de las enzimas de interés unidos a cromóforos o fluoróforos, que permitan la detección de clones positivos mediante visualización, espectrometría o detección de fluorescencia (Kennedy y col., 2011). Como ejemplos de enzimas para las que se utilizan rastreos de este tipo se pueden citar las oxigenasas hidroxilantes, que dan color azul por la generación de índigo a las colonias de los clones que las producen al crecer en medio LB (van Hellemond y col., 2007); y dioxigenasas extradiólicas, que producen color amarillo en las colonias ya crecidas de los clones positivos

al aplicar catecol o un derivado sobre las mismas, debido a su transformación en un compuesto de ese color (Brennerova y col., 2009). También estarían incluidas las celulasas, que proporcionan a los clones que las producen cultivados en medio suplementado con carboximetilcelulosa la capacidad de generar un halo de degradación de este compuesto alrededor de la colonia productora, detectable con una tinción posterior de la placa (Wang y col., 2009); y las proteasas, cuya identificación se puede conseguir también mediante la detección de un halo claro alrededor de la colonia productora tras cultivar los clones en medio LB con leche desnatada (Neveu y col., 2011). Otro ejemplo de actividades detectadas por este tipo de rastreos es la producción de compuestos antimicrobianos. Para ello, sobre colonias crecidas de los clones de la metagenoteca, se siembra un organismo indicador (como *Bacillus subtilis*), y se comprueba si hay clones positivos mediante visualización de zonas de inhibición de crecimiento del organismo indicador sobre la colonia de alguno de los clones (Courtois y col., 2003). Como en el grupo de ensayos anterior, también se pueden emplear hospedadores mutantes cuya complementación con genes de interés permita la detección de un fenotipo característico en determinadas condiciones.

Una vez que se tengan clones capaces de llevar a cabo la actividad de interés, se procede a la identificación de la secuencia del gen responsable, mediante la secuenciación del clon o de algún subclón que mantenga la actividad. En esas secuencias se buscarán, mediante alineamiento con las bases de datos, secuencias homólogas a genes descritos previamente como responsables de dicha actividad, o mejor, secuencias que traducidas sean homólogas a proteínas conocidas. A diferencia de lo que ocurre al analizar secuencias metagenómicas al azar en los rastreos basados en secuencia, en este caso, como se sabe que realmente alguna de las pautas abiertas de lectura codifica una enzima capaz de llevar a cabo la función de interés, si no se identifica ninguna secuencia homóloga a ese tipo de enzimas, se deben subclonar dichas pautas abiertas de lectura hasta encontrar aquella que codifique la enzima en cuestión. Aunque más laborioso, estos resultados son muy interesantes, al permitir el descubrimiento de enzimas responsables de una actividad con secuencia completamente distinta a las descritas con anterioridad, que pueden tener características nuevas con más probabilidad que aquellas con secuencias menos novedosas. El descubrimiento de estas enzimas y la incorporación de sus secuencias a las bases de datos, facilitará y hará más efectivos los rastreos basados en secuencia sobre los que se habló anteriormente. A su vez, podría facilitar la asignación funcional de enzimas catalogadas como proteínas hipotéticas.

Entre las limitaciones de esta técnica se encuentra la necesidad de que exista un método de rastreo adecuado y suficientemente sensible, que frecuentemente son bastante laboriosos, y de que el producto génico sea activo en la bacteria hospedadora. Pero la restricción más importante es la baja probabilidad de expresión de los genes codificantes de dicha enzima en la estirpe heteróloga. De hecho, se ha comprobado que la mayoría de los genes normalmente no se expresan en las bacterias hospedadoras seleccionadas (Ekkers y col., 2012; Uchiyama y Miyazaki, 2009). Generalmente la frecuencia de clones que poseen un determinado tipo de gen es baja, por lo que si además se añade el hecho de que muchos de esos genes no se expresan en la bacteria hospedadora, se ve claramente la importancia de esa limitación.

La especie más utilizada para la realización de los rastreos es *E. coli*, donde se ha estimado por análisis bioinformáticos con 32 genomas que el 40 % de los genes podrían ser detectados por clonación aleatoria (Gabor y col., 2004b). Por tanto, aunque *E. coli* ha permitido el descubrimiento de una gran cantidad de enzimas, el empleo de otras estirpes permitirá la expresión y detección de otras enzimas incapaces de expresarse en *E. coli*. Por otro lado, el desarrollo de sistemas de expresión heteróloga también aumentará las probabilidades de expresión de los genes contenidos en el ADN metagenómico en una estirpe determinada. Entre otros asuntos, en el siguiente apartado se discute sobre los principales problemas que pueden impedir la expresión de genes metagenómicos y sobre posibles mejoras.

3. MEJORAS EN METAGENÓMICA FUNCIONAL

Al margen de sus ventajas, los análisis basados en función presentan una serie de limitaciones, entre las que se encuentran principalmente la falta de un sistema de rastreo adecuado para algunas actividades y, sobre todo, la baja probabilidad de expresión del ADN metagenómico en las bacterias hospedadoras. Estas limitaciones se pueden reducir mediante el desarrollo de nuevos sistemas de rastreo y mediante estrategias encaminadas a forzar la expresión de ese ADN.

3.1 Desarrollo de sistemas de rastreo

La detección de cada actividad requiere sistemas de rastreo específicos. De este modo, las enzimas susceptibles de ser identificadas mediante rastreos funcionales son solo aquellas para las que se haya desarrollado un sistema de detección. En teoría, cualquier ensayo bioquímico que permita la detección de una actividad podría emplearse como ensayo de rastreo. El problema que se presenta es que normalmente es necesario analizar muchos clones para identificar uno positivo, debido a que la frecuencia de clones positivos para una actividad en una metagenoteca suele ser baja. Esto limita en muchas ocasiones los rastreos a aquellos más sencillos, de manera que se permita la identificación de los clones en forma de colonias en placas.

Una alternativa a la realización de los rastreos en placas con colonias es el empleo de placas multipocillo (p. ej. Suenaga y col., 2007), que permite analizar los clones de manera individual o en grupos reducidos en cultivo líquido o incluso trabajar con extractos, lo que amplía el número de posibles ensayos bioquímicos a realizar. El empleo de extractos aumenta la sensibilidad del rastreo, lo que resulta muy útil para la identificación de clones con bajos niveles de actividad y para la detección de actividades producidas por enzimas intracelulares o que no estén siendo secretadas, pero son muy laboriosos e inviables para analizar un elevado número de clones de manera manual. Por todo lo comentado, resulta de gran utilidad el uso de sistemas robotizados (Kennedy y col., 2011), que además permiten realizar réplicas ordenadas de los clones que se estén ensayando en el caso del uso de extractos o de otro tipo de análisis que implique la muerte de las células, a partir de las que se podrá recuperar la biomasa de los clones que resulten de interés. Otra opción en lugar de la realización de extractos en el caso de enzimas extracelulares que presenten problemas de secreción, es la incorporación de un paso de lisis celular empleando detergentes no iónicos como Triton X o Tween, que permite la liberación de la enzima de la bacteria, sin alterar su conformación nativa/activa (Kennedy y col., 2011).

3.2 Estrategias para el aumento de las probabilidades de expresión del ADN metagenómico

La identificación de un clon con una actividad enzimática determinada puede verse dificultada porque el entorno de la bacteria hospedadora no sea adecuado para esa enzima, por ejemplo por faltarle algún cofactor necesario para su actividad o por la potencial toxicidad del

producto génico o de un producto de su actividad para la bacteria hospedadora (Warren y col., 2008). Sin embargo, la principal limitación, como ya se ha comentado en más de una ocasión, reside en la baja probabilidad de expresión del ADN metagenómico en las bacterias hospedadoras.

La expresión génica tiene distintas etapas, cada una de las cuales puede afectar a la cantidad y buen estado de la proteína de interés, influyendo en el nivel de actividad final. La falta de expresión puede deberse a una falta de transcripción eficiente (por la ausencia del promotor en caso de operones multicistronicos, por la incapacidad de reconocimiento del promotor por parte de la maquinaria de la bacteria hospedadora, o por la falta de factores de transcripción o inductores necesarios), a una débil traducción (por un sitio de unión al ribosoma poco eficaz con la maquinaria de expresión de la bacteria hospedadora, por la falta de reguladores de la traducción o por un uso de codones diferente), a una pobre secreción en el caso de proteínas extracelulares (porque los sistemas de secreción de la bacteria hospedadora no puedan secretar la proteína heteróloga), a un plegamiento incorrecto (por la falta de las chaperonas adecuadas) (Streit y Schmitz, 2004) o a otras modificaciones postraduccionales requeridas. Todos estos factores pueden influir en distinto grado en la detección de una actividad determinada, en función de la combinación gen/hospedador (Ekkers y col. 2012), si bien la etapa de transcripción es la más limitante y la que más frecuentemente está regulada en bacterias.

Se pueden distinguir dos estrategias que pueden ser complementarias para conseguir la expresión de al menos parte de esos genes. Por un lado, introducir el material genético en distintas especies puede permitir la expresión de estos genes por sí mismos en una u otra bacteria. Por otro lado, forzar la expresión de manera heteróloga, principalmente desde promotores inducibles incorporados al vector dirigidos hacia el ADN metagenómico, puede proporcionar una forma artificial de expresión de estos genes. La primera estrategia puede servir para mejorar la expresión en sus distintas etapas, mientras que la segunda se centra en aumentar los niveles de transcripción, que en muchos casos es la principal limitación.

3.2.1 Desarrollo de sistemas de expresión homóloga

Mediante los sistemas de expresión homóloga, el ADN metagenómico se transcribirá desde sus propios promotores de forma regulada o no. Se trata de proporcionar unas condiciones adecuadas para que esto sea posible. Las bacterias de procedencia del ADN

metagenómico son múltiples y además son desconocidas, así como la naturaleza de los genes contenidos y su posible regulación, por lo que resulta imposible establecer un contexto y unas condiciones que puedan favorecer la expresión de los genes de interés, salvo la adición al medio de cultivo de moléculas potencialmente inductoras de la actividad que se desea identificar.

En fragmentos largos es más probable que se encuentren elementos reguladores que en determinados casos permitan la expresión de los genes desde sus propios promotores, pero esto no siempre garantiza la expresión propia de los genes. Cuando más cercana sea la bacteria hospedadora a la original de la que procede el ADN metagenómico mayores serán las probabilidades de expresión, por lo que una buena estrategia consiste en realizar los rastreos para la actividad de interés en distintas especies. De este modo, además de poder aumentar la transcripción, se aumentan las probabilidades de que alguna de las estirpes hospedadoras tenga un uso de codones, chaperonas, cofactores y, en su caso, sistema de secreción, adecuados. La falta de un sistema de secreción apropiado en ocasiones se puede suplir con la incorporación de un paso de lisis en los protocolos de rastreo, como se comentó anteriormente.

La especie más empleada para los rastreos funcionales es *E. coli*, debido a que es una especie muy fácil de cultivar y manipular en el laboratorio y con una alta frecuencia de transformación. Además, tiene requerimientos relativamente relajados en relación al reconocimiento de señales de inicio de la traducción (Gabor y col., 2007). En *E. coli* se expresan muchos genes del ADN metagenómico por sí mismos, pero hay muchos más que no son capaces de expresarse (solo el 40 % de las actividades enzimáticas podrían ser detectadas por clonación aleatoria en *E. coli* (Gabor y col., 2004b)). Teniendo esto en cuenta, cada vez es más habitual el uso de otras especies, entre las que se encuentran *Bacillus subtilis*, *Streptomyces lividans* o *Pseudomonas putida*. Puede ser muy interesante emplear en los rastreos para una determinada actividad estirpes en las que se hayan descrito previamente enzimas de ese tipo, por ejemplo del género *Pseudomonas* o *Rhodococcus* para actividades degradadoras de contaminantes orgánicos, o *Streptomyces* para producción de antibióticos (Martínez y col., 2004). Del mismo modo, si se buscan enzimas con características extremas, podría ser ventajoso emplear alguna estirpe extremófila, como *Thermus thermophilus* para detectar enzimas que trabajen a altas temperaturas (Ferrer y col., 2007). La elección de la estirpe para los rastreos tiene como limitación la existencia de vectores adecuados para ella, en los que habrá que construir previamente la metagenoteca, así como la capacidad de esa

estirpe de ser cultivada y manipulada en el laboratorio (se necesita al menos poder introducirle ADN). El empleo de unas u otras especies en los rastreos funcionales hay que tenerlo en cuenta a la hora de seleccionar el vector para la construcción de las metagenotecas, ya que estos deben ser capaces de mantenerse y replicarse en esas especies. Por tanto, el vector seleccionado debe tener un marcador (por lo general un gen de resistencia a algún antibiótico) válido para las estirpes con las que se vaya a trabajar, además de un origen de replicación, o varios, adecuados. Es interesante que los vectores permitan ensayar la metagenoteca en más de una estirpe, bien porque tengan varios orígenes de replicación o uno de amplio espectro, de manera que se aumentan las probabilidades de expresión de un gen concreto, ya sea en uno u otro organismo. En ese caso, resulta muy útil incorporar al vector también un origen de transferencia conjugativa, que por lo general permite la transferencia entre las distintas estirpes de una manera más eficiente que mediante transformación. Hay que valorar los elementos a introducir en el vector, ya que en ocasiones el aumento del tamaño del vector irá en detrimento del tamaño aceptado de ADN metagenómico, como ocurre en el caso de cósmidos y fósidos, debido a que el tamaño final no debe superar las aproximadamente 50 kb capaces de ser empaquetadas por partículas del fago lambda.

Aunque hay estudios en los que se construyó la metagenoteca directamente en otro hospedador, encontrando clones de interés, como Wang y col. (2000), en *S. lividans*, lo más habitual es construir las metagenotecas en *E. coli*, en vectores que permitan su transferencia a otras estirpes. En algunos casos esos vectores se replican autónomamente en los nuevos hospedadores, pero en otros se integran en sitios específicos del cromosoma (Courtois y col., 2003). Algunos trabajos han demostrado la utilidad de seguir empleando *E. coli* como bacteria hospedadora, al haberse encontrado enzimas que no eran capaces de expresarse en las otras bacterias empleadas (Courtois y col., 2003), pero las ventajas de ampliar el rango de bacterias hospedadoras es claro, ya que ha permitido detectar clones que no pudieron detectarse en *E. coli* (Wexler y col., 2005; Craig y col., 2009). Sosio y col. (2000) y Martínez y col. (2004) desarrollaron una serie de vectores BACs y estirpes de *S. lividans*, y de *S. lividans* y *P. putida*, respectivamente, a las que se podía transferir la metagenoteca desde *E. coli*. En el caso de Martínez y col. (2004) comprobaron que las 3 estirpes hospedadoras tenían diferentes capacidades de expresión para algunos genes de la metagenoteca ensayada. Por otro lado, Hain y col (2008) construyeron un nuevo vector tipo fósido de amplio espectro (pUvBBAC) capaz de replicarse tanto en bacterias Gram negativas como Gram positivas. Otro ejemplo es el cósmido de amplio espectro pJWC1 (Craig y col., 2009), en el que construyeron una

metagenoteca y realizaron rastreos de actividad en seis especies: *Agrobacterium tumefaciens*, *Burkholderia graminis*, *Caulobacter vibrioides*, *Escherichia coli*, *Pseudomonas putida* y *Ralstonia metallidurans*; observando un solapamiento mínimo en los grupos de clones a los que proporcionaron acceso cada una (Craig y col., 2010). Y el vector pGNS-BAC, capaz de replicarse en diversas especies Gram negativas, como *Escherichia coli*, *Pseudomonas* spp., *Salmonella enterica*, *Serratia marcescens*, *Vibrio vulnificus* y *Enterobacter nimipressuralis* (Kakirde y col., 2011). También se han desarrollado sistemas que permiten la transferencia de la genoteca al organismo termófilo *Thermus thermophilus* (pCT3FK, Angelov y col., 2009). Algunos de estos vectores de amplio espectro también permiten la regulación del número de copias, lo que puede facilitar la detección de la actividad, como pRS44 (Aakvik y col., 2009).

3.2.2 Desarrollo de sistemas de expresión heteróloga

Mediante los sistemas de expresión heteróloga, se trata de forzar la transcripción del ADN metagenómico desde promotores externos incorporados a los vectores empleados para la construcción de las metagenotecas. El objetivo es maximizar las probabilidades de expresión de cualquier gen presente en las bibliotecas metagenómicas. Se puede emplear como una estrategia alternativa o, mejor, complementaria a la descrita en el apartado anterior. En este caso, también se requiere que el vector para la construcción de metagenotecas contenga una serie de elementos, como promotores dirigidos hacia el sitio de clonación del ADN metagenómico, entre otros. En ocasiones, parte de los elementos necesarios para los sistemas de expresión heteróloga se introducen en el cromosoma de la estirpe hospedadora. Puede ser de utilidad un origen de transferencia conjugativa, para facilitar la transferencia de la metagenoteca a distintas estirpes que, aunque sean de la misma especie, permitan el uso de diferentes sistemas de expresión heteróloga. A continuación se comentan distintos elementos que se pueden emplear, dirigidos a aumentar la tasa de transcripción del ADN metagenómico, ya sea forzando su inicio o dificultando su terminación.

Se pueden incorporar a los vectores promotores heterólogos bacterianos, constitutivos o inducibles, orientados hacia el sitio de clonación del ADN metagenómico. La transcripción desde esos promotores se llevará a cabo por la propia ARN polimerasa de la bacteria hospedadora. Para la selección del promotor heterólogo hay que tener en cuenta el nivel de expresión deseado. Los niveles altos de expresión por lo general aumentarán los niveles de actividad detectados, pero puede ser perjudicial si el producto génico es tóxico para el

hospedador. Mediante el uso de promotores gradualmente inducibles se puede controlar el nivel de expresión en cada momento. Algunos ejemplos se describen en el trabajo de Troeschel y col. (2012), que construyeron una serie de vectores que, entre otros elementos, tenían el promotor inducible por xilosa *p_{xyl}*; y en el de van Hellemond y col. (2007), que utilizaron el vector *pZerO-2*, que lleva un promotor *plac*, inducible con IPTG.

El uso de estos promotores es útil sobre todo para metagenotecas de tamaño de ADN metagenómico pequeño, y en consecuencia, para actividades que solo requieran la expresión de un gen, ya que la transcripción desde el promotor heterólogo solo será capaz de alcanzar el primero o con suerte los primeros genes clonados en el mismo sentido que el promotor, hasta encontrar una señal de terminación de la transcripción. Otra limitación, por tanto, es que solo se podrá inducir la expresión de los genes que se encuentren en la misma orientación que el promotor heterólogo. Este último problema se puede solucionar colocando un promotor a cada lado del sitio de clonación del ADN metagenómico orientados ambos hacia este. Lämmle y col. (2007) construyeron una metagenoteca en el vector *pJOE930* (Altenbuchner y col., 1992), con dos promotores *plac* inducibles convergentes al ADN metagenómico.

Un sistema interesante para transcribir heterológamente fragmentos de ADN metagenómico grandes es el uso de la ARN polimerasa del fago T7. Esta polimerasa es mucho más rápida que la ARN polimerasa bacteriana y es insensible a muchos de los terminadores de la transcripción bacterianos (Studier y Moffatt, 1986; Jeng y col., 1990), siendo por tanto una estrategia válida para la detección de actividades dependientes de más de un gen. Habría que incorporar el promotor del gen *10* del fago T7 (en adelante *pT7*) en el vector, orientado hacia el ADN metagenómico, y el gen de la ARN polimerasa del fago T7, preferentemente en el cromosoma de la estirpe hospedadora. Dos promotores *pT7* convergentes flanqueando el sitio de clonación de fragmentos de ADN metagenómico permitirían la transcripción bidireccional del inserto, incrementando de este modo las posibilidades de expresión de un gen sin reparar en su orientación en el fragmento. Como la transcripción desde el promotor *pT7* puede alcanzar unos niveles muy altos, se deben controlar los niveles de la ARN polimerasa del fago T7, manteniendo la expresión del gen que la codifica en unos niveles bajos (para evitar posibles problemas por productos génicos tóxicos, que llevarían a la pérdida de los clones con esos genes). Hay publicados varios trabajos en los que se construyeron metagenotecas en vectores con promotor *pT7*, como por ejemplo el de Troeschel y col. (2012), que crearon vectores que contenían, entre otros elementos, el promotor *pT7* y una estirpe de *P. putida* productora de la ARN polimerasa del

fago T7; y Rhee y col. (2005), que utilizaron el vector pCC1FOS (Epicentre), que posee un promotor pT7, y la estirpe productora de la ARN polimerasa del fago T7 BL21(DE3) (Novagen).

Estas dos vías que persiguen la mejora de la expresión del ADN metagenómico (desarrollo de sistemas de expresión homóloga y heteróloga) se pueden combinar, con la creación de vectores que contengan promotores heterólogos y que a su vez sean de amplio espectro de hospedadores. Entre otros, está el sistema construido por Troeschel y col. (2012), que consiste en vectores que se pueden mantener en distintas bacterias hospedadoras, incluyendo *E. coli*, *B. subtilis* y *P. putida*, y que contienen los promotores pT7 y *pxyl*.

En esta tesis se han desarrollado dos sistemas distintos con el objetivo de aumentar las probabilidades de expresión de genes contenidos en el ADN metagenómico en estirpes de *E. coli*, combinando elementos víricos y bacterianos. Por un lado se ha introducido en el vector el promotor del fago T7 comentado anteriormente y se ha construido una estirpe productora de la ARN polimerasa de T7. Y por otro, se ha incorporado al vector un promotor bacteriano, *psal*, al que se ha acoplado el sistema de antiterminación N del fago lambda, no habiéndose empleado previamente este sistema para la expresión de metagenotecas. En cuanto al promotor *psal*, es un promotor inducible por salicilato (Yen y Gunsalus, 1982; Schell, 1985; Schell y Wender, 1986), que permite modular los niveles de expresión en función de la concentración de salicilato aportada (Royo y col., 2007). La proteína antiterminadora N del fago lambda modifica a la ARN polimerasa, de manera que la hace insensible a diferentes terminadores de la transcripción, avanzando a través de ellos. Para este sistema se requieren las proteínas Nus de *E. coli* (Nodwell y Greenblatt, 1991; Greenblatt y col., 1998). Se aportó el gen de la proteína antiterminadora N en el cromosoma de la estirpe hospedadora, y el sitio de reconocimiento de esta en la posición adecuada respecto al promotor heterólogo *psal* en el vector. De este modo, la transcripción desde el promotor heterólogo *psal* sujeta a antiterminación por la proteína antiterminadora N podría transcurrir a través de toda la longitud del fragmento metagenómico, aumentando las posibilidades de expresión de un gen sin reparar en su posición dentro del fragmento metagenómico clonado.

4. RESISTENCIAS A ANTIBIÓTICOS BETALACTÁMICOS

El término antibiótico hace referencia a cualquier clase de molécula orgánica que inhibe el crecimiento o mata microorganismos mediante interacciones específicas con sitios de reconocimiento bacterianos (Schmieder y Edwards, 2012). Los antibióticos son fundamentales en tratamientos médicos, pero tienen un tiempo de vida útil limitado, debido a la aparición de organismos resistentes. Por eso es importante el descubrimiento continuo de nuevos antibióticos y conocer los mecanismos de resistencia a estos.

Los antibióticos se distribuyen en distintos grupos. Uno de ellos lo constituyen los antibióticos betalactámicos, entre los que se encuentran las penicilinas, ampicilinas, cefalosporinas y monolactámicos, producidos tanto por bacterias Gram positivas como Gram negativas (Palzkill, 2013). Se caracterizan por poseer un anillo betalactámico en su estructura molecular. Estos antibióticos actúan inhibiendo enzimas transpeptidasas, que son esenciales para la síntesis de la capa de peptidoglicano de la pared celular de la bacteria (Sauvage y col., 2008; Palzkill, 2013), con la consecuente muerte de las células en crecimiento. La alta eficacia y baja toxicidad (Livermore, 1996) de los antibióticos betalactámicos hacen que se encuentren entre los antibióticos más frecuentemente prescritos para humanos y ganado. Este elevado uso ejerce una fuerte presión selectiva para genes que codifiquen resistencia a este tipo de antibióticos (Henriques y col., 2006; Demanèche y col., 2008).

Los genes capaces de proporcionar resistencia a antibióticos, incluidos los betalactámicos, son comunes en ambientes naturales (Allen y col., 2010), y son muy abundantes en suelos (Schmieder y Edwards, 2012), incluso se han identificado genes de resistencia a antibióticos en ambientes sin exposición conocida al uso médico de antibióticos (Allen y col., 2009). De hecho, muchos de estos genes probablemente tengan otro papel principal. Por ejemplo, algunas bombas de eflujo, que pueden tener como función en su ambiente proporcionar tolerancia a metales pesados, permiten a su vez la resistencia a antibióticos (Nies, 2003; Poole, 2005; Allen y col., 2010). La identificación de resistencias a antibióticos se puede llevar a cabo mediante análisis metagenómicos funcionales sencillos consistentes en la detección de crecimiento de clones en medio sólido que contenga el antibiótico en cuestión (Allen y col. 2009; Donato y col., 2010).

Además de la relativamente baja permeabilidad de la membrana externa de las bacterias, hay 3 mecanismos principales que confieren tolerancia a antibióticos betalactámicos (Wilke y col., 2005): la producción de enzimas betalactamasas, que hidrolizan el anillo

betalactámico de este tipo de antibióticos; la utilización de peptidasas de pared celular insensibles a betalactámicos; y la expulsión activa de moléculas de betalactámicos de las células mediante bombas de eflujo. Respecto al último punto, se han descrito bombas de eflujo de la familia de permeasas de resistencia-nodulación-división celular (RND) que contribuyen a la tolerancia de bacterias a antibióticos betalactámicos (Poole, 2005), aunque no que proporcionen resistencia a este tipo de antibióticos por sí mismas. La forma de resistencia más frecuente es la producción de betalactamasas (Henriques y col., 2006). En base a su secuencia de aminoácidos, las betalactamasas se agrupan en cuatro clases: A, B, C y D (Ambler, 1980). Las de las clases A, C y D son serina-betalactamasas, y las de clase B son metalo-betalactamasas dependientes de cinc, y poseen un mecanismo de acción diferente a las anteriores (Palzkill, 2013).

5. DEGRADACIÓN DE CONTAMINANTES AROMÁTICOS

Una parte importante de los contaminantes orgánicos que se encuentran en el medio son hidrocarburos aromáticos. Estos compuestos, por lo general tóxicos y con frecuencia mutagénicos, son difíciles de degradar, debido a la gran estabilidad que les proporcionan los anillos aromáticos que poseen sus moléculas, a lo que en ocasiones hay que sumar las sustituciones que contengan. Sin embargo, hay un elevado número de microorganismos capaces de degradar estos compuestos, ya sea de forma anaerobia, mediante reacciones reductivas, o aerobia, a través de reacciones de oxidación.

La degradación aerobia de hidrocarburos aromáticos (figura 2), implica en primer lugar una desestabilización del anillo mediante reacciones de oxigenación, de manera que en la molécula quedan dos grupos hidroxilo en dos posiciones contiguas del anillo (catecol o derivado catecólico). La formación de los correspondientes catecoles la llevan a cabo monooxigenasas o dioxigenasas hidroxilantes, según aporten uno o dos átomos de oxígeno. Si la molécula tiene ya un grupo hidroxilo, las monooxigenasas introducen otro grupo hidroxilo, liberándose H_2O ; y si la molécula no tiene ninguno, principalmente actúan las dioxigenasas, que incorporan al anillo aromático dos grupos hidroxilo, seguido de una deshidrogenación que elimina dos átomos de hidrógeno y restablece la resonancia en el anillo aromático (Harayama y col., 1992).

Estos intermediarios hidroxilados, sufren una nueva reacción que conduce a la rotura del anillo. Este último paso lo realizan dioxigenasas de rotura del anillo (Harayama y col., 1992), que pueden ser dioxigenasas intradiólicas, cuando abren el anillo entre los dos carbonos que poseen grupos hidroxilo (rotura-*orto*); o extradiólica, cuando la apertura la realizan entre un carbono con grupo hidroxilo y otro que no lo tiene (rotura-*meta*). Además de la estrategia general descrita, en la degradación de los hidrocarburos aromáticos tienen lugar otras reacciones más específicas según la molécula de la que se trate, como la transformación de los posibles sustituyentes que pudiera tener. De este modo, partiendo de un rango muy amplio de compuestos, se converge en unos pocos intermediarios clave, catecol o derivados catecólicos, que a partir de ese momento continúan siendo metabolizados por una de las siguientes dos vías (con un grupo de enzimas asociado a cada una de ellas): la ruta de rotura-*orto* (también llamada ruta del β -ceto adipato) y la ruta de rotura-*meta* (también llamada ruta del α -ceto adipato) (Peng y col., 2008). En el caso de moléculas poliaromáticas, se producen reacciones sucesivas de desactivación y rotura de los anillos. Los compuestos desaromatizados (que son cadenas hidrocarbonadas), sufren otra serie de transformaciones hasta convertirse en intermediarios del metabolismo central de la célula, como piruvato, acetil CoA o succinato, entre otros.

Las oxigenasas, que como se acaba de comentar están implicadas en distintos pasos de la ruta de degradación de hidrocarburos aromáticos, son de gran interés. Conocer mejor las enzimas ya identificadas de este tipo y descubrir otras nuevas, además de tener una evidente utilidad en biorremediación de ambientes contaminados, también es importante en la producción de compuestos de interés en la industria (en el campo textil, agrícola, etc.) y medicina (por la generación de compuestos anticancerígenos y otros medicamentos). Esto es debido a que muchas de estas enzimas son bastante promiscuas, y catalizan la reacción de un amplio rango de sustratos además de sus sustratos nativos; al requerimiento de unas condiciones por lo general moderadas de pH, temperatura y presión para la reacción; a su capacidad para utilizar oxígeno como oxidante (barato y respetuoso con el medio ambiente); y a sus interesantes propiedades de quimio-, regio- y/o enantioselectividad, frente a lo que difícilmente pueden competir los procesos químicos (Fishman y col., 2005; van Hellemond y col., 2007; Kennedy y col. 2011).

Las oxigenasas hidroxilantes o iniciales y las dioxigenasas de rotura del anillo, forman parte, junto con otras, del grupo de enzimas oxigenasas. Las oxigenasas incorporan átomos de oxígeno en la molécula sustrato, para lo cual normalmente necesitan NAD(P)H como donador

de electrones y un cofactor metálico u orgánico (van Hellemond y col., 2007). A su vez, las oxigenasas están incluidas dentro del diverso grupo de enzimas de las oxidorreductasas, caracterizadas por su capacidad de catalizar la transferencia de electrones entre moléculas.

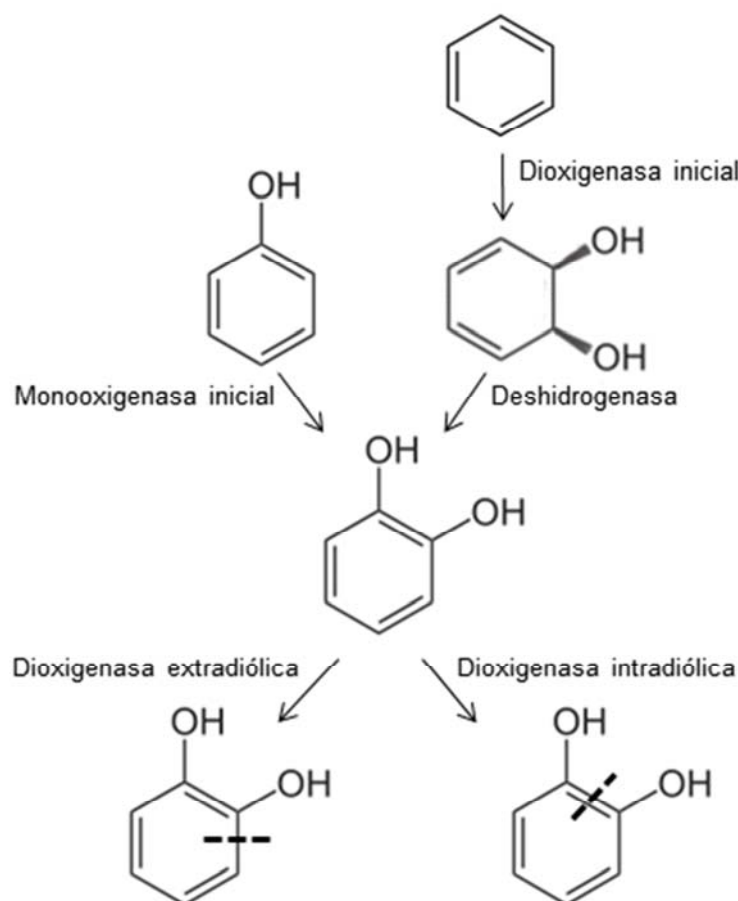


Figura 2. Esquemización de los principales pasos en la degradación aerobia de hidrocarburos aromáticos. Se representa como ejemplo la rotura del anillo de dos hidrocarburos aromáticos sencillos, fenol (a la izquierda) y benceno (a la derecha), y se indican las enzimas implicadas en cada paso.

5.1 Oxigenasas hidroxilantes de anillos aromáticos

Como se ha comentado anteriormente, las oxigenasas hidroxilantes de anillos aromáticos, también llamadas oxigenasas iniciales, son enzimas clave en la degradación de compuestos aromáticos, que catalizan la reacción de incorporación de uno o dos átomos de oxígeno al anillo aromático produciendo su desestabilización. En función de que incorporen uno o dos átomos de oxígeno se denominan mono- o dioxigenasas, respectivamente (figura 2). Existen oxigenasas que pueden llevar a cabo tanto reacciones de monooxigenación como

de dioxigenación, por ejemplo la naftaleno dioxigenasa de *Pseudomonas* sp. 9816-4 (Gibson y col., 1995). Todas ellas emplean oxígeno como cosustrato y NAD(P)H como fuente de poder reductor. Son abundantes en bacterias, aunque se encuentran en todos los grupo de seres vivos (van Hellemond y col., 2007). Tradicionalmente, se han identificado oxigenasas iniciales mediante el estudio de microorganismos capaces de crecer en presencia de compuestos aromáticos. Con el desarrollo de la metagenómica, se ha acelerado el descubrimiento de nuevas enzimas de este tipo. Se pueden realizar ensayos funcionales empleando distintos sustratos cromogénicos y fluorogénicos (Kennedy y col., 2011). El ensayo más habitual es la detección de colonias de color azul por la formación de índigo al cultivar *E. coli* en medio rico LB. Algunas oxigenasas iniciales son capaces de oxidar el indol que se produce desde el triptófano por la triptofanasa endógena de *E. coli*, generando el tinte azul índigo (Ensley y col., 1983, O'Connor y col., 1997).

Se han establecido distintas clasificaciones, atendiendo a distintos criterios, como el número y características de los componentes proteicos u homología de secuencia (Batie y col., 1991; Nam y col., 2001; Kweon y col., 2008). Las oxigenasas iniciales son complejos multiméricos, constituidos por dos o tres proteínas, distinguiéndose la oxigenasa en sí, y las proteínas de la cadena de transporte de electrones (ferredoxina reductasa y, en su caso, ferredoxina), que permiten el trasiego de electrones desde la fuente de poder reductor (NAD(H)P) hasta la oxigenasa. De este modo, la oxigenasa puede estar acompañada por una ferredoxina reductasa, o por una ferredoxina reductasa y una ferredoxina. La oxigenasa inicial es la proteína que realiza la oxigenación del anillo aromático. Existen oxigenasas homoligoméricas (α_n) y heteroligoméricas ($\alpha_n\beta_n$), siendo la subunidad α la que posee el centro catalítico (Peng y col., 2010). Las ferredoxinas de estos sistemas son proteínas pequeñas con un centro sulfoférrico. Aceptan los electrones de la ferredoxina reductasa y se los ceden a la oxigenasa, con la que interaccionan de forma bastante específica. Las ferredoxinas reductasas tienen al menos una flavina, y pueden tener también un centro sulfoférrico. Toman los electrones de la fuente de poder reductor NAD(P)H, y se los ceden a las ferredoxinas o, en caso de no haber ferredoxinas, directamente a la oxigenasa. En el caso de ausencia de la ferredoxina, la reductasa contiene tanto una flavina como un centro sulfoférrico (Mason y Cammack, 1992). Las reductasas no son tan específicas como las ferredoxinas, y en ocasiones su función la puede realizar otra reductasa presente en la célula.

Un tipo particular de oxigenasas iniciales son los citocromos P450. A diferencia de las demás, son hemoproteínas, y deben su nombre a su pico de absorción a 450 nm. Catalizan

reacciones de monooxigenación, de muy variado tipo, sobre un enorme número de sustratos, siendo de utilidad para la producción de medicamentos, síntesis de compuestos fisiológicamente importantes como esteroides y ácidos grasos, biorremediación, degradación de herbicidas e insecticidas, etc. (Hannemann y col, 2007; Kelly y Kelly, 2013). Constituyen el grupo de monooxigenasas más abundantes y diversas, con representantes tanto en microorganismos como en animales y plantas. Por otro lado, hay bacterias en las que no se encuentra este tipo de enzima, como *E. coli*.

El sistema de nomenclatura de los citocromos P450, fue propuesto y desarrollado por Nebert y col. (Nebert y col., 1987; Nebert y col., 1989; Nebert y col., 1991; Nelson y col., 1993; Nelson y col., 1996), clasificándolos en base a la homología de secuencia en familias y subfamilias. En él se establece que los genes P450 se identifican por la abreviatura CYP seguida por un número que indica la familia (proteínas con más del 40 % de identidad de secuencia), una letra que indica una subfamilia (más del 55 % de identidad) y un número que representa el gen individual dentro de la subfamilia, por ejemplo CYP105A3. Existen citocromos P450 cuyas secuencias tienen una identidad menor del 20 % entre ellas. Sin embargo, su estructura general está muy conservada, y se mantiene un mecanismo común de activación de oxígeno (Hannemann y col, 2007). Por otro lado, las regiones más variables se corresponden con las regiones flexibles de reconocimiento de sustrato (Gotoh, 1992), que hace de los citocromos P450 los catalizadores biológicos más versátiles conocidos. Los citocromos P450 requieren, como el resto de oxigenasas iniciales, un sistema de transporte de electrones en el que participan una o dos proteínas más.

La evolución de estas proteínas ha sido muy compleja y una clasificación únicamente basada en secuencia no resulta muy adecuada. Degtyarenko y Archakov (1993), además de construir un nuevo árbol evolutivo de la superfamilia de citocromos P450 utilizando alineamientos múltiples de secuencias consenso, propusieron una clasificación basada en el número de componentes proteicos, según sean uno, dos o tres, y a su vez según sean procariotas o eucariotas. Ha habido otras clasificaciones en función del sistema de transporte de electrones (p. ej. Werck-Reichhart y Feyereisen, 2000), pero la más completa probablemente sea la de Hannemann y col (2007). Realizaron una clasificación en función de la composición y características de los sistemas de transporte de electrones, dividiendo los citocromos P450 conocidos hasta la fecha en 10 clases, en 8 de las cuales se encuentran representantes bacterianos. La mayoría de los citocromos P450 bacterianos pertenecen a la clase I (es un sistema de 3 componentes formado por una ferredoxina reductasa, una

ferredoxina y el citocromo P450), incluyendo el mejor caracterizado (CYT101 o P450cam, de *P. putida*). Todos los citocromos P450 bacterianos descritos hasta el momento requieren al menos una proteína de transporte de electrones, aunque en ocasiones esta se encuentra fusionada al propio citocromo. Como ocurre con el resto de oxigenasas iniciales, algunos citocromos P450 son capaces de recibir los electrones de otras proteínas que no sean las de su propio sistema de transporte de electrones, pero otros no.

5.2 Dioxigenasas extradiólicas

Las dioxigenasas extradiólicas son enzimas importantes en la degradación de compuestos aromáticos, que catalizan la incorporación de dos átomos de oxígeno al anillo una vez que este ha sido desestabilizado por las oxigenasas iniciales (actúan normalmente sobre catecol o derivados catecólicos), produciendo la apertura del anillo entre un carbono con grupo hidroxilo y otro que no lo tiene (rotura-*meta*) (figura 2). Las dioxigenasas extradiólicas son metaloenzimas que requieren la presencia de Fe(II), o en ocasiones Mg(II) o Mn(II), no unido a un grupo hemo. No necesitan un agente reductor externo. Al igual que ocurre con las oxigenasas iniciales, tradicionalmente las dioxigenasas extradiólicas se han estudiado a partir de organismos capaces de crecer en presencia de compuestos aromáticos, y con la metagenómica se ha acelerado el descubrimiento de nuevas enzimas. Se pueden realizar ensayos funcionales para dioxigenasas extradiólicas de manera relativamente sencilla, mediante la adición de catecol o de un derivado catecólico sobre colonias crecidas. Los productos resultantes de la rotura-*meta* de los anillos de estos compuestos presentan color amarillo, pudiéndose identificar visualmente las colonias portadoras de dicha actividad (Brennerova y col., 2009).

Se conocen dioxigenasas extradiólicas de un dominio y de dos dominios. Aunque hay otras clasificaciones de dioxigenasas extradiólicas, como la de Spence y col. (1996), probablemente la más utilizada en la actualidad sea la de Eltis y Bolin (1996), basada en homología de secuencia de aminoácidos. Establecieron dos categorías principales: tipo I y tipo II. Las de tipo I presentan homología de secuencia unas con otras, compartiendo una secuencia consenso en el sitio activo, e incluyen a la mayoría de las dioxigenasas extradiólicas caracterizadas. Dentro del tipo I se establecieron 5 familias (denominadas por un número arábigo) y varias subfamilias (por una letra), por ejemplo I.3.A. Cada secuencia de proteínas de una familia tipo I presenta menos del 22 % de identidad con cualquier secuencia de otra

familia, y las secuencias dentro de la misma subfamilia muestran entre sí identidades mayores del 54 %. Posteriormente, esta clasificación se ha ampliado con la incorporación de nuevas secuencias que han generado subfamilias nuevas (p. ej. Suenaga y col., 2007) e incluso una familia nueva, I.6 (Sakai y col., 2002). Dentro del tipo I, la familia I.1 comprende las enzimas de un solo dominio, como la 2,3-dihidroxibifenil 1,2-dioxigenasa de *Rhodococcus globerulus* P6 (Asturias y col, 1994); encontrándose en el resto de categorías enzimas de dos dominios, como la 1,2-dihidroxi-naftaleno dioxigenasa ThnC de *Sphingopyxis macrogoltabida*, en la familia I.3 (Andújar y col., 2000). Al tipo II pertenecen proteínas diversas que no presentan características comunes definidas, y en principio se asignaron a este tipo las secuencias que no podían ser alineadas con las de tipo I (Eltis y Bolin, 1996). Un ejemplo de este grupo es la protocatecuato 4,5-dioxigenasa de *Sphingomonas paucimobilis*. Las dioxigenasas extradiólicas tipo I y II emplean catecol o derivados catecólicos como sustratos. Posteriormente se incluyeron secuencias en dos nuevos tipos. Las de tipo III son dioxigenasas extradiólicas pertenecientes a la familia de las cupinas, y emplean como sustratos intermediarios de rutas de degradación no catecólicos, como son entre otros gentisato, salicilato, homogentisato o ácidos aromáticos monohidroxilados (Fetzner, 2012; Díaz y col., 2013). Un ejemplo es la gentisato 1,2-dioxigenasa de *Pseudomonas acidovorans* (Harpel y Lipscomb, 1990). Y por último, se ha incorporado un nuevo tipo a la clasificación, en el que se incluiría la enzima dioxigenasa dependiente de Fe^{2+} (NicX) de *Pseudomonas putida* KT2440, cuya secuencia no muestra ninguna similitud significativa con las de las dioxigenasas extradiólicas descritas hasta el momento, y tiene como sustratos compuestos aromáticos N-heterocíclicos (Jiménez y col., 2008). Más recientemente, Steiner y col. (2010) han descrito otras enzimas capaces de romper anillos de ese tipo de molécula, como 1-H-3-hidroxi-4-oxoquinolina 2,4-dioxigenasa de *Pseudomonas putida* 33/1.

6. CELULASAS

La celulosa es la molécula orgánica más abundante del planeta. En su degradación interviene un amplio grupo de enzimas que llevan a cabo distintas reacciones, entre las que destacan hidrolasas como las celulasas. Las celulasas son un grupo de enzimas extracelulares presentes en bacterias y hongos capaces de degradar la celulosa hasta moléculas de glucosa. Hay tres tipos principales de celulasas, requiriéndose todas ellas para dicha degradación:

endoglucanasas, exoglucanasas y β -glucosidasas o celobiasas. Las endoglucanasas rompen enlaces internos en zonas amorfas en la cadena de celulosa, creando nuevos extremos de cadena libres. Las exoglucanasas realizan cortes en los extremos de la cadena, liberando dímeros (celobiosas) o monómeros (glucosas). Por último, las β -glucosidasas o celobiasas hidrolizan las celobiosas, generando moléculas de glucosa (Lynd y col., 2002).

Este proceso de despolimerización de la celulosa tiene múltiples aplicaciones, entre las que destaca la producción de bioetanol (Rubin, 2008), siendo también de interés en las industrias alimentaria, textil y de pulpa y papel (Bhat, 2000; Miettinen-Oinonen y Suominen, 2002).

Existen ensayos que permiten identificar celulasas de cada uno de los tres tipos principales a partir de genotecas o de metagenotecas. Para la identificación de endoglucanasas se emplean sustratos celulíticos amorfos, como la carboximetilcelulosa (CMC), que se aporta al medio sólido de cultivo sobre el que crecen los clones. Una vez crecidas, las placas se tiñen con el indicador rojo Congo, que se une a la CMC (Teather y Wood, 1982) y permite detectar las colonias positivas por la aparición de halos alrededor de las mismas, debido a la falta de tinción por la ausencia de CMC. La identificación de exoglucanasas se puede realizar del mismo modo pero sustituyendo el sustrato amorfo por uno cristalino, como el avicel. Además de los indicados, existen otros sustratos que permiten la identificación de halos sin necesidad de teñir las placas, como por ejemplo el rojo brillante ostazin hidroxietil-celulosa para la detección de endoglucanasas. Por último, la detección de β -glucosidasas se puede llevar a cabo mediante el crecimiento de las colonias en medio mínimo con celobiosa como única fuente de carbono. Aquellas que tengan la enzima de interés serán capaces de obtener glucosa a partir de la celobiosa. Otra opción es la adición al medio de cultivo del sustrato fluorogénico 4-metilumbeliferil- β -D-celobiosido (MUC). Se han identificado varias celulasas mediante metagenómica funcional a través de ensayos como los descritos. Se trata en su mayoría, aunque no exclusivamente, de endoglucanasas, como las encontradas en los trabajos de Ferrer y col. (2005), Kim y col., 2007, de Kim y col. (2008), de Wang y col. (2009), de Nacke y col. (2012) y de Ilmberger y col. (2012), entre otros.

OBJETIVOS

1. Desarrollar y validar nuevos sistemas de expresión heteróloga para mejorar la eficacia de los rastreos funcionales que se realizan con metagenotecas.
2. Identificar clones con actividades de interés.
3. Identificar los genes responsables de las actividades de interés.
4. Caracterizar la actividad de algunos de los clones identificados.

MATERIALES Y MÉTODOS

1. ESTIRPES, PLÁSMIDOS Y OLIGONUCLEÓTIDOS

A continuación se detallan las estirpes bacterianas, los plásmidos y los oligonucleótidos utilizados en esta tesis.

| Estirpe | Características relevantes | Referencia |
|---------------------------------------|--|--------------------------|
| <i>E. coli</i> DH5 α | ϕ 80 <i>dlacZ</i> Δ M15 Δ (<i>lacZYA-argF</i>)U169 <i>recA1 endA1 hsdR17</i> ($r_k^- m_k^+$) <i>supE44 thi-1 gyrA relA1</i> | Hanahan, 1983 |
| <i>E. coli</i> EPI300-T1 ^R | F ⁻ <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) (Str ^R) Φ 80 <i>dlacZ</i> Δ M15 Δ <i>lacX74 recA1 endA1 araD139</i> Δ (<i>ara, leu</i>)7697 <i>galU galK</i> λ^- <i>rpsL nupG trfA tonA dhfr</i> | Epicentre |
| <i>E. coli</i> MPO553 | EPI300-T1 ^R Δ <i>trg::placUV5-atenuador de nasF-gen I</i> | Esta tesis |
| <i>E. coli</i> MPO554 | EPI300-T1 ^R Δ <i>trg::nahR/psal-gen N</i> | Esta tesis |
| <i>E. coli</i> MPO555 | EPI300-T1 ^R Δ <i>trg::nahR/psal-gen N</i> truncado | Esta tesis |
| <i>E. coli</i> NCM631 | HsdS <i>gal lacZDM15:lacI lacUV5:genI</i> (ARN polimerasa de T7) Δ <i>lac-Tn10</i> . | Govantes y Santero, 1996 |

Tabla 1. Estirpes bacterianas utilizadas en esta tesis.

| Plásmido | Características relevantes | Referencia |
|--------------------|--|--------------------------------|
| pBBR1MCS-3 | Plásmido movilizable por conjugación, Tc ^R | Kovach y col., 1995 |
| pBluescript II SK+ | Vector de clonación, Ap ^R | Stratagene |
| pBluescript II KS+ | Vector de clonación, Ap ^R | Stratagene |
| pCC1FOS-CeuI | pCC1FOS (Epicentre) con dos sitios <i>CeuI</i> flanqueando el sitio de clonación <i>Eco72I</i> , Cm ^R | M. Ferrer |
| pCNB4-S2 | pUT/mini-Tn5 <i>nahR/psal</i> → <i>sylS2</i> , Ap ^R , Km ^R | Cebolla y col., 2001 |
| pCP20 | Rep ^{TS} (30°C), Ap ^R , Cm ^R | Cherepanov y Wackernagel, 1995 |

| | | |
|---------|---|--|
| pET-14b | Vector de expresión para superproducción de proteínas, con pT7, Ap ^R | Novagen |
| pGP1-2 | Plásmido que contiene el gen <i>I</i> (ARN polimerasa del fago T7), Km ^R | Tabor y Richardson, 1985 |
| pIZ227 | Derivado pLysE, <i>lacI^q</i> , Cm ^R | Govantes y Santero, 1996 |
| pIZ1016 | Vector de amplio espectro derivado del pBBRMCS-5, con <i>ptac</i> y <i>lacI^q</i> , Gm ^R | Martínez-Pérez y col., 2004 |
| pIZ1110 | Fragmento que contiene la región intergénica <i>thnL-thnM</i> clonado en pLAFR3, Tc ^R | Tesis doctoral de M. Eloísa Andújar Pulido, 2001 |
| pKD3 | OriR γ , Ap ^R , Cm ^R | Datsenko y Wanner, 2000 |
| pKD4 | OriR γ Ap ^R , Km ^R | Datsenko y Wanner, 2000 |
| pKD46 | <i>oriR101</i> , <i>repA101(ts)</i> , <i>araBp-gam-bet-exo</i> , Ap ^R | Datsenko y Wanner, 2000 |
| pMPO27 | Vector de expresión con <i>rrnBT1T2-pm</i> -atenuador de <i>nasF::MCS</i> , origen de replicación de <i>ColE1</i> , Ap ^R | Royo y col., 2005 |
| pMPO200 | Vector de amplio espectro derivado de pBBR1MCS-4 para fusiones traduccionales a <i>lacZ</i> , Ap ^R | García-González y col., 2005 |
| pMPO556 | <i>placUV5</i> clonado en pBluescript II SK+, Ap ^R | Esta tesis |
| pMPO557 | Atenuador de <i>nasF</i> clonado en pMPO556, Ap ^R | Esta tesis |
| pMPO558 | Gen de resistencia a kanamicina de pKD4 clonado en pMPO557, Ap ^R , Km ^R | Esta tesis |
| pMPO559 | Gen de resistencia a cloranfenicol de pKD3 clonado en pGP1-2, Ap ^R , Km ^R , Cm ^R | Esta tesis |
| pMPO561 | Fósmido derivado de pCC1FOS-CeuI con <i>oriT</i> , Cm ^R | Esta tesis |
| pMPO563 | <i>nahR-pnah-psal</i> clonado en pBluescript II SK+, Ap ^R | Esta tesis |
| pMPO564 | Gen de resistencia a cloranfenicol de pKD3 clonado en pMPO563, Ap ^R , Cm ^R | Esta tesis |

| | | |
|---------|---|------------|
| pMPO565 | Gen <i>N</i> del fago lambda clonado en pMPO564, Ap ^R , Cm ^R | Esta tesis |
| pMPO571 | Fósmido derivado de pMPO561 con <i>psal-nut_L</i> clonado aguas arriba del promotor del gen <i>10</i> del fago T7 y del sitio de clonación <i>Eco72I</i> , Cm ^R | Esta tesis |
| pMPO575 | Plásmido derivado de pMPO565 con un cambio en la pauta de lectura en el gen <i>N</i> , Ap ^R , Cm ^R | Esta tesis |
| pMPO578 | Fósmido derivado de pMPO571 con un gen <i>gfp</i> sin promotor con su propia Shine Dalgarno clonado aguas abajo del sitio de clonación <i>Eco72I</i> , Cm ^R | Esta tesis |
| pMPO579 | Fósmido derivado de pMPO571 con un gen <i>gfp</i> sin promotor con la Shine Dalgarno del gen <i>10</i> del fago T7 clonado aguas abajo del sitio de clonación <i>Eco72I</i> , Cm ^R | Esta tesis |
| pMPO580 | Fósmido derivado de pMPO579 con el terminador transcripcional de <i>thnL</i> de <i>Sphingomonas macrogolitabida</i> estirpe TFA clonado en el sitio <i>Eco72I</i> , Cm ^R | Esta tesis |
| pMPO584 | Gen de β-lactamasa del clon ETN1 subclonado en pMPO579, Cm ^R | Esta tesis |
| pMPO585 | Gen de β-lactamasa seguido de la <i>orf</i> conservada del clon ETN1 subclonado en pMPO579, Cm ^R | Esta tesis |
| pMPO586 | Gen de bomba de eflujo del clon TN4 subclonado en pMPO579, Cm ^R , Cb ^R | Esta tesis |
| pMPO587 | Genes <i>dszE</i> , <i>dszA</i> , <i>dszB</i> y <i>dszC</i> del clon UPO21 subclonados en pMPO579, Cm ^R | Esta tesis |
| pMPO588 | Gen <i>dszR</i> del clon UPO21 subclonado en pIZ1016, Gm ^R | Esta tesis |
| pMPO592 | Fragmento final de <i>dszC</i> fusionado a <i>lacZ</i> en pMPO200, Ap ^R | Esta tesis |
| pMPO593 | Fusión traduccional a <i>lacZ</i> y transcripcional a <i>gfp</i> , que contiene el inserto de pMPO587 hasta el | Esta tesis |

| | | |
|----------|--|-----------------------------|
| | penúltimo codón de <i>dszC</i> , Cm ^R | |
| pMPO594 | Fusión traduccional a <i>lacZ</i> y transcripcional a <i>gfp</i> , que contiene el inserto de pMPO587 hasta el tercer codón de <i>dszE</i> , Cm ^R | Esta tesis |
| pMPO595 | Fragmento de pMPO587 que incluye el promotor dependiente de σ^N hasta el 267° codón de <i>dszB</i> clonado en pMPO579, Cm ^R | Esta tesis |
| pMPO596 | Fusión traduccional a <i>lacZ</i> y transcripcional a <i>gfp</i> , que contiene el inserto de pMPO587 desde el nucleótido 302° hasta el penúltimo codón de <i>dszC</i> , Cm ^R | Esta tesis |
| pMPO597 | Fusión traduccional a <i>lacZ</i> y transcripcional a <i>gfp</i> , que contiene el inserto de pMPO587 hasta el 154° codón de <i>dszA</i> , Cm ^R | Esta tesis |
| pMPO634 | Plásmido que contiene el gen <i>gfp</i> , Km ^R | Tomás-Gallardo y col., 2009 |
| pMPO1383 | Gen <i>upoedo11</i> del clon UPO36 subclonado en pET-14b, Ap ^R | Esta tesis |
| pNK736 | Ap ^R , plásmido que contiene <i>placUV5</i> | Simons y col., 1983 |
| pRK2013 | Plásmido auxiliar en conjugaciones triparentales, Km ^R | Figurski y Helinski, 1979 |
| RP4 | (Tn1) Tc ^R IncP-1 Tra ⁺ Cma ⁺ , Km ^R Ap ^R | Holloway, 1979 |

Tabla 2. Plásmidos utilizados en esta tesis.

| Oligonucleótidos | Secuencia 5'-3' |
|------------------|---|
| OriTHpaIFw | CAAGTTAACCTTGCCCTCATCTGTTACG |
| OriTHpaIRev | TCGGTTAACCCAGTCGGTAGATATTCCAC |
| psalnut1 | TAAGGCGCCTTATTGCTGGTGCCCGGCCGGGCGCAATAT TCATGTTGATGATTTATTATATATCGAGTGGTGTATTTAT |

| | |
|----------------------|--|
| | CAATATTGTTTGCTCC |
| psalnut2 | GTCACCTTCATGGTGGTCAGTGCGTCCTGCTGATTAATA ACGATAACGGAGCAAACAATATTGATAAATACACC |
| psalnut3 | GCACTGACCACCATGAAGGTGACGCTCTTAAAAATTA GCCCTGAAGAAGGGCAGCATTCAAAGCAGAAGGCTTTG GGGTGTGTG |
| psalnut4 | TATGGCGCCCCGGAATCGCACTTACGGCCAATGCTTCGT TTCGTATCACACACCCCAAAGCCTTCTGC |
| HindChlFw | CAGGCATGCAAGCTTGAG |
| BstZredFRv | GGTATACCGGCATACAGC |
| SDGFPKpnIsolap | TAGAGGGTACCGCTGAAAATGTAAAGATCCTC |
| KpnISDpT77GFP | TAGAGGGTACCAATAATTTTGTTTAACTTTAAGAAGGAG ATATACATATGAGTAAAGGAGAAGAAGCTTTTC |
| GFPMluIRvsopal | GGAACAACAAGACGCGTGCTGAAGTCAAG |
| GFPMluIFwsolap | CTTGACTTCAGCACGCGTCTTGTAGTTCC |
| GFPXbaI-TFB-PCRsolap | AGGTCTTCTAGATTATTTGTATAGTTCATC |
| Sac-P1 | TATAGAGCTCTGTAGGCTGGAGCTGCTTC |
| Sac-P2 | TATAGAGCTCATATGAATATCCTCCTTAG |
| trgEc-P12 | GGTTTTTGCATCACATCAGGTTGGTTCCGTTATTTGCCT GCATTCTAGGGTGTAGGCTGGAGCTGCTTC |
| trgEc-BSK2 | CGCGAGGTTCTGCCGACACAGAATGTTTGTGCAGACGG AATACATCCACCCCATGATTACGCCAAGCTCG |
| terSacP2-G1 | CCGGGCGCTTTTTTTTTTGCGCGAATTCGATTATAGAGCTC ATATGAATATTTACTAACTGGAAGAGGCAC |
| NotN | TATGCGGCCGCCCCTGGCGGTGATACTG |
| KspN | TAACCGCGGAAAGCCAAGGCCAATATC |
| trgEc-BKS | GGTTTTTGCATCACATCAGGTTGGTTCCGTTATTTGCCT GCATTCTAGGTAACGACGGCCAGTGAGC |
| trgEc-P1 | CGCGAGGTTCTGCCGACACAGAATGTTTGTGCAGACGG AATACATCCACCGTGTAGGCTGGAGCTGCTTC |

| | |
|-------------------------|----------------------------------|
| thnB-6-thnC | GATCATGCAT |
| blactamETN1EcoRIfw | TAAGAATTCACAAGGCGTTGGGGTAGC |
| blactamETN1HindIIIrv | TCAAAGCTTTGCGCATCATGGCTGCAC |
| blactampepETN1HindIIIrv | TCAAAGCTTGCAAATCAGCGGTCAAGG |
| UPOEDO11NdeIfw | TTATCATATGGCCATGACCGGCATTC |
| UPOEDO11BamHIrv | TTATGGATCCTCAGGTGTTGGCAGTGAC |
| EcoKpndszCfw | TTAGAATTCGGTACCATGCGCCAGACCGACAG |
| BamdszCrv | TTAGGATCCCAGGAGGTGAATCCCGGCA |

Tabla 3. Oligonucleótidos utilizados en esta tesis.

2. MEDIOS, CONDICIONES DE CULTIVO Y CONSERVACIÓN

2.1 Medios y condiciones de cultivo

Como medio rico para las distintas estirpes de *E. coli* se utilizó Luria-Bertani (LB) (Sambrook y Russell, 2001).

Como medio mínimo para las estirpes EPI300-T1^R y derivadas se utilizó una variante de M9 (Sambrook y Russell, 2001) suplementado con leucina e isoleucina (al ser auxótrofas de estos aminoácidos).

Los cultivos en medio mínimo con dibenzotiofeno como única fuente de azufre consistieron en la variante del medio M9 en la que además se sustituyó el MgSO₄ por dibenzotiofeno 0,2 mM.

Cuando fue necesario, y para asegurar la selección y mantenimiento de plásmidos, se añadieron los siguientes antibióticos a las concentraciones indicadas: cloranfenicol (Cm) a 12,5-15 µg/ml, ácido nalidíxico (Nal) 15 µg/ml, kanamicina (Km) a 25 µg/ml, ampicilina (Ap) a 100 µg/ml, rifampicina (Rif) 20 µg/ml, carbenicilina (Cb) 100 µg/ml, tetraciclina (Tc) 5 µg/ml y gentamicina (Gm) 10 µg/ml. Otros compuestos que se añadieron al medio cuando fue necesario y sus concentraciones, salvo que se indique lo contrario, son: arabinosa 1 mM, salicilato 1 mM, IPTG 1 mM y X-gal 25 mg/L.

Salvo que se indique lo contrario, los cultivos se incubaron a 37 °C, y con agitación suave (180 r.p.m.) cuando se trataba de cultivos líquidos.

LB: Bactotripton 10 g/L, extracto de levadura 5 g/L y NaCl 5 g/L (y agar 15 g/L para medio sólido).

Variante del medio mínimo M9: 10 % 10X (Na₂HPO₄ 423 mM, KH₂PO₄ 220 mM, NH₄Cl 187 mM y NaCl 86 mM) pH 7,4, CaCl₂ 0,1 mM, MgSO₄ 2 mM, glucosa 0,2 %, leucina 500 µM e isoleucina 300 µM (y agar noble 15 g/L para medio sólido).

L-Leucina (L8000, Sigma-Aldrich). El stock (50 mM) se preparó en agua bidestilada, se autoclavó y se conservó a temperatura ambiente.

L-Isoleucina (I2752, Sigma-Aldrich). El stock (30 mM) se preparó en agua bidestilada, se autoclavó y se conservó a temperatura ambiente.

L-(+)-Arabinosa (A3256, Sigma-Aldrich). El stock (1 M) se preparó en agua bidestilada, se filtró y se conservó a temperatura ambiente.

Salicilato sódico (S3007, Sigma-Aldrich). El stock (1 M) se preparó en agua bidestilada, se filtró y se conservó a 4 °C en oscuridad.

Dibenzotiofeno (347833, Sigma-Aldrich). El stock (0,2 M) se preparó en N,N'-dimetilformamida (en tubo de polipropileno) y se conservó a temperatura ambiente.

2.2 Conservación del material biológico

Las estirpes bacterianas y los plásmidos (contenidos en estas) se conservaron a corto plazo a 4 °C en placas de LB sólido. Para su conservación a largo plazo se congelaron a -80 °C cultivos en LB líquido con glicerol al 15 % (v/v).

3. MANIPULACIÓN DE ADN

3.1 Extracción de ADN plasmídico de *Escherichia coli*

Las preparaciones de ADN plasmídico de las distintas estirpes de *E. coli* empleadas en esta tesis se realizaron siguiendo tres procedimientos distintos. Para obtener gran cantidad de

ADN de gran pureza de fósido para la construcción de metagenotecas y ADN de algunos clones que presentaban dificultad en su extracción para su secuenciación, se empleó el kit Large Construction (Qiagen). Para el resto de situaciones en las que se requería bastante pureza se empleó el kit Nucleospin Plasmid Mini (Macherey-Nagel). Para las extracciones con los dos kits comentados se siguieron las indicaciones del fabricante. Para el resto de aplicaciones se usó el protocolo simplificado de lisis alcalina (Stephen, 1990).

Cuando se llevó a cabo el procedimiento de lisis alcalina la biomasa de 3-5 ml de un cultivo saturado se recogió por centrifugación a 16.100 g durante 30 segundos. Los sedimentos se resuspendieron en 100 μ l de solución I (GTE) y se incubaron 5 minutos en hielo. A continuación se añadieron 200 μ l de solución II recién preparada para la lisis, y tras mezclar por inversión se incubaron 5 minutos en hielo. Tras la incubación se añadieron 150 μ l de solución III y tras mezclar por inversión los tubos se volvieron a incubar en hielo durante 5 minutos. Posteriormente, se sedimentaron los restos celulares por centrifugación 10 minutos a 16.100 g. El sobrenadante se transfirió a otro tubo y se precipitó el ADN añadiendo el mismo volumen (unos 450 μ l) de etanol 96 % (v/v) a -20 °C y centrifugando inmediatamente 10 minutos a 16.100 g. El precipitado se lavó con 1 ml de etanol 70 % (v/v) a -20 °C y se volvió a centrifugar 5 minutos a 16.100 g. Tras esta última centrifugación, el precipitado se secó mediante vacío y se resuspendió en 25-50 μ l de TER dependiendo del número de copias del plásmido por célula. Por último se incubó al menos 30 minutos a 37 °C. Las preparaciones de plásmidos se almacenaron a -20°C.

Solución I (GTE): Glucosa 50 mM, EDTA 10 mM, Tris-HCl 25 mM, pH 8. Se conservó a 4 °C.

Solución II: NaOH 0,2 N, SDS 1 % (p/v). Usar recién preparada.

Solución III: Acetato potásico 3 M ajustado a pH 4,8 con ácido fórmico. Se conservó a 4 °C.

TER: Tris-HCl 10 mM pH 8, EDTA 1 mM, ribonucleasa 0,02 % (p/v). Se conservó a 4 °C.

3.2 Amplificación de ADN mediante la reacción en cadena de la polimerasa (PCR)

Cuando fue necesario amplificar fragmentos de ADN para su posterior clonación, se realizaron reacciones de PCR utilizando el ADN correspondiente como molde. Las reacciones

se llevaron a cabo por lo general en un volumen final de 50 μ l en tubos de 0,2 ml, en presencia de los oligonucleótidos adecuados a una concentración de 0,2 μ M, dNTPs a 0,2 mM cada uno y 1 unidad de DNA polimerasa termorresistente de alta fidelidad Expand High Fidelity Plus (Roche) o i-pfu DNA Polymerase (Intron Biotechnology), en el tampón que suministra la casa comercial, que lleva $MgCl_2$ a una concentración final de 1,5 mM. Como molde se usaron por lo general 1-20 ng de ADN plasmídico, o en ocasiones células resuspendidas procedentes de colonias.

El programa de PCR del termociclador varió en función del tamaño y naturaleza del fragmento a amplificar, pero en líneas generales consistió en lo siguiente: Un primer paso de desnaturalización inicial de 5 minutos a 94 °C. A continuación, 35 ciclos con 3 etapas: 30 segundos a 94 °C (desnaturalización), seguida de 30 segundos a temperatura variable en función de la temperatura de desnaturalización (T_m) de los oligonucleótidos (hibridación) y por último 60-120 segundos por kilobase de ADN a amplificar a 72 °C (extensión). Finalmente, 5 minutos de extensión final a 72 °C.

3.3 Asociación de oligonucleótidos complementarios

Para sintetizar un pequeño fragmento de ADN se empleó un oligonucleótido con la secuencia deseada autocomplementario (otra posibilidad es emplear dos oligonucleótidos complementarios entre sí). Se incubaron 30 μ l del oligonucleótido autocomplementario a 100 μ M (o 15 μ l de cada uno si son dos complementarios entre sí) durante 5 minutos a 85 °C, seguido de 30 minutos a temperatura ambiente. En caso de requerir hacer los extremos del fragmento de doble cadena generado romos, este se trata con Klenow o con la polimerasa de T4 como se indica en el apartado 3.6 de Materiales y Métodos.

3.4 PCR recurrente

Para sintetizar fragmentos de mayor tamaño que los descritos en el apartado anterior (a partir de decenas de pb), se empleó la técnica descrita por Prodromou y Pearl (1992). Esta técnica consiste en amplificar una secuencia a partir de oligonucleótidos solapantes, que en conjunto cubran toda la secuencia a sintetizar, ya sea en una u otra cadena. De este modo, cada uno sirve de molde a otro oligonucleótido con el que solapa (figura 3). Para ello, se diseñaron oligonucleótidos de entre 60-100 pb de longitud con un mínimo de unos 20 pb de

solapamiento. Se llevó a cabo una reacción de PCR con todos los oligonucleótidos, proporcionando los externos correspondientes a las secuencias flanqueantes del fragmento 100 veces más concentrados que los internos (los externos a $1 \mu\text{M}$ y los internos a $0,01 \mu\text{M}$), sin aportar otro ADN molde. Al estar los oligonucleótidos externos más concentrados, en los últimos ciclos se producirá amplificación fundamentalmente con estos, aumentándose la cantidad de producto completo.

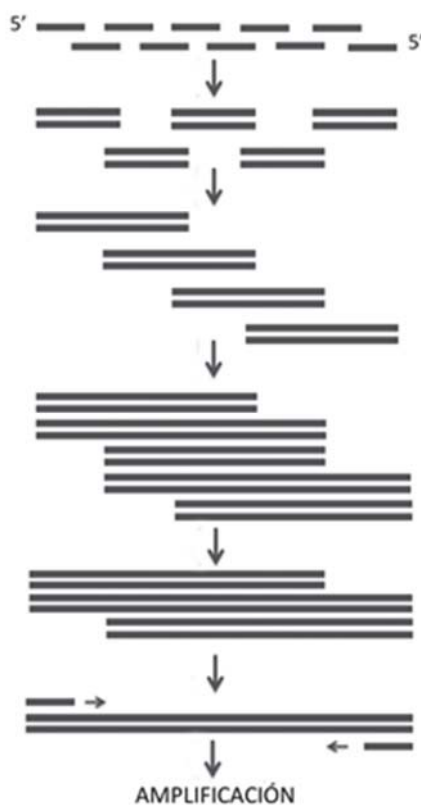


Figura 3. Esquema de una PCR recurrente para sintetizar fragmentos de ADN. (Adaptado de Prodromou y Pearl, 1992)

3.5 Mutagénesis dirigida por PCR (PCR solapante)

Para introducir mutaciones puntuales se utilizó una modificación de la técnica de PCR solapante descrita por Ho y col. (1989) (Camacho y Casadesús, 2005). Esta técnica consta de dos fases de amplificación mediante PCR representadas en la figura 4. En una primera fase se realizaron dos PCR independientes, cada una con un oligonucleótido externo a la región a amplificar y uno interno que contiene la mutación a insertar. Los oligonucleótidos internos deben tener temperaturas de desnaturalización mayores que los externos. Cada oligonucleótido interno contiene las mutaciones centradas en su secuencia flanqueadas por

unos 10-15 nucleótidos de secuencia silvestre, siendo los dos oligonucleótidos mutagénicos complementarios entre sí. Los fragmentos amplificados se aislaron según su tamaño y se purificaron de un gel de agarosa.

En una segunda fase, los dos fragmentos amplificados se diluyeron en función de su concentración y se mezclaron, utilizándose como molde en cantidades aproximadamente equimoleculares para una segunda PCR en la que se añadieron los dos oligonucleótidos externos (no mutagénicos). En este caso la amplificación se realizó en dos pasos consistentes en 10 ciclos iniciales con una temperatura de hibridación más alta para favorecer la extensión desde la zona solapante (donde se había generado la mutación), y así enriquecer en molde mutante, y otros 25 ciclos a temperatura de hibridación más baja para favorecer la extensión desde los oligonucleótidos externos y obtener cantidades suficientes de producto final mutado.

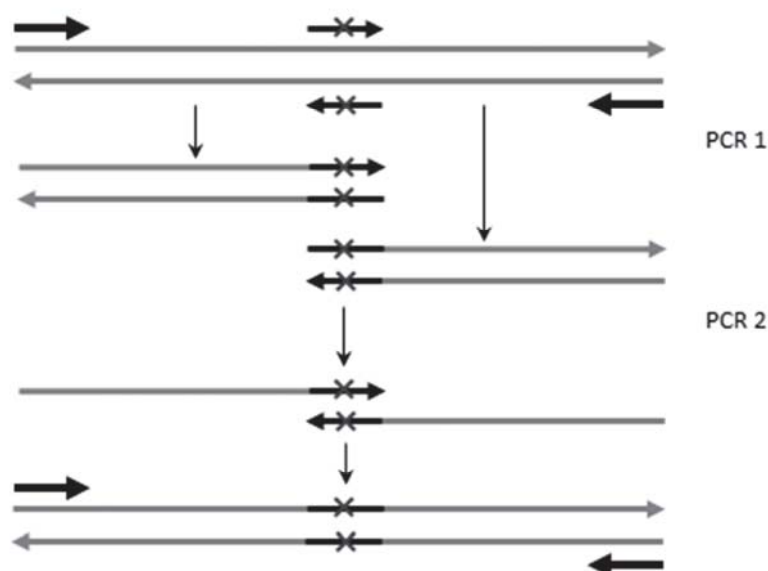


Figura 4. Esquema de una PCR solapante. Las flechas con aspa representan los oligonucleótidos que contienen las mutaciones y las flechas anchas representan los oligonucleótidos externos.

3.6 Digestión y modificación de fragmentos de ADN

Los fragmentos de ADN, procedentes de un plásmido o de una reacción de PCR, se digirieron con las enzimas de restricción adecuadas y el correspondiente tampón de reacción siguiendo las indicaciones del fabricante (Roche, New England Biolabs o Fermentas).

Cuando fue necesario realizar digestiones parciales, se mezcló ADN concentrado, tampón de reacción correspondiente a la enzima y agua bidestilada hasta 100 μ l. La mezcla se dividió en alícuotas de 30, 20, 20, 20 y 10 μ l, y se añadió 0,05-0,2 U de enzima/ μ g de ADN al primer tubo con 30 μ l. A continuación, tras mezclar se fueron pasando 10 μ l de un tubo al siguiente, de manera que al final quedaron todos con 20 μ l. La preparación de las mezclas se realizó en hielo. Todas las muestras se incubaron durante 15 minutos a la temperatura de reacción de la enzima, y finalmente se detuvieron las reacciones con 0,5 μ l de EDTA 0,5 M. Seguidamente se realizó una electroforesis de las muestras en gel de agarosa para determinar qué cantidad de enzima era necesaria para obtener digestiones parciales.

En el caso de requerir la conversión de extremos cohesivos en romos, si se trataba de rellenar los extremos 3' de los fragmentos 5' protuberantes se empleó ADN polimerasa Klenow (Roche) o ADN polimerasa de T4 (Roche), y si había que eliminar algún extremo 3' protuberante se empleó ADN polimerasa de T4 (Roche).

La desfosforilación de los extremos de los vectores para evitar religaciones en las clonaciones, lo que se realizó fundamentalmente con el fósforo pMPO579 para la construcción de metagenotecas, se llevó a cabo con la enzima SAP (Shrimp Alkaline Phosphatase, USB) siguiendo las instrucciones del fabricante.

3.7 Electroforesis de ADN

Para visualizar preparaciones de ADN y separar fragmentos lineales se empleó la técnica de electroforesis de ADN en gel. Mediante esta técnica se comprobó la eficiencia de las extracciones de ADN, se determinó el perfil de fragmentos tras la digestión de plásmidos, se visualizaron productos de PCR, etc. Se llevaron a cabo dos tipos de electroforesis de ADN, de agarosa o de acrilamida, dependiendo del tamaño de las bandas esperadas.

3.7.1 Electroforesis en geles de agarosa

En general, si el fragmento de interés era superior a 100 pb se utilizaron geles de agarosa, cuya concentración osciló entre 0,6 y 2 % (p/v), dependiendo del tamaño de los fragmentos de ADN a separar, en tampón TAE 1X. Como patrón de peso molecular estándar se usó el marcador 1 Kb Plus DNA Ladder (Invitrogen), para fragmentos de hasta 10 kb aproximadamente, o ADN del fago lambda digerido con *HindIII*, para fragmentos de mayor

tamaño. Antes de cargar las muestras en el gel, se les añadió tampón de carga de ADN 6X de manera que este quedara diluido 6 veces. Los geles se sometieron a una diferencia de potencial que varió entre 40 y 135 V.

Para visualizar el ADN tras la electroforesis, los geles se tiñeron en una solución de bromuro de etidio a 1 mg/L en agua destilada. Las bandas de ADN se visualizaron con un transiluminador de radiación ultravioleta (Vilber Lourmat) y se registraron con un analizador de imágenes. En los casos en los que se pretendía recuperar ADN de gel se minimizó la exposición a radiación ultravioleta para evitar daños en el ADN.

TAE 1X: Tris-ácido acético 40 mM, EDTA 10 mM, pH 7,7.

Tampón de carga de ADN 6X: Azul de bromofenol 0,05 % (p/v), xileno cianol 0,05 % (p/v), glicerol 30 % (v/v), Tris-HCl 10 mM pH 8 y EDTANa₂ 1 mM pH 8.

3.7.2 Electroforesis en geles de poliacrilamida

En general, si el fragmento de interés era inferior a 100 pb se emplearon geles verticales no desnaturalizantes de poliacrilamida (acrilamida:bisacrilamida 29:1) al 8 % (p/v), en tampón TBE 0,5X. En este caso los geles se sometieron a una corriente fija de 30 mA. Como patrón de peso molecular estándar se usó el marcador 1 Kb Plus DNA Ladder (Invitrogen). Antes de cargar las muestras en el gel, se les añadió el mismo tampón de carga de ADN 6X descrito para geles de agarosa, y se tiñeron y visualizaron del mismo modo.

TBE 0,5X: Tris-ácido bórico 45 mM, EDTA 1 mM, pH 8.

3.8 Aislamiento y purificación de fragmentos de ADN

Los fragmentos de interés se identificaron mediante visualización en gel de agarosa o poliacrilamida con una lámpara de radiación ultravioleta de baja energía, y se aislaron recortándolos del gel. Para purificar estos fragmentos de DNA se siguieron 2 métodos en función del tipo de gel empleado en la electroforesis. Si el DNA se separó en gel de agarosa, el fragmento se purificó con el kit GFX (GE Healthcare) siguiendo las instrucciones del fabricante. Si el DNA era de menor tamaño y se separó en un gel de poliacrilamida, el fragmento se purificó por difusión a un tampón de elución durante al menos una noche a 37

°C en agitación. A continuación se precipitó el DNA con 0,5 volúmenes de acetato amónico 7,5 M y 2,5 volúmenes de etanol 96 % (v/v), con una incubación de al menos 1 hora a -20 °C. Después se centrifugó 15 minutos a 16.100 g, se lavó con etanol 70 % (v/v), se secó el precipitado y se resuspendió en el volumen deseado de TE.

Tampón de elución: NaAc 0,5 M, EDTA 1 mM.

TE: Tris HCl 10 mM, EDTA 1 mM pH 8.

3.9 Cuantificación de ADN

Para cuantificar las preparaciones de ADN de doble cadena (ADN plásmidico, fragmentos de ADN purificados, productos de PCR, etc.) se midió la absorbancia a 260 nm en un espectrofotómetro Nanodrop ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA), que calculó la concentración aplicando un coeficiente de extinción de $50 \text{ ng cm}^{-1} \mu\text{l}^{-1}$. En los casos en los que las soluciones que contenían el ADN interferían con la medida del espectrofotómetro, se procedió a su estimación aproximada mediante comparación con 100 ng de ADN del fago lambda digerido con *Hind*III en gel de agarosa.

3.10 Clonación de fragmentos de ADN en vectores

Para las clonaciones se obtuvo el ADN vector y el ADN inserto, por lo general, a partir de extracciones de ADN plasmídico o amplificaciones mediante PCR, que frecuentemente requerían reacciones de digestión y en algunos casos modificación de los extremos, como se ha descrito en apartados anteriores.

El fragmento a insertar y el vector se mezclaron normalmente en proporción de 5:1 a 10:1. Las ligaciones se realizaron en tampón quinasa con ATP 1 mM, cuando los fragmentos a ligar tenían ambos extremos cohesivos, o 0,1 mM, cuando al menos uno de los extremos a ligar era romo, en un volumen final de 10 μl . A cada reacción se le añadió 1 unidad de ADN ligasa de T4 (Roche) y se incubó un mínimo de 12 horas a 16 °C.

El producto de la ligación se transformó en la estirpe *E. coli* DH5a competente mediante choque térmico o en la estirpe *E. coli* EPI300-T1^R o derivada electrocompetente mediante electroporación (cuando el vector era el fósido pMPO579 o derivado de este, con el fin de poder obtener mayores cantidades de ADN) (apartado 4.1 de Materiales y Métodos).

Se sembró en placas con medio rico LB con el antibiótico adecuado para la selección del plásmido. Los clones obtenidos se comprobaron mediante restricción enzimática o mediante secuenciación si el fragmento clonado provenía de PCR o se introducían mutaciones puntuales por sustitución de secuencias silvestres.

Tampón quinasa: Tris-HCl 0,5 M pH 7,6, Cl₂Mg 0,1 M, DTT 50 mM y ATP 1 mM / 0,1 mM.

3.11 Secuenciación de ADN y análisis de secuencias

Las reacciones de secuenciación para confirmar la presencia o ausencia de mutaciones en el ADN fueron llevadas a cabo por la empresa Secugen (Madrid), utilizando como molde preparaciones de ADN plasmídico o productos de PCR de alta calidad. El análisis de las secuencias obtenidas se realizó empleando el paquete de herramientas BLAST (Altschul y col., 1997) disponible en la página web del National Center for Biotechnology Information (NCBI) (<http://blast.ncbi.nlm.nih.gov>).

4. TRANSFERENCIA DE ADN A ESTIRPES BACTERIANAS

4.1 Transformación

4.1.1 Transformación de *Escherichia coli* mediante choque térmico

4.1.1.1 Preparación de células competentes

La preparación de células competentes se llevó a cabo mediante dos procedimientos distintos dependiendo de la finalidad de la transformación.

Para la transformación de *E. coli* DH5 α con ligaciones se prepararon células competentes según una variante del método de Inoue y col. (1990), con el que se consigue una alta frecuencia de transformación (de $5 \cdot 10^7$ a 10^9 transformantes por μg de ADN plasmídico). Un cultivo saturado de DH5 α se diluyó a D.O.₆₀₀ de 0,1 en 200 ml de medio SOB y se incubó con agitación a 22 °C hasta alcanzar una D.O.₆₀₀ de 0,5. El cultivo se transfirió rápidamente a hielo, donde se mantuvo durante 10 minutos. A continuación, se recogieron las células mediante centrifugación a 2.500 g a 4 °C durante 10 minutos. El sedimento se resuspendió en 20 ml de TB a 0 °C, se añadieron otros 60 ml de TB a 0 °C y se incubó 10 minutos en hielo.

Se volvieron a recoger las células en las mismas condiciones y se resuspendieron en 20 ml de TB a 0 °C. Se añadieron 1,5 ml de DMSO y tras incubar 10 minutos en hielo se hicieron alícuotas, que se congelaron en nitrógeno líquido y se almacenaron a -80 °C.

Medio SOB: Triptona 2 % (p/v), extracto de levadura 0,5 % (p/v), NaCl 8,5 mM, KCl 1,25 mM. Ajustar el pH a 7,0 con NaOH y autoclavar. Añadir posteriormente MgCl₂ 10 mM y glucosa 0,36 mM.

TB: Pipes (ácido libre) 10 mM, CaCl₂·2H₂O 15 mM, KCl 250 mM. Ajustar el pH a 6,7 con KOH. Añadir MnCl₂·4H₂O 55 mM. Filtrar.

Para transferir plásmidos a las distintas estirpes de *E. coli* se prepararon células competentes según el método descrito por Chung y col. (1989), con pequeñas variaciones. Un cultivo saturado de la estirpe receptora se diluyó a D.O.₆₀₀ de 0,1 en medio LB y se incubó a 37 °C en agitación hasta fase exponencial temprana (D.O.₆₀₀ de 0,4 aproximadamente). Se repartieron alícuotas de 1 ml en tubos, se enfriaron en hielo 5 minutos y se centrifugaron 30 segundos a 16.100 g. Se desechó el sobrenadante, se añadieron 75 µl de LB frío y se volvió a incubar en hielo 5 minutos. Se añadieron 75 µl de TSS 2X y se incubó en hielo otros 5 minutos. A continuación se llevó a cabo la transformación.

TSS 2X: Triptona 0,8 % (p/v), extracto de levadura 0,5 % (p/v), NaCl 85,5 mM, polietilenglicol 8000 25 mM, DMSO 10 % (v/v), MgSO₄·7H₂O 1 M, pH 6,5.

4.1.1.2 Transformación

Para transformar las células competentes se añadió 1-5 µl de una preparación de plásmido (100-300 ng de ADN aproximadamente) o 10 µl en el caso de reacciones de ligación a 100 µl de células competentes. La mezcla se incubó en hielo durante 30 minutos y a continuación se realizó un choque térmico a 42 °C durante 45 segundos. Se devolvió rápidamente al hielo y se añadió 1 ml de LB a temperatura ambiente. Se incubó a 37 °C durante 60-90 minutos, se sembró en placas de medio selectivo (LB con los correspondientes antibióticos) y se incubó toda la noche a 37 °C (salvo en el caso de plásmidos termosensibles, en el que las incubaciones se realizaron a 30 °C).

4.1.2 Transformación de *Escherichia coli* mediante electroporación

4.1.2.1 Preparación de células electrocompetentes

Para la transformación de estirpes de *E. coli* distintas a DH5 α con ligaciones, o para la transferencia de plásmidos en los casos en los que se requerían frecuencias de transformación mayores a las obtenidas con el método del TSS, se prepararon células electrocompetentes. Para la preparación de células electrocompetentes se diluyó un cultivo saturado a D.O.₆₀₀ de 0,1 en medio LB y se incubó a 37 °C en agitación hasta D.O.₆₀₀ de 0,6-0,8. Las células se dividieron en alícuotas de 6 ml y se recogieron por centrifugación a 16.100 g durante 1 minuto a 4 °C. A continuación, se resuspendieron en 3 ml de glicerol 10 % en agua bidestilada (v/v) frío y se centrifugaron a 16.100 g durante 1 minuto a 4 °C. Seguidamente se repitió el proceso resuspendiendo las células en 1 ml de glicerol 10 % frío, y nuevamente en 500 μ l de glicerol 10 % frío. Finalmente, se volvieron a centrifugar y a resuspender en 100 μ l de glicerol 10 % frío, y se incubaron en hielo hasta proceder a la transformación.

4.1.2.2 Transformación

Para transformar células electrocompetentes se añadió 1-5 μ l de una preparación de plásmido (100-300 ng de ADN aproximadamente) libre de sales o 10 μ l en el caso de reacciones de ligación a 100 μ l de células electrocompetentes. La mezcla se transfirió a una cubeta de electroporación con una distancia entre los electrodos de 2 mm, previamente enfriada, y se sometió a una diferencia de potencial de 2,5 kV en un electroporador MicropulserTM (Bio-Rad). Inmediatamente después del pulso se añadió 1 ml de LB a temperatura ambiente a las células. Se incubó a 37 °C durante 60-90 minutos, se sembró en placas de medio selectivo (LB con los correspondientes antibióticos) y se incubó toda la noche a 37 °C (salvo en el caso de plásmidos termosensibles, en el que las incubaciones se realizaron a 30 °C).

4.2 Conjugación

4.2.1 Conjugación triparental

Para transferir plásmidos movilizables con un origen de transferencia conjugativa pero carentes de genes *tra* desde una estirpe a otra de *E. coli* se llevaron a cabo conjugaciones triparentales. Se empleó el plásmido auxiliar pRK2013 (Figurski y Helinski, 1979) que

contiene los genes de transferencia, en la estirpe DH5 α . Se partió de cultivos saturados crecidos en medio LB de las estirpes donadora, auxiliar y receptora. Se diluyeron en medio LB a D.O.₆₀₀ de 0,1, se incubaron a 37 °C hasta una D.O.₆₀₀ aproximada de 0,5 y se mezclaron los cultivos en proporción 1:1:1 en un volumen final de 300 μ l. Se lavó la mezcla con LB fresco para eliminar los antibióticos. Se centrifugó y se eliminó la mayor parte del sobrenadante, y el sedimento de células se resuspendió en el volumen residual de medio. Las células resuspendidas se depositaron en forma de gota (parche de conjugación) en una placa de LB, y se incubaron durante 8 h a 37 °C. Posteriormente se segregó parte del parche en una placa de LB con los antibióticos adecuados para la selección de la transferencia del plásmido (antibiótico al que proporciona resistencia el plásmido y al que proporciona resistencia la estirpe receptora).

Para transferir las metagenotecas (construidas en pMPO579, que contiene *oriT*) de una estirpe a otra de *E. coli* para realizar los rastreos de actividades se llevaron a cabo conjugaciones triparentales, pero con las siguientes diferencias respecto al procedimiento que se acaba de describir. Se tomó una alícuota de la metagenoteca almacenada a -80 °C (de 250 μ l) y se cultivó en un volumen total de 5 ml de LB con Cm 12,5 μ g/ml hasta alcanzar una D.O.₆₀₀ de 0,5 aproximadamente. A continuación se diluyó este cultivo (donador) y cultivos saturados de las estirpes auxiliar y receptora en LB a D.O.₆₀₀ de 0,1, se incubaron a 37 °C hasta una D.O.₆₀₀ aproximada de 0,5 y se mezclaron los cultivos en proporción 1:1:1 en un volumen final de 3 ml (1 ml de cada uno). Tras la incubación del parche de conjugación, este se recogió y se resuspendió en 2 ml de tampón fosfato y se sembraron diluciones con los factores de dilución seleccionados (se indican en el apartado 9 de Materiales y Métodos, según el tipo de rastreo). Las placas tenían el medio de cultivo, antibióticos y demás compuestos necesarios para la selección de la transferencia de los clones de la metagenoteca a la estirpe receptora y la realización del rastreo funcional que se fuera a llevar a cabo (ver apartado 9 de Materiales y Métodos).

Para el cálculo de la frecuencia de conjugación, se realizaron diluciones del parche de conjugación resuspendido en 2 ml tampón fosfato y se sembraron en diferentes placas. Por un lado, para el conteo de transconjugantes, se sembraron placas de LB con el antibiótico al que aporta resistencia el plásmido a transferir y con el antibiótico al que aporta resistencia la estirpe receptora, con factores de dilución 10⁴ y 10⁵. Y por otro, para el conteo de viables, se sembraron placas de LB con el antibiótico al que aporta resistencia la estirpe receptora, con factores de dilución 10⁶ y 10⁷. Tras la incubación de las placas se realizaron los conteos de

colonias y se obtuvo la frecuencia de conjugación dividiendo el número de transconjugantes entre el número de viables.

Tampón fosfato 50X: 302,5 g/L $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ y 50 g/L KH_2PO_4 .

5. PLÁSMIDOS CONSTRUIDOS EN ESTA TESIS

A continuación se detalla la elaboración de los plásmidos construidos en esta tesis.

5.1 Vectores para metagenotecas

pMPO571: Se necesitó la construcción previa del plásmido pMPO561.

pMPO561: Se amplificó mediante PCR el origen de transferencia conjugativa *oriT* a partir del plásmido RP4 (Holloway, 1979), utilizando los oligonucleótidos OriTHpaIFw y OriTHpaIRev (introduciendo un sitio de restricción *HpaI* en cada extremo del fragmento de ADN amplificado). El producto de la PCR (de 326 pb) se cortó con *HpaI*, y se clonó en el sitio *HpaI* único de pCC1FOS-CeuI, generando pMPO561 (de 8.531 pb).

pMPO571: Se sintetizó una secuencia que contenía un promotor *psal* seguido por el sitio *nut_L* del fago lambda mediante PCR recurrente (Prodromou y Pearl, 1992) (apartado 3.4 de Materiales y Métodos), utilizando los oligonucleótidos *psalnut1*, *psalnut2*, *psalnut3* y *psalnut4*. En la reacción de PCR se usaron los mismos oligonucleótidos como molde, poniendo los externos (*psalnut1* y *psalnut4*) a 1 μM y los internos (*psalnut2* y *psalnut3*) a 0,01 μM . El programa de PCR del termociclador empleado fue: 5' a 94 °C; 30'' a 94 °C, 1'30'' a 72 °C (35 ciclos); 7' a 72 °C (se unieron las fases de hibridación y extensión). Primero se clonó este fragmento (de 250 pb) en la diana *EcoRV* del sitio de clonación múltiple de pBluescript II SK+. Posteriormente se obtuvo el fragmento de interés digiriendo con *XbaI* y *HindIII* y haciendo los extremos romos con Klenow (de 296 pb) y se clonó en el sitio *SfoI* único de pMPO561. Se seleccionó un transformante que tenía la orientación adecuada del fragmento *psal-nut_L* en relación con el sitio de clonación de ADN metagenómico, y se denominó pMPO571. El fósido pMPO571 tiene un tamaño de 8.827 pb.

pMPO578: Se necesitó la construcción de 2 plásmidos previos.

Plásmido intermedio 1: Se amplificó parte del fósido pCC1FOS-CeuI que contenía el gen que proporciona resistencia a cloranfenicol y parte de *redF*, utilizando los oligonucleótidos HindChIFw y BstZredFRv (fragmento de 1,4 kb flanqueado por los sitios de *HindIII* y *Bst1107I*) y se clonó este fragmento en el sitio *EcoRV* del sitio de clonación múltiple de pBluescript II SK+.

Plásmido intermedio 2: Se obtuvo la región codificante de una variante del gen *gfp* de *Aequorea victoria* (Cormack, Valdivia y Falkow, 1996) precedida por su propia Shine-Dalgarno a partir del plásmido pMPO634 (Tomás-Gallardo y col., 2009). El gen *gfp* tiene un sitio *Eco72I* (el mismo que el sitio de clonación para el ADN metagenómico en el fósido), que se eliminó mediante PCR solapante (Ho y col., 1989) (apartado 3.5 de Materiales y Métodos), generando un sitio *MluI* pero sin cambiar la secuencia de aminoácidos codificados por el gen. La amplificación se realizó utilizando los oligonucleótidos SDGFPKpnIsolap, GFP $MluI$ Fwsolap, GFP $MluIR$ vsolap y GFPXbaI-TFB-PCRsolap. En la primera fase de PCR se realizó una PCR con los oligonucleótidos SDGFPKpnIsolap y GFP $MluI$ Fwsolap, y otra con GFP $MluIR$ vsolap y GFPXbaI-TFB-PCRsolap, utilizándose pMPO634 como molde en ambas. En la segunda fase de PCR se emplearon los productos de las PCR de la primera fase como molde y los oligonucleótidos SDGFPKpnIsolap y GFPXbaI-TFB-PCRsolap como cebadores. El programa de PCR del termociclador empleado fue: 5' a 94 °C; 30'' a 94 °C, 30'' a 68 °C, 1'15'' a 72 °C (10 ciclos); 30'' a 94 °C, 30'' a 51 °C, 1'15'' a 72 °C (25 ciclos); 7' a 72 °C. Se digirió el producto de la PCR (de 774 pb) con la enzima de restricción *XbaI* (que tenía dianas justo antes de la Shine-Dalgarno y tras el gen *gfp*), se hicieron romos sus extremos con Klenow y se clonó en el plásmido intermedio 1, antes del gen que proporciona resistencia a cloranfenicol, entre los dos sitios *HindIII*, tras hacer los extremos romos con Klenow.

pMPO578: Se extrajo el fragmento de interés, que lleva *gfp* con su Shine-Dalgarno sin sitio *Eco72I* y parte del gen de resistencia a cloranfenicol, del plásmido intermedio 2, digiriendo con *XhoI* y *ScaI*, se hicieron los extremos romos con Klenow, y se clonó en el fósido pMPO571 tras el sitio de clonación de ADN metagenómico, entre los sitios *HindIII* y *ScaI*, haciendo los extremos romos con Klenow, de manera que el gen que proporciona resistencia a cloranfenicol quedó regenerado. El fósido pMPO578 tiene un tamaño de 9.609 pb.

pMPO579: Se necesitó la construcción previa de un plásmido.

Plásmido intermedio 3: Se obtuvo la región codificante de una variante del gen *gfp* de *Aequorea victoria* (Cormack, Valdivia y Falkow, 1996) a partir del plásmido pMPO634 (Tomás-Gallardo y col., 2009). El sitio *Eco72I* del gen *gfp* se mutó creando un sitio *MluI* mediante PCR solapante (Ho y col., 1989) (apartado 3.5 de Materiales y Métodos), sin cambiar la secuencia de aminoácidos codificados por el gen, con los oligonucleótidos KpnISDpT77GFP, GFP_{MluI}Fwsolap, GFP_{MluI}Rvsolap y GFP_{XbaI}-TFB-PCRsolap, quedando las dianas de *KpnI* y *XbaI* en los extremos del fragmento. El extremo 5' del oligonucleótido KpnISDpT77GFP contenía la secuencia Shine-Dalgarno del gen *10* del fago T7 de tal manera que esa secuencia quedó a la distancia adecuada antes de la región codificante del gen *gfp* amplificada. En la primera fase de PCR se realizó una PCR con los oligonucleótidos KpnISDpT77GFP y GFP_{MluI}Fwsolap, y otra con GFP_{MluI}Rvsolap y GFP_{XbaI}-TFB-PCRsolap, utilizándose pMPO634 como molde en ambas. En la segunda fase de PCR se emplearon los productos de las PCR de la primera fase como molde y los oligonucleótidos KpnISDpT77GFP y GFP_{XbaI}-TFB-PCRsolap como cebadores. El programa de PCR del termociclador empleado fue: 5' a 94 °C; 30'' a 94 °C, 30'' a 68 °C, 1'15'' a 72 °C (10 ciclos); 30'' a 94 °C, 30'' a 51 °C, 1'15'' a 72 °C (25 ciclos); 7' a 72 °C. Se digirió el producto de la PCR (de ~800 pb) con las enzimas de restricción *KpnI* y *XbaI*, se hicieron romos sus extremos con la ADN polimerasa de T4 y se clonó en el plásmido intermedio 1 (cuya construcción se explica anteriormente), antes del gen que proporciona resistencia a cloranfenicol, entre los dos sitios *HindIII*, tras hacer los extremos romos con Klenow.

pMPO579: Se extrajo el fragmento de interés, que lleva la Shine-Dalgarno del gen *10* del fago T7, *gfp* sin sitio *Eco72I* y parte del gen de resistencia a cloranfenicol, del plásmido intermedio 3, digiriendo con *XhoI* y *ScaI*, se hicieron los extremos romos con Klenow, y se clonó en el fósido pMPO571 tras el sitio de clonación de ADN metagenómico, entre los sitios *HindIII* y *ScaI*, haciendo los extremos romos con Klenow, de manera que el gen que proporciona resistencia a cloranfenicol quedó regenerado. La estructura del fósido final, denominado pMPO579, se representa en la figura 6. El fósido pMPO579 tiene un tamaño de 9.621 pb.

pMPO580: Es un derivado de pMPO579 con un terminador transcripcional construido para validar el sistema. Se obtuvo un fragmento de unas 2,5 kb correspondientes a la región intergénica entre los genes *thnL* y *thnM* de la estirpe TFA de *Sphingomonas macrogolitabida* (que contiene el terminador de *thnL*) por restricción con *EcoRI* y *KpnI* del plásmido pIZ1110 (tesis doctoral de M. Eloísa Andújar Pulido, 2001). Los extremos del fragmento se hicieron

romos con la ADN polimerasa de T4, y este se clonó en pMPO579, en el sitio *Eco72I* (sitio de clonación del ADN metagenómico). El terminador quedó ubicado aproximadamente a 2,3 kb aguas abajo de *psal* y a 2 kb del promotor pT7, en la orientación adecuada.

5.2 Subclones con gen(es) metagenómico(s)

pMPO584: El gen *blaC* del clon ETN1 (sin la pequeña *orf* que le sigue) se obtuvo mediante PCR utilizando la pareja de oligonucleótidos *blactamETN1EcoRI*fw y *blactamETN1HindIII*rv (introduciendo los sitios de restricción de *EcoRI* y *HindIII* en los extremos del fragmento de ADN amplificado). El producto de PCR, de 1,2 kb, se digirió con las enzimas de restricción *EcoRI* y *HindIII* y se clonó en pMPO579 digerido con las mismas enzimas, de manera que el gen *blaC* quedaba a continuación y en el mismo sentido que *psal*.

pMPO585: El gen *blaC* seguido de una pequeña *orf* del clon ETN1 se obtuvo mediante PCR utilizando la pareja de oligonucleótidos *blactamETN1EcoRI*fw y *blactampepETN1HindIII*rv (introduciendo los sitios de restricción de *EcoRI* y *HindIII* en los extremos del fragmento de ADN amplificado). El producto de PCR, de 1,5 kb, se digirió con las enzimas de restricción *EcoRI* y *HindIII* y se clonó en pMPO579 digerido con las mismas enzimas, de manera que el gen *blaC* quedaba a continuación y en el mismo sentido que *psal*.

pMPO586: Los genes codificantes de los 3 componentes de la bomba de eflujo de los clones TN3 y TN4 se obtuvieron por restricción de TN4 con las enzimas *KpsI* y *PstI*, haciendo los extremos romos con la ADN polimerasa de T4. El fragmento resultante, de 6.739 pb, se clonó en pMPO579 digerido con *PmlI* (isoesquizómero de *Eco72I*, en el sitio de clonación del ADN metagenómico), de manera que los genes de la bomba quedaron tras los promotores heterólogos, en su misma orientación.

pMPO587: Se obtuvo un fragmento de 5,7 kb que contenía los genes *dszE*, *dszA*, *dszB* y *dszC*, flanqueados por unas 800 pb aguas arriba de la región codificante de *dszE* y unas 600 pb aguas abajo del último de los genes, por restricción del clon UPO21 con *CelIII*. Los extremos se hicieron romos con Klenow. Este fragmento se clonó en el fósido pMPO579 digerido con *PmlI* (isoesquizómero de *Eco72I*, en el sitio de clonación del ADN

metagenómico), de manera que los genes quedaron tras los promotores heterólogos, en su misma orientación.

pMPO588: Se obtuvo por restricción del clon UPO21 con *NcoI* y *KpnI* un fragmento de 1,4 kb que contenía el gen *dszR*. Los extremos se hicieron romos con la ADN polimerasa de T4. Este fragmento se clonó en pIZ1016 (Martínez-Pérez y col., 2004) digerido con *HindIII* y hechos los extremos romos con Klenow. De este modo, el gen del regulador quedó tras el promotor *ptac* del vector, en su misma orientación.

pMPO1383: El gen *upoedo11* se obtuvo por PCR del clon UPO36 utilizando la pareja de oligonucleótidos UPOEDO11NdeI_{fw} y UPOEDO11BamHI_{rv} (introduciendo los sitios de restricción de *NdeI*, acoplado con el codón de inicio, y *BamHI* en los extremos del fragmento de ADN amplificado). El producto de PCR, de unas 930 pb, se digirió con las enzimas de restricción *NdeI* y *BamHI* y se clonó en el vector de superproducción pET-14b (Novagen) digerido con las mismas enzimas, de manera que el gen codificante de la dioxigenasa extradiólica quedaba a continuación y en el mismo sentido que el promotor pT7.

5.3 Fusiones traduccionales a *lacZ* y transcripcionales a *gfp*

pMPO593: Se necesitó la construcción previa del plásmido pMPO592.

pMPO592: Un fragmento de unas 650 pb que contenía la última parte del gen *dszC*, dejando fuera el codón de parada, se amplificó por PCR, empleando como molde pMPO587, con los oligonucleótidos EcoKpndszC_{fw} y BamdszC_{rv}. El oligonucleótido EcoKpndszC_{fw} llevaba en el extremo 5' la diana *EcoRI*, a continuación la diana *KpnI* y después los nucleótidos necesarios para que al final en la fusión quedara reconstituida la diana *SphI* de *dszC*. El oligonucleótido BamdszC_{rv} llevaba en el extremo 5' la diana *BamHI*. El producto de amplificación se cortó con *EcoRI* y *BamHI* y se clonó en el vector de fusiones traduccionales pMPO200 (García-González y col., 2005) digerido con las mismas enzimas, de manera que el fragmento final del gen *dszC* quedaba fusionado en fase al gen *lacZ* del vector.

pMPO593: El final del gen *dszC* seguido de *lacZ* se obtuvo por restricción (fragmento de 4,1 kb) cortando pMPO592 con *KpnI* y *SalI* y haciendo romos los extremos con la ADN polimerasa de T4. Ese fragmento se clonó en pMPO587 digerido con *SphI* y hechos los

extremos romos con la ADN polimerasa de T4, de manera que se elimina el final del gen *dszC* y el resto del ADN del inserto metagenómico, sustituyéndolo por el final de ese mismo gen sin incluir el codón de parada y con *lacZ* fusionado en fase.

pMPO594: Un fragmento de 3,5 kb con el gen *lacZ* se obtuvo de pMPO200 (García-González y col., 2005) por restricción con *EcoRI* y *SalI*. Se hicieron los extremos romos con Klenow, y se clonó en pMPO587 digerido con *SphI* y *HindIII*, haciendo los extremos romos con la ADN polimerasa de T4, de manera que se incluía hasta el tercer codón del gen codificante de la reductasa (*dszE*), quedando este en fase con *lacZ*.

pMPO596: Se necesitó la construcción previa del plásmido pMPO595.

pMPO595: Se obtuvo un fragmento de unas 3,3 kb que incluía desde 453 nucleótidos aguas arriba del dinucleótido GG del promotor dependiente de sigma 54 hasta el 267° codón del gen *dszB* por restricción con *SmaI* de pMPO587, y se clonó en el fósido pMPO579 linealizado con *PmlI*.

pMPO596: Por restricción de pMPO593 con *NsiI* y *ScaI* se obtuvo un fragmento de unas 9,1 kb que incluía los cuatro genes, *dszE*, *dszA*, *dszB* y *dszC*, pero dejaba fuera el promotor dependiente de sigma 54, y llegaba por el otro extremo hasta el noveno codón del gen codificante de la resistencia a cloranfenicol del vector. Se clonó en pMPO595 cortado con las mismas enzimas (fragmento de unas 8,8 kb), de manera que la resistencia a cloranfenicol quedaba reconstituida, y el resto de la construcción exactamente igual que en pMPO593 salvo que no tenía el terminador.

pMPO597: Se obtuvo un fragmento de 3,5 kb que contenía el gen *lacZ* de pMPO200 (García-González y col., 2005) por restricción con *EcoRI* y *SalI*, se hicieron los extremos romos con Klenow, y se clonó en pMPO587 digerido con *XhoI* y *SphI*, tras hacer los extremos romos con la ADN polimerasa de T4, de manera que se incluían los primeros 154 codones del gen *dszA*, quedando este en fase con *lacZ*.

6. CONSTRUCCIÓN DE ESTIRPES

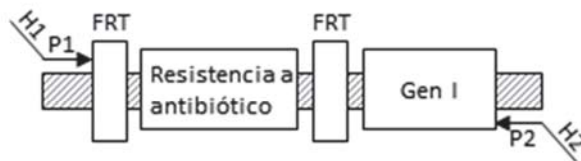
6.1 Introducción de genes en el cromosoma utilizando productos de PCR

Para la construcción de estirpes de *E. coli* derivadas de EPI300-T1^R se utilizó una modificación del método descrito por Datsenko y Wanner (2000). El método original permite la inactivación de genes cromosómicos utilizando productos de PCR. La modificación empleada permite sustituir el gen inactivado por un fragmento que contiene un gen u otros elementos de interés.

Este procedimiento está basado en el sistema de recombinación Red del fago λ . Una de las razones por las que *E. coli* no es transformable con ADN lineal es la presencia de exonucleasas intracelulares que lo degradan. El sistema Red del fago λ incluye los genes γ , β y *exo* que codifican las proteínas Gam, Bet y Exo, respectivamente. Gam inhibe la exonucleasa V del hospedador permitiendo el acceso de Bet y Exo a los extremos de ADN para promover la recombinación.

La estrategia básica consiste en reemplazar una secuencia cromosómica (el gen B en la figura 5) por un fragmento que contiene un gen de resistencia a antibiótico seleccionable junto a la secuencia de interés a introducir en el cromosoma (el gen I en la figura 5), que es generado por PCR utilizando oligonucleótidos que contienen 50 nucleótidos de homología con la secuencia que queremos reemplazar (H1 y H2 en la figura 5). Si las zonas de homología son contiguas se puede introducir el fragmento de interés en el cromosoma sin deletar ninguna secuencia. La recombinación es llevada a cabo por el sistema Red del fago λ , expresado desde un promotor inducible en un plásmido de bajo número de copias y termosensible (pKD46, Ap^R). Tras la selección, el gen de resistencia puede eliminarse usando otro plásmido (pCP20, Ap^R y Cm^R) que expresa la recombinasa FLP del plásmido de 2 μ de *Saccharomyces cerevisiae*, de manera que solo queda en el cromosoma la secuencia de interés. El sistema FLP actúa en las repeticiones directas FRT (*FLP recognition target*) que flanquean la secuencia del gen de resistencia (figura 5). Los plásmidos que contienen los sistemas Red y FLP son termosensibles y pueden curarse fácilmente mediante crecimiento a 37°C.

Paso 1. Amplificación por PCR del gen de resistencia a antibiótico flanqueado por las secuencias FRT y del gen de interés.



Paso 2. Transformación de la estirpe que expresa la recombinasa Red de λ .



Paso 3. Selección de transformantes resistentes al antibiótico.



Paso 4. Eliminación del gen de resistencia a antibiótico usando el sistema FLP.

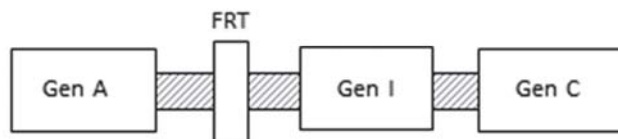


Figura 5. Esquema del proceso de introducción de genes en el cromosoma utilizando productos de PCR. H1 y H2 indican regiones de homología de las secuencias flanqueantes del gen a deletar. P1 y P2 son las regiones de homología de las secuencias que flanquean el gen de resistencia a antibiótico seguido del gen de interés. Modificado de Datsenko y Wanner, 2000.

6.1.1 Preparación del ADN para la sustitución

Los plásmidos usados como molde en las reacciones de PCR fueron construidos previamente, ubicando los genes de resistencia a Cm de pKD3 o a Km de pKD4 junto a la secuencia de interés a introducir en el cromosoma. Los oligonucleótidos utilizados como cebadores se diseñaron de forma que el gen a introducir en el cromosoma quedara en sentido contrario al gen objeto de la deletación. Estos oligonucleótidos contenían 20 nucleótidos de homología con los extremos de la secuencia a introducir (gen de resistencia más secuencia de interés) y en su extremo 5' 50 nucleótidos de homología con la región flanqueante del gen a deletar.

Se llevó a cabo la PCR y el producto fue purificado de gel de agarosa. A continuación, para eliminar el ADN molde utilizado en la PCR que pudiera quedar, el ADN purificado se digirió con *DpnI*, que reconoce selectivamente el ADN metilado, y se purificó de nuevo eluyendo en agua bidestilada.

6.1.2 Transformación de los receptores

Se prepararon células electrocompetentes de la estirpe receptora EPI300-T1^R portando el plásmido pKD46, que expresa el sistema Red del fago λ bajo el control del promotor de *araB* inducible por arabinosa. Para la obtención de células electrocompetentes, cultivos saturados incubados a 30 °C en LB con ampicilina se diluyeron 100 veces en 50 ml de SOB sin magnesio suplementado con ampicilina y L-arabinosa 1mM, y se incubaron a 30 °C hasta alcanzar una D.O.₆₀₀ de 0,6-0,8. Las células se recogieron mediante centrifugación durante 15 minutos a 4.000 r.p.m. a 4 °C y se resuspendieron en 20 ml de glicerol al 10 % frío. Posteriormente se centrifugaron de nuevo 15 minutos a 4.000 r.p.m. a 4 °C y se resuspendieron en 10 ml de glicerol al 10 % frío. Tras una nueva centrifugación, se resuspendieron en 5 ml de glicerol al 10 % frío. Y por último, las células se volvieron a centrifugar y se resuspendieron en 500 μ l de glicerol al 10 % frío.

Se electroporaron alícuotas de 100 μ l de células electrocompetentes con 10 μ l (1-1,5 μ g totales) del ADN purificado tras la digestión con *DpnI*. Tras el pulso con el electroporador, las células se resuspendieron en 1 ml de LB y se incubaron a 37 °C durante 90 minutos. Tras la incubación, se sembró la mitad en placas con medio selectivo (LB con kanamicina o cloranfenicol) a 37°C. El resto se dejó a temperatura ambiente y en caso de no obtener ningún transformante resistente a antibiótico, se sembró tras 24 horas de incubación.

SOB sin magnesio: Medio SOB: Triptona 2 % (p/v), extracto de levadura 0,5 % (p/v), NaCl 8,5 mM, KCl 1,25 mM. Ajustar el pH a 7,0 con NaOH.

6.1.3 Comprobación de candidatos

Las colonias obtenidas se cultivaron en medio sólido LB con Km o LB con Cm, y LB con Ap y se incubaron a 37 °C. Tras 24 horas, se seleccionaron aquellas colonias sensibles a Ap, que habían perdido pKD46, y resistentes al antibiótico Km o Cm. La inserción del nuevo

fragmento que contiene además de la secuencia de interés el gen de resistencia sustituyendo el gen a delecionar, se comprobó mediante PCR con distintas parejas de oligonucleótidos, y posteriormente se comprobó el fenotipo de las distintas estirpes construidas.

6.1.4 Escisión del gen de resistencia a antibiótico

La eliminación del gen de resistencia al antibiótico introducido se realizó transformando la estirpe con el plásmido pCP20, cultivando las células tras la transformación a 30 °C y sembrándolas posteriormente en LB con Ap e incubando a 30 °C. Para eliminar el plásmido pCP20, se segregó un transformante en placa de LB sin antibiótico a 37 °C. La pérdida del plásmido y la escisión de la resistencia (Cm^R o Km^R) se comprobaron observando ausencia de crecimiento en placas de LB con Ap y de LB con Cm/ Km respectivamente. La escisión del gen de resistencia se comprobó además mediante PCR con los oligonucleótidos externos.

7. ESTIRPES CONSTRUIDAS EN ESTA TESIS

A continuación se detalla la elaboración de las estirpes de *E. coli* construidas en esta tesis, mediante la técnica de introducción de genes en el cromosoma utilizando productos de PCR descrita en el apartado 6.1 de Materiales y Métodos. Para cada una de ellas, se explica la construcción de los plásmidos necesarios y a continuación la integración de la secuencia de interés en el cromosoma.

MPO553: Estirpe derivada de EPI300-T1^R que tiene integrado en su genoma un fragmento de ADN de ~3,3 kb con el promotor *placUV5*, atenuador de *nasF* y el gen *1* del fago T7, que codifica la ARN polimerasa de este fago.

Construcción de plásmidos:

pMPO556: El promotor *placUV5* se obtuvo por restricción como un fragmento *EcoRI-BamHI* de unas 95 pb a partir del plásmido pNK736 (Simons y col., 1983), haciendo los extremos de *EcoRI* romos con Klenow, y se clonó en pBluescript II SK+ digerido con *NotI*-Klenow y *BamHI*.

pMPO557: El atenuador de *nasF* se obtuvo por restricción a partir de pMPO27 (Royo y col., 2005b) flanqueado por dos sitios *EcoRI* (fragmento de unas 120 pb) y se clonó en pMPO556 digerido con *EcoRI*, tras *placUV5*.

pMPO558: El gen que proporciona resistencia a kanamicina de pKD4 (Datsenko y Wanner, 2000) se amplificó por PCR usando los oligonucleótidos Sac-P1 y Sac-P2. El fragmento obtenido, de 1.477 pb, se clonó en pMPO557 digerido con *EcoRV*, a continuación del atenuador de *nasF*.

pMPO559: El gen que proporciona resistencia a cloranfenicol de pKD3 (Datsenko y Wanner, 2000) se amplificó por PCR usando los oligonucleótidos Sac-P1 y Sac-P2. El fragmento obtenido, de 1.014 pb, se clonó en pGP1-2 (Tabor y Richardson, 1985) a continuación del gen *I* (ARN polimerasa de T7), tras digerir el plásmido con *BamHI* y hacer los extremos romos con Klenow.

Integración en el cromosoma:

La construcción de interés se integró en el locus *trg* en dos partes porque la construcción completa era demasiado grande para amplificarse mediante una única reacción de PCR y obtener una cantidad suficiente de producto.

La primera parte de la construcción, en pMPO558, se amplificó mediante PCR usando los oligonucleótidos *trgEc-P12* y *trgEc-BSK2*. Los extremos 5' de estos oligonucleótidos son homólogos a los límites de la secuencia genómica que va a reemplazarse por la construcción amplificada. Se llevaron a cabo 6 reacciones iguales de PCR en un volumen de 50 μ l cada una. El producto de PCR obtenido, de 1,8 kb, se digirió con *DpnI*, obteniendo finalmente 2,7 μ g de ADN totales. Se introdujo mediante electroporación en la estirpe EPI300-T1^R que contenía pKD46, para integrar la primera parte de la construcción en el locus *trg* (Datsenko y Wanner, 2000).

La segunda parte de la construcción, en pMPO559, se amplificó mediante PCR usando los oligonucleótidos *trgEc-P12* y *terSacP2-G1*. Los extremos 5' de estos cebadores son homólogos a los límites de la secuencia genómica que va a reemplazarse por la construcción amplificada. Se llevaron a cabo 6 reacciones iguales de PCR en un volumen de 50 μ l cada una. El producto de PCR, de 3,7 kb, se digirió con *DpnI*, obteniendo finalmente 2,4 μ g de ADN totales. Se introdujo mediante electroporación en la estirpe EPI300-T1^R con la primera parte de la construcción ya en el genoma y con el plásmido pKD46, para integrar la segunda

parte de la construcción tras la primera (reemplazando el gen que proporciona resistencia a kanamicina de pKD4) (Datsenko y Wanner, 2000).

Finalmente, se eliminó el gen que proporciona resistencia a cloranfenicol del genoma usando pCP20 (Cherepanov y Wackernagel, 1995 y Datsenko y Wanner, 2000).

MPO554: Estirpe derivada de EPI300-T1^R que tiene integrado en su genoma un fragmento de ADN de ~1,9 kb que contiene el gen *nahR*, su promotor *pnah*, el promotor *psal*, cercano al anterior y colocado de manera divergente, y el gen *N* del fago lambda, en el mismo sentido y a continuación de *psal*.

Construcción de plásmidos:

pMPO563: Un fragmento que contenía el gen *nahR* con su promotor *pnah* y el promotor *psal*, de ~1,2 kb, se obtuvo por restricción con *NotI* y *BamHI* a partir del vector pCNB4-S2 (Cebolla y col., 2001) y se clonó en pBluescript II KS+ digerido con las mismas enzimas de restricción.

pMPO564: El gen que proporciona resistencia a cloranfenicol de pKD3 (Datsenko y Wanner, 2000) se amplificó por PCR usando los oligonucleótidos Sac-P1 y Sac-P2. El fragmento obtenido, de 1014 pb, se clonó en pMPO563 digerido con *EcoRV*, aguas abajo del gen *nahR*.

pMPO565: El gen *N* se amplificó mediante PCR a partir de ADN de lambda usando los oligonucleótidos NotN y KspN (introduciendo los sitios de restricción de *NotI* y *KspI* en los extremos del fragmento de ADN amplificado). El producto de PCR, de unas 614 pb, se digirió con las enzimas de restricción *NotI* y *KspI* y se clonó en pMPO564 parcialmente digerido con *KspI* y completamente digerido con *NotI* (de esta manera, el vector solo perdía un pequeño fragmento fuera de la construcción de interés), aguas abajo de *psal* y en la misma orientación.

Integración en el cromosoma:

La construcción de interés se integró en el locus *trg*. Dicha construcción, en pMPO565, se amplificó mediante PCR usando los oligonucleótidos *trgEc-BKS* y *trgEc-P1*. Los extremos 5' de estos oligonucleótidos son homólogos a los límites de la secuencia genómica que va a reemplazarse por la construcción amplificada. Se llevaron a cabo 6 reacciones iguales de PCR en un volumen de 50 µl cada una. El producto de PCR obtenido, de 2,8 kb, se digirió con *DpnI*, obteniendo finalmente 1,3 µg de ADN totales. Se introdujo mediante electroporación en

la estirpe EPI300-T1^R que contenía pKD46 para integrar la construcción en el locus *trg* (Datsenko y Wanner, 2000).

Finalmente, se eliminó el gen que proporciona resistencia a cloranfenicol del genoma usando pCP20 (Cherepanov y Wackernagel, 1995 y Datsenko y Wanner, 2000).

MPO555: Estirpe derivada de EPI300-T1^R igual que MPO554 pero con un desfase del marco de lectura en el codón 60 del gen *N*, que la hace productora de una proteína N truncada. Se construyó para validar el sistema.

Construcción de plásmidos:

pMPO575: Se hibridó el oligonucleótido *thnB-6-thnC* consigo mismo y se rellenaron los extremos 3' con Klenow creando un inserto con extremos romos de 14 pb. Este inserto se clonó en pMPO565 digerido con *HpaI*, generando un cambio de la pauta de lectura en el codón 60 del gen *N*.

Integración en el cromosoma: Se realizó igual que para MPO554 pero empleando pMPO575 en lugar de pMPO565.

7.1 Selección de mutantes resistentes espontáneos a antibióticos

Se seleccionaron mutantes resistentes espontáneos a los antibióticos rifampicina y ácido nalidíxico de EPI300-T1^R, MPO553 y MPO554. Se empleó rifampicina 20 µg/ml y ácido nalidíxico 15 µg/ml, tras comprobar que a esas concentraciones esas estirpes no presentaban crecimiento.

Se partió de cultivos saturados en LB de cada una de las tres estirpes indicadas. Para la búsqueda de resistentes espontáneos a rifampicina se sembraron 200 µl de cada una de ellas en placas de LB con Rif 20 µg/ml, y se incubaron a 37 °C hasta el día siguiente. Aparecieron aproximadamente entre 50 y 100 colonias en cada placa, y se seleccionó una que presentaba buen aspecto para cada estirpe. Para la búsqueda de resistentes espontáneos a ácido nalidíxico se sembró 1 ml de cultivo centrifugado y resuspendido en menos volumen de cada una de ellas en placas de LB con Nal 15 µg/ml, y se incubaron a 37 °C hasta el día siguiente. Aparecieron entre 2 y 8 colonias en cada placa, y se seleccionó una que presentaba buen aspecto para cada estirpe.

8. CONSTRUCCIÓN DE METAGENOTECAS

8.1 Recogida de muestras

Las muestras de suelo se tomaron de las zonas seleccionadas tras eliminar previamente el suelo más superficial. Se almacenaron en bolsas semicerradas a temperatura ambiente (por considerarse una temperatura similar a la del lugar de procedencia de las muestras) y se procesaron en el menor tiempo posible.

8.2 Separación de la fracción bacteriana

Antes de extraer el ADN, se separó de la muestra total una fracción concentrada de bacterias, mediante un gradiente de densidades empleando la resina Nycodenz (Axis-Shield). En primer lugar, se añadieron 140 ml de *disruption buffer* por cada 40 g de suelo, y se batió la mezcla mediante 3 ciclos de 10 segundos seguidos de 1 minuto en hielo. A continuación, se incubó la mezcla durante varias horas en agitación a 4 °C para favorecer la separación de las células de las partículas del suelo. La suspensión se centrifugó a baja velocidad (a 400 g durante 3 minutos seguido de otra centrifugación a 600 g durante 1 minuto al observarse el sobrenadante aún muy turbio), y se transfirieron unos 25 ml del sobrenadante a un tubo que contenía 10 ml de una solución de la resina Nycondenz a 1,3 g/ml. A partir de ese momento se trabajó tratando de mantener la muestra siempre en frío. El tubo se centrifugó 40 minutos a 10.000 g a 4 °C, y se observó la aparición de una banda blanquecina enriquecida en bacterias en la interfase entre el Nycodenz y la fase superior acuosa. Se eliminó parte de la fase acuosa y se continuó añadiendo sobrenadante y repitiendo el proceso hasta que este se hubo terminado. Al final, se transfirió la banda blanquecina a un tubo limpio (5-10 ml) y se le añadieron 35 ml de tampón PBS 1X. Las células se depositaron por centrifugación a 10.000 g durante 20 minutos. Los sedimentos de células se resuspendieron en 500 µl de tampón TE.

Tampón *disruption buffer*: NaCl 0,2 M, Tris-HCl 50 mM pH8.0. Se preparó en agua destilada y se conservó a temperatura ambiente.

Nycodenz (1002424, Axis-Shield). Se preparó a una concentración de 1,3 g/ml en *disruption buffer*. Es muy difícil de disolver y puede completarse la disolución con el autoclavado. Se conservó a temperatura ambiente en oscuridad.

Tampón PBS (*phosphate buffered saline*) 10 X: NaCl 80 g/L, KCl 2,01 g/L, Na₂HPO₄ 14,2 g/L, KH₂PO₄ 2,72 g/L. Ajustar a pH 7,4.

Tampón TE: Tris-HCl 10 mM pH 8, EDTA 1 mM.

8.3 Extracción del ADN

Partiendo de las células resuspendidas en tampón TE, se extrajo el ADN con el kit GNOME DNA (MP Biomedicals), siguiendo las indicaciones del fabricante. La muestra se mantuvo siempre que fue posible en frío, las mezclas se realizaron por inversión del tubo y se tomaron los volúmenes con la pipeta con suavidad, para evitar la degradación del ADN. Al final se resuspendió el ADN metagenómico en 500 µl de tampón TE.

Se comprobó que se había obtenido suficiente ADN (se obtuvo >20 µg) midiendo la concentración en un espectrofotómetro Nanodrop, y se visualizó en gel de agarosa al 0,7 % (p/v) para comprobar el tamaño y estado de degradación del mismo. Es muy importante que el ADN no se encuentre degradado y que presente tamaños de 30-40 kb. Cuando fue necesario, por haber obtenido fragmentos de mayor tamaño, se procedió a la rotura mecánica mediante pipeteos. Se emplearon membranas Microcon Ultracel YM-100 (Millipore) o Amicon Ultra 100K (Millipore) para concentrar la muestra de ADN cuando fue necesario (lo ideal son concentraciones de en torno a 100 ng/µl).

El ADN obtenido se almacenó a 4 °C, el menor tiempo posible. Se siguieron las indicaciones para la construcción de metagenotecas del protocolo del kit CopyControl™ Fosmid Library Production (Epicentre) con algunas modificaciones, entre las que destaca el empleo de pMPO579 como vector en lugar de pCC1FOS.

8.4 Preparación del ADN

En la preparación del ADN se incluyen la reparación de extremos, purificación de gel de las moléculas de ADN de tamaño adecuado y lavado y concentración del ADN.

La reparación de extremos se llevó a cabo mediante una reacción con la mezcla de enzimas End-Repair (Epicentre), siguiendo las instrucciones del fabricante, empleando unos 5 µg de ADN por reacción.

A continuación, se cargó la totalidad de la muestra en un gel de agarosa de bajo punto de fusión al 1 % (p/v) de 20 cm de longitud, y migró junto a una pequeña alícuota de la muestra cargada en otro pocillo y un marcador de peso molecular de gran tamaño, a 45-60 V durante 14-12 horas a 4 °C. Se separó la parte del gel que contenía los carriles con la alícuota y el marcador, se tiñó en una solución de bromuro de etidio y se visualizó con un transiluminador de radiación ultravioleta, comprobando que la alícuota de la muestra tenía el tamaño adecuado (30-40 kb) y se hizo una marca en el gel a la altura a la que esta se encontraba. Posteriormente, con la ayuda de la marca realizada, se cortó la banda de ADN metagenómico de la otra parte del gel, de manera que este no fue expuesto a radiación ultravioleta. Una vez cortada la banda, se tiñó y visualizó el resto del gel para comprobar que se había recuperado el ADN.

El ADN se extrajo del gel mediante un tratamiento con la enzima GELase (Epicentre), sin emplear tampón, añadiendo 2 µl de enzima por cada 300 µg de gel, e incubando 3 horas a 45 °C. Para el resto de aspectos relacionados con este tratamiento se siguieron las indicaciones del fabricante. La purificación se finalizó con una centrifugación de 15 minutos a 15.000 g a 4 °C y la recuperación del sobrenadante, siendo cuidadosos con no tomar nada del sedimento.

Por último, se lavó y concentró el ADN con las membranas Microcon Ultracel YM-100 (Millipore) o Amicon Ultra 100K (Millipore). Para ello, se redujo el volumen de la muestra, se añadieron 250 µl de agua bidestilada, se volvió a reducir el volumen y a repetir el proceso, y finalmente se concentró hasta un volumen final máximo de 20-30 µl. Se midió la concentración en el espectrofotómetro Nanodrop y se volvió a comprobar el tamaño y estado del ADN mediante su visualización tras migración en gel de agarosa al 0,7 % (p/v).

Agarosa de bajo punto de fusión (8050, Conda Pronadisa).

8.5 Ligación del ADN metagenómico con el vector pMPO579

Para la preparación del vector, se extrajo una gran cantidad de ADN de pMPO579 y se digirió con *PmlI* (New England Biolabs, isoesquizómero de *Eco72I*). A continuación se desfosforilaron los extremos con la fosfatasa SAP (Shrimp Alkaline Phosphatase, USB) y se lavó y concentró todo lo posible (300-600 ng/µl) con membranas Microcon Ultracel YM-100 (Millipore) o Amicon Ultra 100K (Millipore).

Como inserto se empleó el ADN metagenómico preparado como se indica en los apartados anteriores, obteniéndose concentraciones aproximadas de 70-150 ng/ μ l.

Como se obtuvo ADN metagenómico suficiente, se realizaron varias reacciones de ligación con la enzima Fast-link DNA Ligase (Epicentre), siguiendo las indicaciones del fabricante. El vector y el inserto tienen tamaños aproximados de 10 kb y 40 kb respectivamente. Se probaron distintas proporciones vector:inserto, que variaron entre 10:1 y 3:1 aproximadamente (relación referida a número de moléculas), no quedando claro que alguna proporción fuera ventajosa sobre las demás para la obtención de un mayor número de clones. Las ligaciones se pueden almacenar a -20 °C.

8.6 Titulación de la metagenoteca

Mediante la titulación se comprobó si con el material disponible se podía obtener una metagenoteca con un número de clones adecuado. Para ello, se empaquetaron 10 μ l de ligación con extractos de empaquetamiento MaxPlax Lambda Packaging Extracts (Epicentre), a lo que se añadió tampón PD hasta un volumen final de 1 ml. Se tomó una parte del empaquetado para infectar células EPI300-T1^R, cultivadas en LB con MgSO₄ 10 mM y maltosa 0,2 %, según se indica en el protocolo del kit CopyControlTM Fosmid Library Production (Epicentre), conservando el resto a 4 °C el menor tiempo posible (unos días). Para la infección se mezclaron 100 μ l de células EPI300-T1^R preparadas a D.O.₆₀₀ de 0,8-1 y 10 μ l de diluciones del empaquetado con factores de dilución 10, 10¹, 10², 10³, 10⁴ y 10⁵, y se incubaron 1 h a 37 °C. Las células infectadas se sembraron en placas de LB con Cm 12,5 μ g/ml y se incubaron a 37 °C toda la noche, para seleccionar células con fósido. Se contaron las colonias crecidas y se estimó el número de clones que tendría la metagenoteca si se realizara una infección con el resto del empaquetado, teniendo también en cuenta el resto de ligaciones que se tuvieran que no se hubieran empaquetado aún. Cuando el número de clones estimado fue suficiente (a partir de decenas de miles de clones), se procedió al empaquetado del resto de ligaciones y a la infección en masa con todo el material.

Tampón PD (*Phage Dilution*): 10 mM Tris-ClH pH 8.3, 100 mM NaCl, 10 mM MgCl₂

D-(+)-Maltosa monohidratada (M5885, Sigma-Aldrich). El stock (10 %) se preparó en agua destilada y se conservó a temperatura ambiente.

8.7 Empaquetado de las ligaciones

Se empaquetaron el resto de ligaciones que no se hubieran empleado para la titulación, que se habían mantenido a $-20\text{ }^{\circ}\text{C}$, del mismo modo que se ha descrito para realizar dicha titulación.

8.8 Infección en masa y almacenamiento de la metagenoteca

Se unieron los empaquetados recién preparados con los almacenados a $4\text{ }^{\circ}\text{C}$ y se procedió a la infección en masa. Para ello, se mezclaron $100\text{ }\mu\text{l}$ de empaquetado y 1 ml de células de EPI300-T1^R cultivadas en LB con MgSO_4 10 mM y maltosa $0,2\text{ }\%$ a D.O.₆₀₀ de $0,8-1$, en tantos tubos como fue necesario para utilizar todo el volumen de empaquetado disponible. Se incubaron 90 minutos a $37\text{ }^{\circ}\text{C}$ y tras la incubación se unió todo el volumen de células infectadas en un matraz, se aumentó el volumen con medio LB y se añadió Cm de forma que quedara a una concentración final de $12,5\text{ }\mu\text{g/ml}$. El aumento de volumen dependió de la cantidad de células sometidas a infección y del número de clones estimados, teniéndose al final entre 50 y 100 ml de cultivo con la metagenoteca. Una vez aumentado el volumen, se sembraron distintas cantidades en placas con LB y Cm $12,5\text{ }\mu\text{g/ml}$ y se incubaron a $37\text{ }^{\circ}\text{C}$, para saber el número de clones de esa metagenoteca. A continuación, se incubó el resto del cultivo, para que las células infectadas con fósido con ADN metagenómico (clones de la metagenoteca) se multiplicaran, a $37\text{ }^{\circ}\text{C}$ con agitación, durante $10-12\text{ h}$. Tras ese tiempo se había alcanzado una D.O.₆₀₀ de $1,5$ aproximadamente (no interesa que crezca demasiado para que no se enriquezca mucho en aquellos clones que presenten un crecimiento más rápido y se pierdan los de crecimiento lento). Se añadió glicerol a una concentración final de $15\text{ }\%$ y se realizaron alícuotas de $250\text{ }\mu\text{l}$ y de 1 ml , que se congelaron a $-80\text{ }^{\circ}\text{C}$. Las metagenotecas se almacenaron de este modo en masa en la estirpe EPI300-T1^R. Con el número de colonias de las placas sembradas se realizó el cálculo del número de clones de la metagenoteca, que fue del mismo orden que el estimado en las titulaciones previas, pero más exacto.

A continuación se indican algunos datos sobre las construcciones de metagenotecas realizadas en esta tesis.

Metagenoteca de Punta de San García: Se partió de 160 g de suelo, de los que, tras separar la fracción bacteriana por densidad con la resina Nycodenz, se obtuvieron $24\text{ }\mu\text{g}$ de ADN.

Tras el tratamiento (reparación de extremos, purificación de gel, lavado y concentración) quedaron aproximadamente 1,5 µg de ADN (unos 20 µl a 72,8 ng/µl). Se llevaron a cabo 4 reacciones de ligación. Con este material se obtuvo una metagenoteca de 54.000 clones independientes. Siendo el tamaño medio de los insertos de 35 kb, esta metagenoteca contiene alrededor de 1,89 Gb de ADN metagenómico.

Metagenoteca de pilas de refinería de la industria petroquímica: Se partió de 40 g de suelo, de los que, tras separar la fracción bacteriana por densidad con la resina Nycodenz, se obtuvieron 25 µg de ADN. Tras el tratamiento (rotura mecánica mediante pipeteos, reparación de extremos, purificación de gel, lavado y concentración) quedaron aproximadamente 4 µg de ADN (unos 27 µl a 150 ng/µl). Se llevaron a cabo 6 reacciones de ligación. Con este material se obtuvo una metagenoteca de 185.000 clones independientes. Siendo el tamaño medio de los insertos de 35 kb, esta metagenoteca contiene alrededor de 6,5 Gb de ADN metagenómico.

9. RASTREO DE ACTIVIDADES ENZIMÁTICAS

Las metagenotecas construidas se encuentran congeladas a -80 °C en masa en la estirpe EPI300-T1^R. Para realizar los rastreos de las distintas actividades, las metagenotecas se transfirieron por conjugación a otras estirpes.

A todos los clones positivos para cualquier actividad se les realizó un análisis de restricción que permitía identificar los clones diferentes.

El cálculo del número de clones necesarios para poder asegurar con una probabilidad determinada que se habían analizado todos los clones de la metagenoteca se realizó mediante una adaptación de la fórmula de Clarke y Carbon (1976). La fórmula empleada fue:

$$N = (\ln(1-P)) / (\ln(1-f))$$

donde P es la probabilidad, N el número de clones a analizar y f el inverso del número de clones de la metagenoteca. El número de clones analizados se estimó mediante conteos de las colonias crecidas en varias placas y extrapolación al total.

9.1 Rastreo de resistencias a antibióticos

Este rastreo se realizó en mutantes resistentes espontáneos a los antibióticos rifampicina (Rif) y ácido nalidíxico (Nal) de EPI300-T1^R, MPO553 y MPO554. La metagenoteca de Punta de San García se transfirió por conjugación triparental a cada una de estas estirpes como se describe en el apartado 4.2.1 de Materiales y Métodos. Se buscaron clones resistentes a carbenicilina. Se seleccionó la concentración 100 µg/ml de Cb por ser la concentración más baja de las probadas a la que la estirpe EPI300-T1^R no fue capaz de crecer.

Se realizaron diluciones del parche de conjugación resuspendido en 2 ml de tampón fosfato y se sembraron en placas de 9 cm de diámetro: 2-3 placas con factor de dilución 5·10 y otras 2-3 placas con factor de dilución 10². Estas placas contenían medio LB con arabinosa 1 mM (para aumentar el número de copias del fósido), Rif 20 µg/ml o Nal 15 µg/ml (para seleccionar la estirpe receptora) y Cb 100 µg/ml. No se aportó Cm para seleccionar las bacterias con fósido porque con Cb es suficiente (solo crecerán aquellas con fósido que contenga algún gen que les aporte resistencia a este antibiótico). Cuando la estirpe receptora era MPO554, se realizaron además siembras en placas que también contenían salicilato 1 mM (para permitir la expresión heteróloga desde *psal* en el fósido y la antiterminación de la transcripción por la proteína N, apartado 1.1 de Resultados). Las placas se incubaron a 30 °C durante 2 días. Tras ese tiempo, se identificaron colonias de buen tamaño. Se comprobó que los clones positivos tenían fósido y que esos fósidos aportaban resistencia a otra bacteria al transferírsele (para descartar que fueran resistentes espontáneos).

Para esta actividad también se realizó una búsqueda directa en la estirpe EPI300-T1^R en la que está almacenada la metagenoteca. Para ello, se cultivó una alícuota de la metagenoteca de 250 µl como se indica para realizar la conjugación, pero cuando alcanzó D.O.₆₀₀ de aproximadamente 0,5, se sembraron directamente diluciones en placas de 9 cm de diámetro: 10 placas con factor de dilución 10² y 5 placas con factor de dilución 10³. Estas placas contenían medio LB con arabinosa 1 mM y carbenicilina 100 µg/ml, y se incubaron a 30 °C durante 2 días.

En todos los casos, también se realizaron siembras con distintos factores de dilución (10⁴-10⁷) en placas con el mismo medio que el de los rastreos salvo que llevaban Cm en lugar de Cb, con el objetivo de calcular el número de clones que se estaban analizando. Los cálculos indicaron que en cada placa de selección de resistencia a Cb se estaban sembrando del orden de 10⁵-10⁶ bacterias.

9.2 Rastreo de dioxigenasas extradiólicas

Este rastreo se realizó en un mutante resistente espontáneo a ácido nalidíxico (Nal) de MPO554. La metagenoteca de pilas de refinería de la industria petroquímica se transfirió por conjugación triparental a esa estirpe como se describe en el apartado 4.2.1 de Materiales y Métodos.

Se realizaron diluciones del parche de conjugación resuspendido en 2 ml de tampón fosfato y se sembraron en placas de 15 cm de diámetro: 185 placas en total, con factores de dilución $5 \cdot 10^4$ y 10^4 . Estas placas contenían medio LB con arabinosa 1 mM (para aumentar el número de copias del fósido), salicilato 1 mM (para permitir la expresión heteróloga desde *psal* en el fósido y la antiterminación de la transcripción por la proteína N, apartado 1.1 de Resultados), Cm 12,5 $\mu\text{g/ml}$ (para seleccionar el fósido) y Nal 15 $\mu\text{g/ml}$ (para seleccionar la estirpe receptora). Las placas se incubaron a 30 °C durante 2 días y se aplicó catecol 1 % con un spray sobre las colonias crecidas. Tras unos minutos se identificaron colonias que habían cambiado su color a amarillo, por la transformación del catecol en 2-hidroximuconato semialdehído.

Se repitieron los análisis como se han descrito, sembrando aproximadamente 150 placas con los mismos factores de dilución. Pero esta vez, sobre las colonias crecidas se aplicó 2,3-dihidroxibifenilo 0,5 % con un spray. Tras unos minutos se identificaron colonias que habían cambiado su color a amarillo, por la transformación del 2,3-dihidroxibifenilo en 2-hidroxi-6-oxo-6-fenil-hexa-2,4-dienoato.

En ambos casos se detectó el cambio de color de las colonias tras la adición del sustrato correspondiente pasados entre 1 y 30 minutos aproximadamente.

Sembrando diluciones del parche de conjugación con los factores de dilución indicados se obtuvieron 1.000-2.000 y 2.500-4.000 colonias en cada placa respectivamente.

Catecol (135011, Sigma-Aldrich). El stock (1 % (p/v)) se preparó en agua bidestilada. Se conservó en oscuridad a temperatura ambiente.

2,3-dihidroxibifenilo (17403, Sigma-Aldrich). El stock (0,5 % (p/v)) se preparó en etanol 96 %. Se conservó en oscuridad a temperatura ambiente.

9.3 Rastreo de oxigenasas hidroxilantes de anillos aromáticos

Este rastreo se realizó en un mutante resistente espontáneo a ácido nalidíxico (Nal) de MPO554. La metagenoteca de pilas de refinería de la industria petroquímica se transfirió por conjugación triparental a esa estirpe como se describe en el apartado 4.2.1 de Materiales y Métodos.

Se realizaron diluciones del parche de conjugación resuspendido en 2 ml de tampón fosfato y se sembraron en placas de 15 cm de diámetro: unas 200 placas en total, con factores de dilución $5 \cdot 10^4$ y 10^4 . Estas placas contenían medio LB con arabinosa 1 mM (para aumentar el número de copias del fósido), salicilato 1 mM (para permitir la expresión heteróloga desde *psal* en el fósido y la antiterminación de la transcripción por la proteína N, apartado 1.1 de Resultados), Cm 12,5 $\mu\text{g/ml}$ (para seleccionar el fósido) y Nal 15 $\mu\text{g/ml}$ (para seleccionar la estirpe receptora). Las placas se incubaron a 30 °C durante 2 días y a continuación a temperatura ambiente hasta 7 días. Se identificaron colonias azules desde el segundo hasta el séptimo día.

Sembrando diluciones del parche de conjugación con los factores de dilución indicados se obtuvieron 1.000-2.000 y 2.500-4.000 colonias en cada placa respectivamente.

Para este rastreo se emplea como sustrato inicial el triptófano presente en el medio LB, a partir del cual, la triptofanasa endógena de *E. coli* es capaz de producir indol.

9.4 Rastreo de endoglucanasas

Este rastreo se realizó en un mutante resistente espontáneo a ácido nalidíxico (Nal) de MPO554. La metagenoteca de pilas de refinería de la industria petroquímica se transfirió por conjugación triparental a esa estirpe como se describe en el apartado 4.2.1 de Materiales y Métodos. Se empleó carboximetilcelulosa (CMC) como sustrato, que puede ser degradada por endoglucanasas, y rojo Congo como indicador.

Se realizaron diluciones del parche de conjugación resuspendido en 2 ml de tampón fosfato y se sembraron en placas de 15 cm de diámetro: 278 placas en total, con factores de dilución $5 \cdot 10^4$ y 10^4 . Estas placas contenían medio LB con arabinosa 1 mM (para aumentar el número de copias del fósido), salicilato 1 mM (para permitir la expresión heteróloga desde *psal* en el fósido y la antiterminación de la transcripción por la proteína N, apartado 1.1 de Resultados), Cm 12,5 $\mu\text{g/ml}$ (para seleccionar el fósido), Nal 15 $\mu\text{g/ml}$ (para seleccionar la

estirpe receptora) y CMC 1 %. Las placas se incubaron a 30 °C durante 2 días. A continuación, se aplicó una capa de LB-agar de cobertera (8 ml por placa) para inmovilizar las colonias. Una vez solidificada se procedió a la tinción. Para ello, se añadieron sobre cada placa 5 ml de rojo Congo 1 % y se incubaron a temperatura ambiente 30 minutos. Tras ese tiempo, se retiró el rojo Congo, se añadieron sobre cada placa 8 ml de NaCl 1 M y se incubaron a temperatura ambiente 30 minutos. Por último, se retiró el NaCl, y se observaron las placas teñidas en busca de halos alrededor de las colonias. Los halos se deben a que las colonias con actividad endoglucanasa han degradado la CMC, por lo que no puede teñirse con rojo Congo.

Sembrando diluciones del parche de conjugación con los factores de dilución indicados se obtuvieron 700-2.000 y 2.000-4.000 colonias en cada placa respectivamente.

LB-agar de cobertera: Bactotripton 10 g/L, extracto de levadura 5 g/L, NaCl 5 g/L y agar 7,5 g/L.

Rojo Congo (C6277, Sigma-Aldrich). El stock (1 % (p/v)) se preparó en agua destilada (es necesario filtrarlo) y se conservó a temperatura ambiente.

NaCl 1 M. Se preparó en agua destilada y se conservó a temperatura ambiente.

Carboximetilcelulosa (21902, Sigma-Aldrich). Se preparó a 1 % (p/v) aportándola directamente al medio de cultivo junto con los demás componentes.

9.5 Rastreo de β -glucosidasas

Este rastreo se realizó en un mutante resistente espontáneo a ácido nalidíxico (Nal) de MPO554. La metagenoteca de pilas de refinería de la industria petroquímica se transfirió por conjugación triparental a esa estirpe como se describe en el apartado 4.2.1 de Materiales y Métodos. Se buscaron clones capaces de crecer con celobiosa 1 % como única fuente de carbono.

Se realizaron diluciones del parche de conjugación resuspendido en 2 ml de tampón fosfato, tras lavarlo tres veces con ese mismo tampón para eliminar los restos del medio LB, y se sembraron en placas de 9 cm de diámetro: 6 placas con factor de dilución $5 \cdot 10$ y otras 6 placas con factor de dilución 10^2 . Estas placas contenían la variante del medio mínimo M9 descrita en el apartado 2.1 de Materiales y Métodos pero empleando celobiosa 1 % como

única fuente de carbono en lugar de glucosa 0,2 %, con arabinosa 1 mM (para aumentar el número de copias del fósido), salicilato 1 mM (para permitir la expresión heteróloga desde *psal* en el fósido y la antiterminación de la transcripción por la proteína N, apartado 1.1 de Resultados) y Nal 15 µg/ml (para seleccionar la estirpe receptora). No se aportó Cm para seleccionar las bacterias con fósido porque el crecimiento a expensas de celobiosa seleccionará aquellas con fósido que contenga algún gen que les permita obtener glucosa a partir de celobiosa. Las placas se incubaron a 30 °C y se observó la aparición de colonias tras unos 2-5 días.

También se realizaron siembras con factores de dilución 10^4 y 10^5 en placas con medio LB con Cm 12,5 µg/ml, Nal 15 µg/ml, arabinosa 1 mM y salicilato 1 mM, con el objetivo de calcular el número de clones que se estaban analizando. Los cálculos indicaron que en cada placa en la que se sembraron diluciones del parche de conjugación con factores de dilución $5 \cdot 10^4$ y 10^2 se estaban sembrando aproximadamente $3 \cdot 10^5$ y $6 \cdot 10^5$ clones respectivamente.

D-(+)-Celobiosa (22150, Sigma-Aldrich). El stock (10 % (p/v)) se preparó en agua bidestilada, y se conservó a temperatura ambiente.

10. SECUENCIACIÓN DE CLONES Y ANÁLISIS DE SECUENCIAS

La secuenciación de los clones metagenómicos se llevó a cabo con un secuenciador Roche 454 GS FLX Ti (454 Life Sciences, Branford, CT, USA) por la empresa Lifesequencing (Valencia), a partir de preparaciones de ADN de alta calidad de los fósidos. El ensamblaje de las secuencias se realizó mediante Newbler GS De Novo Assembler v.2.3 (Roche). Algunos clones se secuenciaron de manera individual y otros agrupados en *pools*. Posteriormente, se unieron algunos contigs y se completó la secuencia de otros mediante reacciones de secuenciación llevadas a cabo por la empresa Secugen (Madrid).

De gran parte de las secuencias se obtuvieron anotaciones automáticas. Por otro lado, se realizaron comparaciones con las de las bases de datos nucleotide collection, non-redundant protein sequences y UniProtKB/Swiss-Prot empleando el paquete de herramientas BLAST (Altschul y col., 1997) disponible en la página web del National Center for Biotechnology Information (NCBI) (<http://blast.ncbi.nlm.nih.gov>). Se buscaron dominios conservados de las

proteínas de interés en la base de datos Pfam (<http://pfam.sanger.ac.uk>) (Bateman y col., 2004).

Para el alineamiento de secuencias se empleó el programa ClustalX (Thompson y col., 1994) junto con el programa BioEdit (Hall, 1999). La construcción de árboles filogenéticos se realizó con el programa ClustalX utilizando el método Neighbour-Joining y se editaron con el programa TreeView.

11. ANÁLISIS DE PROTEÍNAS

11.1 Superproducción de proteínas

Con el objetivo de superproducir la proteína UPOEDO11 se usó la estirpe NCM631 transformada previamente con el plásmido pIZ227, que contiene un sistema de expresión basado en la RNA polimerasa de T7 inducible con IPTG (Govantes y Santero, 1996). En esta estirpe se introdujo el plásmido que contenía el gen de la proteína a superproducir clonado bajo el promotor pT7 (pMPO1383). Como control se usó el vector original donde se había clonado (pET-14b).

Se diluyeron inóculos saturados de las estirpes transformadas a D.O.₆₀₀ de 0,1 en LB con Cm y Ap. Se incubaron en agitación a 37 °C hasta que la D.O.₆₀₀ alcanzó un valor aproximado de 0,7 y se incubaron en hielo 15 minutos. En ese momento se añadió IPTG a los cultivos a una concentración final de 1 mM y se incubaron a 26 °C durante 12 h. Tras ese tiempo se midió la D.O.₆₀₀ alcanzada.

11.2 Obtención de lisados totales de cultivos bacterianos

Para el análisis de la proteína total de las células se tomó 1 ml de los cultivos de superproducción descritos en el apartado anterior, se centrifugaron 1 minuto a 16.100 g y se congeló el sedimento a -80 °C hasta su utilización. Tras descongelar el sedimento, se resuspendió en 25 µl de agua bidestilada por cada unidad de D.O.₆₀₀ de cultivo y se le añadieron otros 25 µl de tampón de carga 2X. La mezcla se hirvió durante 10 minutos. Los restos celulares se depositaron por centrifugación durante 5 minutos a 16.100 g. Se utilizaron 5 µl del sobrenadante para visualizar las proteínas en gel.

Tampón de carga 2X: Tris-HCl 160 mM pH 8, glicerol 20 % (v/v), SDS 4 %, azul de bromofenol 0,1 %, β -mercaptoetanol 10 % (v/v). Añadir el β -mercaptoetanol justo antes de usar.

11.3 Electroforesis de proteínas

Para comprobar la superproducción de la proteína UPOEDO11 se empleó la electroforesis desnaturalizante en gel de poliacrilamida-SDS. Para ello, se emplearon los lisados totales preparados como se indica en el apartado anterior.

Las proteínas totales fueron analizadas mediante electroforesis en geles verticales desnaturalizantes de poliacrilamida con SDS 0,1 % (p/v), básicamente según el método descrito por Laemmli (1970). La electroforesis de las muestras se llevó a cabo en geles discontinuos consistentes en una parte superior de empaquetamiento y una inferior de separación. La concentración de acrilamida en estos geles fue del 12,5 % en el de separación y del 4,5 % en el gel de empaquetamiento. Como marcador de peso molecular se usó el marcador B de Ready Gel Bands (Bio-Rad). La electroforesis se realizó en tank buffer 1X a una intensidad de corriente de 30 mA.

Para su tinción, el gel se lavó 2 veces con agua destilada durante 10 minutos, se tiñó con EZBlue Gel Staining Reagent (Sigma-Aldrich) durante 20 minutos y después se lavó con agua destilada hasta que las bandas correspondientes a las proteínas teñidas se visualizaron claramente. Para conservar el gel, tras escanearlo, se secó en un secador de geles HoeferTM Slab Gel Dryer CD 2000 (GE Healthcare) durante 1 h a 80 °C.

Gel de empaquetamiento: Tris-HCl 125 mM pH 6,8, SDS 0,1 % (p/v), acrilamida:bisacrilamida (37,5:1) 4,5 % (p/v), persulfato amónico 0,1 % (p/v), TEMED 0,1 % (p/v).

Gel de separación: Tris-HCl 375 mM pH 8,8, SDS 0,075 % (p/v), acrilamida:bisacrilamida (37,5:1) 12,5 % (p/v), persulfato amónico 0,033 % (p/v), TEMED 0,067 % (p/v).

Tank buffer 1X: Trizma base 25 mM, glicina 19,2 mM y SDS 0,1 % (p/v).

12. ANÁLISIS DE LA EXPRESIÓN GÉNICA

12.1 Ensayos de medida del nivel de fluorescencia

Se realizaron medidas de fluorescencia para determinar los niveles de expresión del fósido para la construcción de metagenotecas pMPO579, pMPO578 y el control con terminador pMPO580 en distintas condiciones, utilizando como gen indicador el gen *gfp* sin promotor presente en estos fósidos aguas abajo y en el mismo sentido que los promotores heterólogos. Se utilizaron las estirpes EPI300-T1^R, MPO553, MPO554 y MPO555. A continuación se indican las combinaciones de estirpes y fósidos ensayadas y las condiciones de inducción en cada caso: EPI300-T1^R con pMPO578, sin ningún inductor y con arabinosa 1 mM, y con pMPO579, además con arabinosa 1 mM más salicilato 5 mM. MPO553 con cada uno de los 3 fósidos, sin ningún inductor y con arabinosa 1 mM. MPO554 con cada uno de los 3 fósidos, sin ningún inductor, con arabinosa 1 mM, con salicilato 1 mM y con arabinosa 1 mM más salicilato 1 mM. Y MPO555 con pMPO579 y con pMPO580, sin ningún inductor, con arabinosa 1 mM y con arabinosa 1 mM más salicilato 1 mM.

Se partió de cultivos saturados, que se diluyeron a D.O.₆₀₀ de 0,1 en LB con Cm 12,5 µg/ml, y con arabinosa 1 mM en su caso, y se incubaron a 37 °C en agitación hasta que alcanzaron una D.O.₆₀₀ aproximada de 0,3-0,4. En el caso de inducción con salicilato, este se añadió en este momento a una concentración 5 mM y los cultivos se volvieron a incubar a 37 °C en agitación 6 h. Tras ese tiempo se tomaron 250 µl de los cultivos, se lavaron y resuspendieron en 1 ml de tampón fosfato y de ahí se echaron 200 µl a pocillos de una placa multipocillo Costal 96. Como blanco se empleó tampón fosfato. Cada muestra se midió en 3 pocillos independientes. A las microplacas con las muestras se les midió D.O.₆₀₀ y fluorescencia empleando los filtros de excitación 485BP1 y de emisión EM520, y con un valor de ganancia de 1000, en un lector de microplacas multifuncional POLARstar Omega (BMG LABTECH GmbH, Alemania). Los valores mostrados se corresponden a las unidades de fluorescencia divididas entre la D.O.₆₀₀ de la muestra. En todos los casos se realizaron al menos tres réplicas independientes, mostrando los valores medios y las desviaciones estándar entre dichas réplicas.

Por otro lado, también se realizaron medidas de fluorescencia emitida para determinar los niveles de expresión de las fusiones a *gfp*. Se ensayaron las construcciones pMPO593, pMPO594, pMPO596 y pMPO597 acompañadas todas ellas por un lado del plásmido pMPO588 y por otro del plásmido vacío pIZ1016, en la estirpe EPI300-T1^R y en MPO554

Nal^R. Se partió de cultivos saturados en medio LB con Cm 12,5 µg/ml y Gm 10 µg/ml incubados a 37 °C. Se lavaron 2 veces con tampón fosfato y se diluyeron a una D.O.₆₀₀ aproximada de 0,01 en la variante del medio mínimo M9 descrita en el apartado 2.1 de Materiales y Métodos pero que llevaba además dibenzotiofeno (DBT) 0,2 mM. Se incubaron en agitación a 30 °C 24 h. Tras ese tiempo, se diluyeron a D.O.₆₀₀ de 0,1 en un medio igual pero con MgSO₄ 0,2 mM (menos concentrado que en ese medio), DBT 0,2 mM, IPTG 1 mM y en el caso de la estirpe MPO554 Nal^R también salicilato 1 mM. Se incubaron en agitación a 30 °C durante 6 h. Los fósidos se mantuvieron en bajo número de copias (sin aportar arabinosa).

Una vez transcurrido el tiempo de incubación, se diluyeron a D.O.₆₀₀ de 0,10-0,15 en tampón fosfato. Se echaron 150 µl de las muestras en placas multipocillo opacas (Microfluor) y se midió la emisión de fluorescencia como se indicó anteriormente. En todos los casos se realizaron al menos tres réplicas independientes, mostrando los valores medios y las desviaciones estándar entre dichas réplicas.

Tampón fosfato 50X: 302,5 g/L Na₂HPO₄·12H₂O y 50 g/L KH₂PO₄.

12.2 Ensayos de actividad β-galactosidasa

Los ensayos de actividad β-galactosidasa de las fusiones a *lacZ* se realizaron siguiendo el método descrito por Miller (Miller, 1972) con algunas modificaciones. Se ensayaron las construcciones pMPO593, pMPO594, pMPO596 y pMPO597 acompañadas todas ellas por un lado del plásmido pMPO588 y por otro del plásmido vacío pIZ1016, en la estirpe EPI300-T1^R y en MPO554 Nal^R. Se partió de cultivos saturados en medio LB con Cm 12,5 µg/ml y Gm 10 µg/ml incubados a 37 °C. Se lavaron 2 veces con tampón fosfato y se diluyeron a una D.O.₆₀₀ aproximada de 0,01 en la variante del medio mínimo M9 descrita en el apartado 2.1 de Materiales y Métodos pero que llevaba además dibenzotiofeno (DBT) 0,2 mM. Se incubaron en agitación a 30 °C 24 h. Tras ese tiempo, se diluyeron a D.O.₆₀₀ de 0,1 en un medio igual pero con MgSO₄ 0,2 mM (menos concentrado que en ese medio), DBT 0,2 mM, IPTG 1 mM y en el caso de la estirpe MPO554 Nal^R también salicilato 1 mM. Se incubaron en agitación a 30 °C durante 6 h. Los fósidos se mantuvieron en bajo número de copias (sin aportar arabinosa).

Una vez transcurrido el tiempo de incubación, se diluyeron a D.O.₆₀₀ de 0,10-0,15 en tampón fosfato. Se tomaron 100 µl de las muestras y se mezclaron en hielo con 655 µl de tampón Z con β-mercaptoetanol recién añadido, 30 µl de cloroformo y 15 µl de SDS 0,1 % (p/v). Las mezclas se homogenizaron en el vórtex y se incubaron 10 minutos a 30 °C. A continuación se les añadieron 200 µl de ONPG 13 mM. Se volvieron a homogenizar en el vórtex y se incubaron a 30 °C hasta observar la aparición de un color amarillo en las mezclas, momento en el que se paró la reacción añadiendo 500 µl de Na₂CO₃ 1 M, se volvió a homogeneizar en el vórtex y se colocaron en hielo. Se midió el tiempo de reacción, es decir, el tiempo transcurrido entre la adición de ONPG y la de Na₂CO₃. Las mezclas se centrifugaron 15 minutos a 9.300 g y se midió la absorbancia a 420 nm del sobrenadante en un espectrofotómetro UV-Visible Shimadzu (Kyoto, Japon). La actividad de cada muestra se calculó según la siguiente fórmula:

$$\text{Actividad } \beta\text{-galactosidasa (en Unidades Miller)} = \frac{A_{420}}{V \cdot D.O._{600} \cdot t} \cdot 1000$$

donde A₄₂₀ es la absorbancia a 420 nm, V es el volumen de cultivo usado en el ensayo en ml (normalmente 0,1 ml), D.O.₆₀₀ es la densidad óptica a 600 nm del cultivo en el momento de empezar el ensayo y t es el tiempo de reacción en minutos. En todos los casos se realizaron al menos tres réplicas independientes, mostrando los valores medios y las desviaciones estándar entre dichas réplicas.

Tampón fosfato 50X: 302,5 g/L Na₂HPO₄·12H₂O y 50 g/L KH₂PO₄.

Tampón Z con β-mercaptoetanol: Na₂HPO₄ 60 mM, NaH₂PO₄·H₂O 40 mM, KCl 10 mM, MgSO₄·7H₂O 1 mM, β-mercaptoetanol 0,27% (v/v). Añadir el β-mercaptoetanol justo antes de usar.

ONPG 13 mM: o-nitrofenil-β-D-galactósido, preparado en tampón Z.

13. ENSAYOS DE ACTIVIDAD ENZIMÁTICA DIOXIGENASA EXTRADIÓLICA

Se determinó la actividad dioxigenasa extradiólica sobre diferentes derivados catecólicos midiendo la aparición de los correspondientes productos de rotura del anillo

bencénico (típicamente de color amarillo). Los ensayos se realizaron con células enteras en tampón fosfato Na/K 50 mM pH 6,8 a temperatura ambiente. Los sustratos empleados fueron 2,3-dihidroxibifenilo, catecol, 3-metilcatecol, 4-metilcatecol y 4-clorocatecol.

Se realizaron ensayos con los 58 clones identificados en los rastreos para esta actividad y con pMPO1383, un vector de superproducción en el que se clonó el gen codificante de dioxigenasa extradiólica *upoedo11*. Los ensayos de los 58 clones se llevaron a cabo en la estirpe MPO554 NaI^R. Se partió de cultivos saturados en medio LB con Cm 12,5 µg/ml, arabinosa 1 mM y salicilato 1 mM incubados a 30 °C. El subclón pMPO1383 se sometió a las condiciones de superproducción que se explican en el apartado 11.1 de Materiales y Métodos.

Una vez obtenidos los cultivos de los clones y del subclón, se recogió el volumen de cultivo necesario para tener finalmente la D.O.₆₀₀ deseada para cada caso (que varió de 3 a 10 en función del nivel de actividad que presentaba cada clon). Las células se recogieron por centrifugación durante 1 minuto a 16.100 g a 4 °C, se lavaron en tampón fosfato, se volvieron a centrifugar y se resuspendieron en el volumen final deseado de tampón fosfato (1-2 ml). Se midió la D.O.₆₀₀ y se mantuvieron en hielo hasta que se continuó con el ensayo. Se prepararon placas multipocillo Costal 96 con 100 µl de blanco y de cada muestra. Como blanco se empleó el mismo tampón fosfato. Cada muestra se repartió en 3 pocillos independientes. Una vez que estaban las placas preparadas con las muestras, se fue añadiendo el sustrato a cada uno de los pocillos. Inmediatamente, se realizó una cinética de 20 minutos midiendo cada minuto la absorbancia a la longitud de onda adecuada según el sustrato. Esto se realizó con cada uno de los 5 sustratos seleccionados, empleando para ello muestra procedente de la misma preparación. Las medidas se realizaron en un lector de microplacas multifuncional POLARstar Omega (BMG LABTECH GmbH, Alemania).

La concentración a la que se aportaron los distintos sustratos para el ensayo fue 2,3-dihidroxibifenilo 0,5 mM, catecol 0,5 mM, 3-metilcatecol 0,4 mM, 4-metilcatecol 0,8 mM y 4-clorocatecol 1 mM. Se prepararon stocks de los 5 sustratos a 10 mM, por lo que a cada pocillo con 100 µl de muestra se le añadieron 5 µl, 5 µl, 4 µl, 8 µl o 10 µl del sustrato respectivamente. A continuación se indican la longitud de onda de máxima absorción ($\lambda_{\text{máx}}$) y el coeficiente de extinción (ϵ) del producto de rotura correspondiente para cada sustrato: 2,3-dihidroxibifenilo, $\lambda_{\text{máx}} = 434 \text{ nm}$, $\epsilon = 13,2 \text{ mM}^{-1} \text{ cm}^{-1}$; catecol, $\lambda_{\text{máx}} = 375 \text{ nm}$, $\epsilon = 36 \text{ mM}^{-1} \text{ cm}^{-1}$; 3-metilcatecol, $\lambda_{\text{máx}} = 388 \text{ nm}$, $\epsilon = 13,8 \text{ mM}^{-1} \text{ cm}^{-1}$; 4-metilcatecol, $\lambda_{\text{máx}} = 382 \text{ nm}$, $\epsilon =$

28,1 mM⁻¹ cm⁻¹; y 4-clorocatecol, $\lambda_{\text{máx}} = 379$ nm, $\epsilon = 39,6$ mM⁻¹ cm⁻¹ (Andújar y col., 2000; Heiss y col., 1995).

Para el cálculo de actividad con cada sustrato se averiguó el incremento en la concentración del producto generado por unidad de tiempo, según la ecuación de Lambert-Beer: $A = \epsilon \cdot C \cdot L$, donde A es la absorbancia, ϵ es el coeficiente de extinción en unas condiciones determinadas, C es la concentración y L es el paso de luz en la medida de absorbancia. A partir de ese incremento de concentración, se calculó la cantidad de producto generado por unidad de tiempo, y se refirió a una unidad de D.O.₆₀₀. Al valor de actividad de cada clon o subclón se le restó el valor obtenido con el vector vacío correspondiente. Esos fueron considerados los valores de actividad absoluta. Los valores de actividad relativa de cada clon con cada uno de los sustratos empleados se calcularon considerando 100 % el valor de actividad con el sustrato con el que cada clon presentaba su máxima actividad. En todos los casos se realizaron tres réplicas independientes, mostrando los valores medios y las desviaciones estándar entre dichas réplicas.

Tampón fosfato Na/K: Na₂HPO₄ KH₂PO₄ 50 mM, pH 6,8

Catecol (135011, Sigma-Aldrich). El stock (10 mM) se preparó en agua bidestilada. Se conservó en oscuridad a -20 °C.

2,3-dihidroxibifenilo (17403, Sigma-Aldrich). El stock (10 mM) se preparó en etanol 96 %. Se conservó en oscuridad a -20 °C.

3-metilcatecol (M34006, Sigma-Aldrich). El stock (10 mM) se preparó en agua bidestilada. Se conservó en oscuridad a -20 °C.

4-metilcatecol (M34200, Sigma-Aldrich). El stock (10 mM) se preparó en agua bidestilada. Se conservó en oscuridad a -20 °C.

4-clorocatecol (148520, Sigma-Aldrich). El stock (10 mM) se preparó en etanol 96 %. Se conservó en oscuridad a -20 °C.

14. CITOMETRÍA DE FLUJO

Se realizaron ensayos de citometría de flujo y clasificación de células activadas por fluorescencia (FACS, Fluorescence-Activated Cell Sorting) para determinar si se podían

distinguir y separar bacterias emitiendo distintos niveles de fluorescencia, con cultivos de la estirpe MPO554 con el fósido pMPO579, induciendo la expresión del gen *gfp* desde el promotor heterólogo *psal*.

Se partió de cultivos saturados de la variante resistente a Nal y de la variante resistente a Rif de la estirpe MPO554, ambas con pMPO579, crecidos a 37 °C en medio LB con Cm 12,5 µg/ml. Se diluyeron a D.O.₆₀₀ de 0,1 y se incubaron a 37 °C hasta que llegaron a 0,3. En ese momento se les añadió arabinosa 1 mM, y al cultivo de MPO554 Nal^R también salicilato 5 mM, y se incubaron 6 h a 37 °C. Los dos cultivos se diluyeron en tampón PBS y se mezclaron en distintas proporciones, quedando las bacterias de la población inducida respecto a las de la no inducida en proporción 1:1, 1:10⁴ y 1:10⁵. Las mezclas de los cultivos se sometieron a la técnica de clasificación de células activadas por fluorescencia y se trataron de obtener las bacterias con alto nivel de fluorescencia, o sea, las que habían sido inducidas con salicilato. Las bacterias obtenidas se sembraron en placas, y se comprobó la resistencia a ácido nalidíxico o a rifampicina mediante la realización de estrías de las colonias resultantes en placas con dichos antibióticos. De este modo se identificaron las que realmente provenían del cultivo que había sido inducido, y se calculó el porcentaje de bacterias inducidas entre las bacterias seleccionadas por citometría. Los análisis se realizaron en el citómetro de flujo FACSCalibur (Becton Dickinson). Se realizaron 2 réplicas independientes del ensayo.

Tampón PBS (tampón fosfato salino, *phosphate buffered saline*) 10 X: NaCl 80 g/L, KCl 2,01 g/L, Na₂HPO₄ 14,2 g/L, KH₂PO₄ 2,72 g/L. Ajustar a pH 7,4.

15. ENSAYO COLORIMÉTRICO PARA LA DETECCIÓN DE NAFTOLES

Se realizó un ensayo colorimétrico empleando el compuesto indicador Fast Blue B (Sigma-Aldrich), adaptando el ensayo descrito por Canada y col. (2002) y Wackett y Gibson (1983). Este indicador reacciona con 1-naftol, y en menor medida con 2-naftol, dando lugar a un producto color violeta, que se puede cuantificar mediante medidas de absorbancia a 528 nm. Se ensayaron distintos clones con actividad oxigenasa hidroxilante frente al sustrato indol y el vector vacío pMPO579 en la estirpe MPO554 Nal^R. Se partió de dos cultivos de cada uno saturados en 15 ml de medio LB con Cm 12,5 µg/ml, arabinosa 1 mM y salicilato 1 mM

incubados a 30 °C. Las células se recogieron por centrifugación durante 1 minuto a 16.100 g, se lavaron y se resuspendieron en 1,5 ml de la variante del medio mínimo M9 descrita en el apartado 2.1 de Materiales y Métodos con Cm 12,5 µg/ml, arabinosa 1 mM y salicilato 1 mM, y uno de cada también con naftaleno 1 mM. Se sellaron los tubos y se incubaron a 30 °C durante una noche. Las muestras se centrifugaron durante 5 minutos a 16.100 g. Se tomaron 800 µl de cada sobrenadante y se les añadieron 40 µl de solución Fast Blue B 10 mg/ml. Se incubaron 15-20 segundos a temperatura ambiente y se les añadieron 160 µl de ácido acético glaciar. A continuación se midió la absorbancia a 528 nm en un espectrofotómetro UV-Visible Shimadzu (Kyoto, Japon). Se realizaron tres réplicas independientes, mostrando los valores medios y las desviaciones estándar entre dichas réplicas.

Solución Fast Blue B (D9805, Sigma-Aldrich): 10 mg/ml en agua. Usar recién preparada.

Naftaleno (184500, Sigma-Aldrich). El stock (3 M) se preparó N,N'-dimetilformamida (en tubo de polipropileno) y se conservó a temperatura ambiente.

16. ENSAYOS MEDIANTE HPLC

Se realizaron ensayos mediante cromatografía líquida de alta resolución (HPLC) para detectar la producción de naftoles a partir de naftaleno y de catecol a partir de fenol, adaptando los ensayos descritos por Canada y col. (2002) y Tao y col. (2004). Se ensayaron distintos clones con actividad oxigenasa hidroxilante frente al sustrato indol y el vector vacío pMPO579 en la estirpe MPO554 Nal^R. Se partió de dos cultivos de cada uno saturados en 20 ml de medio LB con Cm 12,5 µg/ml incubados a 30 °C. Se diluyeron a D.O.₆₀₀ de 0,05 en 200 ml de medio LB con Cm 12,5 µg/ml, arabinosa 1 mM y salicilato 1 mM. Al día siguiente, se recogieron las células por centrifugación durante 1 minuto a 16.100 g, se lavaron y se resuspendieron en 2 ml de la variante del medio mínimo M9 descrita en el apartado 2.1 de Materiales y Métodos con Cm 12,5 µg/ml, arabinosa 1 mM y salicilato 1 mM, y uno de cada también con naftaleno 5 mM / fenol 5 mM. Se sellaron los tubos y se incubaron a 30 °C durante 20 h. A continuación, a cada muestra se le añadieron 2 ml de cloroformo y se centrifugaron durante 10 minutos a 4.000 r.p.m. Se recogió la fase superior y se dividió el volumen en dos, se añadió a cada uno 9 volúmenes de etanol y se incubó a -20 °C durante una noche para que se produjera la precipitación. Tras esa incubación, se centrifugó durante 30

minutos a 10.000 r.p.m. a 4 °C. El sobrenadante se traspasó a un tubo y se sometió a liofilización durante una noche. Por último, las muestras con naftaleno se resuspendieron en 150 µl de metanol y las muestras con fenol en 150 µl de agua bidestilada, se centrifugó durante 5 minutos a 10.000 r.p.m. a 4 °C y se recogió el sobrenadante, que se almacenó a -80 °C hasta ser ensayado mediante HPLC.

Para los ensayo mediante HPLC se empleó el equipo HP serie 1100 (Hewlett Packard, Waldbronn, Alemania) con un detector de diodo-array utilizando una columna Microsorb-MV 100-5 C18 250 x 4,6 mm. Para los ensayos con naftaleno se empleó un método isocrático con una fase móvil consistente en 60 % de metanol y 40 % de agua bidestilada con 0,1 % de ácido fórmico, con un flujo de 0,5 ml/minuto. La longitud de onda empleada fue 219 nm. Con patrones de 1-naftol y 2-naftol se determinó el tiempo de retención en esas condiciones para cada uno, siendo de 18 y 16 minutos respectivamente. Para los ensayos con fenol se empleó un método con gradiente que fue de 30 % a 60 % de acetonitrilo en agua bidestilada con 0,1 % de ácido fórmico, durante 8 minutos, seguido de un descenso de 60 % a 30 % de acetonitrilo en el minuto 22. El flujo empleado fue de 0,5 ml/minuto. La longitud de onda empleada fue 276 nm. Con patrón de catecol se determinó el tiempo de retención en esas condiciones, siendo de 8,4 minutos. La cantidad de producto generado se obtuvo a partir del área del pico, empleando factores de conversión calculados con el patrón correspondiente. Para cada ensayo se realizaron tres réplicas independientes, mostrando los valores medios de producto generado a partir de cada sustrato y uno de los cromatogramas obtenidos en cada ensayo.

Naftaleno (184500, Sigma-Aldrich). El stock (3 M) se preparó N,N'-dimetilformamida (en tubo de polipropileno) y se conservó a temperatura ambiente.

Fenol (A1153, Applichem).

1-naftol (N1992, Sigma-Aldrich).

2-naftol (185507, Sigma-Aldrich).

Catecol (135011, Sigma-Aldrich).

RESULTADOS

1. CONSTRUCCIÓN Y VALIDACIÓN DE SISTEMAS DE EXPRESIÓN HETERÓLOGA EN VECTORES PARA METAGENOTECAS

1.1 Construcción de los sistemas de expresión

1.1.1 Construcción de vectores para metagenotecas

Una de las principales limitaciones de la metagenómica funcional, es la dificultad de expresión de los genes contenidos en el ADN metagenómico en las bacterias hospedadoras en las que se realizan los ensayos de búsqueda de actividades. Para aumentar las probabilidades de expresión del ADN metagenómico, lo que incrementa a su vez las posibilidades de descubrimiento de nuevas enzimas de interés, elaboramos nuevos vectores para la construcción de bibliotecas metagenómicas que incorporaban sistemas de expresión heteróloga basados tanto en elementos víricos como bacterianos.

Se partió del fósido comercial para la construcción de metagenotecas pCC1FOS (Epicentre), que es uno de los vectores de este tipo más utilizados. Puede albergar fragmentos de ADN metagenómico con un tamaño medio de unas 35 kb, y se empaqueta eficazmente en partículas lambda. El vector original es replicativo en *Escherichia coli* y tiene un gen que proporciona resistencia a cloranfenicol. Posee 2 orígenes de replicación. Por una parte, tiene todo el sistema de replicación proveniente del plásmido F de *E. coli* de forma que el fósido se replica desde un origen de replicación (*ori2*) de copia única, y un sistema de partición que le confiere la gran estabilidad característica del plásmido F original. El segundo origen de replicación presente en el fósido es *oriV* proveniente de RK2 (Thomas y col., 1981), que permite un alto número de copias. Para que haya replicación desde este *oriV* es necesaria la proteína de replicación TrfA, codificada por *trfA*, gen que no está presente en el fósido sino insertado en el cromosoma de la bacteria hospedadora *E. coli* EPI300-T1^R (Epicentre). La adición de una solución de inducción (arabinosa 1 mM) al cultivo de esta estirpe induce la expresión de *trfA*, incrementando el número de copias del fósido (Protocolo del kit CopyControl™ Fosmid Library Production, Epicentre). Este doble origen de replicación del fósido proporciona la ventaja de la estabilidad de la metagenoteca en *E. coli*, al estar como copia única en condiciones normales, y del aumento de la probabilidad de detección de actividades en los rastreos funcionales al aumentar la dosis génica con el alto número de copias del fósido. Además, pCC1FOS contiene el promotor del gen *10* del fago T7 (en

adelante pT7), previo al sitio de clonación del ADN metagenómico, que aprovechamos para su expresión en *E. coli*.

Al fósido se le han incorporado una serie de elementos que le proporcionan varias ventajas respecto al fósido original:

- Origen de transferencia conjugativa *oriT* de un vector de amplio espectro de hospedador, que permite que las metagenotecas construidas en este vector puedan ser transferidas eficazmente entre diferentes estirpes de *E. coli* mediante conjugación triparenteral (Figurski y Helinski, 1979).

- Promotor inducible por salicilato (*psal*) seguido del sitio de utilización de la proteína antiterminadora N del fago lambda (*nut_L*), previo al sitio de clonación del ADN metagenómico. Esto constituye un sistema de expresión heteróloga que permite la transcripción del ADN metagenómico clonado en el vector desde uno de los extremos, desde el promotor regulable *psal*. Este promotor está sujeto a control positivo por parte de NahR, el activador tipo LysR responsable de activar la transcripción de los operones de degradación de naftaleno en respuesta a la presencia del inductor salicilato, un intermediario de la ruta de degradación (Yen y Gunsalus, 1982; Schell, 1985; Schell y Wender, 1986). Por otro lado, el sitio *nut* colocado a continuación de *psal*, permite el ensamblaje del complejo de antiterminación del fago lambda mediado por la proteína antiterminadora N sobre la molécula de ARN polimerasa que ha generado el ARN mensajero (Nodwell y Greenblatt, 1991; Greenblatt y col., 1998). Para que se produzca la expresión heteróloga desde *psal* y la antiterminación de esos transcritos es necesaria una estirpe que contenga tanto el sistema de regulación de este promotor como el gen codificante de la proteína N. De este modo, la transcripción iniciada desde *psal* puede continuar insensible a las posibles señales terminadoras que pueda encontrarse en el ADN metagenómico, gracias a la antiterminación mediada por la proteína N.

- Gen *gfp* carente de su propio promotor, colocado tras el sitio de clonación del ADN metagenómico. La presencia de un gen indicador, como es el codificante de la proteína fluorescente verde (GFP), por una parte va a permitir estudiar el comportamiento del vector para validar las mejoras en la expresión realizadas, y por otra va a permitir detectar sistemas de regulación presentes en las metagenotecas construidas que respondan a una molécula determinada (tecnología SIGEX, Uchiyama y col., 2005), lo que puede ser muy útil para desarrollar biosensores que detecten esa molécula.

La incorporación de estos elementos se produjo sobre una modificación previa del fósido pCC1FOS, denominada pCC1FOS-CeuI. Esta modificación previa consistió en la incorporación de un sitio de reconocimiento de la endonucleasa *CeuI* a cada lado del sitio de clonación del ADN metagenómico (*Eco72I*) (Manuel Ferrer, comunicación personal). El objetivo era poder separar el fragmento de ADN metagenómico completo del vector, mediante la digestión con la enzima *CeuI*, debido a que su sitio de reconocimiento, de 26 pb, se iba a encontrar con muy baja frecuencia en el ADN metagenómico.

Partiendo de pCC1FOS-CeuI se construyeron tres fósidos, uno con el origen de transferencia conjugativa y el promotor *psal* seguido del sitio *nut* (pMPO571) y otros dos que además llevan el gen *gfp* sin promotor (pMPO578 y pMPO579).

Para la construcción del vector pMPO571, en primer lugar se incorporó el origen de transferencia conjugativa *oriT*. El *oriT* se amplificó mediante la reacción en cadena de la polimerasa (PCR) a partir del plásmido RP4 (Holloway, 1979) y se clonó en pCC1FOS-CeuI, generando pMPO561. Posteriormente, se introdujo un promotor *psal* seguido por el sitio *nut_L* del fago lambda. La secuencia que contenía estos dos elementos se sintetizó mediante PCR recurrente (Prodromou y Pearl, 1992) y se clonó en pBluescript II SK+ (Stratagene). A continuación se obtuvo el fragmento de interés *psal-nut_L* por restricción y se clonó en pMPO561, de manera que quedó aguas arriba y en la orientación adecuada en relación con el sitio de clonación de ADN metagenómico, dando lugar a pMPO571 (figura 6). El fósido pMPO571 tiene un tamaño de 8.827 pb, cerca de 700 pb más que el fósido original pCC1FOS (8.139 pb). Estas construcciones se explican en detalle en el apartado 5.1 de Materiales y Métodos.

Para la construcción de los vectores pMPO578 y pMPO579 se realizaron modificaciones adicionales al fósido pMPO571, consistentes en la incorporación del gen *gfp* carente de su propio promotor. Se construyeron dos variantes de esta modificación: una que mantenía el sitio de unión del ribosoma del propio gen *gfp* (pMPO578) y otra en la que se sustituyó el sitio de unión del ribosoma del gen *gfp* por el del gen *IO* del fago T7 (pMPO579), por ser este último más eficiente, lo que podría incrementar la sensibilidad de la detección de fluorescencia.

Para la construcción de pMPO578, en primer lugar se amplificó por PCR parte del fósido pCC1FOS-CeuI que contenía el gen que proporciona resistencia a cloranfenicol y parte de *redF*, y se clonó este fragmento en pBluescript II SK+ (plásmido intermedio 1). A

continuación, se obtuvo la región codificante de una variante del gen *gfp* de *Aequorea victoria* (Cormack y col., 1996) precedida por su propia Shine-Dalgarno a partir del plásmido pMPO634 (Tomás-Gallardo y col., 2009). Ese gen *gfp* tiene un sitio *Eco72I* (el mismo que el sitio de clonación para el ADN metagenómico en el fósido), que se eliminó mediante PCR solapante (Ho y col., 1989), generando un sitio *MluI* pero sin cambiar la secuencia de aminoácidos codificados por el gen. El gen *gfp* con el sitio *Eco72I* mutado se clonó en el plásmido intermedio 1, antes del gen que proporciona resistencia a cloranfenicol, generando de este modo el plásmido intermedio 2. Por último, se extrajo la construcción de interés del plásmido intermedio 2 por restricción (*gfp* con su Shine-Dalgarno sin sitio *Eco72I* y parte del gen de resistencia a cloranfenicol) y se clonó en el fósido pMPO571, sustituyendo una parte tras el sitio de clonación de ADN metagenómico, de manera que el gen que proporciona resistencia a cloranfenicol quedó regenerado. El fósido resultante pMPO578 tiene un tamaño de 9.609 pb, cerca de 1,5 kb más que el fósido original pCC1FOS (8.139 pb). Estas construcciones se explican en detalle en el apartado 5.1 de Materiales y Métodos.

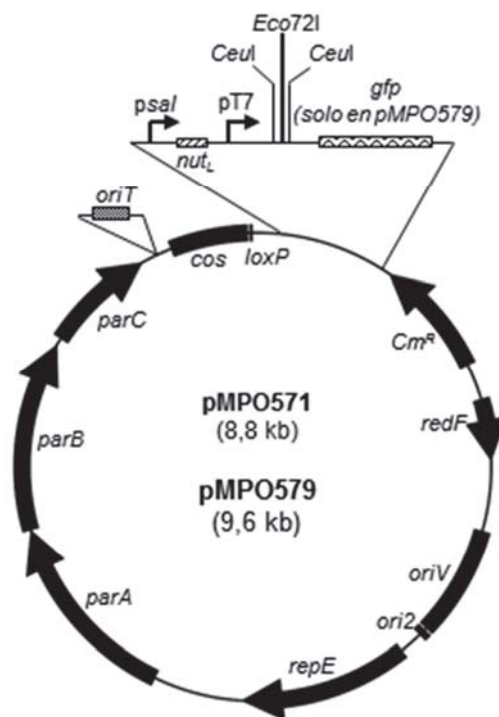


Figura 6. Fósidos derivados de pCC1FOS-CeuI: pMPO571 y pMPO579. Se resaltan los elementos incorporados en pCC1FOS-CeuI durante esta tesis, incluyendo *oriT*, *psal*, sitio *nut_L* y *gfp* sin promotor.

La construcción de pMPO579 se realizó de forma similar a la de pMPO578, pero uno de los oligonucleótidos empleados para la amplificación del gen *gfp* contenía en el extremo 5'

la secuencia Shine-Dalgarno del gen *10* del fago T7 de tal manera que esa secuencia quedó a la distancia adecuada antes de la región codificante del gen *gfp* amplificada. La estructura del fósido final, denominado pMPO579, se representa en la figura 6. Estas construcciones se explican en detalle en el apartado 5.1 de Materiales y Métodos. El fósido pMPO579 tiene un tamaño de 9.621 pb, frente a las 8.139 pb del fósido original pCC1FOS. Con ese incremento de tamaño de solo cerca de 1,5 kb se le han introducido un origen de transferencia conjugativa, un promotor *psal* seguido del sitio de utilización de la proteína antiterminadora N y un gen *gfp* sin promotor.

1.1.2 Construcción de estirpes hospedadoras

Para el aprovechamiento tanto del promotor pT7, ya presente en pCC1FOS, como del promotor *psal* seguido por el sitio de reconocimiento de la proteína antiterminadora N, fue necesaria la construcción de dos estirpes derivadas de EPI300-T1^R, que mantuviesen la capacidad de aumentar el número de copias del fósido en condiciones de inducción pero en las que estos promotores heterólogos pudieran ser utilizados para dirigir la transcripción del ADN metagenómico. Estas estirpes son MPO553, productora de la ARN polimerasa del fago T7; y MPO554, productora del activador NahR, necesario para la transcripción desde *psal*, y de la proteína antiterminadora N del fago lambda. Estos sistemas de expresión heteróloga se esquematizan en la figura 7.

La estirpe MPO553 tiene integrado en su genoma un fragmento de ADN que contiene el promotor *placUV5*, el atenuador de *nasF* y el gen *I* del fago T7, que codifica la ARN polimerasa de este fago. Esta estirpe carece del represor transcripcional LacI por lo que la transcripción a partir de *placUV5* es constitutiva. Sin embargo, la mayor parte de la transcripción iniciada en *placUV5* termina en el atenuador de *nasF*, una estructura terminadora de la transcripción procedente del operón de asimilación de nitrato de *Klebsiella pneumoniae* (Chai y Stewart, 1999) que ha sido utilizada en nuestro laboratorio para reducir los niveles basales de expresión de diversos vectores de expresión regulada (Royo y col., 2005b; Medina y col., 2011) y solo una pequeña fracción llega a transcribir el gen *I*, expresándose este en bajos niveles. De este modo, MPO553 tiene la ARN polimerasa del fago T7 además de su ARN polimerasa y, por tanto, se puede transcribir el ADN metagenómico desde el promotor pT7 presente en el fósido. La ARN polimerasa de T7 es mucho más rápida que la ARN polimerasa bacteriana y es insensible a muchos de los terminadores de la

transcripción bacterianos (Studier y Moffatt, 1986; Jeng y col., 1990), incrementando así la posibilidad de transcribir el ADN metagenómico.

Para la construcción de esta estirpe se integró en el genoma de la estirpe EPI300-T1^R un fragmento de ADN de aproximadamente 3,3 kb que contenía el promotor *placUV5*, atenuador de *nasF* y el gen *I* del fago T7. Dicha construcción se realizó en dos partes, que se integraron una tras otra en el cromosoma como se indica más adelante, al ser la construcción completa demasiado grande para realizarlo en un único paso.

Para la primera parte de la construcción, el promotor *placUV5* se obtuvo por restricción a partir del plásmido pNK736 (Simons y col., 1983), y se clonó en pBluescript II SK+, dando lugar a pMPO556. A continuación, el atenuador de *nasF* se obtuvo por restricción a partir de pMPO27 (Royo y col., 2005b) y se clonó en pMPO556, tras *placUV5*, generando de ese modo pMPO557. El gen que proporciona resistencia a kanamicina de pKD4 (Datsenko y Wanner, 2000) se amplificó por PCR y se clonó en pMPO557, a continuación del atenuador de *nasF*, produciendo así pMPO558. Para la segunda parte de la construcción, se amplificó por PCR el gen que proporciona resistencia a cloranfenicol a partir de pKD3 (Datsenko y Wanner, 2000) y se clonó en pGP1-2 (Tabor y Richardson, 1985), tras el gen *I* (ARN polimerasa de T7), generando pMPO559.

La construcción se integró en el locus *trg* de la estirpe EPI300-T1^R (en orientación opuesta a la de este gen), reemplazando prácticamente toda su secuencia codificante, utilizando una modificación del método descrito por Datsenko y Wanner, 2000. En primer lugar, se integró la parte de la construcción presente en pMPO558 en el locus *trg* de la estirpe EPI300-T1^R. A continuación, se integró el resto de la construcción, presente en pMPO559, en la estirpe EPI300-T1^R con la primera parte de la construcción ya en el genoma, reemplazando el gen que proporciona resistencia a kanamicina de pKD4. Finalmente, se eliminó el gen que proporciona resistencia a cloranfenicol de pKD3 del genoma. Los detalles de la construcción se indican en el apartado 7 de Materiales y Métodos. En la figura 7 se muestra un esquema de la integración genómica en la estirpe MPO553.

La estirpe MPO554 tiene integrado en su genoma un fragmento de ADN que contiene el gen *nahR*, su promotor *pnah*, el promotor *psal*, cercano al anterior y colocado de manera divergente, y el gen *N* del fago lambda, en el mismo sentido y a continuación de *psal*. NahR es el activador que, en presencia del inductor salicilato, activa la transcripción a partir del promotor *psal*, induciéndose así la producción de la proteína antiterminadora N. A su vez,

NahR activa en presencia de salicilato la transcripción a partir del promotor *psal* presente en el fósido pMPO579 y la proteína N impide la terminación de la transcripción que discurre a través del sitio *nut* hacia el ADN metagenómico.

Para la elaboración de esta estirpe se integró en el genoma de la estirpe EPI300-T1^R un fragmento de ADN de aproximadamente 1,9 kb que contenía el gen *nahR*, su promotor *pnah*, el promotor *psal* y el gen *N* del fago lambda. Para obtener dicha construcción, el fragmento que contenía el gen *nahR* con su promotor *pnah* y el promotor *psal* se obtuvo por restricción a partir del vector pCNB4-S2 (Cebolla y col., 2001) y se clonó en pBluescript II KS+ (Stratagene), generando así pMPO563. A continuación, se amplificó por PCR el gen que proporciona resistencia a cloranfenicol a partir de pKD3 (Datsenko y Wanner, 2000) y se clonó en pMPO563, aguas abajo del gen *nahR*, produciendo así pMPO564. Entonces, se amplificó por PCR el gen *N* a partir de ADN de lambda y se clonó en pMPO564, aguas abajo de *psal* y en la misma orientación, dando lugar de este modo a pMPO565.

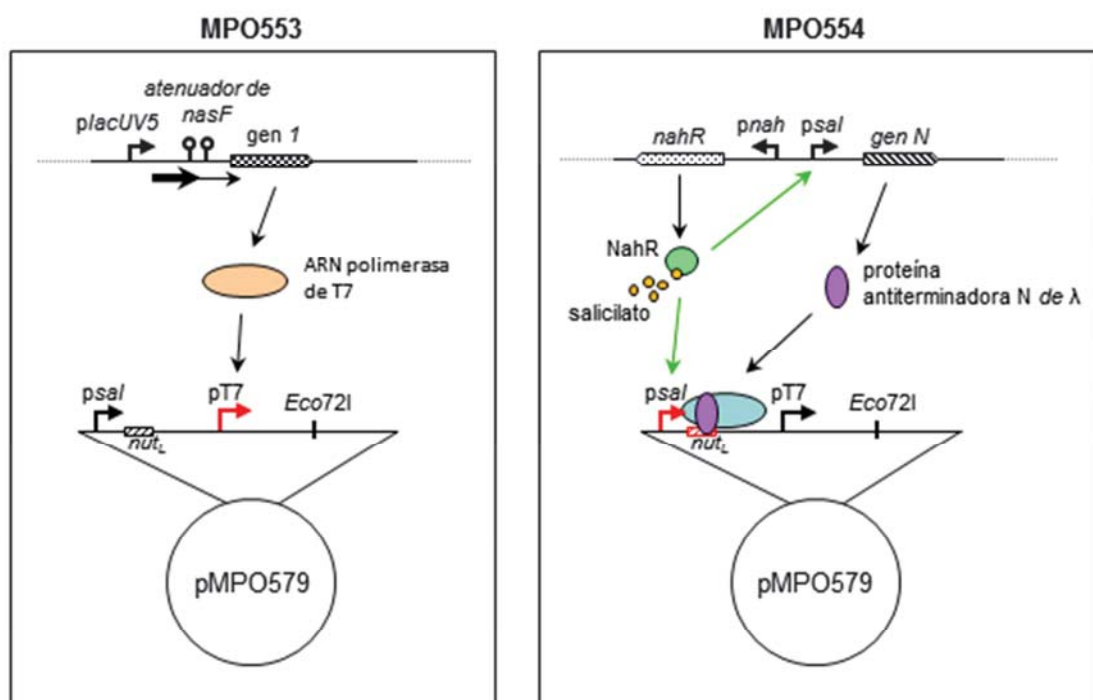


Figura 7. Estirpes especializadas derivadas de la estirpe de *E. coli* EPI300-T1^R: MPO553 y MPO554, y esquema del sistema de inducción heteróloga que permite cada una de ellas. Se indican los elementos del fragmento de ADN que ha sido integrado en el locus *trg* y el modo en el que se produciría la expresión heteróloga en el fósido en cada una de las estirpes. MPO553 es la estirpe productora de la ARN polimerasa de T7. MPO554 es la estirpe productora de la proteína antiterminadora N y de NahR (el activador necesario para la transcripción desde el promotor *psal* en presencia de salicilato por parte de la ARN polimerasa bacteriana). El sitio *Eco72I* indicado en el fósido es el sitio de clonación del ADN metagenómico.

De forma similar a la construcción de la estirpe MPO553, el inserto en pMPO565 se integró en el locus *trg* de la estirpe EPI300-T1^R (en orientación opuesta a la de este gen), reemplazando prácticamente toda su secuencia codificante. Finalmente, se eliminó el gen que proporciona resistencia a cloranfenicol de pKD3 del genoma. Los detalles de la construcción se indican en el apartado 7 de Materiales y Métodos. En la figura 7 se muestra un esquema de la integración genómica en la estirpe MPO554.

Con el objetivo de poder transferir los fósmidos de una estirpe a otra por conjugación, se seleccionaron mutantes resistentes espontáneos a los antibióticos rifampicina (Rif) y ácido nalidíxico (Nal) de EPI300-T1^R, MPO553, MPO554.

A continuación se recopilan las ventajas teóricamente aportadas con el nuevo sistema de expresión (combinación de fósrido pMPO579 y estirpes especializadas MPO553 o MPO554):

- ✓ Transferencia de metagenotecas de una estirpe de *E. coli* a otra mediante conjugación.
- ✓ Expresión heteróloga de los genes metagenómicos:
 - desde el promotor del gen *10* del fago T7, por la ARN polimerasa del fago T7, en la estirpe MPO553.
 - desde el promotor inducible por salicilato, por la ARN polimerasa bacteriana, cuya terminación en terminadores potenciales de la transcripción del ADN metagenómico estará reducida por la proteína antiterminadora N del fago lambda, en la estirpe MPO554.
- ✓ Detección de la expresión génica justo detrás del ADN metagenómico mediante GFP, que permite la identificación de sistemas reguladores a través de la tecnología SIGEX (Uchiyama y col., 2005).

1.2 Validación de los fósmidos de expresión

1.2.1 Análisis de la capacidad de conjugación del fósrido

La presencia de *oriT* en los nuevos fósmidos debería permitirles ser transferidos eficazmente por conjugación de una estirpe a otra. Para comprobarlo, se realizaron

conjugaciones triparentales (Figurski y Helinski, 1979) utilizando EPI300-T1^R portando pMPO579 como estirpe donadora, mutantes espontáneos resistentes a rifampicina y a ácido nalidíxico de EPI300-T1^R como estirpes receptoras y DH5 α portando el plásmido auxiliar pRK2013 como la estirpe auxiliar de conjugación. Como control, se intentó transferir del mismo modo el fósido pCC1FOS-CeuI, que no posee *oriT*. Y por último, se transfirió un plásmido movilizable por conjugación bien conocido, como es pBBR1MCS-3 (Kovach y col., 1995). En este caso, se transfirió el plásmido desde la estirpe DH5 α . Se calculó la frecuencia de conjugación en cada caso (estimada como la relación entre transconjugantes de la estirpe receptora y el total de viables de la estirpe receptora). El fósido pMPO579 se transfirió de manera muy eficaz a las distintas estirpes receptoras (con una frecuencia de conjugación en torno a 10^{-1}), siendo los resultados similares a los obtenidos con pBBR1MCS-3, mientras que el plásmido pCC1FOS-CeuI no pudo transferirse mediante conjugación (frecuencia de conjugación de $<10^{-7}$).

1.2.2 Medida de niveles de expresión desde el vector

Como ya se ha comentado, el nuevo sistema permitiría la expresión heteróloga del ADN presente en el fósido desde el promotor pT7 en la estirpe MPO553, y desde el promotor *psal* junto con la antiterminación de la transcripción proporcionada por la proteína antiterminadora N en la estirpe MPO554. Para comprobar que el sistema funcionaba, se midieron los niveles de expresión en distintas condiciones, utilizando como gen indicador el gen *gfp* sin promotor presente en el fósido aguas abajo y en el mismo sentido que los promotores heterólogos. Los niveles de fluorescencia detectados en las distintas condiciones, indicarían la magnitud de la transcripción que discurre por el sitio de clonación de ADN metagenómico al gen *gfp*.

En primer lugar, se compararon los niveles de fluorescencia de los fósidos pMPO578 (que mantenía el sitio de unión del ribosoma del propio gen *gfp*) y pMPO579 (en el que se sustituyó el sitio de unión del ribosoma del gen *gfp* por el del gen *10* del fago T7). Como se observa en la figura 8, en condiciones de inducción, ya fuera aumentando el número de copias del fósido añadiendo la solución de inducción de Epicentre (arabinosa) en la estirpe MPO553, o aportando además salicilato 5 mM, que induce la transcripción desde el promotor heterólogo *psal* en la estirpe MPO554, los niveles de fluorescencia alcanzados fueron alrededor de 3 veces mayores con pMPO579 que con pMPO578. Por ese motivo, se decidió

trabajar a partir de ese momento con pMPO579. La respuesta a las distintas condiciones de inducción se trata con más detalle a continuación.

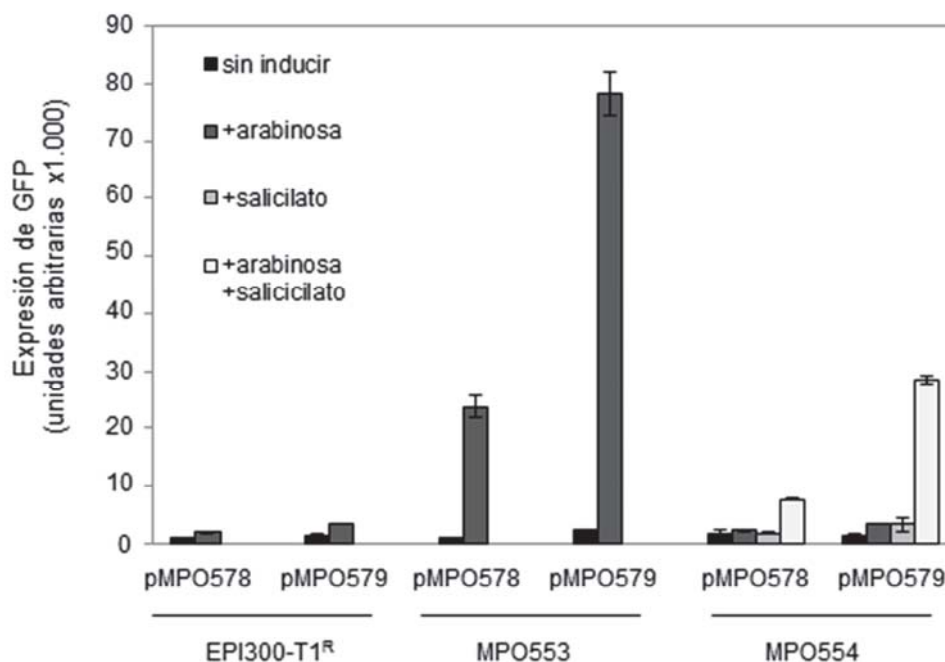


Figura 8. Expresión del gen *gfp* en los fósmidos pMPO578 y pMPO579 en las estirpes EPI300-T1^R, MPO553 y MPO554. Se comparan los niveles de expresión de GFP en ambos fósmidos, sin inducción y en las condiciones de inducción posibles en las distintas estirpes: con arabinosa (que aumenta el número de copias del fósrido) y salicilato (que permite la inducción heteróloga desde *psal* en la estirpe MPO554). Los valores representados son la media de al menos 3 réplicas. Las barras de error se corresponden con la desviación estándar de los datos.

Previamente a la realización de las siguientes pruebas, se llevaron a cabo dos nuevas construcciones que permitirían comprobar si la transcripción iniciada en los promotores heterólogos podría realizarse a través de terminadores de la transcripción, y si la proteína N realmente produce antiterminación de la transcripción.

Por un lado, se construyó el plásmido pMPO580, derivado de pMPO579 con un fragmento de ADN que contiene un terminador de la transcripción introducido en el sitio de clonación del ADN metagenómico (*Eco72I*). Para ello se obtuvo un fragmento de unas 2,5 kb correspondientes a la región intergénica entre los genes *thnL* y *thnM* de la estirpe TFA de *Sphingomonas macrogolitabida* (que contiene el terminador de *thnL*) por restricción del plásmido pIZ1110 (Andújar, 2001). Ese fragmento se clonó en pMPO579 en el sitio *Eco72I*. El terminador quedó ubicado aproximadamente a 2,3 kb aguas abajo de *psal* y a 2 kb del promotor pT7 en la orientación adecuada.

Y por otro lado, se construyó la estirpe MPO555, que es igual a MPO554, salvo por la diferencia de que MPO555 tiene un desfase del marco de lectura en el codón 60 del gen *N*, que la hace productora de una proteína *N* truncada (aunque la activación transcripcional mediante *NahR* se mantendría). Para la creación de esta estirpe se integró en el genoma de la estirpe EPI300-T1^R un fragmento de ADN que contenía la misma construcción que se integró para MPO554 pero con el gen *N* truncado: el gen *nahR*, su promotor *pnah*, el promotor *psal*, cercano al anterior y colocado de manera divergente, y el gen *N* del fago lambda truncado, en el mismo sentido y a continuación de *psal*.

Para obtener dicha construcción, se le introdujo al gen *N* en el plásmido pMPO565 descrito anteriormente un pequeño fragmento de ADN de 14 pb, como se describe en el apartado 7 de Materiales y Métodos, generando un cambio de la pauta de lectura en el codón 60 (de un total de 133 aminoácidos que posee la proteína) de dicho gen. El plásmido resultante fue pMPO575. A partir de aquí la estirpe se construyó igual que MPO554 pero empleando pMPO575 en lugar de pMPO565.

Como se comentó anteriormente, los niveles de transcripción desde los promotores heterólogos presentes en el vector modificado pMPO579 fueron medidos aprovechando el gen indicador *gfp* sin promotor localizado en ese fósido aguas abajo del sitio de clonación del ADN metagenómico.

La expresión del gen *gfp* en la estirpe EPI300-T1^R/ pMPO579 fue baja (similar a la fluorescencia intrínseca de EPI300-T1^R) y el aumento del número de copias del plásmido añadiendo arabinosa 1 mM no dio como resultado un aumento sustancial de la expresión (figura 9). En la estirpe especializada MPO553, que produce de manera constitutiva niveles bajos de ARN polimerasa de T7, los niveles de expresión también fueron muy bajos cuando el número de copias del plásmido era bajo (en ausencia de arabinosa). Sin embargo, los niveles de expresión incrementaron drásticamente (35 veces) cuando se aumentó el número de copias del plásmido mediante la adición de arabinosa. Este resultado muestra claramente que el promotor pT7 en pMPO579 es funcional y puede dirigir la transcripción del ADN metagenómico clonado en este vector, y que la estirpe especializada MPO553 produce ARN polimerasa de T7, aunque a niveles lo suficientemente bajos como para ser casi indetectable cuando el número de copias del plásmido es bajo, lo que permite suponer que esta expresión heteróloga en ausencia de arabinosa no va a ser deletérea para la bacteria hospedadora. El nivel de expresión de *gfp* desde el plásmido pMPO580, que contenía un terminador de la transcripción, fue incluso superior que el de pMPO579, lo que indica que este terminador no

puede terminar la transcripción realizada por la ARN polimerasa de T7 desde el promotor pT7 (figura 9).

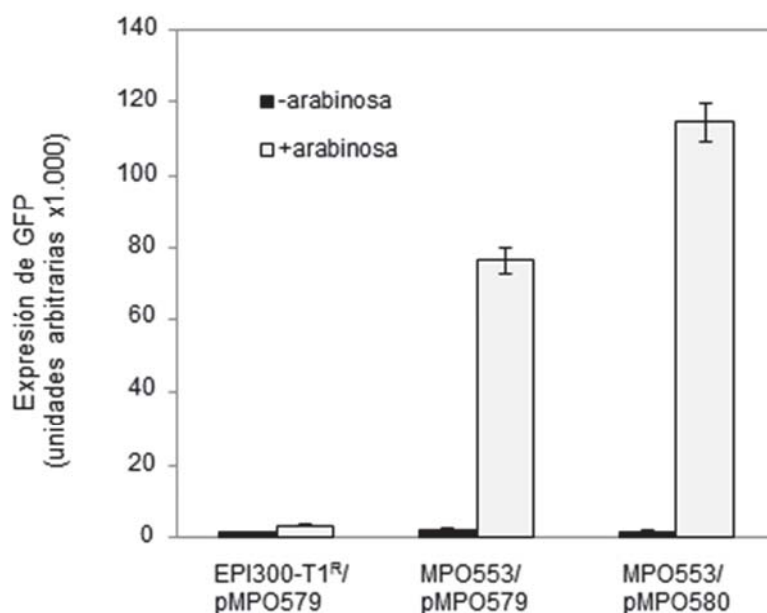


Figura 9. Expresión del gen *gfp* en el fósido sin terminador y cuando lo precede un terminador de la transcripción, en la estirpe MPO553. Por un lado se comparan los niveles de expresión de GFP en el fósido pMPO579 en la estirpe MPO553 (productora de la ARN polimerasa de T7) y en la original EPI300-T1^R. Y por otro, se muestra el efecto de un terminador entre el promotor pT7 y el gen *gfp* (pMPO580) en MPO553. Todo esto en ausencia y presencia de arabinosa (que aumenta el número de copias del fósido). Los valores representados son la media de al menos 3 réplicas. Las barras de error se corresponden con la desviación estándar de los datos.

En la estirpe especializada MPO554, que produce el activador transcripcional NahR para la expresión desde el promotor *psal*, tanto el del fósido para transcribir *gfp* como el del cromosoma para dirigir la síntesis de la proteína antiterminadora N, el nivel basal de expresión de *gfp* a partir de pMPO579 fue también bajo. El aumento del número de copias del plásmido por adición de arabinosa 1mM aumentó los niveles de expresión solo 2 veces. La inducción de la transcripción desde *psal* por adición de salicilato 5 mM resultó en un aumento moderado similar en los niveles de expresión. Sin embargo, cuando se añadió tanto arabinosa como salicilato al medio de cultivo se obtuvieron niveles de expresión altos (20 veces superiores a los niveles alcanzados sin añadir ninguno de los inductores) (figura 10). Este resultado indica que el promotor *psal* en pMPO579 es funcional y puede dirigir la transcripción del ADN metagenómico clonado en este vector, y que la estirpe especializada

MPO554 produce el activador transcripcional NahR requerido para la transcripción a partir de *psal*.

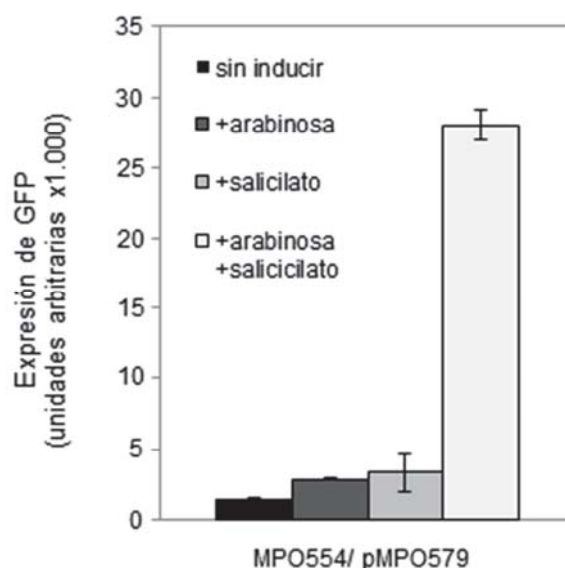


Figura 10. Expresión del gen *gfp* en la estirpe MPO554 en distintas condiciones de inducción. Se muestra la respuesta de la expresión de GFP en la estirpe MPO554 (productora de NahR, necesaria para la transcripción desde *psal*) a arabinosa (que aumenta el número de copias del fósido) y a salicilato (que permite la inducción heteróloga desde *psal*). Los valores representados son la media de al menos 3 réplicas. Las barras de error se corresponden con la desviación estándar de los datos.

Los niveles de expresión de *gfp* a partir de pMPO579 en la estirpe MPO555 en presencia de salicilato fueron algo menores que los de la estirpe MPO554 (figura 11), pero indican claramente que la transcripción a partir de *psal* puede inducirse también en esta estirpe. Sin embargo, esta expresión a partir de pMPO580, que tiene el terminador transcripcional de *thnL*, no pudo inducirse en la estirpe MPO555, que carece de la proteína N (figura 11), lo cual demuestra que el terminador de la transcripción introducido es funcional. Por el contrario, en la estirpe MPO554, que produce la proteína N, la expresión a partir de pMPO580 fue alta. Este resultado indica claramente que la transcripción a partir de *psal* en pMPO580 se termina en su mayoría antes de alcanzar el gen *gfp*, en el terminador transcripcional clonado, pero que la transcripción puede continuar hasta el gen *gfp* si también se induce la producción de la proteína antiterminadora N mediante la adición de salicilato (en la estirpe MPO554). Es interesante el hecho de que la transcripción de *gfp* desde pMPO580 en MPO554, que expresa tanto el gen *nahR* como el gen *N*, es casi tan alta como la obtenida con pMPO579 en esa estirpe, y que esta transcripción desde pMPO579 en MPO554 es incluso

algo superior a la obtenida por este fósido en la estirpe MPO555 carente de proteína N, lo que sugiere que la proteína N pueda estar mejorando ligeramente la procesividad de la ARN polimerasa incluso en ausencia de un terminador definido.

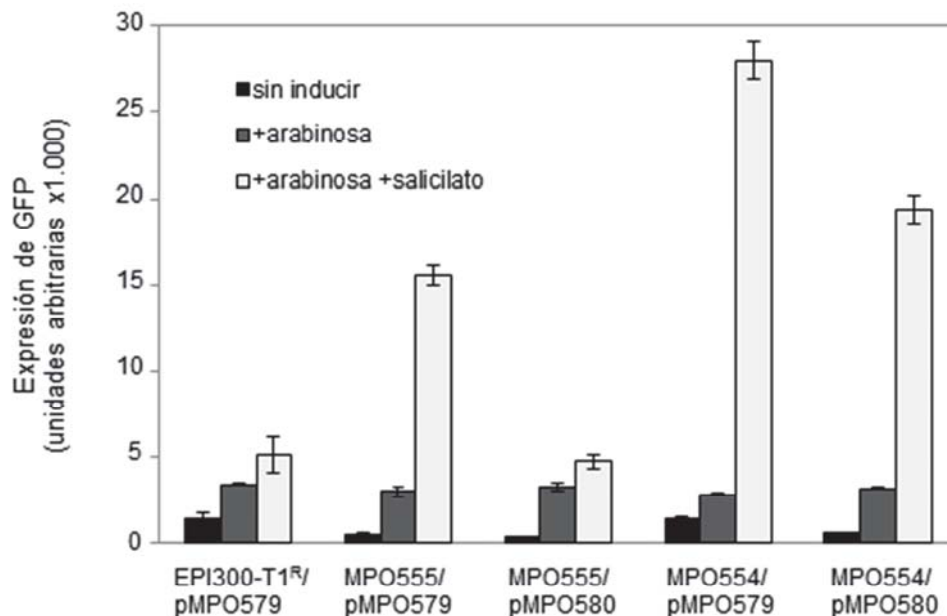


Figura 11. Expresión del gen *gfp* en el fósido sin terminador y cuando lo precede un terminador de la transcripción, en presencia y ausencia de la proteína antitermiadora N. Se muestra el efecto de un terminador entre el promotor *psal* y el gen *gfp* (pMPO580) en comparación con el fósido sin terminador (pMPO579) en la estirpe MPO554 (productora de NahR, necesaria para la transcripción desde *psal*, y de la proteína antiterminadora N), en comparación con el efecto en la estirpe MPO555 (productora de NahR y de una proteína antiterminadora N truncada). Todo esto con el fósido en bajo número de copias (sin arabinosa) y en alto número de copias sin y con salicilato. Los valores representados son la media de al menos 3 réplicas. Las barras de error se corresponden con la desviación estándar de los datos.

1.2.3 Identificación de distintos niveles de fluorescencia y enriquecimiento de clones mediante la tecnología SIGEX

Para averiguar si el fósido permitiría detectar sistemas inducibles por algún sustrato gracias al gen *gfp* sin promotor, se realizó una simulación mediante ensayos de citometría de la tecnología SIGEX (Uchiyama y col., 2005), en la que se utilizó el promotor *psal* del fósido como si fuera un posible promotor que pudiera estar presente en algún clon con ADN metagenómico y que fuera inducible por algún sustrato. Para ello, se prepararon dos cultivos con el fósido pMPO579, uno inducido con salicilato 5 mM y otro no (ambos con arabinosa

1 mM, para que el fósido se encontrara en alto número de copias), y se mezclaron en distintas proporciones, quedando las bacterias de la población inducida respecto a las de la no inducida en proporción 1:1, 1:10⁴ y 1:10⁵. En el cultivo inducido, el fósido se encontraba en la variante resistente a ácido nalidíxico (Nal^R) de la estirpe MPO554 mientras que en el no inducido en la variante resistente a rifampicina (Rif^R), de manera que posteriormente se podrían distinguir las bacterias procedentes de cada cultivo mediante comprobación de la resistencia a esos antibióticos. A continuación, las mezclas de los cultivos se sometieron a la técnica de clasificación de células activadas por fluorescencia (FACS, Fluorescence-Activated Cell Sorting) y se trataron de obtener las bacterias con alto nivel de fluorescencia, o sea, las que habían sido inducidas con salicilato. Las bacterias obtenidas se sembraron en placas, y se comprobó la resistencia a ácido nalidíxico o a rifampicina de las colonias resultantes para identificar las que realmente provenían del cultivo que había sido inducido. Como era de esperar, todas las colonias seleccionadas por citometría fueron resistentes a uno u otro antibiótico. Los resultados del ensayo mostraron que de la mezcla inicial 1:1 el 100 % de las bacterias seleccionadas por citometría eran inducidas (Nal^R), de la mezcla 1:10⁴ el 30 % resultaron ser de la población inducida y de la mezcla 1:10⁵ el 15 %, habiéndose enriquecido la mezcla unas 10⁴ veces en este último caso (tabla 4). Por tanto, extrapolado a lo que ocurriría con una metagenoteca, estos resultados indicarían que la tecnología SIGEX podría usarse para identificar sistemas reguladores que respondan a un inductor particular en las bibliotecas metagenómicas construidas en el fósido pMPO579.

| MPO554 Nal ^R / pMPO579 : MPO554 Rif ^R / pMPO579 | Porcentaje antes de la separación celular | Porcentaje después de la separación celular |
|--|--|--|
| 1:1 | 50 % | 100 % |
| 1:10.000 | 0,01 % | 30 % |
| 1:100.000 | 0,001 % | 15 % |

Tabla 4. Porcentaje de bacterias inducidas en una población antes y después de FACS. Incremento de la proporción de bacterias inducidas analizadas por citometría de flujo, comenzando con un cultivo en el que se mezclan MPO554 Nal^R/ pMPO579 inducida y MPO554 Rif^R/ pMPO579 no inducida en cocientes diferentes (1:1, 1:10.000 y 1:100.000, siendo siempre la población inducida la minoritaria).

1.2.4 Construcción de una metagenoteca de suelo de Punta de San García

Ya se había comprobado que los sistemas de expresión heteróloga incorporados al nuevo fósido y estirpes especializadas funcionaban, pero era interesante demostrar que al realizar un rastreo de una determinada actividad, estos nuevos sistemas permitían la detección de un mayor número de clones que el sistema original.

Se construyó una metagenoteca con una muestra de suelo de una zona costera que había sido contaminada en repetidas ocasiones con vertidos de hidrocarburos, en Punta de San García, Algeciras (Cádiz), según se explica en el apartado 8 de Materiales y Métodos. En la figura 12 se muestran imágenes de la zona y del suelo recogido. Se partió de 160 g de suelo, de los que se separó la fracción bacteriana por densidad con la resina Nycodenz, y se purificó el ADN, obteniéndose 24 μg de ADN. Ese ADN se trató convenientemente, quedando finalmente alrededor de 1,5 μg de ADN. Con ese material se construyó una metagenoteca usando el fósido pMPO579. Mediante siembras en placas de LB con Cm 12,5 mg/L, antibiótico al que aporta resistencia el fósido, se calculó que esta contenía aproximadamente 54.000 clones independientes. Se realizó un análisis de restricción a una serie de clones seleccionados al azar, comprobando que tenían inserto y que su tamaño superaba las 20 kb. El tamaño medio de los fragmentos de ADN metagenómico clonados se estimó en 35 kb aproximadamente, de manera que la metagenoteca contenía alrededor de 1,89 Gb de ADN metagenómico. La metagenoteca se mantuvo en la estirpe EPI300-T1^R congelada a -80 °C en masa.



Figura 12. Imágenes de la zona costera de la que se tomó la muestra de suelo para la construcción de la metagenoteca.

1.2.5 Rastreos comparativos de resistencias a antibióticos con y sin la utilización de los sistemas de expresión heteróloga

Para comprobar si el nuevo fósido y las estirpes especializadas permitirían la detección de un mayor número de clones que el sistema original, se empleó la metagenoteca de Punta de San García, construida en el fósido pMPO579. Se realizaron rastreos comparativos en la estirpe original EPI300-T1^R, en la que no se podrían aprovechar los sistemas de expresión heteróloga; en MPO553, en la que se daría expresión heteróloga desde el promotor pT7 por la ARN polimerasa de T7; y en MPO554, en la que se daría expresión desde *psal* y esa transcripción sería capaz de superar los terminadores gracias a la proteína antiterminadora N del fago lambda.

Los suelos son ricos en microorganismos que producen antibióticos betalactámicos, incluso en ausencia de presión antropogénica (Demanèche y col., 2008; Allen y col., 2009). Por eso, y por ser un rastreo sencillo (mediante selección directa en placas con el antibiótico), lo que nos permitiría realizar las repeticiones necesarias, se eligió la resistencia a antibióticos, concretamente a carbenicilina (Cb), como la actividad de interés para ser identificada en los clones de la metagenoteca transferidos a las distintas estirpes. Resultaba interesante estimar la abundancia de genes que proporcionaban esta resistencia en una muestra de suelo y cuántos de estos genes no podrían haberse detectado en bibliotecas metagenómicas convencionales sin los sistemas de expresión heteróloga.

En primer lugar, se realizó una búsqueda directa con el cultivo de EPI300-T1^R que contenía la metagenoteca de clones resistentes a carbenicilina (Cb^R), sembrando en placas de LB que contenían 100 mg/L de carbenicilina y arabinosa 1 mM (para que el fósido estuviera en alto número de copias). Esto resultó en el aislamiento de solo unas pocas colonias diminutas tras 2 días de incubación con una frecuencia mucho menor que 1/54.000. Esto indica que ninguno de los clones de la metagenoteca confiere eficientemente resistencia a Cb 100 mg/L a EPI300-T1^R.

Por otro lado, se transfirió la biblioteca metagenómica mediante conjugación triparental desde EPI300-T1^R a los resistentes espontáneos a ácido nalidíxico y a rifampicina de las estirpes EPI300-T1^R, MPO553 y MPO554, empleando DH5 α / pRK2013 como estirpe auxiliar. Se obtuvieron unas frecuencias de conjugación altas, siempre superiores al 6 %, independientemente de la estirpe receptora. Estos datos indican que los clones metagenómicos pueden transferirse a otras estirpes con una frecuencia similar a la del vector pMPO579, a

pesar de las aproximadamente 35 kb de ADN metagenómico que porta cada clon. Se sembraron diluciones del parche de la conjugación directamente en placas de LB que contenían 100 mg/L de carbenicilina, el antibiótico al que proporciona resistencia la estirpe receptora (ácido nalidíxico o rifampicina) y arabinosa 1 mM. Cuando la estirpe era MPO554, se realizaron además siembras en placas que también contenían salicilato 1 mM.

La tabla 5 muestra el número de clones resistentes a carbenicilina (Cb^R) detectados por cada millón de transconjugantes que habían recibido un clon metagenómico, para cada una de las estirpes. Se obtuvo un número aproximadamente 6 veces mayor de transconjugantes resistentes a carbenicilina con las estirpes especializadas, que permitían la expresión heteróloga bien a partir del promotor pT7 o bien a partir del promotor *psal* con la antiterminación de la proteína N, en comparación con la estirpe EPI300-T1^R original. Además, en el caso de MPO554, este aumento dependía de la adición de salicilato 1 mM, que inducía la activación de la transcripción a partir de *psal* por NahR y la antiterminación por parte de la proteína N.

| | EPI300-T1 ^R | MPO553 | MPO554 | |
|-----------------------------|------------------------|--------|--------------|--------------|
| | | | - Salicilato | + Salicilato |
| Receptores Rif ^R | - | 64 | - | 28 |
| Receptores Nal ^R | 12 | 60 | 15 | 67 |

Tabla 5. Número de clones resistentes a carbenicilina detectados en las diferentes estirpes por cada millón de transconjugantes.

Para cada estirpe receptora se analizaron más de un millón de transconjugantes, por lo que la probabilidad de que en cada una hayamos analizado todos los clones de la metagenoteca es prácticamente del 100 % (según adaptación de la fórmula de Clarke y Carbon, 1976).

Por lo general, se detectaron más clones en las estirpes resistentes a Nal que en las resistentes a Rif, sobre todo en ausencia de expresión heteróloga.

Una vez comprobado que con las estirpes especializadas aparecen clones resistentes a Cb con mayor frecuencia, lo siguiente era saber si además, las estirpes especializadas permiten detectar más clones distintos que la original, es decir, si hay clones que expresen la actividad de interés en las estirpes especializadas y no en la original. Para ello, se analizó el

patrón de restricción de los fósidos que conferían resistencia a carbenicilina presentes en 100 transconjugantes: 10 de EPI300-T1^R, 40 de MPO553 y 50 de MPO554. El resultado indicó que en estos rastreos se aislaron un total de 6 clones diferentes (figura 13). Todos los transconjugantes de EPI300-T1^R Cb^R tenían el mismo fósido, ETN1, cuyo patrón de restricción coincidía con el de los clones aislados directamente desde la metagenoteca (en EPI300-T1^R, sin conjugación). Este fósido también se encontró entre los transconjugantes derivados de las estirpes especializadas. El uso de MPO553, la estirpe que produce ARN polimerasa de T7, permitió la identificación de 3 fósidos adicionales, TN2, TN3 y TN4. Estos fósidos también se identificaron entre los transconjugantes derivados de MPO554, la estirpe especializada que permitía la expresión desde *psal* y la antiterminación de N. Y esta estirpe permitió la identificación de otros 2 fósidos adicionales, N5 y N6, que no se identificaron usando las demás estirpes. Por tanto, de estos resultados se extrae la conclusión de que las estirpes especializadas permiten detectar un número mayor de clones distintos resistentes a Cb.

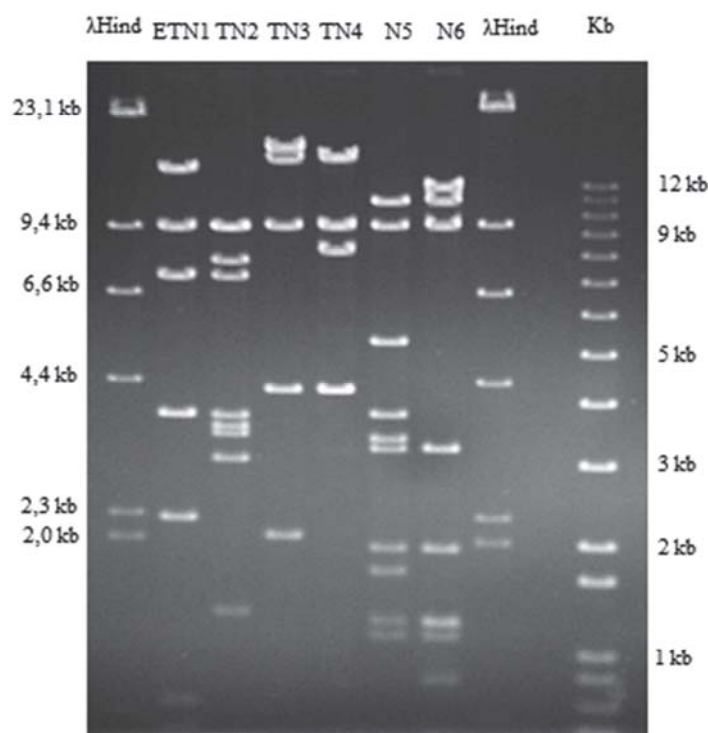


Figura 13. Patrones de restricción de los 6 clones que confieren resistencia a carbenicilina. Los fósidos se digirieron con *Bam*HI, que realiza cortes en el fósido a ambos extremos del inserto (la banda de aproximadamente 9,3 kb es fósido sin inserto). Los marcadores de peso molecular de ADN empleados son ADN de lambda digerido con *Hind*III (indicado como λ Hind) y Kb Plus DNA Ladder (Invitrogen) (indicado como Kb).

Para confirmar la capacidad de cada fósido de conferir Cb^R a cada una de las estirpes receptoras, y descartar así que la ausencia de algunos de los clones en algunas de las estirpes fuera casualidad, se transfirieron los 6 fósidos de vuelta a las estirpes EPI300-T1^R Rif^R y Nal^R. Desde ahí, se volvieron a pasar por conjugación triparental a las 3 estirpes resistentes espontáneas a Nal y a Rif, empleando la estirpe auxiliar DH5 α / pRK2013.

| | | Receptores | | | |
|--|------|---|-------------------------|-------------------------|-------------|
| | | EPI300-T1 ^R Rif ^R | MPO553 Rif ^R | MPO554 Rif ^R | |
| | | | | -Salicilato | +Salicilato |
| Donador: | ETN1 | 10 ⁻⁵ | 1 | 10 ⁻⁵ | 0,8 |
| EPI300-T1 ^R Nal ^R | TN2 | <10 ⁻⁵ | 0,8 | <10 ⁻⁵ | 1 |
| | TN3 | 10 ⁻⁵ | 0,8 | <10 ⁻⁵ | 1 |
| | TN4 | <10 ⁻⁵ | 1 | <10 ⁻⁵ | 0,9 |
| | N5 | <10 ⁻⁵ | <10 ⁻⁵ | 10 ⁻⁴ | 1 |
| | N6 | <10 ⁻⁵ | 10 ⁻⁵ | <10 ⁻⁵ | 0,9 |
| | | | | | |

| | | Receptores | | | |
|--|------|---|-------------------------|-------------------------|-------------|
| | | EPI300-T1 ^R Nal ^R | MPO553 Nal ^R | MPO554 Nal ^R | |
| | | | | -Salicilato | +Salicilato |
| Donador: | ETN1 | 0,1 | 0,07 | 0,3 | 0,5 |
| EPI300-T1 ^R Rif ^R | TN2 | <10 ⁻⁵ | 0,05 | <10 ⁻⁵ | 1 |
| | TN3 | <10 ⁻⁵ | 0,2 | <10 ⁻⁵ | 1 |
| | TN4 | <10 ⁻⁵ | 0,2 | <10 ⁻⁵ | 0,9 |
| | N5 | <10 ⁻⁵ | <10 ⁻⁵ | 0,2 | 1 |
| | N6 | <10 ⁻⁵ | <10 ⁻⁵ | 0,1 | 1 |
| | | | | | |

Tabla 6. Transferencia por conjugación de los clones resistentes a carbenicilina a las distintas estirpes. Los valores representan la frecuencia de transconjugantes resistentes a carbenicilina respecto a los transconjugantes que han recibido el fósido.

Los 6 fósidos se transfirieron eficazmente a cada una de las 3 estirpes receptoras, con frecuencias siempre superiores al 5 %. Tal como se muestra en la tabla 6, solo el fósido ETN1 fue capaz de conferir Cb^R a la estirpe EPI300-T1^R y solo a la variante Nal^R. Además, la expresión del gen que confiere resistencia a carbenicilina en ese fósido está aparentemente limitada, ya que solo una fracción minoritaria de los transconjugantes mostró resistencia y pudo generar una colonia. En las variantes Nal^R de las estirpes especializadas también se detectó que la expresión de la Cb^R desde este fósido no era lo suficientemente buena como

para que la mayoría de los transconjugantes fuesen Cb^R aunque en la estirpe MPO554 este fósido aportó resistencia a Cb más eficientemente en presencia de salicilato. La ventaja del uso de los sistemas de expresión heteróloga en este clon se puso mucho más claramente de manifiesto con el uso de las variantes Rif^R en las que se observó que la expresión propia no confirió Cb^R mientras que la expresión heteróloga con cualquiera de los dos sistemas permitió conferir Cb^R a la mayoría de los transconjugantes. La expresión heteróloga basada en la ARN polimerasa de T7 permitió que los fósidos ETN1, TN2, TN3 y TN4 proporcionaran Cb^R a la estirpe MPO553. A su vez, la expresión heteróloga basada en *psal* y la antiterminación N permitió que además de estos, los clones N5 y N6 confiriesen también Cb^R a la estirpe MPO554 en presencia de salicilato. Estos resultados, usando los fósidos aislados, concuerdan completamente con los obtenidos usando toda la biblioteca metagenómica e indican claramente que al menos 6 clones de la biblioteca contienen genes que aportan resistencia a carbenicilina, y que solo uno de estos 6 fósidos expresa el gen de resistencia por sí mismo lo suficiente para conferir Cb^R en la estirpe EPI300-T1^R.

Por tanto, estos resultados muestran el potencial de la utilización de estos sistemas de expresión heteróloga para detectar clones que contienen genes con funciones de interés.

2. CARACTERIZACIÓN DE CLONES CON RESISTENCIAS A CARBENICILINA

En la metagenoteca de San García se identificaron 6 clones distintos con capacidad de resistir 100 mg/L de carbenicilina, denominados ETN1, TN2, TN3, TN4, N5 y N6. Estos clones presentaron distinto patrón de restricción (figura 13), lo que indica que su secuencia de ADN metagenómico no es la misma.

Recapitulando, el clon ETN1 es el único que aporta resistencia a la estirpe EPI300-T1^R, es decir, el único en el que el gen que proporciona la resistencia a Cb es capaz de expresarse por sí mismo. Los otros 5 clones requieren expresión heteróloga para dar resistencia a Cb. Los clones TN2, TN3 y TN4 expresan la resistencia tanto desde el promotor pT7 (en la estirpe MPO553) como desde el promotor *psal* asociado a la antiterminación de la proteína N (en la estirpe MPO554 en presencia de salicilato). Y los clones N5 y N6 requieren que la expresión se dé desde el promotor *psal* con la antiterminación de la proteína N. Con objeto de establecer los determinantes genéticos que conferirían esta resistencia así como su organización en el contexto genómico de cada clon para entender la forma en que se había producido su expresión heteróloga, se procedió a la caracterización de todos los clones.

2.1 Secuenciación de los clones con resistencias a carbenicilina

Los 6 clones con resistencias a Cb se secuenciaron individualmente por el sistema de secuenciación masiva de Roche (454). En dos de ellos posteriormente hubo que unir *contigs* o completar la secuencia mediante algunas reacciones de secuenciación, pero finalmente se obtuvo la secuencia metagenómica completa de todos ellos. Esas secuencias se depositaron en la base de datos GenBank con los códigos de acceso: JX406851 (ETN1), JX406852 (TN2), JX406853 (TN3), JX406854 (N5) y JX406855 (N6).

El análisis de secuencia de cada fósido reveló varios genes que podrían conferir resistencia a carbenicilina, cuya localización y orientación se muestran en la figura 14. Resultó que, por un lado, los clones TN3 y TN4 comparten secuencia, de hecho el inserto de TN4 estaría incluido en el de TN3, en la misma orientación. Y por otro lado, los clones N5 y N6 también comparten parte del inserto, pero en sentido contrario. Estas parejas de clones comparten los genes identificados como posibles responsables de la resistencia a Cb.

Las secuencias de las posibles proteínas relacionadas con la resistencia a carbenicilina de estos clones se recogen en el Anexo.

2.2 Betalactamasa de clase C y bomba de eflujo del clon ETN1

En el clon ETN1 se identificó un gen codificante de una betalactamasa de clase C (*blaC*) que se encuentra en orientación opuesta a los promotores heterólogos. Las proteínas homólogas a las que más se parece esta betalactamasa pertenecen a diferentes especies de *Pseudomonas*, presentando valores de identidad de 70-89 % con decenas de proteínas. Al menos las anotaciones de las proteínas más similares se han realizado por homología de secuencia, sin que conste que se haya confirmado la actividad. Inmediatamente aguas abajo del gen *blaC* se encuentra una pauta abierta de lectura (*orf*, *open reading frame*) que codificaría una proteína de 103 aminoácidos y que contiene 4 nucleótidos solapados con el extremo 3' del gen de la betalactamasa. Ese solapamiento podría ser indicativo de acoplamiento traduccional, sugiriendo que la pequeña *orf* podría estar relacionada con la función de la betalactamasa. Además, esa *orf* está conservada, sobre todo en distintas especies de *Pseudomonas*, asociada también a genes codificantes de betalactamasas de clase C. Sin embargo, hasta el momento no se ha descrito la existencia de proteínas asociadas a las betalactamasas que puedan contribuir a su función de resistencia al antibiótico.

Para tratar de averiguar si la pequeña *orf* que sigue al gen codificante de la betalactamasa estaba implicada en la resistencia, se subclonó el gen *blaC* sin y con la pequeña *orf* en pMPO579, de manera que quedaba a continuación y en el mismo sentido que *psal*. Los insertos se obtuvieron mediante amplificación por PCR como se indica en el apartado 5.2 de Materiales y Métodos. El subclón con el gen *blaC* sin la pequeña *orf* se denominó pMPO584 y el subclón con el gen *blaC* seguido de la pequeña *orf* pMPO585.

Se comparó el crecimiento de ambos subclones, junto con el de ETN1 y pMPO579 (como controles positivo y negativo respectivamente), en la estirpe MPO554 NaI^R, en presencia de Cb 100 mg/L y arabinosa 1 mM, y en las mismas condiciones pero añadiendo también salicilato 1 mM. Aunque algo más tolerantes que la estirpe con el vector vacío (pMPO579), los subclones pMPO584 y pMPO585 no proporcionan un buen crecimiento en Cb 100 mg/L, lo que indica que la betalactamasa de clase C no es capaz por sí sola de conferir toda la resistencia a Cb proporcionada por el fósforo original (ETN1 permite crecer en

presencia de carbenicilina bastante mejor que los subclones). No se observó diferencia entre los 2 subclones, por lo que no se aclaró nada en cuanto a la pequeña *orf* conservada que sigue al gen *blaC*. Sin embargo, parece que en el fósido ETN1 hay algo que falta en los subclones que proporciona resistencia a Cb, o que al menos participa en ella.

En ETN1 también se identificaron los genes que codifican los 3 componentes de una bomba de eflujo de la familia de permeasas de resistencia-nodulación-división celular (RND) en la misma dirección que los promotores heterólogos, posiblemente formando un operón: una proteína transportadora de membrana interna, una proteína periplásmica o de fusión de membrana y una proteína de membrana externa o porina. Ese fósido también contiene un cuarto componente que codifica una proteína adicional de membrana externa. Las 4 proteínas mencionadas cuentan con decenas de secuencias homólogas pertenecientes al género *Pseudomonas*, con identidades superiores al 80 % (hasta 96-99 %). Se ha descrito que las bombas de eflujo contribuyen a la tolerancia de bacterias a diferentes compuestos tóxicos tales como compuestos aromáticos o antibióticos (Nikaido, 2011), y más concretamente antibióticos betalactámicos (Poole, 2005), por lo que es posible que esta bomba de eflujo de ETN1 pueda contribuir también a la resistencia a Cb, sobre todo si se tienen en cuenta los resultados obtenidos con los clones TN3 y TN4 (ver más adelante). La expresión heteróloga de los genes de esta bomba, localizados a unas 14 kb de los promotores heterólogos, es lo que explicaría que este clon confiera una mejor resistencia a Cb cuando los sistemas de expresión heterólogos están activos.

2.3 Betalactamasa de clase B del clon TN2

En el clon TN2 se localizó el gen codificante de una betalactamasa de clase B (*blaB*) 6 kb aguas abajo de los promotores heterólogos, en la misma orientación. La secuencia con la que esta proteína presenta una mayor identidad (67 %) está anotada como proteína hipotética, seguida de muchas otras betalactamasas de clase B con identidades de en torno al 50 %. Estas proteínas pertenecen a distintas especies, pero son mayoritarias las del género *Acinetobacter*. Con menor identidad, también se encuentran proteínas homólogas de organismos no cultivados: LRA-8, Amo1 y LRA-2 (con identidades del 38 al 32 % y E values todavía significativos). En los 3 casos se trata de betalactamasas de clase B identificadas en análisis funcionales con antibióticos betalactámicos de metagenotecas de un suelo de Alaska (Allen y

col., 2009), en el caso de LRA-8 y LRA-2, y de un suelo de un manzano de Wisconsin (Donato y col., 2010), en el caso de Amo1. También se identificó el gen codificante de un regulador de betalactamasa (*blaR*), concretamente de un activador transcripcional tipo LysR, justo aguas arriba de *blaB*, en orientación opuesta, lo que sugiere que este es un sistema inducible de resistencia a antibióticos betalactámicos. A pesar de la existencia del activador *blaR* en el clon, este gen *blaB* no puede inducirse y expresarse de manera autónoma en *E. coli* en las condiciones de crecimiento, ya que el antibiótico betalactámico utilizado es carbenicilina, un antibiótico diseñado para ser eficaz incluso frente a bacterias con resistencias inducibles, que se caracteriza por no atravesar la membrana interna, (Li y col., 1994) evitando así su reconocimiento por parte del regulador intracitoplásmico. Sin embargo, *blaB* sí parece poder ser transcrito desde cualquiera de los promotores heterólogos presentes en el vector.

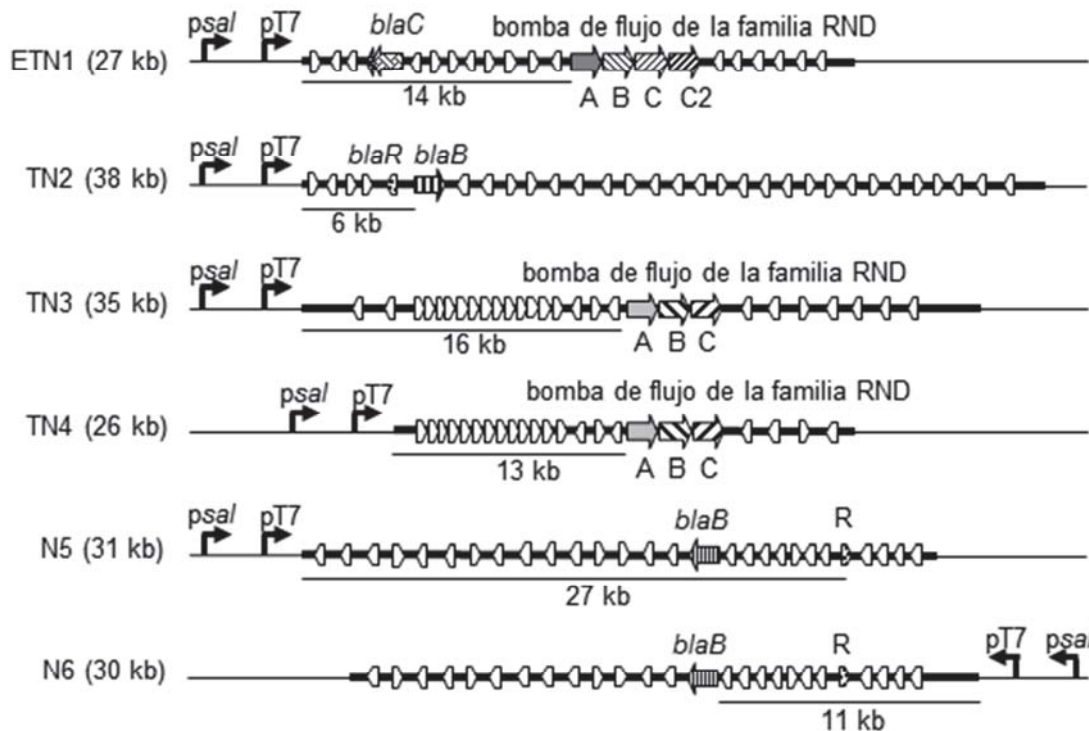


Figura 14. Esquema de los clones que confieren resistencia a carbenicilina. Se indican las pautas abiertas de lectura y sus orientaciones, resaltándose aquellas que podrían estar involucradas en la resistencia a carbenicilina. Las abreviaturas de la figura significan lo siguiente: *blaC*: gen codificante de betalactamasa de clase C; *blaB*: gen codificante de betalactamasa de clase B; *blaR*: gen codificante de regulador de betalactamasa; R: gen codificante de regulador transcripcional putativo. Los genes no están representados a escala.

2.4 Bomba de eflujo de los clones TN3 y TN4

La secuencia metagenómica del clon TN4 está incluida en la del clon TN3, y se encuentran en la misma orientación. Ninguno de estos fósmidos contenía genes que parecieran codificar una enzima betalactamasa. Los únicos genes que parecía que pudieran estar involucrados en la resistencia a Cb en estos fósmidos eran los codificantes de los componentes de una bomba de eflujo tipo RND y se localizaban de forma contigua y posiblemente formando un operón en la misma orientación de los promotores heterólogos, a 16 kb o 13 kb, en los clones TN3 y TN4 respectivamente. Esas proteínas cuentan con bastantes secuencias homólogas con identidades superiores al 80 % (hasta 96-99 %), pertenecientes al género *Achromobacter*, las más cercanas, seguidas de *Bordetella* y algunas otras.

Para determinar si la bomba era realmente la responsable por sí misma de la resistencia a Cb se subclonaron exclusivamente los genes codificantes de la bomba de eflujo en el mismo fósrido pMPO579. El fragmento de interés con los genes de los 3 componentes de la bomba se obtuvo por restricción del fósrido TN4 como se indica en el apartado 5.2 de Materiales y Métodos, clonándose en la misma diana en la que se inserta el ADN metagenómico, de manera que los genes de la bomba quedaron tras los promotores heterólogos, en su misma orientación. El subclón resultante se denominó pMPO586.

Se comparó el crecimiento del subclón con el de TN3, TN4 y pMPO579 (como controles positivos y negativo respectivamente), en la estirpe MPO554 NaI^R, en presencia de Cb 100 mg/L y arabinosa 1 mM, y en las mismas condiciones pero añadiendo también salicilato 1 mM. El subclón es capaz de proporcionar resistencia a Cb 100 mg/L en presencia de salicilato pero no en su ausencia, tal y como ocurre con los fósridos originales, creciendo incluso algo mejor que con TN3 y TN4. Esto indica que esta bomba de eflujo es capaz de proporcionar por sí misma resistencia a 100 mg/L de carbenicilina si se expresa de manera adecuada desde un promotor heterólogo en *E. coli*.

2.5 Betalactamasa de clase B de los clones N5 y N6

Los clones N5 y N6 comparten la mayor parte de sus insertos de ADN metagenómico, que se encuentran en orientaciones opuestas en cada fósrido. Los 2 clones permiten a la

estirpe MPO554 Nal^R crecer con Cb 100 mg/L en presencia de arabinosa y salicilato, aunque también se da algo de crecimiento sin salicilato, sobre todo con N5. Esto cuadra con lo que se observa en la tabla 6. Es decir, con salicilato crecen mejor, pero sin salicilato también pueden acabar creciendo. Parece que la expresión autónoma del gen responsable de la resistencia no es suficientemente eficiente, y la transcripción heteróloga desde el promotor *psal* ayuda a proporcionar la resistencia completa a Cb 100 mg/L en *E. coli*. En ambos clones se identificó un gen codificante de una betalactamasa de clase B, denominado *blaB* en la figura 14. La proteína codificada por este gen presenta una identidad $\leq 50\%$ con otras proteínas de las bases de datos, entre las que se encuentran varias betalactamasas de clase B identificadas en metagenotecas de un suelo de Alaska (Allen y col., 2009), que se comprobó que conferían resistencia a antibióticos betalactámicos en *E. coli*: LRA-19, LRA-12, LRA-17, LRA-3, LRA-7 y LRA-9 (con identidades entre el 43 y 45 %). Estas 6 proteínas junto con las 2 que presentan homología con la betalactamasa del clon TN2, constituyen el total de betalactamasas de clase B identificadas por Allen y col. en ese trabajo.

En N6, el gen codificante de la betalactamasa de clase B comentado, se localiza a 11 kb aguas abajo de los promotores heterólogos y se transcribe en la misma orientación, lo que explicaría la mayor resistencia que confiere el fósido cuando el sistema heterólogo de antiterminación está activo. Curiosamente, en N5 ese gen se encuentra localizado en la orientación opuesta, aunque N5 también requiere inducción con salicilato para conferir o, al menos, mejorar bastante la resistencia a Cb. La única explicación posible es que la transcripción heteróloga desde *psal* en N5 tenga como resultado la expresión de un gen distinto a *blaB* y en orientación contraria a este, que también pudiera participar en la resistencia a Cb en N5. Solo hay otras 6 *orfs* predichas en este clon que cumplan estos requisitos, es decir, que estén en la misma orientación que *psal*. Los análisis BLAST de las secuencias predichas de aminoácidos de 5 de ellas indicaron similitud a proteínas no caracterizadas o hipotéticas por lo que no proporcionaron pista alguna en relación con su potencial función. Sin embargo, existe una pauta abierta de lectura, denominada R en la figura 14, cuyo análisis BLAST mostró similitud a reguladores transcripcionales putativos. Esta proteína tiene un dominio similar a dominios HTH (*helix-turn-helix*, hélice-giro-hélice), capaces de unirse a ADN, que están presentes en muchas proteínas involucradas en la regulación de la expresión génica. Este candidato a gen activador, que podría a su vez activar la transcripción de *blaB*, se encuentra localizado 27 kb aguas abajo del promotor *psal* de manera que la resistencia a carbenicilina conferida por el clon N5 podría estar indicando una

expresión heteróloga a gran distancia y a través de al menos 5 terminadores de la transcripción, lo que pondría de manifiesto el gran potencial que el sistema de expresión inducible por salicilato tiene para expresar genes de bibliotecas metagenómicas. Alternativamente, podría haber un gen de algún pequeño ARN regulador no detectado en nuestro análisis de secuencia, que pudiera ser procesado adecuadamente a partir del transcrito primario heterólogo generado desde *psal* y que su función fuese promover la expresión postranscripcional de *blaB*.

3. IDENTIFICACIÓN DE ACTIVIDADES DE INTERÉS

3.1 Construcción de una metagenoteca de suelo con historia de contaminación

Debido al interés del grupo de investigación en actividades relacionadas con la degradación de contaminantes orgánicos, para la construcción de una metagenoteca se eligió un suelo con historia reciente de contaminación. Se tomó una muestra de suelo de una pila de refinería de la industria petroquímica (Palos de la Frontera, Huelva). En la figura 15 se muestran imágenes de la zona y del suelo recogido. Ese suelo había sido contaminado con hidrocarburos procedentes de dicha industria, y sometido a un proceso de descontaminación (consistente en una adición inicial de carbono y posteriores volteos mensuales con palas durante 6 meses) que se consideró finalizado aproximadamente un mes antes de la toma de la muestra. En ese momento se esperaba que el suelo estuviera enriquecido en bacterias con capacidad de degradación de contaminantes. Se cogió la muestra de suelo cribado eliminando previamente el suelo más superficial. La metagenoteca de pila de refinería de la industria petroquímica se construyó usando como vector el fósido pMPO579, según se explica en el apartado 8 de Materiales y Métodos. Se partió de 40 g de suelo, de los que se separó la fracción bacteriana por densidad con la resina Nycodenz, y se purificó el ADN, obteniéndose aproximadamente 25 μg de ADN metagenómico. Ese ADN se trató convenientemente, quedando finalmente unos 4 μg de ADN. Con ese material se construyó una metagenoteca de unos 185.000 clones independientes, cálculo que se realizó a partir de siembras en placas de LB con Cm 12,5 mg/L, antibiótico al que aporta resistencia el fósido. Se realizó un análisis de restricción a una serie de clones seleccionados al azar, comprobando que tenían inserto y



Figura 15. Imágenes de la pila de la refinería de la industria petroquímica de la que se tomó la muestra de suelo para la construcción de la metagenoteca.

que su tamaño superaba las 20 kb. El tamaño medio de los fragmentos de ADN metagenómico clonados fue aproximadamente de 35 kb, de manera que contenía unas 6,5 Gb de ADN metagenómico. La metagenoteca se mantuvo en la estirpe EPI300-T1^R congelada a -80 °C en masa.

3.2 Rastreo de dioxigenasas extradiólicas

Se buscaron clones con actividad dioxigenasa extradiólica en la metagenoteca procedente de la pila de refinería de la industria petroquímica. Se decidió realizar los rastreos utilizando la estirpe MPO554 Nal^R, con objeto de maximizar la identificación de estos clones, ya que en esa estirpe se daría expresión heteróloga desde el promotor *psal* del fósido y esa transcripción podría superar los posibles terminadores que eventualmente pudiera encontrarse gracias a la proteína antiterminadora N del fago lambda. Además, fue la estirpe que permitió aislar un mayor número de clones resistentes a carbenicilina (apartado 1.2.5 de Resultados). Como la metagenoteca se mantenía en la estirpe EPI300-T1^R, previo a la búsqueda de clones con la actividad se realizó una conjugación triparental con la estirpe EPI300-T1^R con la metagenoteca como donadora, MPO554 Nal^R como receptora y DH5 α / pRK2013 como la estirpe auxiliar de conjugación. Se sembraron diluciones del parche de conjugación en placas de LB con los antibióticos necesarios para seleccionar transconjugantes (cloranfenicol y ácido nalidíxico), arabinosa 1 mM (para aumentar el número de copias del fósido) y salicilato 1 mM (para permitir la expresión heteróloga desde *psal* en el fósido y la antiterminación de la transcripción por la proteína N). El parche de conjugación, resuspendido en 2 ml de tampón fosfato, se sembró en placas de 15 cm de diámetro con factores de dilución de $5 \cdot 10^4$ y de 10^4 , de manera que se obtuvieron 1.000-2.000 y 2.500-4.000 colonias en cada placa, respectivamente. Las variaciones en el número de colonias de una conjugación a otra se deben a diferencias en las frecuencias de conjugación, que se encontraron en torno a 7-10 % en las distintas repeticiones realizadas. Conviene que las colonias queden separadas unas de otras, de manera que se pueda identificar y recuperar la colonia que presente la característica de interés.

El procedimiento de rastreo de la actividad entre las colonias transconjugantes ya crecidas consistió en aplicarles por encima con un espray catecol al 1 % o 2,3-dihibroxibifenilo al 0,5 %. Estos compuestos son dos sustratos distintos, uno con un solo

anillo aromático y el otro con dos, cuyos anillos aromáticos dihidroxilados son susceptibles de ser abiertos por una enzima con actividad dioxigenasa extradiólica (figura 16A). Los clones que expresaran genes codificantes de esta actividad y que además fueran capaces de transformar el sustrato correspondiente producirían 2-hidroximucónico semialdehído o 2-hidroxi-6-oxo-6-fenil-hexa-2,4-dienoato. Estos productos de rotura extradiólica son amarillos, y darían este color a las colonias que los produjeran, haciendo por tanto a los clones con esta actividad distinguibles del resto (figura 16B). El catecol ha sido utilizado previamente para detectar este tipo de actividad en colonias de bacterias con genes procedentes de organismos degradadores (Kulakov y col., 1998) y en bibliotecas metagenómicas (Brennerova y col., 2009).

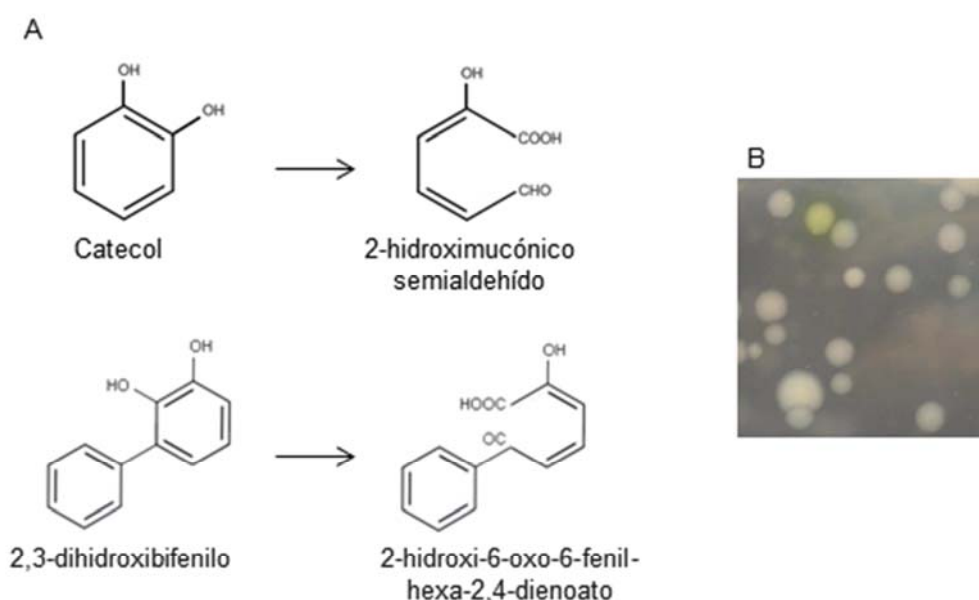


Figura 16. A. Esquema de las reacciones de dioxigenación extradiólica sobre catecol y 2,3-dihidroxibifenilo, generando los correspondientes productos de rotura. B. Colonia capaz de producir 2-hidroximucónico semialdehído a partir de catecol junto a otras sin la actividad en el rastreo de dioxigenasas extradiólicas.

Se analizaron unas 260.000 colonias distribuidas en 185 placas mediante la detección de color amarillo en colonias, pasados unos minutos tras la adición de catecol. Se identificaron 28 colonias amarillas, que se analizaron posteriormente por restricción, encontrándose 17 patrones de restricción distintos. Los clones que mantenían fósmidos distintos identificados usando catecol como sustrato se denominaron UPO33-UPO49.

Se analizaron otras 260.000 colonias distribuidas en unas 150 placas del mismo modo, pero en este caso se utilizó 2,3-dihidroxibifenilo en lugar de catecol. Con este sustrato se

identificaron 67 colonias amarillas, entre las cuales se detectaron 44 patrones de restricción distintos. Tres de estos clones tenían patrones de restricción iguales a los clones UPO35, UPO42 y UPO48 previamente identificados utilizando catecol, pero el resto eran diferentes. Estos últimos se denominaron UPO50-UPO72 y UPO74-UPO91. Por tanto, con los dos sustratos en su conjunto, se obtuvieron 58 clones distintos con actividad dioxigenasa extradiólica.

A pesar de que el catecol es el sustrato que ha sido utilizado previamente con más frecuencia en distintos laboratorios para identificar y caracterizar colonias con actividad dioxigenasa extradiólica, tras analizar el mismo número de clones con cada sustrato, nuestros datos indican que se obtuvieron más clones con actividad dioxigenasa extradiólica usando 2,3-dihidroxibifenilo como sustrato indicador.

Para asegurar con una probabilidad del 90 % que se habían analizado todos los clones de esta metagenoteca (que tenía 185.000 clones) habría que analizar unos 426.000 clones (según adaptación de la fórmula de Clarke y Carbon, 1976). En estos rastreos se analizaron unos 260.000 clones con cada sustrato, por lo que es muy probable que en la metagenoteca hubiera bastantes más clones independientes con actividad dioxigenasa extradiólica. De hecho, el poco grado de repetición de los patrones de restricción, 58 distintos entre 95 colonias amarillas, indica que el rastreo no fue saturante y que hubieran podido aparecer clones distintos si se hubieran rastreado más colonias transconjugantes.

3.3 Rastreo de oxigenasas hidroxilantes de anillos aromáticos

Con la misma metagenoteca de la pila de refinería de la industria petroquímica se realizaron también rastreos para identificar clones con actividad oxigenasa hidroxilante de anillos aromáticos (también llamada oxigenasa inicial). De forma similar, los rastreos también se realizaron en este caso en la estirpe MPO554 Nal^R, ya que es la que permite la antiterminación mediada por la proteína N y fue la que permitió aislar un mayor número de clones resistentes a carbenicilina. Como la metagenoteca se mantenía en la estirpe EPI300-T1^R, de nuevo se realizó una conjugación triparental antes de la búsqueda de clones con la actividad oxigenasa hidroxilante, tal y como se describe en el apartado anterior, sembrando diluciones del parche de conjugación en el mismo medio y con los mismos factores de dilución, obteniendo densidades de colonias por placa y frecuencias de conjugación similares.

La producción de índigo a partir de indol permite la detección de bacterias *E. coli* con actividad oxigenasa hidroxilante por su coloración azul (Ensley y col., 1983; O'Connor y col., 1997). Para ello, basta con cultivar las bacterias en medio rico LB, del que se emplea el triptófano como sustrato inicial. A partir del triptófano, la triptofanasa endógena de *E. coli* produce indol, incoloro, que pueden emplear como sustrato muchas oxigenasas hidroxilantes. Estas oxigenasas producen una doble hidroxilación del anillo heterocíclico del indol, dando lugar a un dihidrodiol inestable, que es químicamente deshidratado para formar indoxil. Sobre el indoxil se produce una oxidación espontánea y una dimerización a índigo (Royo y col., 2005a), siendo también posible la formación del heterodímero indirrubina, de color rojizo (Lim y col., 2005). Estas reacciones se esquematizan en la figura 17.

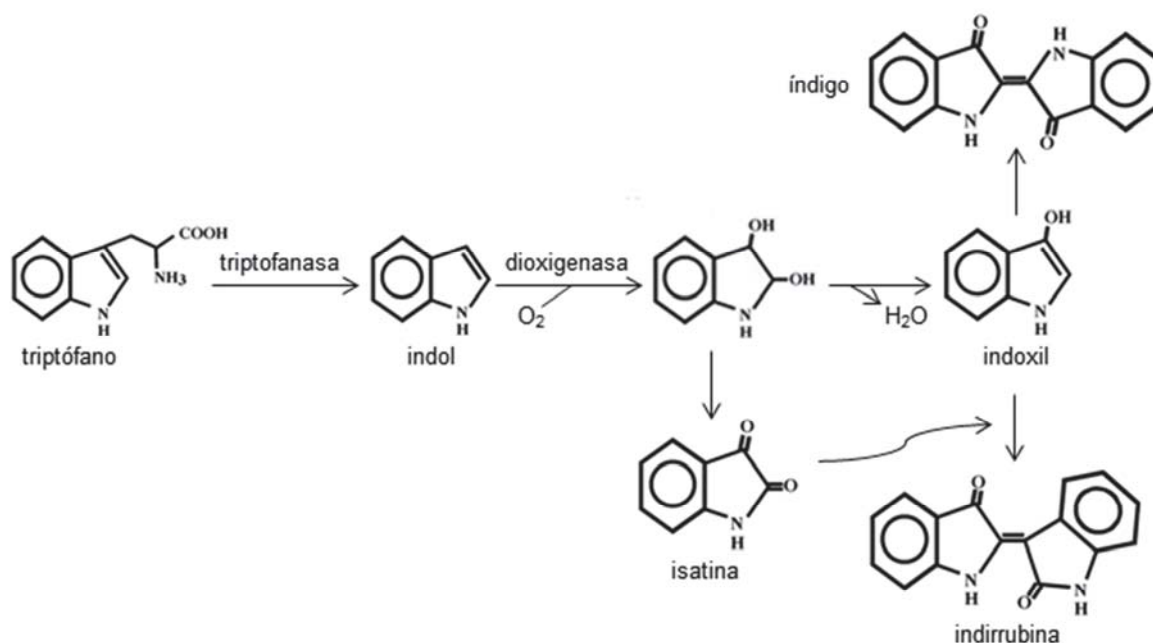


Figura 17. Esquema de las reacciones que tienen lugar para la producción de índigo a partir de triptófano. Se indica el nombre de algunos de los intermediarios, así como las reacciones catalizadas por la triptofanasa y la oxigenasa hidroxilante (representada en el esquema por una dioxigenasa, aunque la transformación hasta indoxil también se puede llevar a cabo de forma directa por una monooxigenasa). El resto de reacciones ocurren de manera espontánea (modificado de Royo y col., 2005a, y Kim y col., 2003).

Este ensayo no puede llevarse a cabo en cualquier estirpe, al requerirse la producción de una triptofanasa como la que se encuentra en *E. coli*, ausente en otras especies. Además, la producción de índigo requiere que las bacterias *E. coli* con los genes codificantes de oxigenasas hidroxilantes crezcan en medio rico LB, no siendo válido el empleo de medio mínimo a menos que este se suplemente con L-triptófano o indol (Hart y col., 1992). Es un

rastreo sencillo que ha permitido la identificación de diversas monooxigenasas y dioxigenasas hidroxilantes capaces de emplear el indol como sustrato, debido a que la oxigenación del indol hasta indoxil se puede llevar a cabo tanto por una dioxigenación seguida de una deshidratación como por una monooxigenación de forma directa (Figura 17), tanto en genotecas procedentes de bacterias degradadoras (Kasai y col., 2003) como en bibliotecas metagenómicas (van Hellemond y col., 2007).

De este modo, los clones que expresaran genes codificantes de oxigenasas iniciales capaces de emplear indol como sustrato darían colonias de color azul, distinguibles por esta característica del resto (figura 18).



Figura 18. Colonia productora de índigo junto a otras sin la actividad en el rastreo de oxigenasas iniciales.

Se analizaron unas 300.000 colonias distribuidas en aproximadamente 200 placas en la búsqueda de colonias azules por la producción de índigo. Las placas, en las que se cultivaron las bacterias en medio LB, se incubaron 2 días a 30 °C y a continuación a temperatura ambiente. Se observaron las placas diariamente durante una semana, detectando la aparición de colonias azules desde el segundo hasta el séptimo día. En total se identificaron 46 colonias azules. El análisis de restricción de los fósidos contenidos en estos transconjugantes, reveló la existencia de 34 clones distintos (con distinto patrón de restricción). Estos se denominaron UPO07-UPO31, UPO92-UPO94 y UPO106-UPO111. A esos 34 clones se les añadieron otros 2 clones con actividad oxigenasa inicial, que habían sido identificados en el rastreo de dioxigenasas extradiólicas pero que presentaban las dos actividades. Estos son UPO41 y UPO55. Por tanto, al final se disponía de 36 clones distintos con aparente actividad oxigenasa inicial. La mayoría de estos clones producían un color azul evidente, mientras que en algunos se observaba un color oscuro más parecido a marrón, pero fueron también seleccionados. Se ha descrito que bacterias con actividad oxigenasa inicial pueden dar lugar a colonias color morado o marrón en lugar de azul, por la producción conjunta de índigo (azul) y de su

isómero estructural indirrubina (rojo-rosa), que se pueden producir en distinta proporción (Hart y col., 1992; Lim y col., 2005).

Para asegurar con una probabilidad del 90 % que se habían analizado todos los clones de esta metagenoteca (que tenía 185.000 clones) habría que analizar unos 426.000 clones (según adaptación de la fórmula de Clarke y Carbon, 1976). En estos rastreos se analizaron unos 300.000 clones, por lo que es muy probable que en la metagenoteca hubiera más clones con actividad oxigenasa inicial. De nuevo, la poca repetición de los patrones de restricción en los clones, 34 patrones distintos de 46 colonias independientes, puso de manifiesto que el rastreo no fue saturante.

3.4 Rastreo de endoglucanasas

A través de una colaboración con la empresa Abengoa Research se realizó una búsqueda de clones con actividad endoglucanasa. Aunque no se trata de una actividad implicada en la degradación de contaminantes, se empleó la metagenoteca de la pila de refinería de la industria petroquímica, por ser una metagenoteca construida a partir de suelo con un buen número de clones que en principio podría ser válida también para la búsqueda de otras actividades. Para este rastreo se utilizó de nuevo la estirpe MPO554 NaI^R, que había dado buenos resultados anteriormente en la búsqueda de clones con otras actividades. Como la metagenoteca se mantenía en la estirpe EPI300-T1^R, se realizó una conjugación triparental antes de la búsqueda de clones con la actividad de interés, tal y como se describe en apartados anteriores, con la diferencia de que al medio LB se le aportó carboximetilcelulosa (CMC) 1 %, como sustrato para el rastreo. El parche de conjugación, resuspendido en 2 ml de tampón fosfato, se sembró en placas de 15 cm de diámetro con factores de dilución de $5 \cdot 10^4$ y de 10^4 , de manera que se obtuvieron 700-2.000 y 2.000-4.000 colonias en cada placa, respectivamente, y frecuencias de conjugación en torno a 4-9 % en las distintas repeticiones realizadas. En este caso, que las colonias queden lo suficientemente separadas unas de otras es incluso más importante que en los rastreos anteriores, de manera que se pueda identificar un halo alrededor de las colonias positivas y recuperar la colonia que presente dicha característica.

El procedimiento de rastreo consistió en la tinción con rojo Congo de placas con las colonias transconjugantes crecidas que habían sido incubadas a 30 °C durante 2 días en medio

LB con CMC. El rojo Congo se une a la CMC (Teather y Wood, 1982), de manera que se distinguirá un halo por falta de tinción sobre el fondo rojo alrededor de aquellas colonias capaces de degradar este compuesto por la producción de endoglucanasas, haciendo por tanto a los clones con esta actividad distinguibles del resto (figura 19). Antes de la tinción de las placas se incluyó un paso adicional consistente en la inmovilización de las colonias aplicando una capa de agar de cobertera. Esto facilita por una parte la visualización de los halos, al no quedar el medio cubierto por la biomasa disgregada de las colonias, y por otra la recuperación de biomasa de los clones positivos. Un procedimiento similar ha sido empleado con anterioridad para detectar esta actividad en metagenotecas (Rees y col., 2003).

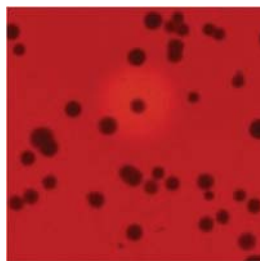


Figura 19. Colonia formadora de un halo en la tinción con rojo Congo por la degradación de la CMC junto a otras sin la actividad en el rastreo de endoglucanasas.

Se analizaron unas 550.000 colonias distribuidas en 278 placas. Mediante el rastreo descrito, se identificaron 15 colonias con halo. El halo detectado en ocasiones era bastante grande y nítido, mientras que en otros casos era más pequeño y presentaba menos contraste con el fondo. Además, a veces la existencia de colonias muy juntas hacía dudar sobre cuál era la colonia positiva. Estos clones dudosos se aislaron y se comprobó su actividad de manera independiente. Posteriormente, se analizaron todos los clones productores de halo por restricción, encontrándose 8 clones con distinto patrón, que se denominaron UPO95-UPO102. Estos clones están siendo caracterizados por la empresa.

Como se ha comentado en apartados anteriores, para asegurar con una probabilidad del 90 % que se habían analizado todos los clones de esta metagenoteca (que tenía 185.000 clones) habría que analizar unos 426.000 clones. En estos rastreos se analizaron unos 550.000 clones, por lo que la probabilidad de que se hayan analizado todos sería mayor del 90 %. Sin embargo, solo se observó repetición de dos de los patrones de restricción, correspondiéndose a dos de los tres clones que producen un halo de mayor tamaño. Por tanto, no se puede descartar que entre los clones analizados haya pasado desapercibido algún otro clon positivo que diera lugar a un halo menos evidente.

3.5 Rastreo de β -glucosidasas

Este rastreo se realizó en el mismo marco de colaboración con la empresa Abengoa Research comentado en el apartado anterior. Se empleó del mismo modo la metagenoteca de la pila de refinería de la industria petroquímica y la estirpe MPO554 Nal^R, por lo que el primer paso fue conjugar la metagenoteca a esta estirpe. Los clones productores de β -glucosidasas serán capaces de obtener glucosa a partir de celobiosa, por lo que podrán crecer en un medio que contenga celobiosa como única fuente de carbono. Se sembraron diluciones del parche de conjugación en placas de la variante del medio mínimo M9 descrita en el apartado 2.1 de Materiales y Métodos, pero empleando celobiosa 1 % como única fuente de carbono en lugar de glucosa 0,2 %, con ácido nalidíxico para seleccionar la estirpe receptora, arabinosa 1 mM y salicilato 1 mM. El empleo de celobiosa como única fuente de carbono es suficiente para seleccionar transconjugantes (solo aquellos que presenten la actividad de interés). El parche de conjugación, resuspendido en 2 ml de tampón fosfato, se lavó para eliminar los restos de LB y se sembró en placas de 9 cm de diámetro con factores de dilución de $5 \cdot 10^2$ y de 10^2 , estimándose posteriormente que en cada placa se estaban analizando del orden de $3 \cdot 10^5$ y $6 \cdot 10^5$ clones respectivamente, obteniéndose una frecuencia de conjugación de en torno al 4 %. Las placas con medio mínimo indicado se incubaron a 30 °C y se observó la aparición de colonias tras unos 2-5 días.

Se analizaron más de 5 millones de transconjugantes en un total de 6 placas. Se detectaron 48 colonias, que tras ser analizadas por restricción resultó que se correspondían con 3 patrones distintos. De los 3 patrones, un patrón apareció 45 veces, otro 2 y otro solo 1 vez, lo que indica un clarísimo sesgo que no se observó en los patrones de restricción de los clones positivos para las otras actividades. Muy posiblemente esto se deba al proceso selectivo de identificación y estará relacionado con la distinta probabilidad de que un transconjugante con un fósido de cada tipo sea capaz de crecer y generar una colonia en esas condiciones. Esos 3 clones distintos se denominaron UPO103-UPO105. Habiendo analizado tal cantidad de transconjugantes, la probabilidad de que hayamos analizado todos los clones de la metagenoteca es prácticamente del 100 %.

4. CARACTERIZACIÓN DE CLONES CON ACTIVIDAD DIOXIGENASA EXTRADIÓLICA

En la metagenoteca de pilas de refinería de la industria petroquímica se identificaron 58 clones distintos con actividad dioxigenasa extradiólica (EDO), denominados UPO33-UPO72 y UPO74-UPO91.

Esos clones se identificaron al presentar la actividad en la estirpe MPO554 Nal^R en presencia de arabinosa y salicilato, es decir, con el fósido en alto número de copias, con expresión heteróloga del ADN metagenómico desde el promotor *psal* del fósido y con la antiterminación de la transcripción proporcionada por la proteína N. Se realizó una prueba con algunos de los clones tratando de establecer si también se hubieran detectado como positivos, mostrando color amarillo, durante el rastreo de actividad dioxigenasa extradiólica en ausencia de salicilato, para lo cual los genes codificantes tendrían que estar expresándose desde sus propios promotores o desde la expresión basal del promotor *psal*. De este análisis resultó que aproximadamente una tercera parte de ellos solo presentó actividad en presencia de salicilato, otro tercio presentó actividad en ambos casos pero las colonias daban un color amarillo bastante más intenso en presencia de salicilato y el resto no presentó diferencia aparente con y sin el inductor. De esta información se concluye que de haber realizado los rastreos con el sistema original (únicamente con la expresión propia de los genes metagenómicos), muy probablemente se habrían identificado bastantes menos clones con la actividad de interés.

4.1 Secuenciación de los clones con actividad dioxigenasa extradiólica

53 de los 58 clones con actividad dioxigenasa extradiólica se enviaron a un servicio de secuenciación masiva, en 3 *pools* diferentes, junto con otros 7 clones portadores de otras actividades, obteniendo las secuencias metagenómicas en *contigs* a priori no asignados a ningún clon concreto. En la tabla 7 se recopila información sobre cada una de estas tandas de secuenciación.

En cada tanda se enviaron unos 20 clones, con tamaño medio de inserto de 35 kb, por lo que se esperarían aproximadamente 700 kb de secuencia de inserto en cada tanda. Teniendo en cuenta la información de la tabla 7, se dispondría de la mayor parte de la secuencia de los insertos.

| | n° <i>contigs</i> | tamaño <i>contigs</i> | total secuencia | n° genes identificados |
|---------|-------------------|-----------------------|-----------------|------------------------|
| tanda 1 | 53 | 611-38.849 nt | 551 kb | 522 |
| tanda 2 | 105 | 552-40.901 nt | 684 kb | 707 |
| tanda 3 | 93 | 619-39.546 nt | 710 kb | 735 |

Tabla 7. Información asociada a cada tanda de secuenciación por *pools*.

4.1.1. Identificación de genes de dioxigenasas extradiólicas y asignación a clones

Del análisis de las secuencias de los *contigs* de las 3 tandas de secuenciación, en principio se obtuvieron 22 secuencias homólogas a genes completos codificantes de dioxigenasas extradiólicas. A esos 22 genes se los nombró como *upoedo1-upoedo21* y *upoedo23*. Además, se identificó otra secuencia homóloga a genes codificantes de dioxigenasas extradiólicas que no estaba completa (se encontraba en un extremo del inserto de ADN metagenómico, faltándole el final), a la que se denominó *upoedo22*. En el fósido, la secuencia codificante empalmaría con la secuencia del vector a continuación, generándose una proteína con 265 aminoácidos originales seguidos de 32 aminoácidos codificados por la secuencia del vector. De acuerdo con los alineamientos con las proteínas más parecidas de las bases de datos se estimó que a *upoedo22* le faltarían aproximadamente los 10 aminoácidos finales propios de la proteína y que podría dar lugar a una EDO funcional.

Al haber sido secuenciados por grupos, fue necesario asignar a clones concretos cada una de esas secuencias homólogas a genes codificantes de EDOs. Para ello, se diseñaron oligonucleótidos que hibridaran en esos genes y se realizaron reacciones de PCR con todos los clones de la tanda de secuenciación correspondiente. Aquellos clones con los que se obtuvo amplificación, se confirmaron mediante secuenciación con uno de los oligonucleótidos empleados en la PCR. De este modo, esos 22 genes codificantes de dioxigenasas extradiólicas, se asignaron a 27 clones, y la secuencia del gen incompleto a otro clon (tabla 8), teniendo así esos 28 clones no solo la secuencia del gen en cuestión asignada sino la de todo el *contig* correspondiente. Varios de los genes no estaban completos en sus *contigs*, y hubo que obtener el resto de la secuencia mediante posteriores reacciones de secuenciación. Se

observó que esto ocurrió cuando varios genes pertenecientes a la misma tanda de secuenciación eran muy parecidos, y que por tanto probablemente se debía a la incapacidad del programa de ensamblaje de *contigs* de determinar cuáles eran las uniones adecuadas.

| clon | gen | clon | gen | clon | gen | clon | gen |
|--------------|-----------------|-------|------------------|--------------|-----------------|-------|-----------------|
| UPO33 | <i>upoedo20</i> | UPO47 | <i>upoedo14</i> | UPO62 | | UPO77 | |
| UPO34 | | UPO48 | | UPO63 | | UPO78 | <i>upoedo8</i> |
| UPO35 | <i>upoedo12</i> | UPO49 | <i>upoedo3</i> | UPO64 | <i>upoedo1</i> | UPO79 | |
| UPO36 | <i>upoedo11</i> | UPO50 | <i>upoedo3</i> | | <i>upoedo2</i> | UPO80 | |
| UPO37 | <i>upoedo7</i> | UPO51 | <i>upoedo22*</i> | UPO65 | | UPO81 | |
| UPO38 | <i>upoedo16</i> | UPO52 | <i>upoedo5</i> | UPO66 | | UPO82 | |
| UPO39 | | UPO53 | | UPO67 | <i>upoedo6</i> | UPO83 | |
| UPO40 | | UPO54 | | UPO68 | <i>upoedo23</i> | UPO84 | |
| UPO41 | <i>upoedo19</i> | UPO55 | <i>upoedo5</i> | UPO69 | | UPO85 | |
| UPO42 | <i>upoedo10</i> | UPO56 | | UPO70 | | UPO86 | |
| UPO43 | <i>upoedo15</i> | UPO57 | <i>upoedo4</i> | UPO71 | | UPO87 | <i>upoedo23</i> |
| UPO44 | | UPO58 | <i>upoedo5</i> | UPO72 | <i>upoedo5</i> | UPO88 | |
| UPO45 | <i>upoedo17</i> | UPO59 | | UPO74 | | UPO89 | <i>upoedo7</i> |
| | <i>upoedo18</i> | UPO60 | <i>upoedo7</i> | UPO75 | | UPO90 | <i>upoedo9</i> |
| UPO46 | <i>upoedo13</i> | UPO61 | <i>upoedo21</i> | UPO76 | | UPO91 | |

Tabla 8. Relación de clones con actividad dioxigenasa extradiólica y gen o genes posibles responsables de dicha actividad identificados en cada uno de ellos. Los clones que no fueron enviados a secuenciar se indican en gris. Los clones de los que se conocía la secuencia completa se indican en negrita. *El gen denominado como *upoedo22* no está completo en el inserto metagenómico, acabando su *orf* con secuencia del vector.

Por tanto, se asignó una secuencia a 28 de los 58 clones con actividad dioxigenasa extradiólica. Todos ellos tienen asignado al menos un gen codificante de enzimas con ese tipo de actividad (tabla 8) mientras que quedan 30 clones sin tener asignado un posible gen codificante de actividad EDO. A dos de esos clones (UPO45 y UPO64) se les asignaron 2 genes a cada uno; y por otro lado, 4 de los genes (*upoedo3*, *upoedo5*, *upoedo7* y *upoedo23*) fueron asignados a varios clones cada uno (tabla 8). Los clones que comparten genes, en

algunos casos comparten parte del inserto, aunque no completo, y en otros casos presentan secuencia metagenómica distinta, coincidiendo solo en algunos fragmentos que incluyen el gen en cuestión. De 8 de los clones con secuencia asignada conocemos la secuencia completa (tabla 8), pero de otros 20 solo parte, que en algunos casos será la mayor parte del inserto pero en otros son solo algunas kb, pudiendo tener más genes codificantes de dioxigenasas extradiólicas no identificados.

Algunas de las proteínas metagenómicas se pueden clasificar claramente en algún tipo concreto de dioxigenasa extradiólica, mientras que otras presentan una identidad baja con alguna proteína previamente catalogada en uno de esos tipos. Pero en principio parece que entre las dioxigenasas extradiólicas de la metagenoteca de las pilas de refinería de la industria petroquímica se encuentran una mayoría de catecol-2,3-dioxigenasas (UPOEDO3, UPOEDO4, UPOEDO5, UPOEDO8, UPOEDO11, UPOEDO12, UPOEDO13, UPOEDO14, UPOEDO15, UPOEDO16, UPOEDO19, UPOEDO20 y UPOEDO23), 2,3-dihidroxibifenil-1,2-dioxigenasas (UPOEDO7, UPOEDO9, UPOEDO10, UPOEDO18 y UPOEDO21), protocatecuato-4,5-dioxigenasas (UPOEDO1, UPOEDO2, UPOEDO6 y UPOEDO22) y una 2,3-dihidroxifenilpropionato-1,2-dioxigenasa (UPOEDO17).

Las secuencias de las proteínas de estos clones, homólogas a dioxigenasas extradiólicas de las bases de datos, se recogen en el Anexo.

Posteriormente, se identificaron otros 12 genes completos codificantes de proteínas homólogas a EDOs y otras 3 secuencias parciales que no se ha confirmado si están completas en el inserto correspondiente. En esta tesis no se trabajó con esas secuencias, pero su futura asignación a clones del mismo modo que se hizo con las demás, probablemente permita conocer la secuencia de la EDO producida por gran parte de los 25 clones con esta actividad enviados a secuenciar a los que aún no se les ha asignado ningún gen.

4.1.2. Análisis filogenético de las dioxigenasas extradiólicas identificadas

Se realizó un estudio filogenético con las secuencias de las 22 proteínas dioxigenasas extradiólicas identificadas más la codificada por el gen incompleto que acaba con secuencia acoplada del vector, todas ellas asignadas a clones. Se adscribieron secuencias a los dos tipos generales de enzimas dioxigenasas extradiólicas descritos por Eltis y Bolin (1996): tipo I y tipo II.

Dentro del tipo I, hay descritas seis familias, I.1-6 (Eltis y Bolin, 1996; Sakai y col., 2002). Las nuevas proteínas se encuentran distribuidas en tres de esas seis familias distintas (I.2, I.3 y I.4) (figuras 20 y 21). Varias de las secuencias metagenómicas son lo suficientemente diferentes a las demás como para catalogarlas en nuevas subfamilias, teniendo en cuenta que las secuencias de aminoácidos dentro de una misma subfamilia deben tener >54 % de identidad unas con otras (Eltis y Bolin, 1996).

Ninguna de las nuevas proteínas pertenece a la familia I.1, que incluye las EDOs descritas de un solo dominio. En las demás categorías se encuentran enzimas con dos dominios.

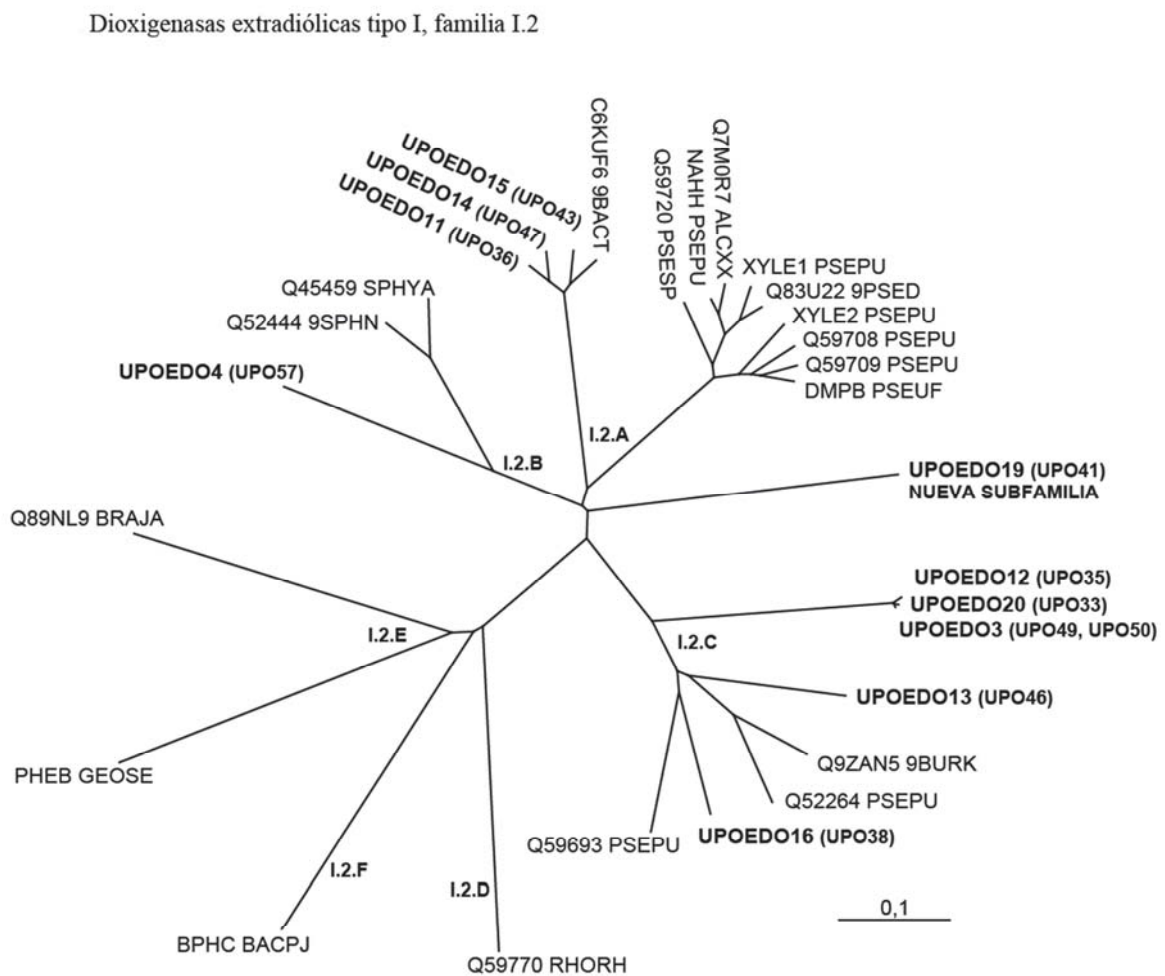


Figura 20. Análisis filogenético de las secuencias metagenómicas de dioxigenasas extradiólicas tipo I, subfamilia I.2. Árbol filogenético resultante del alineamiento de las secuencias de las proteínas de dioxigenasas extradiólicas identificadas en la metagenoteca de pilas de refinería de la industria petroquímica con otras secuencias del mismo tipo previamente identificadas. La escala representa la distancia expresada como diferencia en aminoácidos por posición. Los símbolos en las ramas indican lo siguiente: por ejemplo I.2.A, tipo (I), familia (2) y subfamilia (A). Las secuencias metagenómicas identificadas en este trabajo se muestran en

negrita, acompañadas del clon en el que se encuentran. El resto de secuencias empleadas en el dendrograma y sus códigos de acceso en las bases de datos GenBank, NCBI Reference Sequence o UniProtKB/Swiss-Prot se indican a continuación: BPHC BACPJ: Manganese-dependent 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Bacillus* sp. (Q8GR45); Q59770 RHORH: Catechol 2,3-dioxygenase, *Rhodococcus rhodochrous* (Q59770); HEB GEOSE: Metapyrocatechase, *Geobacillus stearothermophilus* (P31003); Q89NL9 BRAJA: Metapyrocatechase, *Bradyrhizobium diazoefficiens* (Q89NL9); Q59693 PSEPU: Catechol 2,3-dioxygenase II, *Pseudomonas putida* (Q59693); Q9ZAN5 9BURK: Catechol 2,3-dioxygenase, *Comamonas* sp. (Q9ZAN5); Q52264 PSEPU: 3-methylcatechol 2,3-dioxygenase, *Pseudomonas putida* (Q52264); Q52444 9SPHN: Catechol 2,3-dioxygenase, *Sphingomonas agrestis* (Q52444); Q45459 SPHYA: Catechol 2,3-dioxygenase, *Sphingobium yanoikuyae* (Q45459); XYLE2 PSEPU: Metapyrocatechase, *Pseudomonas putida* (Q04285); Q59708 PSEPU: Catechol 2,3-dioxygenase, *Pseudomonas putida* (Q59708); DMPB PSEUF: Metapyrocatechase, *Pseudomonas* sp. (P17262); Q59709 PSEPU: Catechol 2,3-dioxygenase, *Pseudomonas putida* (Q59709); NAHH PSEPU: Metapyrocatechase, *Pseudomonas putida* (P08127); Q59720 PSESP: Catechol 2,3-dioxygenase, *Pseudomonas* sp. (Q59720); Q7M0R7 ALCXX: Catechol 2,3-dioxygenase, *Alcaligenes xylosoxydans* (Q7M0R7); XYLE1 PSEPU: Metapyrocatechase, *Pseudomonas putida* (P06622); Q83U22 9PSED: Catechol 2,3 dioxygenase, *Pseudomonas* sp. (Q83U22); C6KUF6 9BACT: Catechol 2,3-dioxygenase, uncultured bacterium CLON2H2 (C6KUF6).

Diez de las nuevas proteínas estarían incluidas en la familia I.2 (figura 20). De esas diez, nueve estarían distribuidas en 3 de las 7 subfamilias descritas hasta el momento, que son I.2.A-G (Eltis y Bolin, 1996; Hatta y col., 2003; Suenaga, y col., 2007). De este modo, UPOEDO11, UPOEDO14 y UPOEDO15 pertenecerían a la subfamilia I.2.A; UPOEDO4 a la subfamilia I.2.B; y UPOEDO3, UPOEDO12, UPOEDO13, UPOEDO16 y UPOEDO20 a la subfamilia I.2.C. Por otro lado, UPOEDO19 presenta una identidad $\leq 47\%$ con cualquiera de las secuencias de las bases de datos adscritas a alguna subfamilia (incluyendo secuencias de las subfamilias I.2.F-G, no incluidas en el árbol filogenético), por lo que se encontraría en una subfamilia nueva denominada I.2.H. Habría que comentar que no sería el único representante de esa nueva subfamilia, al presentar valores de identidad altos con otras proteínas de la base de datos non-redundant protein sequences no adscritas a ninguna familia. Se trata de varias catecol-2,3-dioxygenasas, destacando una de *Methyloversatilis universalis* con la que presenta un 84 % de identidad.

Dioxigenasas extradiólicas tipo I, familias I.3, I.4, I.5 y I.6

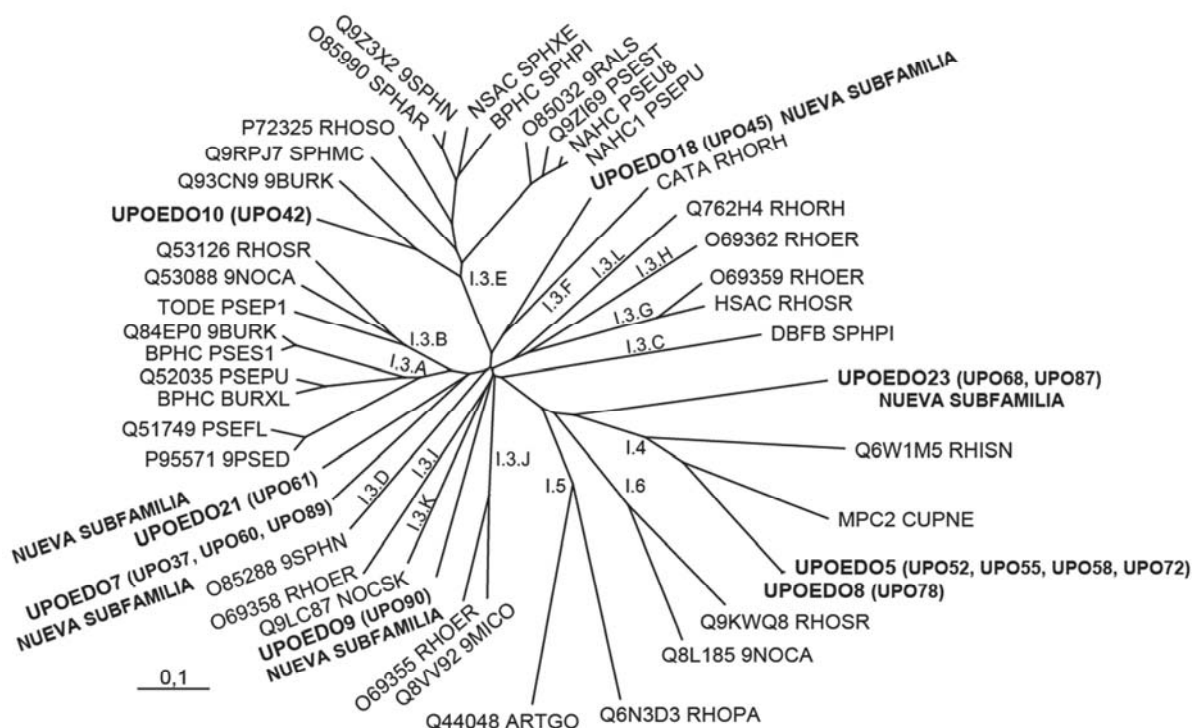


Figura 21. Análisis filogenético de las secuencias metagenómicas de dioxigenasas extradiólicas tipo I, subfamilias I.3, I.4, I.5 y I.6. Árbol filogenético resultante del alineamiento de las secuencias de las proteínas de dioxigenasas extradiólicas identificadas en la metagenoteca de pilas de refinería de la industria petroquímica con otras secuencias del mismo tipo previamente identificadas. La escala representa la distancia expresada como diferencia en aminoácidos por posición. Los símbolos en las ramas indican lo siguiente: por ejemplo I.3.A, tipo (I), familia (3) y subfamilia (A). Las secuencias metagenómicas identificadas en este trabajo se muestran en negrita, acompañadas del clon en el que se encuentran. El resto de secuencias empleadas en el dendrograma y sus códigos de acceso en las bases de datos GenBank, NCBI Reference Sequence o UniProtKB/Swiss-Prot se indican a continuación: Q9RPJ7 SPHMC: 1,2-dihydroxynaphthalene dioxygenase ThnC, *Shingopyxis macrogoltabida* (AAD52962.1); O85990 SPHAR: Dihydroxy naphthalene/biphenyl dioxygenase, *Shingomonas aromaticivorans* (AAD04006.1); Q9Z3X2 9SPHN: 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Shingobium chungbukense* (AAC95319.1); O85032 9RALS: 1,2-dihydroxynaphthalene dioxygenase, *Ralstonia* sp. (AAD12614.1); Q9ZI69 PSEST: 1, 2-dihydroxynaphthalene dioxygenase, *Pseudomonas stutzeri* (AAD02140.1); O85288 9SPHN: Extradiol dioxygenase, *Shingomonas* sp. (AAC38621.1); BPHC SPHPI: Biphenyl-2,3-diol 1,2-dioxygenase, *Shingomonas paucimobilis* (AAA25678.1); P72325 RHOSO: 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus* sp. (BAA18940.1); NSAC SPHXE: 1,2-dihydroxynaphthalene dioxygenase, *Shingobium xenophagum* (AAB06725.2); NAHC PSEU8: 1,2-dihydroxynaphthalene dioxygenase, *Pseudomonas* sp. (AAA16130.1); NAHC1 PSEPU: 1,2-dihydroxynaphthalene dioxygenase, *Pseudomonas putida* (AAA91577.1); P95571 9PSED: 3-isopropylcatechol-2,3-dioxygenase, *Pseudomonas* sp. (AAB36672.1); BPHC BURXL: Biphenyl-2,3-diol 1,2-dioxygenase, *Burkholderia xenovorans* (CAA46910.1); Q53088 9NOCA:

2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus* sp. (AAB07755.1); TODE PSEP1: 3-methylcatechol 2,3-dioxygenase, *Pseudomonas putida* (AAA26010.1); Q52035 PSEPU: 2,3-dihydroxybiphenyl dioxygenase, *Pseudomonas putida* (AAA25756.1); Q53126 RHOSR: 2,3-dihydroxybiphenyl dioxygenase, *Rhodococcus* sp. (BAA06872.1); BPHC PSES1: Biphenyl-2,3-diol 1,2-dioxygenase, *Pseudomonas* sp. (AAA25750.1); Q51749 PSEFL: 3-isopropylcatechol dioxygenase, *Pseudomonas fluorescens* (BAA07080.1); CATA RHORH: Metapyrocatechase, *Rhodococcus rhodochrous* (Q53034); Q8VV92 9MICO: 2,2',3-trihydroxybiphenyl 1,2-dioxygenase, *Terrabacter* sp. (Q8VV92); O69355 RHOER: 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus erythropolis* (O69355); Q762H4 RHORH: 2, 3-dihydroxybiphenyl 1, 2-dioxygenase, *Rhodococcus rhodochrous* (Q762H4); O69362 RHOER: 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus erythropolis* (O69362); O69359 RHOER: 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus erythropolis* (O69359); HSAC RHOSR: Iron-dependent extradiol dioxygenase, *Rhodococcus* sp. (Q9KWQ5); Q9LC87 NOCSK: Extradiol dioxygenase, *Nocardioides* sp. (Q9LC87); O69358 RHOER: 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus erythropolis* (O69358); Q84EP0 9BURK: 2,3-dihydroxybiphenyl dioxygenase, *Cupriavidus oxalaticus* (Q84EP0); Q93CN9 9BURK: Extradiol dioxygenase DbtC, *Burkholderia* sp. (AAK96189.1); MPC2 CUPNE: Metapyrocatechase 2, *Cupriavidus necator* (P17296); Q6W1M5 RHISN: Metapyrocatechase, *Rhizobium* sp. (Q6W1M5); Q9KWQ8 RHOSR: 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus* sp. (Q9KWQ8); Q8L185 9NOCA: Extradiol dioxygenase, *Rhodococcus* sp. (Q8L185); DBFB SPHPI: 2,2',3-trihydroxybiphenyl dioxygenase, *Sphingomonas paucimobilis* (CAA51364.1); Q6N3D3 RHOPA: Putative catechol 2,3-dioxygenase, *Rhodopseudomonas palustris* (Q6N3D3); Q44048 ARTGO: 3,4-dihydroxyphenylacetate 2,3-dioxygenase, *Arthrobacter globiformis* (Q44048).

Otras 5 de las nuevas EDOs pertenecerían a la familia I.3 (figura 21). Aunque por claridad no se encuentran todas incluidas en el árbol filogenético, hasta el momento de han descrito 14 subfamilias en I.3, I.3.A-N (Eltis y Bolin, 1996; Kulakov y col., 1998; Taguchi y col., 2004; Suenaga y col., 2007). UPOEDO10 pertenecería a la subfamilia I.3.E. UPOEDO9 formaría una nueva subfamilia, denominada I.3.O, al presentar una identidad ≤ 44 % con cualquiera de las proteínas de las bases de datos. Por otro lado, UPOEDO18 también pertenecería a una nueva subfamilia, que se denominaría I.3.P, ya que el mayor valor de identidad que presenta con alguna proteína de las bases de datos es del 51 %. Lo mismo ocurriría con UPOEDO7, que sería el único componente hasta el momento de la nueva subfamilia I.3.Q, ya que no se observan valores de identidad con proteínas de las bases de datos > 53 %. Y por último, UPOEDO21 inauguraría la nueva subfamilia I.3.R, siguiendo los mismos criterios.

También habría nuevas proteínas en la familia I.4 (figura 21). Se trataría de UPOEDO5, UPOEDO8 y probablemente también de UPOEDO23. UPOEDO23 se encuentra bastante alejada filogenéticamente de las secuencias adscritas a esta familia (≤ 28 % de identidad), por

lo que constituiría una subfamilia nueva dentro de I.4, en la que hasta la fecha no se han establecido subfamilias.

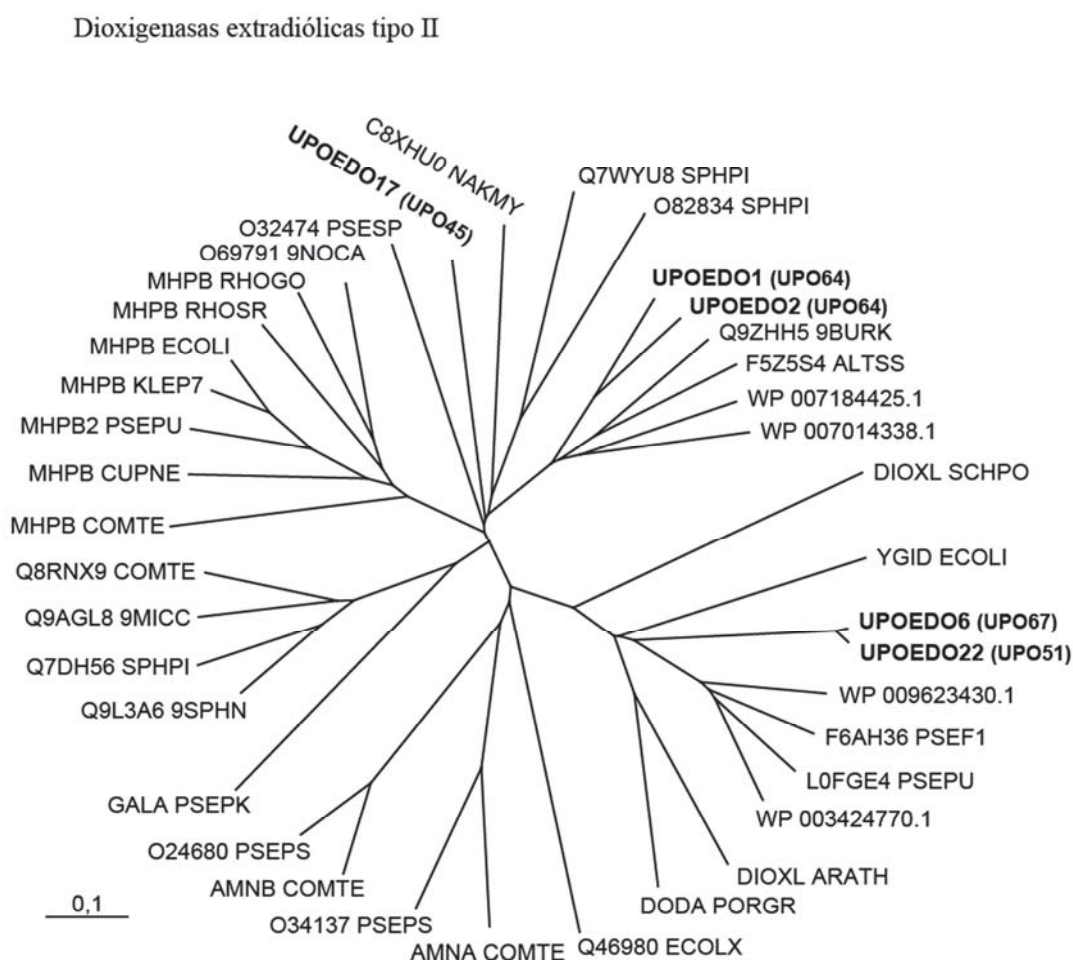


Figura 22. Análisis filogenético de las secuencias metagenómicas de dioxigenasas extradiólicas tipo II. Árbol filogenético resultante del alineamiento de las secuencias de las proteínas de dioxigenasas extradiólicas identificadas en la metagenoteca de pilas de refinera de la industria petroquímica con otras secuencias del mismo tipo previamente identificadas. La escala representa la distancia expresada como diferencia en aminoácidos por posición. Las secuencias metagenómicas identificadas en este trabajo se muestran en negrita, acompañadas del clon en el que se encuentran. El resto de secuencias empleadas en el dendrograma y sus códigos de acceso en las bases de datos GenBank, NCBI Reference Sequence o UniProtKB/Swiss-Prot se indican a continuación: Q9ZHH5 9BURK: Extradiol dioxygenase PhnC, *Burkholderia sartisoli* (AAD09870); F5Z5S4 ALTSS: Protocatechuate 4,5-dioxygenase subunit beta, *Alteromonas* sp. (YP_004468881.1); C8XHU0 NAKMY: Extradiol ring-cleavage dioxygenase class III protein subunit B, *Nakamurella multipartita* (YP_003203382.1); WP 007184425.1: Protocatechuate 4,5-dioxygenase subunit beta, *Hydrocarboniphaga effusa* (WP_007184425.1); WP 007014338.1: Protocatechuate 4,5-dioxygenase subunit beta, *Novosphingobium pentaromativorans* (WP_007014338.1); MHPB KLEP7: 2,3-dihydroxyphenylpropionate 1,2-dioxygenase, *Klebsiella pneumoniae* (A6TAC8.1); MHPB2 PSEPU: 2,3-dihydroxyphenylpropionate 1,2-dioxygenase 2, *Pseudomonas putida* (Q9F9U5.1);

MHPB RHOSR: 2,3-dihydroxyphenylpropionate 1,2-dioxygenase, *Rhodococcus* sp. (Q0SJD2.1); MHPB RHOGO: Dihydroxyphenylpropionate 1,2-dioxygenase, *Rhodococcus globerulus* (AAB81314); MHPB COMTE: Dihydroxyphenylpropionate 1,2-dioxygenase, *Comamonas testosteroni* (BAA82879); F6AH36 PSEF1: Extradiol ring-cleavage dioxygenase class III protein subunit B, *Pseudomonas fulva* (YP_004474756.1); L0FGE4 PSEPU: Extradiol ring-cleavage dioxygenase class III protein subunit B, *Pseudomonas putida* (YP_007228315.1); WP 003424770.1: Catalytic LigB subunit of aromatic ring-opening dioxygenase, *Pseudomonas syringae* (WP_003424770.1); WP 009623430.1: Catalytic LigB subunit of aromatic ring-opening dioxygenase, *Pseudomonas* sp. (WP_009623430.1); YGID ECOLI: Uncharacterized protein YgiD, *Escherichia coli* (P24197.3); DIOXL ARATH: 4,5-DOPA dioxygenase extradiol-like protein, *Arabidopsis thaliana* (Q949R4.1); DODA PORGR: 4,5-DOPA dioxygenase extradiol, *Portulaca grandiflora* (Q7XA48.1); DIOXL SCHPO: 4,5-DOPA dioxygenase extradiol-like protein, *Schizosaccharomyces pombe* (O74741.2); AMNA COMTE: 2-aminophenol 1,6-dioxygenase alpha subunit, *Comamonas testosteroni* (AAT35227.1); AMNB COMTE: 2-aminophenol 1,6-dioxygenase beta subunit, *Comamonas testosteroni* (AAT35226.1); O34137 PSEPS: 2-aminophenol-1,6-dioxygenase alpha subunit, *Pseudomonas pseudoalcaligenes* (AAB71525.1); O24680 PSEPS: 2-aminophenol-1,6-dioxygenase beta subunit, *Pseudomonas pseudoalcaligenes* (AAB71524.1); O32474 PSESP: CarBb catalytic subunit of meta-cleavage enzyme, *Pseudomonas* sp. (BAA21731.1); Q7WYU8 SPHPI: 3-O-methylgallate 3,4-dioxygenase, *Sphingomonas paucimobilis* (BAC79261); O69791 9NOCA: Extradiol dioxygenase, *Rhodococcus* sp. (CAA06875.1); Q9L3A6 9SPHN: FldU protein, *Sphingomonas* sp. (CAB87561.1); GALA PSEPK: GalA Gallate dioxygenase, *Pseudomonas putida* (Q88JX5); Q46980 ECOLX: Homoprotocatechuate dioxygenase, *Escherichia coli* (CAA86042); MHPB ECOLI: MhpB 3-(2,3-dihydroxyphenylpropionate)1, 2-dioxygenase, *Escherichia coli* (CAA70748); Q7DH56 SPHPI: Beta subunit of protocatechuate 4,5-dioxygenase, *Sphingomonas paucimobilis* (BAB88743); O82834 SPHPI: OH-DDVA oxygenase, *Sphingomonas paucimobilis* (BAA75884); MHPB CUPNE: 2,3-dihydroxyphenylpropionate/2,3-dihydroxicinnamic acid 1,2-dioxygenase, *Cupriavidus necator* (P17295); Q9AGL8 9MICC: Protocatechuate 4,5-dioxygenase, *Arthrobacter keyseri* (AAK16524.1); Q8RNX9 COMTE: protocatechuate 4,5-dioxygenase beta subunit, *Comamonas testosteroni* (AAM09637).

Por último, cuatro de las nuevas proteínas más la codificada por el gen incompleto resultan ubicarse en el grupo de enzimas tipo II (figura 22), que son menos abundantes y estudiadas que las de tipo I (Eltis y Bolin, 1996; Suenaga y col., 2007). Se trata de UPOEDO1, UPOEDO2, UPOEDO6, UPOEDO17 y UPOEDO22. Eltis y Bolin (1996) establecieron 2 familias en este grupo (II.1 y II.2), a partir de solo 3 secuencias. Aunque posteriormente se ha incorporado alguna secuencia más siguiendo este criterio (Kulakov y col., 1998), los estudios filogenéticos que agrupan un mayor número de enzimas de este tipo no parecen atender a estas categorías (por ejemplo Nogales y col., 2005), y en el presente trabajo no se ha tratado de asignar las nuevas secuencias a dichas familias. UPOEDO1 y UPOEDO2 se parecen más entre sí que a ninguna de las bases de datos, con las que presentan identidades ≤ 57 y 56 %, respectivamente. Lo mismo ocurre con UPOEDO6 y UPOEDO17,

que presentan identidades con las secuencias de las bases de datos $\leq 57\%$ en ambos casos. La enzima de este tipo más novedosa sería UPOEDO17, ya que no se encuentra en las bases de datos ninguna secuencia con la que presente una identidad $>39\%$. UPOEDO17 parece ser una 2,3-dihidroxifenilpropionato-1,2-dioxigenasa, y las otras 4 protocatecuato-4,5-dioxigenasas.

4.2 Ensayos de actividad dioxigenasa extradiólica con distintos sustratos

Los 58 clones con actividad dioxigenasa extradiólica se ensayaron para determinar la actividad relativa que poseen con distintos sustratos: 2,3-dihidroxibifenilo, catecol, 3-metilcatecol, 4-metilcatecol y 4-clorocatecol.

Los ensayos se llevaron a cabo en la estirpe MPO554 Nal^R, donde se indujo la expresión heteróloga desde el promotor *psal* del fósido que precede al ADN metagenómico (en presencia de salicilato 1 mM) y con el fósido en alto número de copias (en presencia de arabinosa 1 mM), ya que en muchos de estos clones no se detecta actividad de otro modo. En algunos clones, los genes codificantes de dioxigenasas extradiólicas se expresarían desde sus propios promotores, pero en otros se daría expresión, por completo o en parte, desde el promotor heterólogo *psal*. Por tanto, como los niveles de expresión de estos genes en los distintos clones pueden variar significativamente en función de su orientación y posición en el inserto, debido a la expresión heteróloga desde *psal*, los valores de actividad absoluta no tienen mucho significado y en lugar de representar estos se calcularon y representaron los valores de actividad relativa de cada clon con cada uno de los sustratos empleados (considerando la actividad con el sustrato con el que cada clon presenta su máxima actividad como el 100 %). Los resultados se muestran en la figura 23.

Como ya se ha comentado anteriormente, el gen asignado al clon UPO51 (*upoedo22*) no está completo, de manera que los últimos aminoácidos de la proteína se encuentran sustituidos por otros codificados por secuencia del vector, aunque probablemente no le falte mucho de su secuencia original, por lo que puede que sea activa. Los valores de actividad absoluta de ese clon fueron bajos (datos no mostrados), pero del mismo nivel que otros clones. Por otro lado, a ese clon solo se le habían asignado 2 kb de secuencia, por lo que también puede que la actividad se deba a otro gen que se encuentre en el resto del ADN metagenómico.

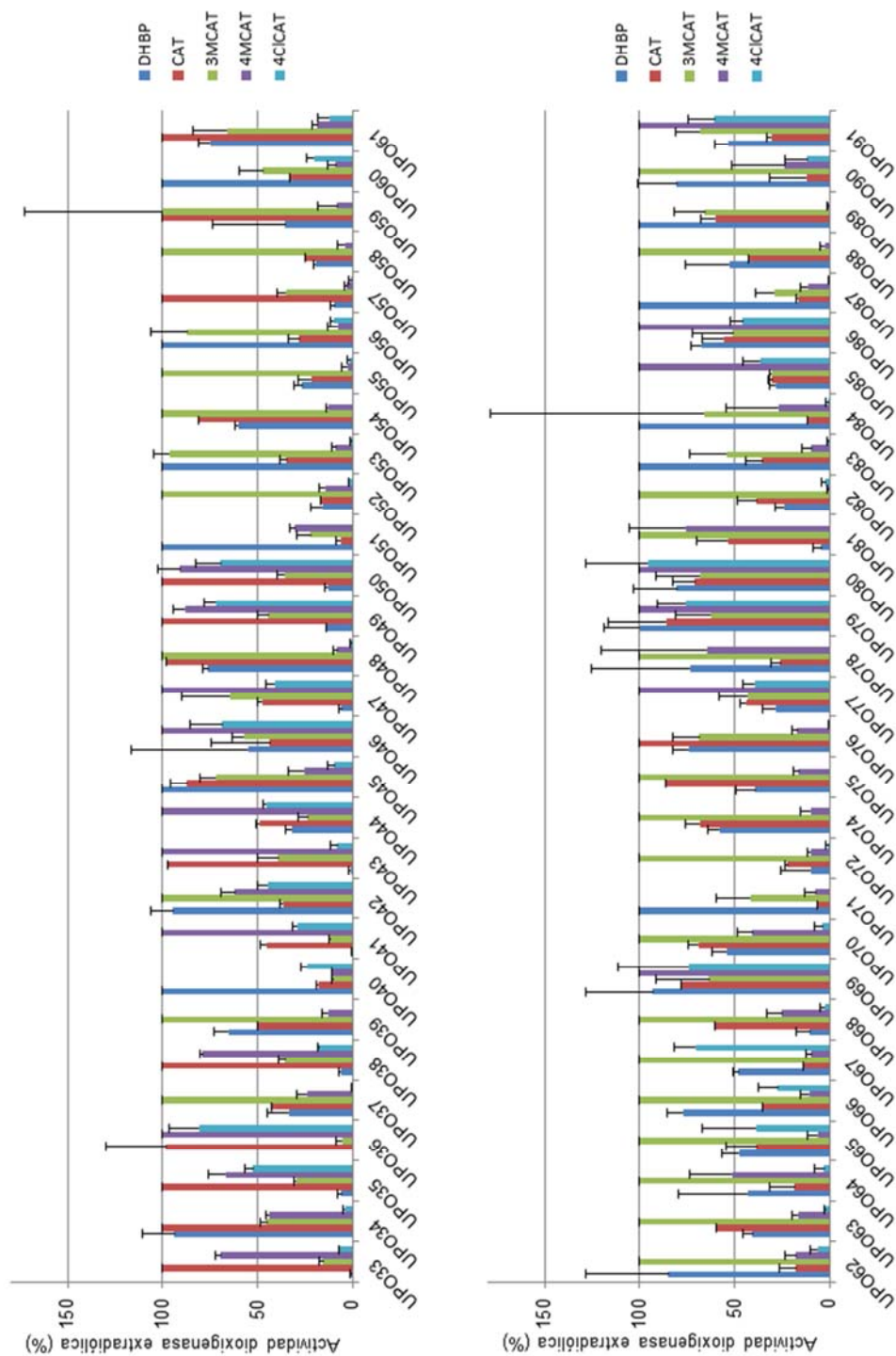


Figura 23. Actividad dioxigenasa extradiólica en los clones metagenómicos. Representación de la actividad dioxigenasa extradiólica de los 58 clones con 5 sustratos: 2,3-dihidroxibifenilo (2,3DHBP), catecol (CAT), 3-metilcatecol (3MCAT), 4-metilcatecol (4MCAT) y 4-clorocatecol (4CICAT). Se muestran los valores de actividad relativa con cada sustrato, considerando 100 % para cada clon el valor con el sustrato con el que presenta su máxima actividad. Los valores representados son la media de 3 réplicas. Las barras de error se corresponden con la desviación estándar de los datos.

Se tienen los valores de actividades relativas de los 58 clones con 5 sustratos (figura 23), y también las secuencias de las enzimas producidas por algunos de esos clones y las relaciones filogenéticas entre ellas y con otras de las bases de datos (figuras 20, 21 y 22). Se intentaron establecer relaciones entre estos datos, pero con las limitaciones de que se desconoce la secuencia de los genes codificantes de dioxigenasas extradiólicas de muchos de los clones y además al menos en algunos clones se identificaron dos genes (tabla 8), pudiendo ser el responsable de la actividad detectada alguno de los dos o una mezcla de ambos. A esto hay que añadir, al haber realizado los ensayos con los clones, en lugar de con subclones de los genes, que no se puede asegurar que la actividad se deba únicamente al gen detectado, sobre todo con los clones de los que no se conoce la secuencia completa, aunque es lo más probable.

Se ha descrito que las EDOs de la familia I.2 muestran preferencias por sustratos monocíclicos, mientras que las de la I.3 lo hacen por sustratos bicíclicos (Eltis y Bolin, 1996). Se comprobó si las nuevas proteínas incluidas en esas familias compartían esa característica. Por un lado, los 11 clones que contienen las 10 EDOs pertenecientes a la familia I.2 (uno de los genes se repite en 2 clones) presentaron los valores más altos de actividad con catecol o con 4-metilcatecol, ambos sustratos monocíclicos, mostrando con 2,3-dihibroxibifenilo (el único sustrato bicíclico utilizado) el valor mínimo en casi todos los casos (en el clon con UPOEDO13 fue el segundo menor valor), pudiendo por tanto decir que se cumple la tendencia descrita anteriormente. El gen repetido es el codificante de UPOEDO3, que se encuentra en los clones UPO49 y UPO50, compartiendo estos el mismo patrón de actividades.

Por otro lado, de los 7 clones que contienen las 5 EDOs pertenecientes a la familia I.3 (uno de los genes se repite en 3 clones), 6 de ellos presentaron valores altos con 2,3-dihidroxibifenilo (siendo su máxima actividad o la segunda mayor, siempre superior al 75 % respecto a la máxima), pero en casi todos los clones se mostraron también valores bastante altos con al menos otro de los sustratos (monocíclicos). La excepción fue el clon UPO37, que produce la enzima UPOEDO7 y que sin embargo mostró su segundo menor valor con 2,3-dihidroxibifenilo (34 % respecto a la máxima). Sin embargo, los clones UPO60 y UPO89, que producen la misma enzima UPOEDO7, mostraron su máximo valor con este sustrato. Estas diferencias se podrían explicar por la posibilidad de que los clones UPO60 y UPO89 contengan otro gen codificante de una EDO con distinta especificidad de sustrato no identificado, al no conocerse gran parte de la secuencia de sus insertos metagenómicos (de UPO37 sí se conoce la secuencia completa). También habría que indicar, que uno de los clones que produce una de las EDOs de la familia I.3, produce también otra proteína

catalogada como de tipo II. Se trata del clon UPO45, que contiene los genes codificantes de UPOEDO18 (familia 3 del tipo I) y de UPOEDO17 (tipo II), por lo que no se podría atribuir el patrón de actividad detectado a ninguna de ellas, pudiendo deberse a una mezcla de ambas.

Al observar los patrones de actividad de los 7 clones que producen las 3 EDOs de la familia I.4 (uno de los genes se repite en 4 clones y otro en 2) se encontraron patrones significativamente distintos entre clones que producen diferentes EDOs. Uno de los genes repetidos es el codificante de UPOEDO5, que se encuentra en los clones UPO52, UPO55, UPO58 y UPO72, siendo los patrones de actividad de estos cuatro clones muy similares. Y el otro es el gen codificante de UPOEDO23, que se encuentra en los clones UPO68 y UPO87, teniendo en este caso estos clones patrones de actividad diferentes, lo que podría deberse a la existencia de algún otro gen codificante de una EDO con distinta especificidad de sustrato no identificado en alguno de los clones, al no conocerse la secuencia completa de ninguno de ellos.

Por último, habría que indicar que los 4 clones que producen las 5 EDOs pertenecientes al tipo II (dos de los genes se encuentra en el mismo clon) presentaron patrones de actividad muy diferentes. El clon UPO64 produce dos de estas proteínas (UPOEDO1 y UPOEDO2), por lo que no se podría atribuir el patrón de actividad detectado a ninguna de ellas, ya que puede deberse a una mezcla de ambas. Lo mismo ocurre con el clon UPO45 comentado anteriormente, que contiene los genes codificantes de UPOEDO17 (tipo II) y de UPOEDO18 (familia 3 del tipo I).

Comparando el perfil de actividad con los sustratos que se utilizaron para los rastreos, catecol y 2,3-dihidroxibifenilo, se encontró que la mayoría (12 de 17) de los clones identificados con catecol (UPO33-UPO49), presentó más actividad con catecol que con 2,3-dihidroxibifenilo, lo que indica que el catecol utilizado como sustrato para el rastreo puede sesgar la especificidad relativa de las EDOs que puedan encontrarse de esta manera. Sin embargo, entre los detectados con 2,3-dihidroxibifenilo (UPO35, UPO42, UPO48, UPO50-UPO72 y UPO74-UPO91) no se apreció un sesgo similar (24 clones presentaron mayor actividad con 2,3-dihidroxibifenilo y 20 con catecol). De hecho, la mayoría de los clones presentó su máxima actividad con un sustrato distinto a catecol y a 2,3-dihidroxibifenilo. Como se muestra en la figura 24, el 3-metilcatecol fue el mejor sustrato para la mayoría de clones (24 clones), seguido por el 4-metilcatecol (13 clones), mientras que el catecol y el 2,3-dihidroxibifenilo estaban prácticamente igualados (10 y 11 clones, respectivamente), y ninguno de los clones presentó su máxima actividad con 4-clorocatecol. De todas formas, la

mayoría de estos clones (31 del total de 58) mostraron una actividad frente a 2,3-dihidroxibifenilo muy significativa, por encima del 50 %, del mismo modo que ocurre con casi la mitad de los clones (26 del total de 58) frente a catecol. Teniendo en cuenta que no se ha apreciado un sesgo significativo y que analizando un mismo número de colonias se identificaron más clones con actividad dioxigenasa extradiólica utilizando 2,3-dihidroxibifenilo que catecol (44 frente a 17), la incorporación del 2,3-dihidroxibifenilo como sustrato en este tipo de rastreos se plantea como una buena opción. A su vez, otros sustratos como el 3-metilcatecol, que es el mejor sustrato para la mayor parte de los clones, podría ser una molécula adecuada para realizar este tipo de rastreos funcionales. Quizá, la utilización de una mezcla de sustratos podría ser la mejor alternativa.

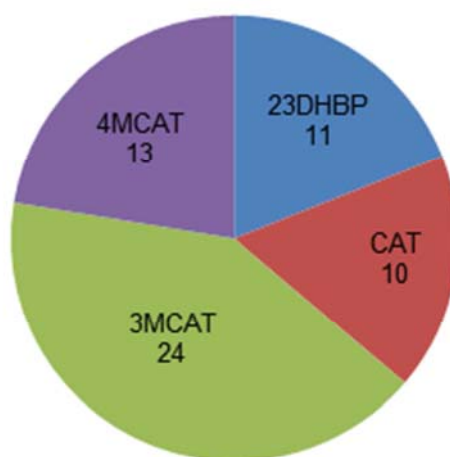


Figura 24. Representación del número de clones que presenta su máxima actividad dioxigenasa extradiólica con cada sustrato. Los ensayos se realizaron con 5 sustratos: 2,3-dihidroxibifenilo (2,3DHBP), catecol (CAT), 3-metilcatecol (3MCAT), 4-metilcatecol (4MCAT) y 4-clorocatecol (4CICAT), con 58 clones.

Trabajos previos en los que se ha comparado la actividad de distintas enzimas EDO con diferentes sustratos, han mostrado una baja preferencia por el sustrato 4-clorocatecol (por ejemplo, Hatta y col., 2003; Suenaga, y col., 2007). Por su escasez entre las bacterias degradadoras conocidas y entre las actividades EDO identificadas por aproximación metagenómica, consideramos la actividad frente a 4-clorocatecol una actividad interesante ya que puede ser importante para la degradación eficaz de hidrocarburos aromáticos clorados. En nuestro caso, ninguno de los clones en estudio presentó su máxima actividad con 4-clorocatecol (figura 24). Sin embargo, 2 de ellos mostraron más del 80 % de su máxima actividad con este sustrato. Uno es UPO80 (96 %), del que no se conoce nada de su secuencia; y el otro es UPO36 (81 %), que sí tiene asignada secuencia. Aunque no se conoce la secuencia

completa del inserto en UPO36, se tienen 35,7 kb, por lo que debe de faltarle poco, siendo improbable que contenga dos genes codificantes de la actividad EDO. Por ello, el gen codificante de EDO asignado (*upoedo11*) sería probablemente el responsable de la actividad. Se decidió comprobar si efectivamente ese gen presenta un alto porcentaje de actividad con 4-clorocatecol respecto a su máximo.

4.3 Subclonación del gen *upoedo11*

El gen *upoedo11* se subclonó para confirmar que es el responsable del patrón de actividades con distintos sustratos visto en el clon UPO36. Además de saber si la proteína codificada por ese gen presentaba una preferencia de sustrato alta por 4-clorocatecol, se pretendía averiguar si en términos absolutos esa actividad era alta en condiciones de superproducción, ya que la actividad detectada en general y en particular con 4-clorocatecol fue muy baja: 0,685 nmol/(min·D.O.₆₀₀). El gen *upoedo11* se obtuvo por PCR y se clonó en el vector de superproducción pET-14b (Novagen) como se indica en el apartado 5.2 de Materiales y Métodos, de manera que el gen codificante de la dioxigenasa extradiólica quedaba a continuación y en el mismo sentido que el promotor pT7. El subclón resultante se denominó pMPO1383.

El subclón pMPO1383 se transfirió a la estirpe superproductora de *E. coli* NCM631/pIZ227 (Govantes y Santero, 1996), y en condiciones de superproducción de la proteína (con una inducción de los cultivos a D.O.₆₀₀ en torno a 0,7 con IPTG 1 mM a 26 °C y 12 h en agitación a 180 r.p.m., como se indica en el apartado 11.1 de Materiales y Métodos), se ensayó con los mismos 5 sustratos que los 58 clones (2,3-dihidroxibifenilo, catecol, 3-metilcatecol, 4-metilcatecol y 4-clorocatecol). Los resultados del ensayo de 3 réplicas indicaron que el gen *upoedo11* codifica una proteína que, al ser ensayada con estos 5 sustratos, presenta su máxima actividad con 4-metilcatecol, y un 80 % del valor de esa actividad con 4-clorocatecol. En términos absolutos la actividad con 4-clorocatecol fue 1,177 nmol/(min·D.O.₆₀₀), más elevada que en el clon UPO36 pero que continúa siendo muy baja. Mediante visualización en gel de poliacrilamida-SDS se comprobó que la inducción había sido efectiva y se había producido superproducción de la proteína (figura 25), deduciéndose que tenía una actividad específica muy baja o que se había producido inactiva en su mayor parte, por lo que no se continuó estudiando la proteína.

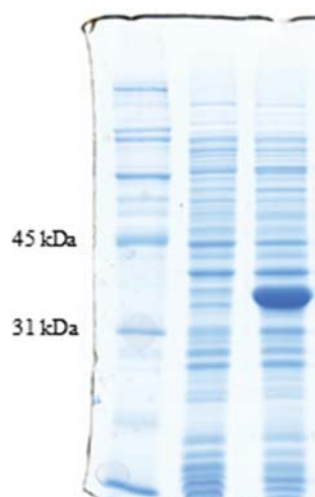


Figura 25. Visualización en gel de poliacrilamida-SDS de la superproducción de la proteína UPOEDO11. Proteínas totales de un cultivo de la estirpe superproductora *E. coli* NCM631/ pIZ227 con el vector vacío pET-14b (segundo carril) o con el plásmido pMPO1383, que contiene el gen *upoedo11* (tercer carril). El marcador de peso molecular (primer carril) empleado es marcador B de Ready Gel Bands (Bio-Rad). La proteína UPOEDO11 tiene un peso de 33,88 kDa.

5. CARACTERIZACIÓN DE CLONES CON ACTIVIDAD OXIGENASA HIDROXILANTE DE ANILLOS AROMÁTICOS

En la metagenoteca de pilas de refinería de la industria petroquímica se identificaron 36 clones distintos con posible actividad oxigenasa hidroxilante de anillos aromáticos (también llamada oxigenasa inicial), denominados UPO07-UPO31, UPO41, UPO55, UPO92-UPO94 y UPO106-UPO111. Esos clones se identificaron porque sus colonias eran de color azul debido a su capacidad de transformar el triptófano del medio LB para producir finalmente índigo. Algunos de los clones presentaban un color azul muy evidente, mientras que en otros, de 7 a 9 clones, el color azul era distinto puesto que se trataba de un color más oscuro virando hacia el marrón, lo que hacía dudar sobre si la coloración se debía realmente a la producción de índigo. Sin embargo, puesto que el color podría deberse a una producción significativa de indirrubina (color rojizo) además de índigo, se mantuvieron como clones potencialmente codificantes de oxigenasas hidroxilantes.

Los rastreos se realizaron en la estirpe MPO554 NaI^R en presencia de salicilato, es decir, con expresión heteróloga del ADN metagenómico desde el promotor *psal* del fósido y con la antiterminación de la transcripción proporcionada por la proteína N. 29 de los 36 clones considerados positivos en el rastreo de producción de índigo se transfirieron a las distintas estirpes: EPI300-T1^R (estirpe original, en la que no se puede producir expresión heteróloga), MPO553 (en la que el ADN metagenómico se puede expresar heterológamente desde el promotor pT7) y MPO554, y se comprobó mediante visualización de colonias si se producía color azul. Los 29 clones produjeron color azul en la estirpe MPO554 en presencia de salicilato, 22 de ellos en la estirpe MPO553, y 18 de los clones en EPI300-T1^R, los mismos que en MPO554 sin aportar salicilato. Por tanto, solo 18 de los 29 clones fueron capaces de expresar los genes responsables de la actividad en niveles detectables desde sus propios promotores, y de haber realizado los rastreos con el sistema original (únicamente con la expresión propia de los genes metagenómicos), muy probablemente se habrían identificado bastantes menos clones con la actividad de interés.

5.1 Secuenciación de clones con actividad oxigenasa

Se secuenciaron 25 de los 36 clones identificados con aparente actividad oxigenasa hidroxilante. Por un lado se secuenciaron 17 individualmente, y por otro 8 en 3 *pools*, junto con otros 52 clones portadores de otras actividades, en su mayoría dioxigenasa extradiólica.

Por tanto, las secuencias metagenómicas de estos últimos 8 clones se obtuvieron en *contigs* a priori no asignados a ningún clon concreto. En la tabla 7 se recopila información sobre cada una de estas tandas de secuenciación en *pools*. Entre los 25 clones secuenciados se encontraban 2 de los clones de los que se tenían dudas sobre su producción de índigo por su color marrón (UPO29 y UPO31).

5.1.1 Identificación de genes de oxigenasas y asignación a clones

En el análisis de las secuencias de los *contigs* de las 3 tandas de secuenciación en *pools*, se identificaron muchas más secuencias homólogas a genes codificantes de oxigenasas iniciales de las esperadas, ya que solo 8 de los fósmidos que habían sido detectados como productores de índigo se habían incluido en la secuenciación en *pools*. De hecho, tras analizar las secuencias de los *contigs* con genes codificantes de dioxigenasas extradiólicas asignados a 28 clones que presentaban esa actividad (apartado 4.1.1 de Resultados), se observó que, efectivamente, en 20 de esos 28 clones había también genes codificantes de oxigenasas iniciales. Sin embargo, solo 2 de esos clones presentan ambas actividades (UPO41 y UPO55), lo que indicaría que buena parte de los genes de oxigenasas hidroxilantes no están expresándose en *E. coli*, no son funcionales o no son capaces de reconocer el indol como sustrato. Otras decenas de *contigs* tenían también genes codificantes de oxigenasas que podrían corresponder a cualquiera de los 8 fósmidos que proporcionaron color azul a las colonias, pero en vista a lo comentado también podrían pertenecer a fósmidos que solo proporcionaron actividad dioxigenasa extradiólica a los que no se les hubiera asignado secuencia del inserto completo. No se trataron de asignar todos los genes codificantes de oxigenasas iniciales identificados en los *contigs* a algún fósrido, sino que nos centramos en los que codifican citocromos P450.

Se identificaron 7 citocromos P450 en los *contigs* de las secuenciaciones por *pools*. Para adscribirlos a clones concretos, se diseñaron oligonucleótidos que pudieran amplificar un fragmento de esos genes y se realizaron PCR con todos los clones de la tanda de secuenciación correspondiente, se hubiera detectado o no en ellos actividad oxigenasa inicial. Aquellos clones con los que se obtuvo amplificación, se confirmaron mediante secuenciación con uno de los oligonucleótidos empleados en la PCR. De esas secuencias se obtuvieron 6 citocromos P450 completos y otro truncado (con un codón de parada aparecido antes del final original de la proteína), que se asignaron a 6 clones (varios genes fueron asignados a un

mismo clon y varios clones compartían un mismo gen). Solo uno de esos clones presenta actividad oxigenasa inicial: UPO28, que contiene 3 citocromos P450. Los otros no presentan dicha actividad, y sin embargo se les asignó un gen codificante de citocromo P450 a cada uno: UPO60, UPO61, UPO37, UPO87 y UPO89. UPO37 y UPO89 comparten el mismo citocromo, y UPO60 posee un gen truncado. La secuencia de varios de estos genes no estaba completa en sus respectivos *contigs* originales, y hubo que obtener el resto de la secuencia mediante posteriores reacciones de secuenciación. Al igual que con los genes codificantes de dioxigenasas extradiólicas (apartado 4.1.1 de Resultados), se observó que esto ocurrió cuando varios genes pertenecientes a la misma tanda de secuenciación eran muy parecidos, probablemente por la incapacidad del programa de ensamblaje de *contigs* de discriminar cuáles eran las uniones adecuadas.

Por tanto, de los 8 clones con actividad oxigenasa que fueron secuenciados en *pools*, solo se le pudo asignar secuencia y posible gen o genes responsables de la actividad a 3 de ellos (UPO28, UPO41 y UPO55) (tabla 9). A UPO28 se le asignó secuencia que contenía genes codificantes de citocromos P450, como se acaba de comentar. A UPO41 y UPO55, que presentaban tanto actividad oxigenasa inicial como dioxigenasa extradiólica, se les habían asignado *contigs* que contenían genes codificantes de dioxigenasas extradiólicas (apartado 4.1.1 de Resultados), y tras el análisis de estos, se identificaron también genes de una oxigenasa inicial en cada caso. De uno de los 3 clones indicados se obtuvo la secuencia completa (UPO41), pero de los otros 2 no.

Por otro lado, de los 17 clones secuenciados individualmente, se obtuvo la secuencia completa de 14 de ellos (UPO08-UPO14, UPO19, UPO21, UPO23-UPO26 y UPO29), mientras que de los otros 3 estaba incompleta (UPO16, UPO30 y UPO31) (tabla 9). Se identificó al menos un posible gen responsable de la actividad en todos los clones salvo en 2 (UPO29 y UPO31). Del clon UPO31 no se tenía la secuencia completa por lo que el gen codificante de la actividad podría estar localizado en la región de secuencia desconocida. Sin embargo, del clon UPO29 conocíamos la secuencia completa. Los dos clones compartían secuencia parecida y el único gen que pudiera potencialmente codificar una actividad oxidante en estos fósmidos era uno, exactamente igual en ambos clones, codificante de una proteína con similitud a sarcosina oxidasas, enzimas implicadas en la desmetilación oxidativa de sarcosina (N-metilglicina). Es posible que estas enzimas oxidasen algún sustrato o intermediario metabólico, produciendo la coloración oscura detectada en la colonia.

| clon | modo de secuenciación | secuencia asignada |
|-------|-----------------------|----------------------------------|
| UPO08 | individual | 27 kb (completa) |
| UPO09 | individual | 36 kb (completa) |
| UPO10 | individual | 33 kb (completa) |
| UPO11 | individual | 31 kb (completa) |
| UPO12 | individual | 39 kb (completa) |
| UPO13 | individual | 35 kb (completa) |
| UPO14 | individual | 41 kb (completa) |
| UPO16 | individual | 33 kb |
| UPO19 | individual | 42 kb (completa) |
| UPO21 | individual | 34 kb (completa) |
| UPO23 | individual | 33 kb (completa) |
| UPO24 | individual | 41 kb (completa) |
| UPO25 | individual | 36 kb (completa) |
| UPO26 | individual | 39 kb (completa) |
| UPO28 | en <i>pool</i> | 2 kb + 24 kb (2 <i>contigs</i>) |
| UPO30 | individual | 36 kb |
| UPO41 | en <i>pool</i> | 25 kb (completa) |
| UPO55 | en <i>pool</i> | 3 kb |

Tabla 9. Relación de clones con actividad oxigenasa hidroxilante detectada y con posible gen responsable de la actividad asignado, e información sobre su secuenciación. Se indican los clones con actividad oxigenasa hidroxilante detectada por formación de colonias de color azul (producción de índigo) y con secuencia asignada con al menos un gen codificante de posible oxigenasa responsable de la actividad. Se muestra el modo de secuenciación (si esta ha sido individual o si el clon formaba parte de un *pool* de secuenciación), la longitud de secuencia asignada, y cuando esta constituye la totalidad del inserto se indica “completa”. A menos que se indique lo contrario, la cantidad de secuencia indicada se corresponde con un solo *contig*. No están incluidos los clones con dudosa producción de índigo UPO29 y UPO30, en los que no se identificaron oxigenasas hidroxilantes. Tampoco se incluyen los clones a los que se les asignó gen codificante de un citocromo P450 pero en los que no se observó actividad oxigenasa hidroxilante (UPO60, UPO61, UPO37, UPO87 y UPO89. UPO37 y UPO89).

Recapitulando, en total, el análisis de las secuencias permitió identificar 18 genes que en principio podían ser responsables de la actividad oxigenasa inicial, adscribiéndose a 18 clones (algunos clones contenían más de un gen y varios genes se encontraban en más de un clon). De 14 de esos clones se obtuvo la secuencia completa, y de 4 no completa (tabla 9). A continuación se recopila la información sobre los genes codificantes de oxigenasas iniciales y los clones en los que se identificaron, considerando solo los genes de este tipo que se localizaron en clones que presentaban actividad productora de índigo.

5.1.1.1 Dioxigenasas

De todos los 18 genes codificantes de oxigenasas putativas, solo 4 de ellos son homólogos a genes codificantes de dioxigenasas. Estos genes parecen codificar concretamente las subunidades α de dioxigenasas.

Hay 2 grupos de clones con genes codificantes de subunidades α de dioxigenasas. Por un lado UPO08 y UPO11 y por otro lado UPO12, UPO13, UPO14 y UPO25 (figura 26). Los clones de cada uno de los 2 grupos se podría decir que comparten gran parte del inserto, con solo algunos nucleótidos diferentes que no afectan a los posibles genes responsables de la actividad. Además de los genes codificantes de las subunidades α , se aprecian en estos clones genes codificantes de otras posibles oxigenasas, como se explica más adelante. Los dos grupos de fósidos son bastante parecidos entre sí, manteniendo la misma distribución de los genes de interés y mostrando similitud entre las proteínas codificadas por estos. Concretamente, los genes de interés, indicados en la figura 26, en uno y otro grupo de clones presentan identidades de alrededor del 90 %. En cada grupo de clones se encuentran 2 genes codificantes de subunidades α de dioxigenasas. Estas son similares a otras de dioxigenasas que contienen subunidades α y β , por lo que se esperaría que las acompañaran genes de subunidades β , pero no se encontraron en ninguno de los fósidos. Puesto que se conoce la secuencia completa de todos estos clones (tabla 9), el no haber encontrado ningún gen homólogo a los codificantes de subunidades β en ninguno de ellos indica que o bien no están o, menos probable, son tan distintos a los existentes en las bases de datos que no se reconocen como tales. En base a esto, y suponiendo por tanto que no hay genes codificantes de subunidades β en estos clones, es muy probable que las subunidades α de esos clones no sean funcionales por sí mismas. En cada grupo de clones se identificaron además 2 genes homólogos a monooxigenasas putativas.

Teniendo en cuenta la distribución de los genes indicados en la figura 26, probablemente la primera de las monooxigenasas putativas o la primera de las dioxigenasas subunidad α (las únicas que se encuentran presentes en UPO25) sea la responsable de la actividad oxigenasa inicial en estos clones (sin descartar que alguno de los otros genes indicados en la figura 26 también pudiera proporcionar la misma actividad). Como se ha comentado anteriormente, la dioxigenasa subunidad α es poco probable que sea funcional por sí misma. Por tanto, lo más probable es que el responsable sea el otro gen, cuyo producto génico presenta homología con monooxigenasas (identidad ≤ 43 % con monooxigenasas putativas). La monooxigenasa a la que más se parecen de las que han sido caracterizadas es a la estireno monooxigenasa StyA de *Pseudomonas* sp., con la que presentan una identidad del 27 y 26 % (esa monooxigenasa es ligeramente diferente en cada uno de los dos grupos de clones). En estos clones también se identificaron genes codificantes de componentes de sistemas de transporte de electrones, no mostrados en la figura 26.

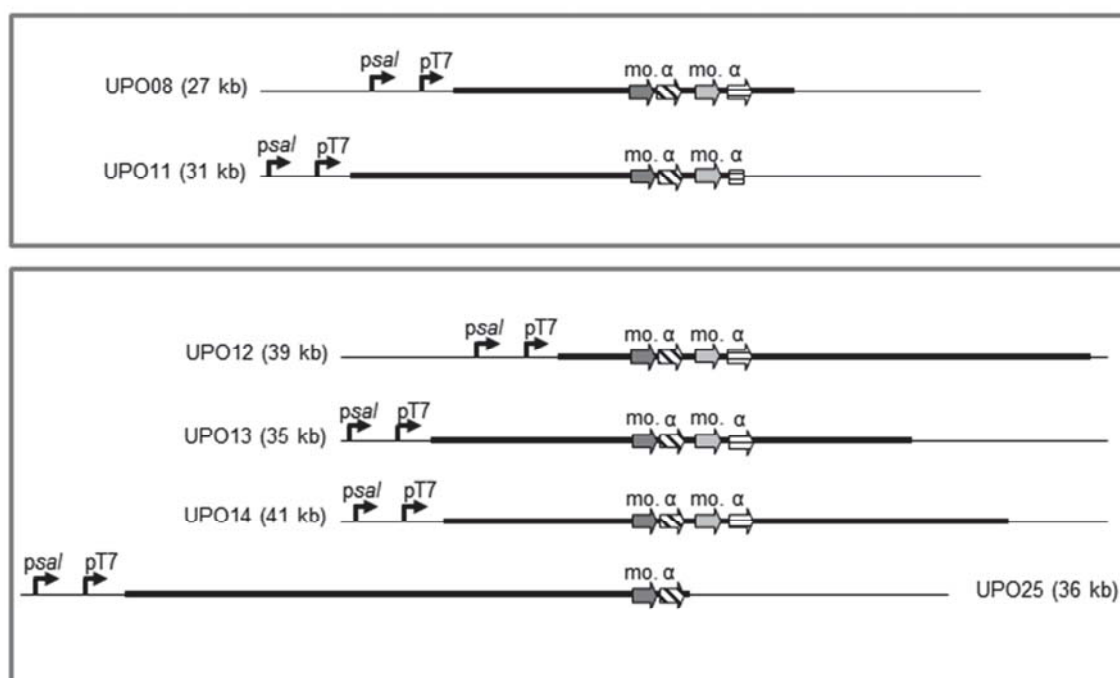


Figura 26. Esquema de los clones con genes codificantes de dioxigenasas subunidad α . Dentro de cada recuadro se incluyen los clones que comparten parte de la secuencia, salvo por algunas diferencias en nucleótidos puntuales. Los clones de un recuadro comparten la estructura génica con los del otro, y los genes resaltados en cada uno de los recuadros presentan una identidad de alrededor del 90 % con el gen equivalente en el otro recuadro. Se resaltan posibles genes involucrados en la actividad oxigenasa inicial. Las abreviaturas de la figura indican lo siguiente: mo.: gen codificante de monooxigenasa putativa; α : gen codificante de dioxigenasa subunidad α . Los genes no están representados a escala.

5.1.1.2 Monooxigenasas

De los 18 genes codificantes de oxigenasas putativas identificados en los clones con dicha actividad, 14 de ellos son homólogos a genes codificantes de monooxigenasas y los otros 4 restantes corresponden a las subunidades α de dioxigenasas ya comentadas. Entre esos 14 genes se encuentran los 2 codificantes de las monooxigenasas de las que se habla en el apartado anterior como posibles responsables de la actividad (la primera de las indicadas en la figura 26 en cada grupo de clones). Los distintos tipos de monooxigenasas en función de su mayor similitud de secuencia con otras monooxigenasas de las bases de datos, así como los fósidos que las codifican, se recogen en la tabla 10.

| Monooxigenasas iniciales | Clones |
|---|----------------------------|
| Estireno monooxigenasa 1 | UPO08, UPO11 |
| Estireno monooxigenasa 2 | UPO12, UPO13, UPO14, UPO25 |
| Alcanosulfonato monooxigenasa 1 | UPO09, UPO19 |
| Alcanosulfonato monooxigenasa 2 | UPO23 |
| Dimetilnilina monooxigenasa | UPO26 |
| 2,4-diclorofenol 6-monooxigenasa | UPO55 |
| Fenol 2-monooxigenasa 1, componentes P0-P5 | UPO16 |
| Fenol 2-monooxigenasa 2, componentes P0-P5 | UPO41 |
| Dibenzotiofeno-5,5-dióxido monooxigenasa | UPO21 |
| Dibenzotiofeno monooxigenasa | UPO21 |
| Citocromo P450 1 (UPOCYTP1) + ferredoxina A | UPO10, UPO24 |
| Citocromo P450 2 (UPOCYTP2) + ferredoxina B | UPO28, UPO30 |
| Citocromo P450 3 (UPOCYTP3) + ferredoxina C | UPO28, UPO30 |
| Citocromo P450 4 (UPOCYTP4) + ferredoxina C | UPO28 |

Tabla 10. Monooxigenasas codificadas por genes de clones con actividad oxigenasa y clones en los que se ubican.

En cada uno de los clones indicados en la tabla 10 se identificó al menos un gen codificante de una monooxigenasa inicial. En uno de ellos, UPO55, del que solo se conocen unas 3 kb de la secuencia, del gen codificante de una monooxigenasa identificado (2,4-

diclorofenol 6-monooxigenasa) no se obtuvo la secuencia completa, pero puede que el gen sí esté completo en el inserto porque esta secuencia es interna, no forma parte de un extremo del inserto de este fósido. Por otra parte, al conocer tan poco de su secuencia puede que en ese clon haya además otro(s) gen(es) codificante(s) de otra(as) oxigenasa(s). Del resto de estos clones se conoce la secuencia completa o casi completa, por lo que resulta improbable que haya otros genes codificantes de oxigenasas adicionales y, por tanto, los genes identificados, o al menos uno de ellos en los casos en los que se asignaron varios genes, muy probablemente sean los responsables de la actividad oxigenasa inicial. En el caso de los clones UPO08, UPO11, UPO12, UPO13, UPO14 y UPO25 sí se identificaron además genes de subunidades α de dioxigenasas iniciales, pero por los motivos expuestos en el apartado anterior, probablemente la monooxigenasa indicada en la tabla 10 sea la responsable de la actividad.

En los clones UPO09, UPO19, UPO23 y UPO26, además de las monooxigenasas indicadas se identificaron ferredoxinas reductasas, que probablemente constituyan su sistema de transporte de electrones. Y en los clones UPO08, UPO11, UPO12, UPO13, UPO14 y UPO25 se encontraron una reductasa y una ferredoxina fusionada a un transportador tipo RND como posibles componentes del sistema de transporte de electrones.

A diferencia de otras monooxigenasas, la fenol 2-monooxigenasa es una enzima multicomponente (Shingler y col., 1989). En los 2 clones en los que se identificaron genes de este tipo (UPO16 y UPO41) se encuentran los genes de los 6 componentes requeridos para la transformación por parte de la bacteria de fenol en catecol (P0-P5, Nordlund y col., 1990), incluyendo el sistema de transporte de electrones. En el clon UPO41, además está presente el gen codificante de una catecol 2,3-dioxigenasa, localizado justo aguas abajo (a 196 pb) del componente P5 de la fenol 2-monooxigenasa, en el mismo sentido, presuntamente implicada en esta ruta de degradación de fenol, cuya actividad extradiol dioxigenasa frente a diferentes sustratos también se estudió (*upoedo19*, apartado 4.2 de Resultados). Los componentes proteicos P0-P5 de la fenol 2-monooxigenasa de UPO41 presentan identidades que se encuentran entre el 40 y el 64 % con los codificados por los genes *dmpKLMNOP* de *Pseudomonas* sp. CF600 (Shingler y col., 1989). Por otro lado, UPOEDO19 presenta una identidad del 47 % con la enzima catecol 2,3-dioxigenasa de la ruta de degradación de fenol descrita por Shingler y col. (1992) de esta estirpe codificada por el gen *dmpB*. En este mismo clon encontramos a su vez genes codificantes de una 2-hidroxiacetato semialdehído deshidrogenasa, localizado 2,7 kb aguas abajo del gen codificante de la catecol 2,3-dioxigena; de una 2-oxopent-4-enoato hidratasa, separado 12 pb del anterior; y de una 4-oxalocrotonato

descarboxilasa, solapando 4 pb con el anterior, encontrándose todos estos genes en el mismo sentido. Además, en el mismo sentido que estos genes pero 2,1 kb aguas arriba del primero de los componentes de la fenol 2-monooxigenasa, se encuentra el gen codificante de una 4-oxalocrotonato tautomerasa. Con todos estos genes mencionados, se podría obtener 4-hidroxi-2-oxoalerato a partir de fenol, faltando solo el gen codificante de la 4-hidroxi-2-oxoalerato aldolasa para la obtención de acetaldehído y piruvato, según la ruta de rotura-*meta* de degradación de fenol (rama de oxalocrotonato) descrita en Sala-Trepap y col. (1972) y Shingler y col. (1989).

Por otro lado, en otro de los clones (UPO21) se encuentra la ruta de desulfuración de dibenzotiofeno completa, al estar presente el gen codificante de la 2-hidroxibifenilo-2-sulfinato desulfinaasa, además de las 2 monooxigenasas indicadas en la tabla 10 (Denome y col., 1994; Gray y col., 1996). Junto a los genes de la ruta, se identificó el gen codificante de una reductasa, que podría constituir el sistema de transporte de electrones de las monooxigenasas.

Se identificaron 4 genes codificantes de citocromos P450 (tabla 10). En todos los casos estos genes van acompañados de genes codificantes de ferredoxinas, aunque no de ferredoxinas reductasas. Teniendo en cuenta que las ferredoxinas reductasas suelen ser bastante inespecíficas, los electrones necesarios para la reacción productora de índigo podrían obtenerse a partir de ferredoxinas reductasas de la propia bacteria hospedadora. Los clones UPO10 y UPO24 comparten ADN metagenómico y contienen uno de estos citocromos, mientras que los clones UPO28 y UPO30 comparten otros 2 cada uno. Las ferredoxinas acompañantes de estos 3 citocromos P450 distintos son a su vez distintas. Por último, en el clon UPO28 se encontró otro gen codificante de citocromo P450, acompañado también por un gen codificante de ferredoxina, en este caso igual a una de las otras dos ferredoxinas identificadas en este fósido. Puesto que no se conoce la secuencia completa de UPO28 y UPO30 (tabla 9) es posible que aun haya más genes codificantes de citocromos P450. Al haber varios genes codificantes de citocromos P450 en un mismo clon (UPO28 y UPO30), no se puede asegurar que todos presenten actividad productora de índigo.

Las secuencias de las proteínas de estos clones, homólogas a monooxigenasas hidroxilantes de las bases de datos, se recogen en el Anexo.

5.1.2 Análisis filogenético de los citocromos P450

Se realizó un estudio filogenético con las secuencias de los 4 citocromos P450 codificados por genes de los clones con actividad oxigenasa inicial (tabla 10), más las de los otros 3 citocromos P450 de los clones que tenían actividad dioxigenasa extradiólica pero no productora de índigo (apartado 5.1.1). Aquellos presentes en clones con actividad oxigenasa inicial son UPOCYTP1, UPOCYTP2, UPOCYTP3 y UPOCYTP4 (tabla 10). Por otro lado, las proteínas de este tipo presentes en clones en los que no se ha detectado dicha actividad se denominaron UPOCYTP5 (en el clon UPO61), UPOCYTP6 (en UPO87) y UPOCYTP7 (en UPO37 y UPO89). En otro de los clones sin actividad oxigenasa detectada también se identificó un gen codificante de un citocromo P450 que se encuentra truncado (con un codón de parada aparecido tras 352 aminoácidos, estimándose que deben de faltar unos 89, al comparar con las proteínas más parecidas de las bases de datos). Este gen truncado, denominado UPOCYTP8 (en UPO60), es igual a otro de los genes hasta el codón de parada aparecido, y no se tuvo en cuenta para los siguientes análisis. No se comprobó que los distintos citocromos P450 fueran funcionales, pero esto es bastante probable en aquellos pertenecientes a clones con actividad, aunque en los casos en los que hay varias proteínas de este tipo en un mismo clon, puede que la actividad solo se deba a una de ellas. En los demás casos puede que no fueran funcionales, que les faltara el sistema de transporte de electrones, que no se estuvieran expresando o que no reconocieran indol como sustrato. En los clones UPO37 y UPO89, los genes codificantes de citocromos P450 se encuentran en el sentido de los promotores heterólogos del vector, pero en los clones UPO61 y UPO87 se encuentran en sentido contrario, por lo que no se estaría consiguiendo transcripción heteróloga. En ninguno de estos 5 clones sin actividad productora de índigo se identificaron genes de los componentes del sistema de transporte de electrones, aunque solo se conocía la secuencia completa del inserto de uno de estos clones (UPO37) (tabla 8).

Los citocromos P450 constituyen un grupo de proteínas muy heterogéneo, con representantes no solo en microorganismos sino también en animales y plantas. En el árbol filogenético realizado (figura 27), se observa que la heterogenidad de la secuencia de estas proteínas no siempre se ajusta a la adscripción taxonómica del organismo, ya que por ejemplo se encuentran secuencias de hongos (CYP55A1 de *Fusarium oxysporum*) más parecidas a secuencias de bacterias que a secuencias de otros hongos. A su vez, secuencias de la misma especie pueden ser muy diferentes (como CYP51 y CYP125A1 de *Mycobacterium tuberculosis*). En cualquier caso, las 7 secuencias de citocromos P450 identificados en el

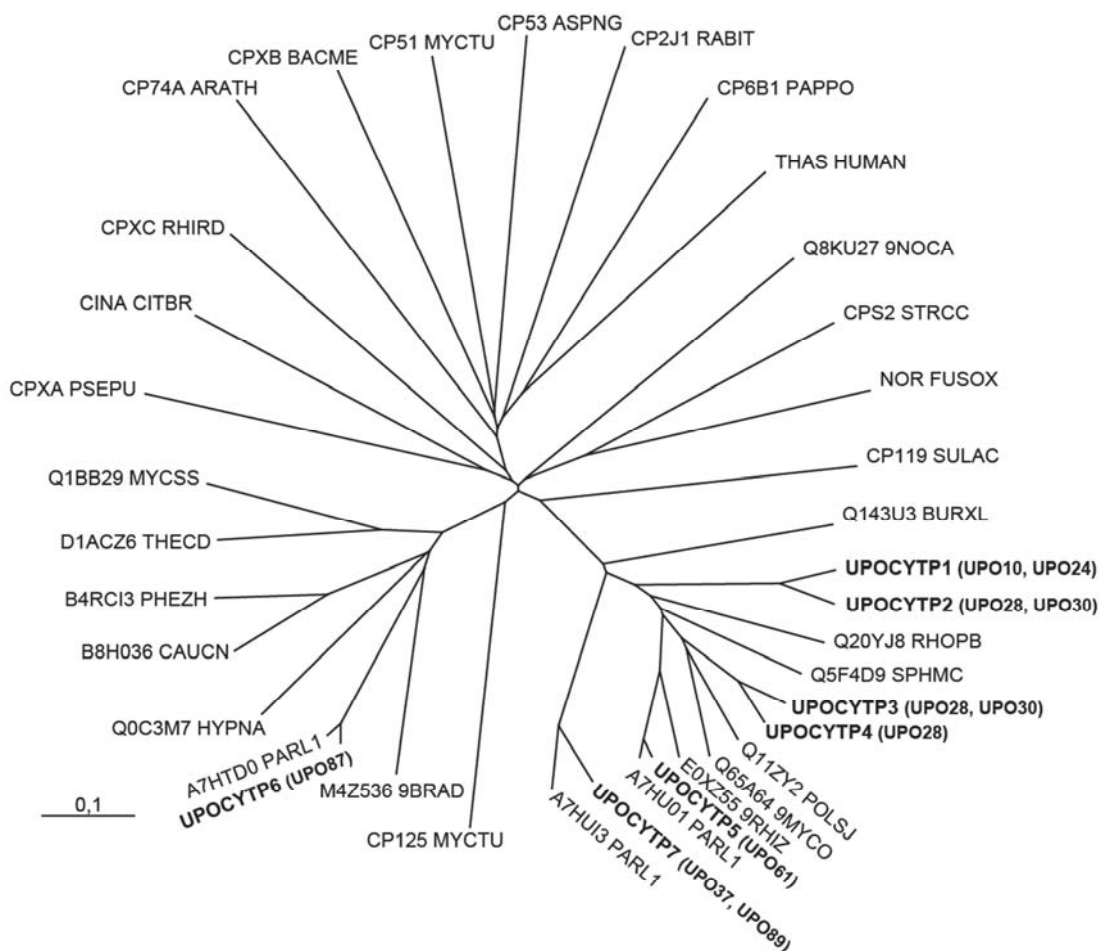


Figura 27. Análisis filogenético de las secuencias metagenómicas de citocromos P450.

Árbol filogenético resultante del alineamiento de las secuencias de las proteínas de citocromos P450 identificadas en la metagenoteca de pilas de refinería de la industria petroquímica con otras secuencias del mismo tipo previamente identificadas. La escala representa la distancia expresada como diferencia en aminoácidos por posición. Las secuencias metagenómicas identificadas en este trabajo se muestran en negrita, acompañadas del clon en el que se encuentran. El resto de secuencias empleadas en el dendrograma y sus códigos de acceso en las bases de datos GenBank, NCBI Reference Sequence o UniProtKB/Swiss-Prot se indican a continuación: CPXA PSEPU: CYP101, *Pseudomonas putida* (P00183); CPS2 STRCC: CYP105A3, *Streptomyces carbophilus* (Q59831); CINA CITBR: CYP176A1, *Citrobacter braakii* (Q8VQF6); CP119 SULAC: CYP119, *Sulfolobus acidocaldarius* (Q55080); CP51 MYCTU: CYP51, *Mycobacterium tuberculosis* (P0A512); Q8KU27 9NOCA: CYP116B2, *Rhodococcus* sp. NCIMB 9784 (Q8KU27); CPXB BACME: CYP102A1, *Bacillus megaterium* (P14779); NOR FUSOX: CYP55A1, *Fusarium oxysporum* (P23295); CP74A ARATH: CYP74A, *Arabidopsis thaliana* (Q96242); CP6B1 PAPPO: CYP6B1, *Papilio polyxenes* (Q04552); CP53 ASPNG: CYP53, *Aspergillus niger* (P17549); CP2J1 RABIT: CYP2J1, *Oryctolagus cuniculus* (P52786); CPXC RHIRD: CYP103, *Rhizobium radiobacter* (P24466); THAS HUMAN: CYP5A1, *Homo sapiens* (P24557); CP125 MYCTU: CYP125A1, *Mycobacterium tuberculosis* (P63709); Q5F4D9 SPHMC: CYP153A7, *Sphingopyxis macrogolittabida* (Q5F4D9); Q11ZY2 POLSJ: Cytochrome P450, *Polaromonas* sp. (Q11ZY2); Q65A64 9MYCO: Cytochrome P450, *Mycobacterium* sp. (Q65A64); E0XZ55 9RHIZ:

Cytochrome P450, uncultured *Rhizobiales bacterium* (E0XZ55); A7HU01 PARL1: Cytochrome P450, *Parvibaculum lavamentivorans* (A7HU01); Q20YJ8 RHOPB: Cytochrome P450, *Rhodopseudomonas palustris* (Q20YJ8); Q143U3 BURXL: Putative cytochrome P450, *Burkholderia xenovorans* (Q143U3); A7HUI3 PARL1: Cytochrome P450, *Parvibaculum lavamentivorans* (A7HUI3); A7HTD0 PARL1: Cytochrome P450, *Parvibaculum lavamentivorans* (A7HTD0); M4Z536 9BRAD: Cytochrome P450, *Bradyrhizobium oligotrophicum* (M4Z536); B8H036 CAUCN Cytochrome P450, *Caulobacter crescentus* (B8H036); Q0C3M7 HYPNA: Cytochrome P450 family protein, *Hyphomonas neptunium* (Q0C3M7); B4RCI3 PHEZH: Cytochrome P450 family protein, *Phenylobacterium zucineum* (B4RCI3); D1ACZ6 THECD: Cytochrome P450, *Thermomonospora curvata* (D1ACZ6); Q1BB29 MYCSS: Cytochrome P450, *Mycobacterium* sp. (Q1BB29).

ADN metagenómico se agrupan esencialmente con secuencias de citocromos P450 identificadas en bacterias. Se observa que cinco de estas siete (UPOCYTP1-5) se encuentran relativamente cercanas unas a otras (con 60-89 % de identidad entre ellas). Las proteínas UPOCYTP1 y UPOCYTP2 se parecen más entre sí que a ninguna otra proteína de las bases de datos, y lo mismo ocurre con UPOCYTP3 y UPOCYTP4. Ambas parejas coinciden en la proteína de las bases de datos con la que presentan una mayor identidad, que es un citocromo P450 de *Polaromonas* sp. (64-79 % de identidad). UPOCYTP5, que muestra una identidad de 61-70 % con las secuencias UPOCYTP1-4, sí presenta más similitud a otras proteínas de las bases de datos. A la que más se parece, mostrando una identidad del 96 %, es a un citocromo P450 de *Parvibaculum lavamentivorans*.

La proteína UPOCYTP7, se encuentra en una rama más alejada de la misma agrupación filogenética y muestra una identidad de 48-55 % con las proteínas UPOCYTP1-4. La proteína con mayor similitud de la base de datos es otro citocromo P450 de *Parvibaculum lavamentivorans*, distinto al citocromo P450 de organismos de esta especie comentado previamente, con quién presenta una identidad del 87 %. Las siguientes proteínas de las bases de datos se encuentran filogenéticamente más alejadas (≤ 59 % de identidad). La secuencia más distante, que se localiza en otra rama del árbol completamente diferente, es la de UPOCYTP6 y presenta una identidad con las demás de solo el 30-32 %. Esta proteína presenta una identidad del 96 % con un citocromo P450 de *Parvibaculum lavamentivorans*, distinto a los 2 citocromos P450 de organismos de esta especie comentados con anterioridad, estando las siguientes proteínas de las bases de datos filogenéticamente bastante más alejadas (≤ 56 % de identidad).

5.2 Detección de monooxigenasas productoras de 1-naftol

Hasta el momento, la actividad oxigenasa inicial detectada en los clones había sido la transformación de indol en índigo, de interés como tinte textil. Queríamos conocer si alguna de las monooxigenasas metagenómicas producidas por estos clones es capaz de catalizar otra reacción de interés. Se seleccionó la producción de 1-naftol a partir de naftaleno, que es de interés en la fabricación de herbicidas, insecticidas y medicamentos. Las monooxigenasas capaces de reconocer el naftaleno como sustrato pueden dar lugar a 1-naftol o 2-naftol, en función de la posición en la que se lleve a cabo la oxigenación.

5.2.1 Ensayo colorimétrico para la detección de naftoles

En primer lugar, se llevó a cabo un ensayo colorimétrico empleando el compuesto indicador Fast Blue B (Sigma-Aldrich) (adaptación de Canada y col., 2002; Wackett y Gibson, 1983). Este indicador reacciona con 1-naftol, y en menor medida con 2-naftol, dando lugar a un producto color violeta, que se puede cuantificar mediante medidas de absorbancia a 528 nm.

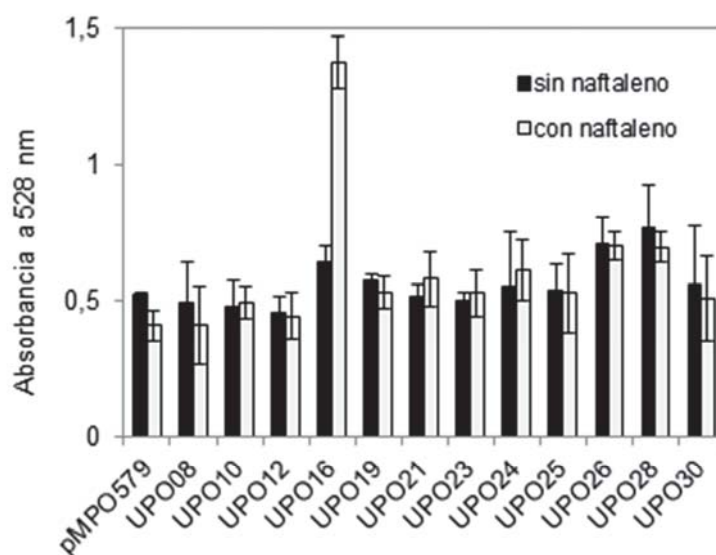


Figura 28. Representación de los resultados del ensayo colorimétrico empleando el indicador Fast Blue B para detectar la producción de naftoles. Se muestran los valores de absorbancia medida a 528 nm del ensayo con los sobrenadantes de cultivos de distintos clones y del fósforo vacío (pMPO579) en ausencia y presencia de naftaleno. Los valores representados son la media de 3 réplicas. Las barras de error se corresponden con la desviación estándar de los datos.

Se ensayaron la mayoría de los clones en los que se había detectado algún gen de oxigenasa inicial, junto con el fósido vacío (pMPO579). Los ensayos se llevaron a cabo en la estirpe MPO554 NaI^R, con cultivos en la variante del medio mínimo M9 descrita en el apartado 2.1 de Materiales y Métodos, con Cm 12,5 mg/L, arabinosa 1 mM (con el fósido en alto número de copias) y salicilato 1 mM (induciendo la expresión heteróloga del ADN metagenómico desde el promotor *psal* del fósido), en ausencia y presencia de naftaleno 1 mM. El protocolo se detalla en el apartado 15 de Materiales y Métodos. De todos los fósidos ensayados, solo uno, el clon UPO16, mostró una absorbancia a 528 nm muy superior al resto cuando fue cultivado en presencia de naftaleno, como se muestra en la figura 28, lo que indica que la monooxigenasa producida por este clon probablemente es capaz de transformar el naftaleno en naftoles.

5.2.2 Ensayo mediante cromatografía líquida de alta resolución para la detección de 1-naftol

El ensayo colorimétrico previo indicó que probablemente el clon UPO16 es capaz de transformar el naftaleno en naftoles. Para comprobar si el producto de la reacción es 1-naftol, 2-naftol o una mezcla de ambos, así como para poder cuantificar esa producción de una manera más exacta de lo que se podría hacer con el ensayo colorimétrico, se realizaron ensayos mediante cromatografía líquida de alta resolución (HPLC).

Para ello, se cultivaron células del clon UPO16 (en la estirpe MPO554 NaI^R) en la variante del medio mínimo M9 descrita en el apartado 2.1 de Materiales y Métodos, con Cm 12,5 mg/L, arabinosa 1 mM y salicilato 1 mM, tanto en presencia como en ausencia de naftaleno 5 mM, durante 20 h. Se prepararon las muestras y se realizaron los análisis mediante HPLC (el protocolo se detalla en el apartado 16 de Materiales y Métodos). Los cromatogramas obtenidos mostraron la aparición de un pico cuyo espectro de absorción y tiempo de retención (18 minutos) coinciden con el del patrón de 1-naftol (figura 29). De la monooxigenación del naftaleno también se podría producir 2-naftol, pero el tiempo de retención del patrón de este compuesto en las condiciones empleadas en el ensayo es de 16 minutos, por lo que el pico obtenido se corresponde solo a 1-naftol. El pico atribuido a 1-naftol no aparece cuando las células se cultivan en ausencia de naftaleno ni cuando el ensayo se realiza sustituyendo el clon UPO16 por el fósido vacío (pMPO579) ni por alguno de los demás clones (datos no mostrados).

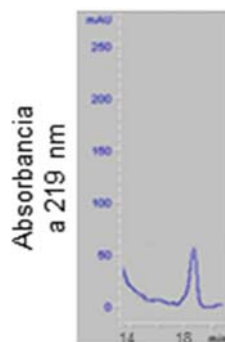


Figura 29. Producción de 1-naftol por células del clon UPO16 cultivadas en presencia de naftaleno. Las muestras se prepararon como se indica en el apartado 16 de Materiales y Métodos, y se analizaron mediante HPLC detectando la absorbancia a 219 nm. El pico que aparece a los 18 minutos en el cromatograma se corresponde con el 1-naftol producido. Se muestra la imagen de una de las 3 réplicas realizadas.

A partir del área del pico de 1-naftol se calculó que el clon UPO16 produjo 28,1 μmol de 1-naftol a partir de 5 mmol de naftaleno. Este valor es del mismo orden al que se calcula a partir del ensayo colorimétrico si se asume que toda la señal era atribuible a 1-naftol (29,3 μmol de 1-naftol a partir de 1 mmol de naftaleno).

Puesto que la única monooxigenasa identificada en la secuencia del clon UPO16 es una fenol 2-monooxigenasa, se asumió que la actividad de producción de 1-naftol era debida a esta enzima. Este clon posee genes homólogos a los de los 6 componentes requeridos para la transformación por parte de la bacteria de fenol en catecol (P0-P5, Nordlund y col., 1990). Se decidió comprobar la capacidad del clon de transformar fenol en catecol, y cuantificar esa producción para compararla con la de 1-naftol.

Para ello se realizó un ensayo mediante HPLC, preparándose las muestras del mismo modo que para el ensayo de producción de 1-naftol, con la diferencia de que a los cultivos se les añadió fenol 5 mM en lugar de naftaleno (el protocolo se detalla en el apartado 16 de Materiales y Métodos). Los cromatogramas obtenidos mostraron la aparición de un pico cuyo espectro de absorción y tiempo de retención (8,4 min) coinciden con el del patrón de catecol (figura 30). El pico atribuido a catecol no aparece cuando las células se cultivan en ausencia de fenol ni cuando el ensayo se realiza sustituyendo el clon UPO16 por el fósido vacío (pMPO579) (datos no mostrados).

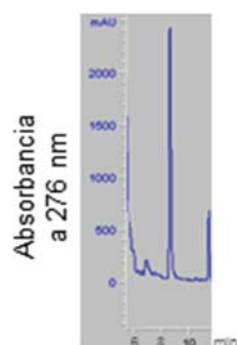


Figura 30. Producción de catecol por células del clon UPO16 cultivadas en presencia de fenol. Las muestras se prepararon como se indica en el apartado 16 de Materiales y Métodos, y se analizaron mediante HPLC detectando la absorbancia a 276 nm. El pico que aparece a los 8,4 minutos en el cromatograma se corresponde con el catecol producido. Se muestra la imagen de una de las 3 réplicas realizadas.

A partir del área del pico de catecol se calculó que el clon UPO16 produjo 4,0 mmoles de catecol a partir de 5 mmoles de fenol. Al comparar la actividad monooxigenasa de este clon teniendo como sustrato naftaleno o fenol, se observa que transforma mucho más eficientemente el segundo de estos sustratos, de manera que la actividad que presenta con naftaleno es, como mucho, el 0,7 % de la que presenta con fenol.

5.3 Ruta de desulfuración de dibenzotiofeno

En el clon UPO21 se identificaron genes homólogos a los de la ruta de desulfuración de dibenzotiofeno (DBT), codificantes de las enzimas dibenzotiofeno-5,5-dióxido monooxigenasa, 2-hidroxibifenilo-2-sulfinato desulfinasas y dibenzotiofeno monooxigenasa (Denome y col., 1994; Gray y col., 1996). Los productos de estos genes junto con una FMN reductasa codificada aguas arriba de ellos, que podría constituir el sistema de transporte de electrones hacia las monooxigenasas, serían suficientes para generar a partir de DBT 2-hidroxibifenilo y sulfito, que podría ser utilizado por *E. coli* como fuente de azufre (figura 31).

Estos genes son muy similares a los descritos en *Rhodococcus* sp. IGTS8 *dszA*, *dszB* y *dszC* (con identidades de las proteínas codificadas del 82 %, 74 % y 79 % respectivamente), por lo que a partir de ahora los denominaremos del mismo modo: *dszA*, *dszB* y *dszC*. Son contiguos y formarían un operon *dszABC*, en el que los genes parecen estar acoplados tradicionalmente, ya que solapan entre sí. Los genes *dszA*, *dszB* y *dszC* de *Rhodococcus* sp.

IGTS8 se encuentran en el mismo orden y también forman un operón. Sin embargo, el solapamiento entre los genes es distinto en ambos casos: en *Rhodococcus* sp. IGTS8 los genes *dszA* y *dszB* solapan 4 nucleótidos y los genes *dszB* y *dszC* están separados 10 nucleótidos, y en el clon UPO21 *dszA* y *dszB* solapan 37 nucleótidos y *dszB* y *dszC* solapan en la forma ATGA. Otra diferencia es que en UPO21, justo aguas arriba de estos genes y en la misma orientación se encuentra el gen codificante de una FMN reductasa (NAD(P)H), que llamaremos *dszE*. En *Rhodococcus* sp. IGTS8, no se encuentra un gen homólogo a este en esa posición. De hecho, la reductasa que asocian al sistema, codificada por el gen *dszD*, se encuentra alejada de los genes *dszABC* (Gray y col., 1996). La reductasa identificada en UPO21 no se parece a la codificada por *dszD*, y presenta un 44 % de identidad con una FMN reductasa de *Pseudomonas putida* KT2440. La distancia que separa a *dszE* y *dszA* es de 4 nucleótidos, lo suficientemente pequeña para considerar que este gen también formaría parte del operón comentado anteriormente, tratándose en ese caso de 4 genes que constituirían un operón transcrito desde un promotor localizado aguas arriba del gen *dszE*.

Las monooxigenasas indicadas, o al menos una de las dos, probablemente sean las responsables de la transformación de indol en índigo, al no haberse identificado ningún otro gen codificante de oxigenasas hidroxilantes en la secuencia de este clon, que conocemos por completo. De hecho, se han descrito otras dibenzotiofeno monooxigenasas capaces de catalizar la transformación de indol en índigo (Coco y col., 2001; Furuya y col., 2004). Sin embargo, nos encontramos de nuevo con una paradoja similar a la observada previamente en el clon N5, codificante de resistencia a carbenicilina (apartado 2.5 de Resultados): para detectar color azul debido la producción de índigo, se necesita la adición de salicilato para inducir la transcripción desde el promotor *psal* en la estirpe MPO554; pero todos estos genes *dsz* se encuentran en sentido contrario a los promotores heterólogos presentes en el vector (figura 32). De forma similar al caso del clon N5, interpretamos que sería de esperar que hubiese otro gen, en sentido del promotor *psal* del fósido, que estuviese implicado de alguna manera en la expresión de los genes *dsz* o en la actividad de alguno de sus productos génicos.

El dibenzotiofeno (DBT) es uno de los componentes minoritarios de los combustibles fósiles. Contiene azufre, y durante su combustión este se libera a la atmósfera, provocando problemas de contaminación como la lluvia ácida (Monticello y Finnerty, 1985). Aunque se realizan procesos químicos de hidrosulfurización para eliminar el azufre de los combustibles fósiles, el DBT es resistente a este tratamiento (Gallardo y col., 1997). La desulfuración biológica del DBT por la ruta descrita en la figura 31 ofrece la posibilidad de

liberar el azufre en forma de sulfito, sin que la molécula pierda su valor como combustible, por lo que esta ruta de desulfuración podría tener una aplicación potencial si pudiera expresarse y fuese funcional en una bacteria con las características adecuadas. Por ello se quiso comprobar si las enzimas codificadas por los genes del clon UPO21 eran capaces de llevar a cabo la desulfuración de DBT. Esto se podía averiguar de forma directa mediante una prueba de crecimiento de la estirpe MPO554 con el clon UPO21 en un medio mínimo con

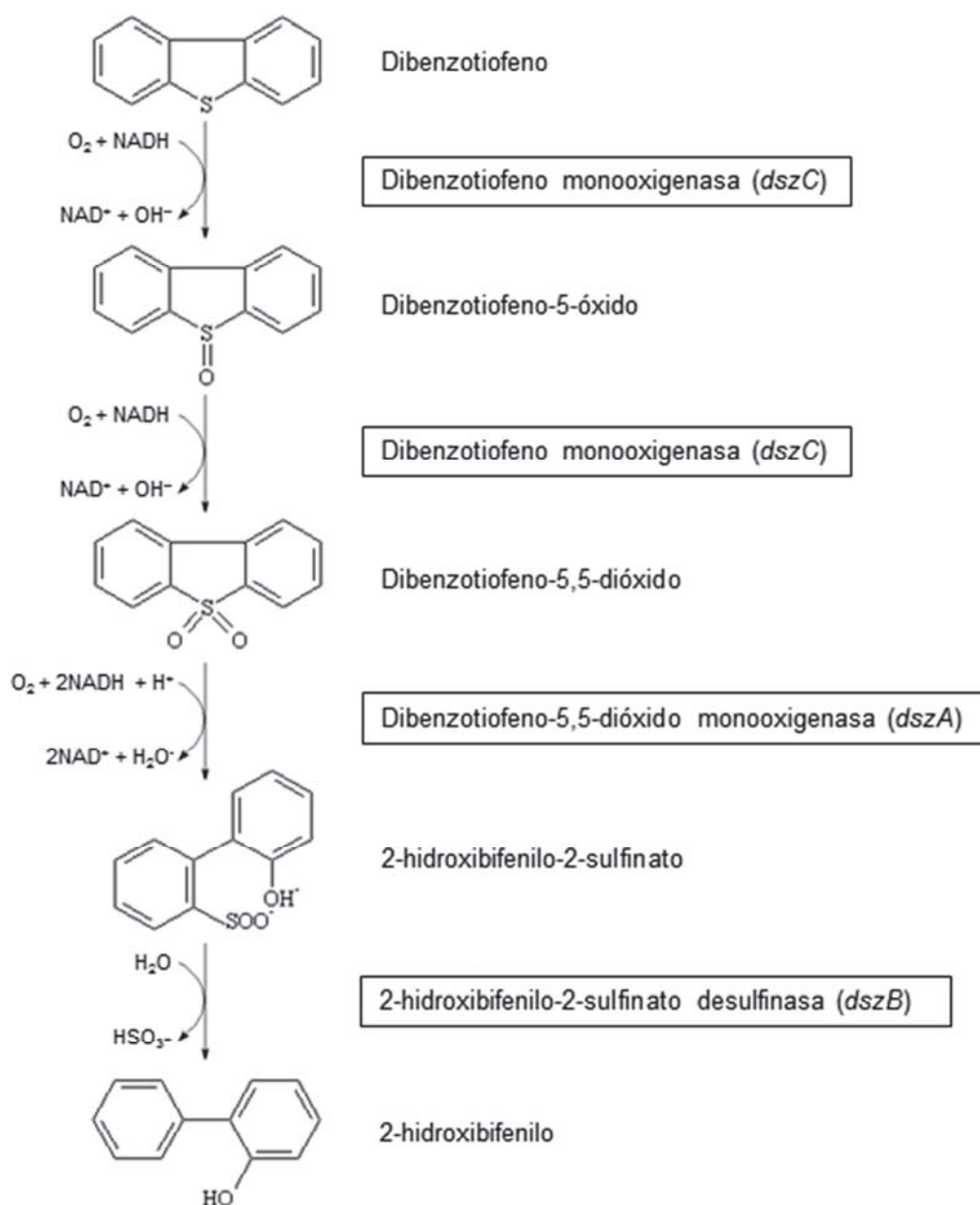


Figura 31. Ruta de desulfuración de dibenzotiofeno. Se recuadra la enzima implicada en cada paso junto al gen codificante entre paréntesis. Adaptación de <http://umbbd.ethz.ch/index.html>, University of Minnesota. Biocatalysis / Biodegradation Database.

DBT como única fuente de azufre. Si se produce la desulfuración de DBT, este se transforma en 2-hidroxibifenilo, liberándose sulfito (figura 31), que las bacterias pueden emplear como fuente de azufre.

Se cultivó la estirpe MPO554 Nal^R con el fósido UPO21 en la variante del medio mínimo M9 descrita en el apartado 2.1 de Materiales y Métodos pero empleando DBT 0,2 mM como única fuente de azufre en lugar de MgSO₄, a 30 °C con agitación, y se observó que fue capaz de crecer (llegando tras 6 días a una D.O.₆₀₀ de 0,6-0,8). El crecimiento usando DBT como fuente de azufre fue dependiente de la adición de salicilato 1 mM, que permite la expresión heteróloga del ADN metagenómico desde el promotor *psal* presente en el fósido y la antiterminación de esta transcripción por la proteína antiterminadora N, o arabinosa 1 mM, que induce el número de copias del fósido (tabla 11). Por tanto, este clon proporciona la capacidad de desulfurar dibenzotiofeno con una eficacia suficiente como para permitir el crecimiento bacteriano usando DBT como fuente de azufre. El crecimiento con DBT sin necesidad de añadir salicilato para inducir *psal* indica que los genes necesarios para ello pueden expresarse por sí solos en la bacteria hospedadora, si bien el nivel de expresión propio es lo suficientemente bajo como para necesitar al menos un incremento del número de copias codificantes para conseguir crecer. Curiosamente, para detectar el color azul acumulado por la conversión de indol en índigo en las colonias de esta misma estirpe (tras 2-3 días de incubación a 30 °C en medio LB) se necesitaron ambas cosas: expresión desde *psal* con la antiterminación de N y alto número de copias del fósido (tabla 11).

5.3.1 Subclonación de los genes de la ruta de desulfuración de dibenzotiofeno

Con el objetivo de confirmar que los genes que habíamos identificado (*dszA*, *dszB* y *dszC*) eran suficientes para la desulfuración de dibenzotiofeno, estos se subclonaron. Junto con los genes codificantes de las dos monooxigenasas y de la desulfinasas, se subclonó también el gen codificante de la reductasa que se ubica inmediatamente aguas arriba de estos (*dszE*), por si fuera necesaria para transferir los electrones a estas monooxigenasas. Se obtuvo un fragmento de 5,7 kb por restricción del clon UPO21 como se indica en el apartado 5.2 de Materiales y Métodos. Dicho fragmento contenía los 4 genes mencionados flanqueados por unas 800 pb aguas arriba de la región codificante de la reductasa, que contendría el promotor propio, y unas 600 pb aguas abajo del último de los genes. Este fragmento se clonó en el fósido pMPO579 en el mismo sitio en el que se inserta el ADN metagenómico, de manera

que los genes de la ruta quedaron tras los promotores heterólogos, en su misma orientación. El subclón resultante se denominó pMPO587.

Se comprobó si pMPO587 proporciona a MPO554 NaI^R la capacidad de transformar indol en índigo y de desulfurar DBT, en las mismas condiciones en las que lo hizo el clon UPO21, pero no se pudo apreciar color azul en las colonias. Tampoco se consiguió crecimiento con DBT como única fuente de azufre, en ninguna condición, es decir, incluso intentando maximizar la expresión induciendo con salicilato e incrementando el número de copias con arabinosa (tabla 11). También se probó este clon en EPI300-T1^R, MPO553 (desde donde se podría dar expresión heteróloga desde el promotor pT7 del fósido) y otras estirpes de *E. coli*, con el mismo resultado negativo. Nos planteamos como posibilidades que hubiera algún elemento que estuviera impidiendo la expresión de esos genes, tanto desde su propio promotor como heterológamente desde los promotores pT7 y *psal* del fósido; o que faltara algún gen necesario, ya fuera para que se produjera la actividad o para la expresión de los otros genes.

5.3.1.1 Activador transcripcional de promotores dependientes de σ^N

Tras examinar la secuencia del clon UPO21, se identificó un gen homólogo al codificante de un activador de promotores dependientes de σ^N (que denominaremos *dszR*), ubicado aguas abajo de los genes de la ruta de desulfuración de DBT y en la misma orientación que los promotores heterólogos del fósido. Este tipo de activadores se unen al ADN unas 80-150 pb aguas arriba del promotor que regulan (Bush y Dixon 2012) y contactan con la holoenzima ARN polimerasa unida al factor sigma alternativo σ^N , activando la transcripción desde los promotores que esta forma de ARN polimerasa reconoce.

Los promotores dependientes de σ^N se caracterizan por una secuencia consenso TGGCACG-N₄-TTGC localizada en las posiciones -24 y -12 respecto al inicio de la transcripción (Barrios y col., 1999). Aguas arriba del primero de los 4 genes, de *dszE*, se identificó a su vez una secuencia TGGCAtG-N₄-cTGC que se ajusta mucho a la de los promotores dependientes de σ^N (figura 32). Dentro de la región codificante del gen *dszE*, se encontró otro posible promotor dependiente de σ^N : TGGaCaa-N₄-gTGC, que podría ser el promotor de otro posible operón que formaran únicamente los genes *dszA*, *dszB* y *dszC*. Teniendo en cuenta estos datos, era muy posible que los genes subclonados necesitasen ese regulador para la transcripción desde su propio promotor, ya fuera uno u otro de los indicados,

lo que encajaría con la dependencia de salicilato para la detección de las actividades, ya que en UPO21 este gen se encuentra en el sentido de *psal*. Por otra parte, esto no explicaría que los genes subclonados en pMPO587 no sean capaces de expresarse desde los promotores heterólogos del fósido. En cualquier caso, si el activador fuese realmente el elemento necesario, se esperaría que al aportar el regulador al subclón pMPO587 los genes de la ruta se pudieran expresar, ya que el promotor dependiente de σ^N se encuentra incluido en la secuencia no codificante subclonada, y el sitio de unión del regulador muy probablemente se localizaría también en ese fragmento.

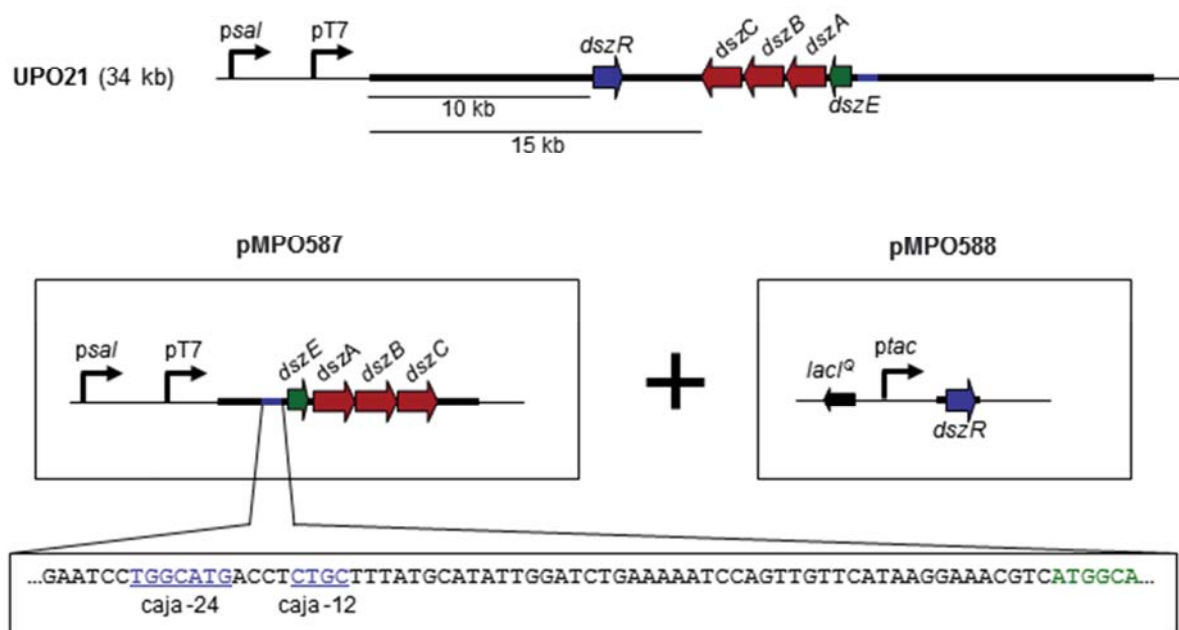


Figura 32. Esquema del clon con los genes de la ruta de desulfuración de dibenzotiofeno (UPO21) y de los subclones con los genes necesarios para que se produzca la actividad (pMPO587 y pMPO588). Se indican los genes involucrados en dicha actividad. Los genes representados codifican las siguientes proteínas: *dszR*: regulador de promotores dependientes de σ^N ; *dszA*: dibenzotiofeno-5,5-dióxido monooxigenasa; *dszB*: 2-hidroxibifenilo-2-sulfinato desulfinaasa; *dszC*: dibenzotiofeno monooxigenasa; *dszE*: FMN reductasa. Se amplía la región que contiene el promotor dependiente de σ^N , indicando en azul las cajas -24 y -12 de dicho promotor, y en verde el principio de la secuencia codificante del gen *dszE*. Aunque no está representado en la figura, se identificó otro posible promotor dependiente de σ^N en la región codificante de *dszE*. Los genes y demás elementos del esquema no están representados a escala.

Los activadores de promotores tipo σ^N poseen por lo general tres dominios: un dominio central que presenta la actividad ATPasa necesaria para la activación, un dominio en el extremo carboxilo-terminal que ejerce el reconocimiento y unión al ADN y un dominio en el extremo amino-terminal implicado en la recepción de las señales pertinentes. La mayoría de

los activadores de promotores tipo σ^N presentan los tres dominios comentados, pero algunos no tienen el dominio regulador del extremo amino-terminal, entre los que se encuentran PspF (de *E. coli*), y HrpR y HrpS (de *Pseudomonas syringae*). La actividad de estos activadores se regula negativamente por otras proteínas que interactúan con ellas inhibiéndolas (Bush y Dixon 2012).

El activador de promotores tipo σ^N identificado en el clon UPO21 carece del dominio amino-terminal regulador que le permite responder a señales ambientales. Al comparar su secuencia de aminoácidos con las de las bases de datos, el mayor valor de identidad (59 %) lo presenta con un regulador de este tipo en *Pseudomonas entomophila*, también carente del dominio regulador.

Se decidió subclonar el gen del regulador de promotores dependientes de σ^N (*dszR*), para averiguar si en presencia de dicho regulador se producía expresión de los genes *dsz* desde su propio promotor. En lugar de incorporarlo en pMPO587 se decidió subclonarlo en otro vector en el que su expresión pudiera regularse independientemente del resto de genes. Se obtuvo el gen del regulador por restricción del clon UPO21 como se indica en el apartado 5.2 de Materiales y Métodos. El fragmento se clonó en pIZ1016 (Martínez-Pérez y col., 2004), dando lugar a pMPO588. El gen del regulador quedó tras el promotor *ptac* del vector, en su misma orientación, siendo su expresión inducible con IPTG.

Se volvió a comprobar si se daban las actividades de producción de índigo y de desulfuración de DBT (mediante crecimiento en medio mínimo con DBT como única fuente de azufre), aportando a la estirpe MPO554 NaI^R el subclón con los genes de la ruta de desulfuración (pMPO587) y el plásmido con el regulador de promotores dependientes de σ^N (pMPO588). Los ensayos se acompañaron de los controles negativos consistentes en la sustitución de cada uno de los vectores por el equivalente vector vacío. En presencia del regulador se detectó color azul por producción de índigo tras 2-3 días de incubación a 30 °C, siempre y cuando se añadiera arabinosa al medio, es decir, con el fósforo que contiene los genes *dszEABC* en alto número de copias. El aporte de IPTG, que induce la expresión del regulador de promotores dependientes de σ^N , y de salicilato, que induce la expresión desde *psal* con la antiterminación de N en el fósforo, no mejoraron la actividad en niveles detectables. Por otro lado, en presencia del regulador, también se observó crecimiento en medio mínimo con DBT 0,2 mM como única fuente de azufre, a 30 °C con agitación, alcanzando tras 3 días a una D.O.₆₀₀ de 0,6-0,8. En este caso, el crecimiento fue dependiente de la adición de IPTG 1 mM, que induce la expresión del activador de promotores

dependientes de σ^N clonado bajo *ptac*, no siendo suficiente con el incremento del número de copias del fósido con los genes *dszEABC* que proporciona la arabinosa (el crecimiento mejora ligeramente en presencia de arabinosa siempre que se haya aportado IPTG). Como sería de esperar, el aporte de salicilato no supone ninguna ventaja (de hecho, su aporte además de IPTG reduce los valores de D.O.₆₀₀ alcanzados) (tabla 11).

| | sin inductores | +Arab | +Salic | +Arab +Salic | +IPTG | +IPTG +Arab | +IPTG +Salic |
|---|-------------------|-------|--------|-----------------|-------|----------------|-----------------|
| Producción de índigo (colonias azules) | | | | | | | |
| UPO21 | No | No | No | Sí | | | |
| pMPO587 | No | No | No | No | | | |
| pMPO587 +pMPO588 | No | Sí | No | Sí | No | Sí | No |
| Desulfuración de DBT (crecimiento con DBT) | | | | | | | |
| UPO21 | No | Sí | Sí | Sí | | | |
| pMPO587 | No | No | No | No | | | |
| pMPO587 +pMPO588 | No | No | No | No | Sí | Sí | Sí* |

Tabla 11. Resumen de las condiciones en las que se produce la actividad de desulfuración de dibenzotiofeno en comparación con la de producción de índigo en presencia de los genes *dsz*. Se indica con Sí o No la presencia o ausencia de cada actividad mediante su detección por producción de colonias azules en medio rico LB o de crecimiento en medio mínimo con DBT como única fuente de azufre, según el caso. La concentración empleada de cada uno de los 3 inductores (arabinosa, salicilato e IPTG) fue de 1 mM. Las abreviaturas de la tabla significan lo siguiente: Arab: arabinosa; Salic: salicilato. Los ensayos se realizaron en la estirpe MPO554 NaI^R con los vectores indicados en cada caso. Las casillas sombreadas indican condiciones no probadas. Aunque no se muestren en la tabla, también se realizaron los ensayos con pMPO579 (fósido sin ADN metagenómico), pMPO579 +pMPO588 y pMPO587 +pIZ1016 (vector sin gen codificante de activador transcripcional), no detectándose ninguna de las actividades en ningún caso. *El crecimiento en presencia de IPTG junto con salicilato fue bastante peor que con IPTG sin salicilato.

Por tanto, se confirma que al menos alguna de las monooxigenasas codificadas por los genes subclonados en pMPO587 es capaz de transformar el indol en índigo y que las enzimas codificadas por los genes subclonados en pMPO587 aportan a la bacteria la capacidad de desulfurar DBT. Sin embargo, sorprendentemente, no se consigue expresar heterológicamente esos genes en el subclón pMPO587, sino que es necesario el aporte de un activador

transcripcional dependiente de σ^N , lo que indica que la transcripción se inicia desde su propio promotor, que aparentemente es dependiente de σ^N .

Llama la atención que los resultados no son plenamente coincidentes en los dos contextos. Por una parte, en el caso de UPO21 parece que se necesita menos expresión del regulador para crecer con DBT como fuente de azufre que para detectar colonias azules, ya que el crecimiento con DBT se puede conseguir si simplemente se incrementa el número de copias (con arabinosa) mientras que el color azul debido al índigo solo es detectable en las colonias cuando además se induce la transcripción desde *psal* (con la adición de salicilato). Por otra parte, en el caso del subclón pMPO587 no ocurre lo mismo, ya que para generar color azul no es necesario inducir la expresión del activador (con IPTG), aunque sí aumentar el número de copias, mientras que para conseguir crecimiento a expensas de DBT como fuente de azufre sí se necesita IPTG.

5.3.2 Caracterización de la expresión de los genes de la ruta de desulfuración de dibenzotiofeno

Al margen de los diferentes requerimientos para la producción de índigo y desulfuración de DBT, lo que sí parece claro es que los genes de la ruta de desulfuración de DBT no pueden expresarse por transcripción heteróloga desde el promotor *psal* del fósido en el subclón pMPO587, y que para su expresión se necesita la presencia del activador transcripcional dependiente de σ^N . Este hecho resulta sorprendente ya que *psal* y los genes *dsz* están en la misma orientación, la distancia que los separa es relativamente pequeña (hay 1,2 kb entre *psal* y el gen *dszE*) y además, la transcripción desde *psal* estaría sujeta a antiterminación por la proteína N.

Para tratar de explicar la imposibilidad de expresar heterológamente los genes de la ruta en pMPO587 desde el promotor *psal*, se volvió a examinar la secuencia de ADN metagenómico. Al principio del inserto, aguas arriba del promotor dependiente de σ^N , se encontró una ristra de Ts seguida de una estructura secundaria, que podría constituir una posible región terminadora de la transcripción, pero que está en sentido contrario a *psal* y a los genes *dszEABC*. Resulta improbable que esta secuencia sea capaz de terminar la transcripción desde los promotores heterólogos, ya que tendría que funcionar en ese sentido y además ser capaz de impedir la transcripción con el sistema de antiterminación de la proteína antiterminadora N. Otra posibilidad es que esa secuencia constituya una señal de terminación

de la transcripción de un ARN interferente (ARNi), que al ser complementario a la secuencia de los genes de la ruta estuviera impidiendo su expresión, por ejemplo promoviendo la degradación del transcrito o evitando su traducción (Fozo y col., 2008) si la transcripción proviene desde los promotores heterólogos localizados aguas arriba, aunque no si proviene del promotor dependiente de σ^N localizado más abajo. Una transcripción desde el promotor dependiente de σ^N podría dar lugar a la expresión de esos genes, quizá porque el ARNi no pueda asociarse con ese transcrito que se inicia más abajo o por algún cambio en la estructura del ARN que evite esa regulación negativa por parte del supuesto ARNi.

Teniendo en cuenta las incógnitas que se nos planteaban en cuanto a la expresión de los genes de la ruta de desulfuración de DBT, se construyeron una serie de fusiones traduccionales y transcripcionales con distintos fragmentos del inserto de pMPO587. Las fusiones se construyeron en el fósrido pMPO579, que ya contenía un gen *gfp* sin promotor aguas abajo del sitio de clonación del ADN metagenómico que serviría como gen indicador para las fusiones transcripcionales. Introduciendo *lacZ* en fase con los distintos fragmentos de la secuencia a ensayar, esas construcciones serían también fusiones traduccionales. De este modo, se realizaron las construcciones indicadas en la figura 33, que serían a la vez fusiones traduccionales a *lacZ* y transcripcionales a *gfp*. La construcción de estas fusiones se detalla en el apartado 5.3 de Materiales y Métodos.

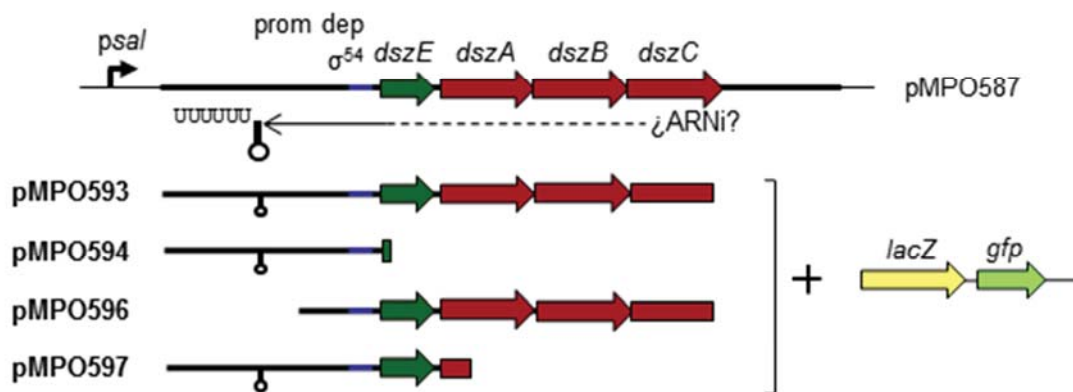


Figura 33. Esquema de las fusiones para el estudio de la expresión de los genes de la ruta de desulfuración de dibenzotiofeno. Se indican los genes involucrados y se representa el fragmento incluido en cada una de las 4 fusiones (pMPO593, pMPO594, pMPO596 y pMPO597). Se trata de fusiones traduccionales a *lacZ* y a su vez transcripcionales a *gfp*. La abreviatura prom dep σ^{54} indica promotor dependiente de σ^{54} . Los genes representados codifican las siguientes proteínas: *dszE*: FMN reductasa; *dszA*: dibenzotiofeno-5,5-dióxido monooxigenasa; *dszB*: 2-hidroxibifenilo-2-sulfonato desulfinaasa; *dszC*: dibenzotiofeno monooxigenasa. Los genes y demás elementos del esquema no están representados a escala.

El inserto en pMPO593 incluye el posible terminador y trata de incluir el posible ARNi. Para su construcción fueron necesarios dos pasos. Primero, se amplificó por PCR, empleando como molde pMPO587, un fragmento que contenía la última parte del gen *dszC*, dejando solo fuera el codón de parada, y se clonó en el vector de fusiones traduccionales pMPO200 (García-González y col., 2005), dando lugar a pMPO592, en el que el fragmento final del gen *dszC* quedaba fusionado al undécimo codón del gen *lacZ* del vector. A continuación, se obtuvo el fragmento que contenía el final del gen *dszC* seguido de *lacZ* por restricción de pMPO592 y se clonó en pMPO587, de manera que se elimina el final del gen *dszC* y el resto del ADN del inserto metagenómico, sustituyéndolo por el final de ese mismo gen sin incluir el codón de parada y con *lacZ* fusionado en fase, dando lugar a pMPO593.

La construcción en pMPO594 consiste en una fusión traduccional *lacZ* al tercer codón del gen codificante de la FMN reductasa (*dszE*) e incluye el posible terminador. El gen *lacZ* se obtuvo por restricción de pMPO200 (García-González y col., 2005) y se clonó en pMPO587, tras eliminar un fragmento que contenía la parte distal de *dszE* y el resto del operon *dsz*, de manera que el principio del gen *dszE* quedara en fase con *lacZ*, generando así pMPO594.

La construcción en pMPO596 tiene la misma fusión a *lacZ* que pMPO593 pero elimina 301 bases del extremo del inserto aguas arriba del promotor del operon *dsz*, excluyendo la posible estructura terminadora localizada en dirección contraria al operón *dsz*. Sin embargo, se mantiene el promotor dependiente de σ^N y muy probablemente los sitios de unión del regulador. Para su construcción fueron necesarios dos pasos. Primero, se obtuvo por restricción de pMPO587, un fragmento que incluía desde 453 nucleótidos aguas arriba del dinucleótido GG del promotor dependiente de σ^N hasta el 267º codón del gen *dszB*, y se clonó en el fósido pMPO579 linealizado con *PmlI*, dando lugar a pMPO595. A continuación, se obtuvo por restricción de pMPO593 un fragmento que incluía los cuatro genes de la construcción pero dejando fuera el promotor dependiente de σ^N , y por el otro extremo parte del vector, hasta el noveno codón del gen codificante de la resistencia a cloranfenicol. Este fragmento se clonó en pMPO595, de manera que la resistencia a cloranfenicol quedaba reconstituida, y el resto de la construcción exactamente igual que en pMPO593 salvo que no tenía el terminador. El plásmido resultante es pMPO596.

La construcción en pMPO597 consiste en una fusión traduccional *lacZ* al gen *dszA*, que incluye la posible estructura terminadora localizada aguas arriba del operón *dsz*. Es una fusión intermedia entre pMPO593 y pMPO594. El gen *lacZ* se obtuvo por restricción de pMPO200

(García-González y col., 2005) y se clonó en pMPO587, tras eliminar un fragmento, de manera que se incluían los primeros 154 codones del gen *dszA*, quedando este en fase con el undécimo codón de *lacZ*, dando lugar a pMPO597.

Cada una de las 4 fusiones construidas se ensayó en las estirpes MPO554 Nal^R y EPI300-T1^R, acompañadas todas ellas por un lado del plásmido con el regulador de promotores dependientes de σ^N , pMPO588, y por otro del plásmido vacío pIZ1016. En la estirpe EPI300-T1^R solo se veía la expresión desde el promotor dependiente de σ^N , y en MPO554 Nal^R con salicilato, además, la expresión desde el promotor *psal*. Los fósmidos se mantuvieron en bajo número de copias (sin aportar arabinosa). Inicialmente, para los ensayos en los que se aportaba el regulador, los cultivos se incubaron antes del ensayo en medio mínimo con DBT y sulfato, con solo DBT o con solo sulfato, con objeto de estudiar su posible inducción por DBT y represión por sulfato, mientras que en los que no se aportaba el regulador los cultivos se incubaron siempre con DBT y sulfato. Sin embargo, se observó que no parecía existir ninguna regulación en función de la fuente de azufre y que DszR podía inducir incluso en presencia de sulfato y ausencia de DBT. Con objeto de simplificar los resultados, los que se muestran en la figura 34 se realizaron en las mismas condiciones de crecimiento, con azufre y DBT. Se partió de preinóculos en la variante del medio mínimo M9 descrita en el apartado 2.1 de Materiales y Métodos, que llevaba además DBT 0,2 mM, y se diluyeron a D.O.₆₀₀ de 0,1 en un medio igual pero con MgSO₄ 0,2 mM, DBT 0,2 mM, IPTG 1 mM y en el caso de la estirpe MPO554 Nal^R también salicilato 1 mM. Se incubaron en agitación a 30 °C durante 6 h. Para las fusiones transcripcionales a *gfp* se cuantificó la fluorescencia emitida, y para las traduccionales a *lacZ* (gen de la β -galactosidasa) se cuantificó la actividad β -galactosidasa.

Al comparar los niveles de expresión basal del operón (sin *dszR* en la figura 34, gráfica superior) con aquellos alcanzados al aportar el activador de promotores dependientes de σ^N (con *dszR*), por lo general no se observan aumentos significativos en los niveles de transcripción, aunque hay un aumento más claro, de dos veces, en pMPO597 (fusión a *dszA*). Sin embargo, los niveles de la fusión traduccional a *dszA*, y más aun los de las fusiones a *dszC*, aumentaron drásticamente. Esto indicaría que curiosamente este activador transcripcional de promotores dependientes de σ^N estaría mejorando los niveles de traducción, más que los de transcripción. La construcción con la fusión traduccional a *dszE* no mostró activación por DszR a nivel transcripcional ni traduccional, indicando que si se induce es a un nivel muy bajo.

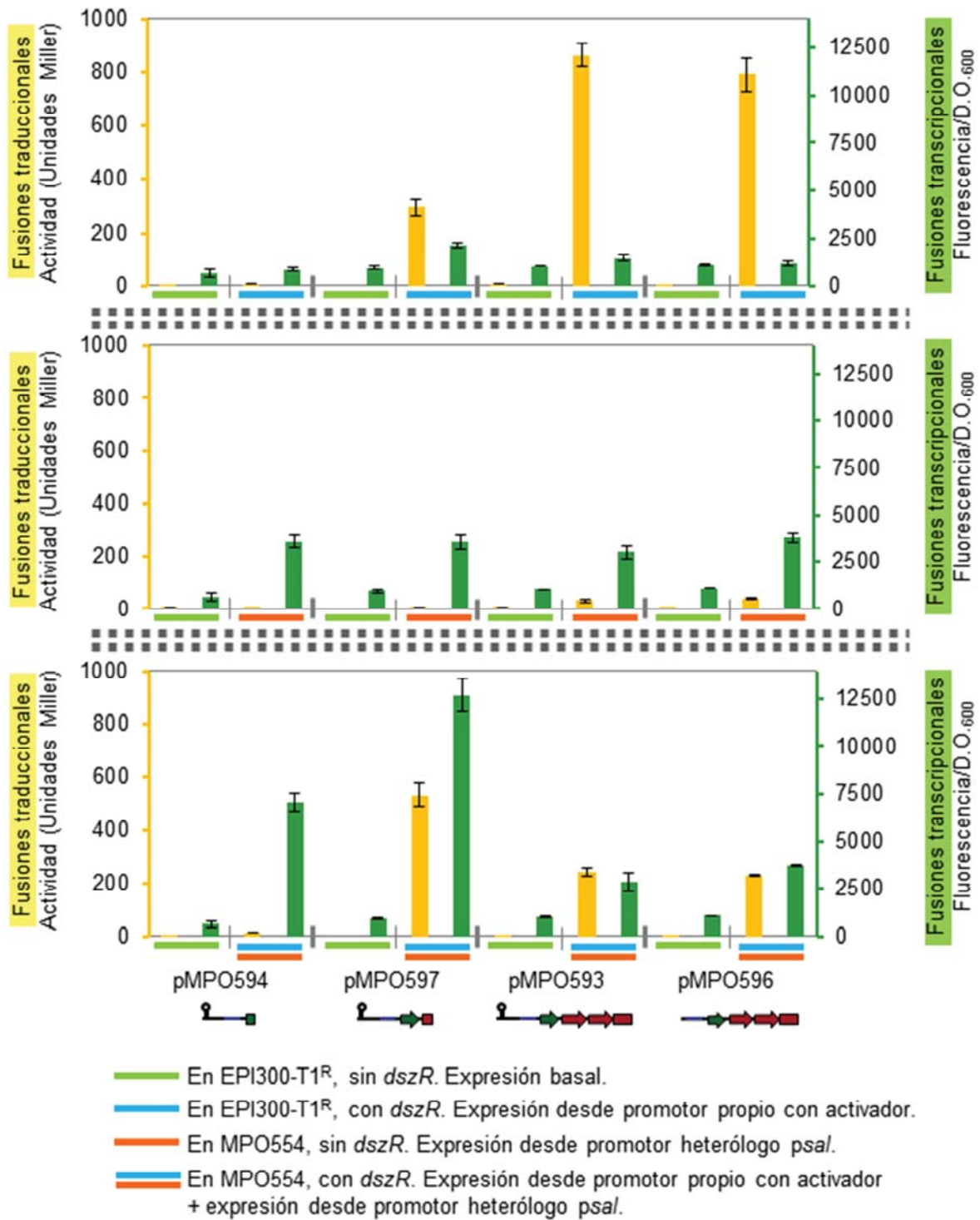


Figura 34. Expresión de las fusiones de la ruta de desulfuración de dibenzotiofeno. En cada una de las tres gráficas se muestra, para cada una de las cuatro fusiones, en columnas amarillas la actividad β -galactosidasa de las fusiones traduccionales a *lacZ*, en unidades Miller (eje de la izquierda), y en columnas verdes, los valores de fluorescencia de las fusiones transcripcionales a *gfp* respecto a la D.O.₆₀₀ de los cultivos ensayados (eje de la derecha). En cada gráfica se comparan los niveles de expresión basal del operón *dsz* (en la estirpe EPI300-T1^R con pIZ1016, sin *dszR*) con los niveles de expresión: desde ese mismo promotor en presencia del activador transcripcional de promotores dependientes de σ^N (EPI300-T1^R con pMPO588, con *dszR*), en la

gráfica superior; desde el promotor heterólogo *psal* del fósido (MPO554 Nal^R con pIZ1016, sin *dszR*), en la gráfica intermedia; y desde el promotor propio del operón en presencia del activador más desde el promotor *psal* (MPO554 Nal^R con pMPO588, con *dszR*), en la gráfica inferior. Los valores representados son la media de al menos 3 réplicas. Las barras de error se corresponden con la desviación estándar de los datos.

Por otro lado, al comparar los niveles de expresión basal del operón con aquellos alcanzados desde el promotor heterólogo *psal* en presencia de salicilato (figura 34, gráfica intermedia), se observan unos aumentos en los niveles de transcripción significativos en todos los casos, lo que indicaría que se está produciendo transcripción desde *psal*. Esta transcripción heteróloga se da a niveles superiores a los obtenidos por el activador DszR, aproximadamente al doble, con ligeras variaciones de una construcción a otra. Sin embargo, los niveles de las fusiones traduccionales no correlacionaron en absoluto con este incremento de transcripción, obteniéndose unos niveles de actividad β -galactosidasa entre 20 y 300 veces inferiores a los obtenidos cuando DszR activa, dependiendo de la construcción. Esto indicaría que los transcritos producidos desde *psal* se traducen muy ineficazmente, lo que concuerda con que induciendo con salicilato no conseguimos detectar la actividad de producción de índigo ni de desulfuración de dibenzotiofeno.

Por último, se comparan los niveles resultantes de la suma de la expresión desde el promotor propio del operón en presencia del activador más la procedente del promotor *psal* (figura 34, gráfica inferior) con el resto de condiciones. Respecto a los niveles basales, se observa un aumento de la transcripción evidente en todos los casos, pero mucho más acusado en pMPO594 (fusión a *dszE*) y en pMPO597 (fusión a *dszA*). Si se comparan los niveles de expresión desde los dos promotores en conjunto con los de cada uno de ellos independientemente, en las fusiones a *dszA* y a *dszE* se observan niveles de transcripción claramente mayores cuando se puede dar transcripción desde los dos promotores que desde cada uno de ellos por separado, llegando a ser incluso superiores a la suma de los promotores individuales, sobre todo en el caso de la fusión a *dszA*. Sin embargo, en las fusiones a *dszC*, la transcripción desde los dos promotores en su conjunto, aunque es algo superior a los niveles alcanzados desde el promotor propio del operón en presencia del activador, es similar a la alcanzada desde *psal* en ausencia de DszR, no observándose por tanto un efecto sumatorio. Respecto a la traducción, en la fusión a *dszE* se observan unos niveles de expresión detectables pero bajísimos, lo que indica que esta proteína se produce a unos niveles muy bajos. En la fusión a *dszA* se produce un aumento importante en los niveles de traducción respecto a los producidos con los transcritos de esos dos promotores de manera independiente.

Sin embargo, si comparamos los niveles de expresión de las fusiones traduccionales (unidades Miller) y las transcripcionales (fluorescencia) de las gráficas superior e inferior, se observa claramente que la traducibilidad de los transcritos producidos conjuntamente desde los dos promotores es menor que la de los transcritos producidos desde su propio promotor exclusivamente. La explicación más plausible es que existe una mezcla de transcritos pero que solo los producidos desde el propio promotor activado por DszR son los que se traducen eficazmente, mientras que los producidos desde *psal* no, tal como indican los datos de la gráfica intermedia. Si hacemos el mismo tipo de comparación con los datos de expresión de las dos fusiones a *dszC* se observa el mismo fenómeno, que puede explicarse de la misma forma: hay una mezcla de transcritos pero los producidos desde *psal* no se traducen. Y resulta que en este caso se aprecia incluso un descenso de aproximadamente 3,5 veces en los niveles de actividad β -galactosidasa al comparar los niveles de expresión desde su propio promotor (gráfica superior) con los obtenidos cuando *psal* está también activo (gráfica inferior). En este caso, debemos asumir que la mayor parte de esta mezcla de transcritos proviene de *psal* y que, de alguna manera, la transcripción desde *psal* está interfiriendo con la activación mediada por DszR.

Un hecho sorprendente es que, tanto cuando se inducía el promotor *psal*, como cuando se activaba la transcripción desde su propio promotor con DszR, los niveles de fluorescencia incrementaron muy poco respecto a los niveles basales de expresión. Cuando se estuvo validando la expresión desde los promotores heterólogos del fósido, ya se observó que para obtener niveles altos de inducción tanto desde el promotor de T7 como desde *psal* era necesario incrementar el número de copias con arabinosa (figura 8). Aunque no tenemos una explicación racional para este fenómeno, la realización de los ensayos en las mismas condiciones pero en presencia de arabinosa podrían mostrar con más claridad la actividad transcripcional de DszR y mejorar los niveles inducidos de transcripción tanto desde *psal* como desde su propio promotor, de manera que además se pueda analizar mejor la traducibilidad de los transcritos producidos desde cada promotor.

Recopilando, de esta información puede deducirse que la presencia del activador que permite la expresión desde el promotor propio del operón, supone un incremento principalmente en los niveles de traducción, a pesar de ser una proteína homóloga a un regulador transcripcional de promotores tipo σ^N , aunque habría que comprobar si este fenómeno se repite realizando los ensayos con el fósido en alto número de copias, que de algún modo se ha observado que permite detectar bastante mejor incrementos en la

transcripción cuando esta se induce. Por otro lado, la inducción del promotor *psal* se ve reflejada en un aumento en los niveles de transcripción, que se refleja a su vez en un ligero aumento en los niveles de traducción en las fusiones a *dszC* pero no en la fusión a *dszA*. La relación entre los niveles de traducción y los de transcripción es mucho mayor cuando los transcritos proceden del promotor propio que desde *psal*, lo que indicaría una traducción mucho menos eficiente del ARN mensajero (ARNm) procedente de *psal*. Otro aspecto curioso es que cuando se activa a la vez la transcripción desde los dos promotores, en las fusiones a *dszA* y a *dszE*, sí se observa un aumento en la transcripción, y en la de *dszA* también en la traducción, con respecto a la de cada promotor independiente; pero en las fusiones a *dszC* no ocurre lo mismo, obteniéndose niveles de traducción inferiores a los alcanzados cuando todo el ARNm procede del promotor propio del operón. En todos los casos, la traducibilidad respecto a la transcripción alcanzada es más baja cuando se tiene transcripción desde los dos promotores que exclusivamente desde el propio. Esto indicaría que muy probablemente existan otros elementos de regulación en este sistema que no se están teniendo en cuenta y que jugaría un papel importante una regulación postranscripcional.

La eliminación del posible terminador de la transcripción no se vio reflejada en un aumento significativo de los niveles de transcripción, presentando las fusiones pMPO593 y pMPO596 un comportamiento muy similar, por lo que no es probable que este sea el responsable de la inexistente o reducida expresión heteróloga de estos genes. De hecho, el comportamiento similar de ambas fusiones en todas las condiciones indicaría que la eliminación de ese fragmento no tiene efecto a ningún nivel en la expresión.

En cuanto a la posible existencia de un ARNi que terminara en la supuesta señal de terminación de la transcripción detectada que pudiera estar impidiendo la traducción de los transcritos de estos genes, los datos no parecen apoyar esta hipótesis, ya que al inducir la transcripción desde *psal*, no se observaron relaciones mayores de los niveles de traducción respecto a los de transcripción en las fusiones al gen codificante de la reductasa (pMPO597) ni al de *dszA* (pMPO594) que en las fusiones al gen *dszC* (pMPO593).

Estos experimentos indican que el activador aportado mejora los niveles de expresión, pero no aclaran cuál de los dos posibles promotores dependientes de σ^N , el que precede a *dszE* o a *dszA*, es el que permite la expresión de estos genes.

Los niveles de traducción de la FMN reductasa que acompaña a los genes *dszABC* fueron bajísimos. No se confirmó que se requiriera esa enzima para la transferencia de los

electrones a las monooxigenasas, por lo que es posible que esa función la cumpla otra reductasa producida por la bacteria.

Aparte de para los ensayos de medida de niveles de expresión, las fusiones construidas se aprovecharon también para comprobar si en ellas se mantenían las actividades de producción de índigo y de desulfuración de dibenzotiofeno (mediante crecimiento en medio mínimo con dibenzotiofeno como única fuente de sulfato), y no se mantuvo ninguna de las dos actividades en ninguna de las fusiones. Las fusiones pMPO593 y pMPO597 tienen los genes de la ruta completos, solo que en *dszC*, sustituyendo al codón de parada, se encuentra fusionado el gen *lacZ*. Esto indica que la enzima dibenzotiofeno monooxigenasa queda inactiva con la fusión, y que esta enzima es necesaria además de para la desulfuración de dibenzotiofeno para la producción de índigo, como ya se había descrito en otros sistemas, (Coco y col., 2001; Furuya y col., 2004).

DISCUSIÓN

1. CONSTRUCCIÓN Y VALIDACIÓN DE SISTEMAS DE EXPRESIÓN HETERÓLOGA PARA METAGENÓMICA FUNCIONAL

La metagenómica es una herramienta muy potente para el estudio de la diversidad microbiana y el descubrimiento de nuevas enzimas y productos, basada en el aislamiento de ADN ambiental y su clonación en vectores que se introducen en hospedadores heterólogos, constituyendo metagenotecas. El análisis de estas metagenotecas se puede realizar mediante aproximaciones basadas en secuencia o en función. Las estrategias basadas en secuencia permiten la identificación de nuevas enzimas mediante la secuenciación del ADN metagenómico y su comparación con las secuencias almacenadas en las bases de datos, estando por tanto restringido al descubrimiento de nuevas enzimas con cierto grado de similitud con otras previamente conocidas. Por otro lado, las estrategias basadas en función permiten la identificación de enzimas completamente novedosas, pero presentan como limitación principal la dificultad de la expresión de los genes que las codifican en las bacterias hospedadoras heterólogas. La expresión génica es un proceso que comprende varias etapas, incluyendo la transcripción, traducción y modificaciones postraduccionales, siendo la transcripción el proceso más frecuentemente regulado, limitando en muchas ocasiones la expresión del ADN metagenómico en bacterias heterólogas. Por ese motivo, en esta tesis nos planteamos el desarrollo de nuevos sistemas que aumentaran las probabilidades de expresión del ADN metagenómico, mediante mejoras en la transcripción.

Una clara posibilidad para incrementar la probabilidad de que un gen de una metagenoteca se exprese es transferir la metagenoteca a cuantas especies distintas de bacterias sea posible. Realizando rastreos funcionales en las distintas bacterias se incrementa la probabilidad de identificar un gen codificante de la actividad de interés ya que la maquinaria de expresión génica de bacterias no emparentadas filogenéticamente puede ser lo suficientemente diversa para que alguna de ellas pueda expresar el gen de interés. Con este propósito se han diseñado diversos vectores que pueden ser transferidos a un amplio rango de bacterias Gram negativas y también a alguna Gram positiva. Sin embargo, ello obliga a realizar rastreos funcionales en cada una de las distintas bacterias y, en cualquier caso, el gen en cuestión solo puede ser expresado desde sus propias señales, lo que sigue limitando las posibilidades de expresión, sobre todo de genes regulados positivamente por el concurso de un activador.

Un procedimiento alternativo o complementario consiste en utilizar para la construcción de la metagenoteca vectores de expresión que puedan forzar la transcripción del ADN metagenómico desde promotores del propio vector. Ya existen vectores creados con este objetivo de mejora de la expresión, mediante la incorporación de promotores heterólogos inducibles, como los vectores pEBP con el promotor *pxyl* (Troeschel y col.; 2012) o el vector pJOE930 (Altenbuchner y col., 1992) con promotores *plac* empleado para la construcción de una metagenoteca por Lämmle y col. (2007). La efectividad de estos sistemas normalmente está limitada a fragmentos de ADN de pequeño tamaño, debido a que la expresión heteróloga proporcionada solo puede alcanzar al primer operón del inserto metagenómico a causa de su parada en las señales de terminación de la transcripción presentes al final de los operones. Teniendo en cuenta que en bacterias hay poco ADN recombinante y que los operones se encuentran muy juntos, el uso de insertos de pequeño tamaño obliga a construir bibliotecas metagenómicas con un número de clones muy elevado para tener una representatividad aceptable e impide a su vez el descubrimiento de actividades o funciones que requieran la participación de varios genes. Por este motivo, se pensó que acoplar sistemas de antiterminación a la transcripción heteróloga sería una aportación de utilidad para la metagenómica funcional, al permitir aumentar las probabilidades de expresión de fragmentos de ADN metagenómico de tamaño medio (decenas de kb). De este modo, se desarrollaron dos sistemas de expresión, que permiten mejorar la transcripción del ADN metagenómico de metagenotecas construidas en un nuevo fósido, pMPO579 (derivado de pCC1FOS, Epicentre), los cuales se encuentran esquematizados en la figura 7. El primero de ellos está basado en la ARN polimerasa del fago T7, cuyo promotor se encuentra en el fósido orientado hacia el sitio de clonación del ADN metagenómico. Se requiere la estirpe de *E. coli* MPO553, productora de dicha ARN polimerasa de manera constitutiva en niveles bajos. La ARN polimerasa del fago T7 posee una velocidad de transcripción mucho más elevada que la bacteriana y sus pausas durante la elongación son menores, lo que probablemente sea responsable de que muchos de los terminadores de la transcripción bacterianos sean mucho menos eficaces con esta ARN polimerasa (Studier y Moffatt, 1986; Jeng y col., 1990). Hay publicados algunos trabajos sobre vectores con promotor pT7 empleados para la construcción de metagenotecas, como el de Rhee y col. (2005), que utilizaron el vector pCC1FOS (Epicentre), que posee un promotor pT7, y la estirpe de *E. coli* productora de la ARN polimerasa del fago T7 BL21(DE3) (Novagen), a la que transfirieron la metagenoteca por transformación, y en la que el vector se puede replicar en copia única. Por otro lado, Troeschel y col. (2012) construyeron el vector pEBP18, que tiene un promotor pT7 y que

emplearon en estirpes de *E. coli* y de *P. putida* que producen la ARN polimerasa de T7 de manera inducible, estimando el número de copias de este vector en unas 37 y 5 copias respectivamente en dichas estirpes. En este último trabajo simplemente utilizaron el vector descrito para expresar proteínas heterólogas, pero contemplan la posibilidad de emplearlo como vector para la construcción de metagenotecas. En ambos casos se permitiría la construcción de metagenotecas con tamaño de inserto similar al que puede abarcar pMPO579, y los vectores también pueden ser empaquetados por partículas de fago lambda gracias a los sitios *cos*. Como mejora de estas características, en nuestro sistema se aporta un origen de transferencia conjugativa al vector, que permite la transferencia de los clones metagenómicos de una estirpe a otra mediante conjugación, de manera más eficaz de lo que se podría hacer mediante transformación. Por otro lado, en la estirpe especializada desarrollada en este trabajo la ARN polimerasa de T7 se produce de manera constitutiva, con un sistema desreprimido desde el promotor *placUV5* pero con unos niveles de expresión significativamente menores que en la estirpe superproductora BL21 utilizada anteriormente. Ello se debe a la presencia del atenuador de *nasF* entre el promotor y la región codificante de la ARN polimerasa de T7 que, de acuerdo a lo previamente descrito, debe reducir sus niveles de transcripción 8-10 veces (Royo y col., 2005; Medina y col., 2011).

El segundo de los sistemas consiste en la expresión heteróloga inducible por salicilato desde el promotor *psal* (Yen y Gunsalus, 1982; Schell, 1985; Schell y Wender, 1986) introducido en el fósido orientado hacia el sitio de clonación del ADN metagenómico, acompañado del sitio de utilización de la proteína antiterminadora N del fago lambda (Nodwell y Greenblatt, 1991; Greenblatt y col., 1998). En este caso se requiere la estirpe de *E. coli* MPO554, productora del activador NahR, necesario para que se produzca la inducción de la expresión desde *psal*, y de la proteína antiterminadora N, cuya expresión también es inducible por salicilato. De este modo, mediante la adición de salicilato, se inducirá la transcripción desde el promotor *psal* por la ARN polimerasa bacteriana, a la que se asociará la proteína antiterminadora N, evitando que se detenga en señales de terminación de la transcripción. Este sistema, que permite regular no solo la iniciación de la transcripción sino además su terminación en respuesta a la misma molécula inductora, es completamente novedoso, ya que la antiterminación de la proteína N del fago lambda no se ha empleado con anterioridad para la expresión de genotecas o metagenotecas. De este modo, teniendo metagenotecas construidas en el fósido pMPO579, con insertos metagenómicos de tamaño medio de 30-40 kb, se pueden realizar análisis funcionales de las mismas en una u otra estirpe

especializada, forzando la expresión de los genes mediante transcripción heteróloga sujeta a antiterminación, en MPO554, o mediante una ARN polimerasa de fago insensible a señales de antiterminación bacterianas, en MPO553.

A su vez, el número de copias del vector puede ser tanto alto como bajo, lo que permite modular los niveles de expresión de los genes desde este vector entre 20 y 35 veces, dependiendo del sistema de expresión (figuras 9 y 10). Controlando por un lado el número de copias del fósido (inducible por arabinosa) y por otro los niveles de expresión heteróloga, con los dos sistemas descritos se pueden alcanzar distintos niveles de expresión de un mismo gen. Esto puede ser de utilidad en función de si interesa producir grandes cantidades de una proteína por no disponer de sistemas de rastreo muy sensibles o, por el contrario, mantener los niveles de producción de la proteína en niveles más bajos por ser esta, o el producto de su actividad, tóxica para la bacteria. Desde el promotor *psal* se pueden modular los distintos niveles de expresión en función de la concentración de salicilato aportada (Royo y col., 2007), pero, en términos generales, con estos sistemas se podrían alcanzar tres niveles de expresión: bajos, con el fósido en bajo número de copia en cualquiera de las dos estirpes, ya sea MPO553 o MPO554 inducida con salicilato; medios, con el fósido en alto número de copia en la estirpe MPO554 inducida con salicilato; y altos, con el fósido en alto número de copia en la estirpe MPO553 (figuras 9 y 10).

La ventaja que aporta el uso de los nuevos sistemas se ha comprobado experimentalmente con metagenotecas mediante la búsqueda de distintas actividades enzimáticas: resistencia al antibiótico carbenicilina, dioxigenasa extradiólica, oxigenasa hidroxilante de anillos aromáticos y endoglucanasa, al permitir en cada uno de esos casos la identificación de clones capaces de llevar a cabo la actividad que no eran capaces de hacerlo, al menos en niveles detectables, únicamente con la expresión propia de los genes correspondientes. Con la resistencia a carbenicilina se llevó a cabo un estudio comparativo en profundidad, realizando una búsqueda de clones de una metagenoteca con dicha actividad en la estirpe EPI300-T1^R (estirpe que no permite expresión heteróloga alguna, de la que derivan las demás), MPO553 y MPO554. Los resultados indicaron que solo un clon fue capaz de expresar la resistencia a carbenicilina por sí mismo, siendo el único identificado en EPI300-T1^R. Gracias a la expresión heteróloga basada en la ARN polimerasa de T7, en la estirpe MPO553 se pudieron identificar, además del clon mencionado, otros tres. Y por último, la expresión heteróloga basada en *psal* y la antiterminación N permitió que además de estos cuatro clones, otros dos confiriesen también resistencia a carbenicilina a la estirpe MPO554

en presencia de salicilato. Claramente, el uso del sistema de expresión heteróloga acompañado de la antiterminación de la transcripción, permitió la identificación de cinco clones con la actividad de interés que hubieran pasado desapercibidos de haber contado solo con la expresión propia del ADN metagenómico en la misma bacteria hospedadora. Aunque los números de clones positivos son demasiado bajos para hacer estadística, estos representan un claro incremento del número de clones positivos con un único rastreo funcional, lo que muestra las ventajas del uso de la expresión heteróloga en comparación a los vectores de amplio espectro que requieren hacer múltiples rastreos funcionales en distintas especies bacterianas (Craig y col., 2010).

Por otro lado, para la búsqueda de enzimas codificantes del resto de actividades se empleó solo la estirpe MPO554, pero posteriormente se comprobó que en una parte importante de los clones identificados no se podía detectar la actividad en cuestión solo con la expresión propia. El porcentaje de clones que no fueron capaces de proporcionar la actividad de interés sin expresión heteróloga varió aproximadamente entre un 25 y 45 % de los clones según la actividad rastreada. Por otro lado, una parte de esos clones que sí mostraban el fenotipo de interés con la expresión propia del ADN metagenómico, lo hacían con una clara menor intensidad que cuando se inducía la expresión heteróloga, requiriéndose más tiempo de incubación para detectar colonias azules en el caso de la actividad oxigenasa inicial, dando lugar a un amarillo menos intenso como consecuencia de la actividad dioxigenasa extradiólica, y produciendo halos menos evidentes en los clones codificantes de endoglucanasas. Ello sugiere que una buena parte de estos clones probablemente tampoco hubieran sido detectados entre los cientos de miles de colonias rastreadas durante la búsqueda de cada actividad, si no se hubiese utilizado el sistema de expresión heteróloga.

Teniendo en cuenta la dependencia de expresión heteróloga para la detección de la actividad de interés en ciertos clones, se trató de relacionar esta con la orientación y posición de los posibles genes responsables respecto a los promotores heterólogos. De muchos de estos clones se conoce al menos parte de su secuencia, habiéndoseles asignado como mínimo un gen como posible responsable de la actividad. Y de algunos de ellos se conoce suficiente secuencia como para poder saber la orientación y posición de esos genes en el inserto. Analizando esta información, se identificaron 15 clones (5 clones con actividad de resistencia a carbenicilina, 4 dioxigenasa extradiólica y 6 oxigenasa inicial) que reúnen esas dos características: requieren expresión heteróloga y en ellos se ha identificado al menos un posible gen necesario para la actividad del que además se conoce su posición y orientación en

el inserto. En la mayoría de estos clones, el posible gen responsable se ubica, como es de esperar, en la misma orientación que los promotores heterólogos.

Solo en dos de ellos, N5 y UPO21, los genes codificantes se encuentran en sentido contrario. En el caso de UPO21, clon con actividad oxigenasa inicial y con todos los genes necesarios para la desulfuración de dibenzotiofeno, los genes codificantes de las proteínas implicadas en dichas actividades se encuentran en sentido contrario a los promotores heterólogos, pero se identificó el gen de un activador transcripcional de promotores tipo σ^N en la misma orientación que los promotores del fósido a unas 10 kb de distancia. Mediante la subclonación de este gen y la de los demás genes implicados, se confirmó la necesidad de este regulador, y por tanto, se explicó la necesidad de expresión heteróloga en el clon UPO21 para la detección de la actividad. En el caso de N5, clon resistente a carbenicilina, se localizó un gen codificante de una betalactamasa en sentido contrario a los promotores heterólogos. Una posible explicación sería que alguna de las *orfs* orientadas en sentido adecuado pudiera estar influyendo en la expresión o actividad de la betalactamasa. Puede que esa *orf* fuera una que mostró similitud con reguladores transcripcionales putativos, localizada a 27 kb del inicio del inserto, en la misma orientación que los promotores heterólogos del fósido, aunque esta posibilidad no se confirmó. Si fuese así, este dato mostraría la eficiencia del sistema de transcripción heteróloga, que permitiría la expresión de un gen localizado a tanta distancia del promotor.

En los otros 13 clones, los posibles genes responsables de la actividad correspondiente identificados se encuentran en la misma orientación que los promotores heterólogos, a distancias que van desde 4 a 33 kb, estando en 11 de los clones a más de 10 kb. Esto da una idea del alcance de la transcripción desde los promotores heterólogos asociados a antiterminación de los nuevos sistemas de expresión desarrollados. Se han subclonado los genes o grupos de genes implicados en las actividades de interés de 3 clones cuya actividad responde a expresión heteróloga y en los 3 casos se ha confirmado que los genes subclonados efectivamente eran los codificantes de la actividad y que su expresión heteróloga en el clon correspondiente era necesaria para la actividad en cuestión. Se trata de los genes codificantes de la bomba de eflujo de los clones TN3 y TN4, ubicados a 16 y 13 kb del inicio del inserto en el clon correspondiente, del gen codificante de dioxigenasa extradiólica *upoedo11* de UPO36, localizado a 18 kb, y el gen *dszR* de UPO21, codificante del activador de promotores tipo σ^N necesario para la expresión de los genes de la ruta de desulfuración de dibenzotiofeno, localizado a 10 kb. Con estos datos se puede asegurar que la expresión heteróloga de estos

sistemas alcanza al menos a genes ubicados a 18 kb de los promotores heterólogos, como ocurre en el clon UPO36, cuya transcripción discurre a través de 9 *orfs* dispuestas en dirección contraria y de 9 *orfs* en el mismo sentido, con los consiguientes terminadores de la transcripción que se puedan encontrar en un fragmento de ADN con esas características. Probablemente merezca la pena destacar también que en el clon TN3 esa transcripción transcurre a través de 14 *orfs* que se encuentran en el mismo sentido antes de los genes codificantes de la bomba de eflujo. La transcripción heteróloga probablemente alcance a distancias bastante mayores, ya que en el clon N5 la transcripción alcanza presumiblemente a un gen ubicado a 27 kb y en el clon UPO25, que presenta actividad oxigenasa inicial, el posible gen responsable de la actividad de interés se sitúa a 33 kb de los promotores heterólogos. Con este clon se puede producir índigo tanto en la estirpe MPO553 como en la MPO554 en presencia de salicilato, lo que indica que los dos sistemas son efectivos a esa distancia tan larga. Sin embargo, estos genes no se han subclonado, y aunque el análisis de la totalidad de la secuencia disponible para cada clon indica que con una elevada probabilidad esos genes son los responsables de la actividad detectada, no se puede afirmar con rotundidad.

Por otra parte, aunque los ensayos realizados con el vector vacío indicaron que la expresión desde pT7 en MPO553 alcanza niveles unas 3 veces mayores que desde *psal* en MPO554, se ha observado que en varios clones se puede obtener expresión heteróloga de los genes de interés exclusivamente desde *psal* con el sistema de antiterminación de la proteína N, no logrando la transcripción desde pT7 alcanzar esos genes. Es el caso de los clones N5 y N6, expresando resistencia a Cb, y de UPO21 y UPO92, con oxigenasas iniciales. Esto indica que puede haber algunas regiones que no se consigan transcribir por la ARN polimerasa de T7 y sí por la bacteriana desde *psal* con el sistema de antiterminación N, lo que pone de manifiesto una mayor eficiencia de este nuevo sistema de expresión heteróloga frente al sistema dependiente de T7 previamente utilizado en vectores para la construcción de metagenotecas.

La búsqueda de varias de las actividades llevadas a cabo en este proyecto también se ha visto beneficiada por la ventaja de las metagenotecas de insertos de cierto tamaño consistente en la posibilidad de detección de actividades que requieren la expresión de varios genes. Se trata de la identificación de dos bombas de eflujo implicadas en la resistencia a carbenicilina. Una de ellas está compuesta por las proteínas codificadas por 3 genes, que abarcaría 5,8 kb, o por 4 genes que abarcarían 7,4 kb, en caso de requerir además una proteína de membrana externa adicional codificada por un gen ubicado a continuación de los otros 3, en el clon

ETN1. La otra estaría codificada por 3 genes, que ocupan 5,8 kb, en los clones TN3 y TN4. Si se hubieran empleado metagenotecas de tamaño de inserto pequeño, las enzimas identificadas como responsables de esta actividad probablemente se habrían visto limitadas a betalactamasas, como ha ocurrido con análisis metagenómicos funcionales previos en busca de resistencias a antibióticos betalactámicos (Sommer y col., 2009). Otro ejemplo es la identificación de 4 genes codificantes de citocromos P450 en clones con actividad oxigenasa inicial, que van acompañados de ferredoxinas (1 de ellos en los clones UPO10 y UPO24, 2 en el clon UPO30 y estos 2 últimos y otro más en UPO28). De los componentes de los sistemas de transporte de electrones que requieren las oxigenasas, las ferredoxinas son bastante específicas, por lo que a diferencia de lo que ocurre en muchas ocasiones con las reductasas, de no encontrarse en el inserto metagenómico difícilmente va a poder suplirse su actividad con la de otra enzima propia de la bacteria hospedadora. También se puede mencionar la identificación de 2 enzimas multicomponentes fenol 2-monooxigenasas distintas, cada una codificada por 6 genes, que abarcan 4,7 kb en ambos casos, en los clones UPO16 y UPO41. Uno de estos clones, UPO41, presentó también otra actividad implicada en la degradación de compuestos aromáticos, como es la actividad dioxigenasa extradiólica, y en su secuencia se identificaron más genes implicados en la degradación del fenol, pudiéndose en teoría conseguir la obtención de 4-hidroxi-2-oxovalerato a partir de fenol, faltando solo el gen codificante de la 4-hidroxi-2-oxovalerato aldolasa para la obtención de acetaldehído y piruvato. Además, se identificó otra ruta completa, la ruta de desulfuración de DBT, en el clon UPO21. En esta ocasión se requieren 3 genes, que ocupan 3,7 kb, o 4 genes si también fuera necesario el gen codificante de la FMN reductasa ubicado justo aguas arriba de los otros 3 genes, que abarcarían 4,3 kb. Por otro lado, para la expresión de estos genes, se requirió un activador codificado por un gen ubicado a 3,4 kb de los otros. Por tanto, para incluir todo lo necesario para que la desulfuración de DBT se pueda llevar a cabo, es necesario un fragmento de unas 9 kb.

2. OTROS FACTORES IMPORTANTES PARA LA IDENTIFICACIÓN DE ACTIVIDADES DE INTERÉS

A la hora de realizar una búsqueda de enzimas de interés, lo primero es disponer de una metagenoteca adecuada. Para ello, se debe seleccionar la muestra a partir de la cual extraer el

ADN en función de las actividades que se vayan a buscar posteriormente. En nuestro caso, debido al interés del grupo de investigación en actividades de degradación de contaminantes, se seleccionaron dos suelos con historial de contaminación por hidrocarburos.

Antes de proceder a la extracción del material genético de las muestras hay que decidir el vector en el que se van a construir las metagenotecas, ya que el tratamiento de la muestra puede variar en función de esa decisión. En este punto hay que tener en cuenta distintos factores, como el tamaño de los fragmentos de ADN metagenómico, el número de copias del vector o las posibles bacterias hospedadoras. Los fósquidos constituyen probablemente los vectores más apropiados para construir metagenotecas con el objetivo de realizar rastreos funcionales por el adecuado equilibrio de sus características. Por un lado, su replicación es extraordinariamente estable y aunque tenga un número de copia muy bajo, la utilización de un sistema de replicación adicional permite incrementar su número de copias. Por otro lado, se puede aprovechar un sistema eficaz de introducción de fragmentos de ADN relativamente grandes en las bacterias hospedadoras a través de empaquetamiento en partículas del fago lambda y transfección, que permite obtener un número de clones equivalente a la clonación de pequeños fragmentos de ADN en plásmido mediante transformación. A estas características deseables se le han sumado en esta tesis la posibilidad de transferir la metagenoteca por conjugación y la de expresar heterológamente genes que se encuentren muy lejos de los extremos de los fragmentos de ADN, como se ha comentado previamente, lo que supone una mejora sustancial respecto a los vectores existentes con anterioridad.

Entre los rastreos realizados, destacan los de enzimas capaces de llevar a cabo los dos pasos clave en la degradación de compuestos aromáticos: mono- o dioxigenasas hidroxilantes de anillos aromáticos y dioxigenasas extradiólicas. Teniendo en cuenta resultados anteriores, que indicaban que la transcripción de los genes metagenómicos es un problema muy importante que limita la eficacia del análisis metagenómico funcional, y que este puede ser solventado, al menos en parte, utilizando los sistemas de expresión derivados de fagos que se han desarrollado en esta tesis, estos rastreos se realizaron en la estirpe especializada MPO554, que permitía la inducción de la transcripción desde el promotor heterólogo *psal* asociada a la antiterminación aportada por la proteína N.

El rastreo de dioxigenasas extradiólicas se basa en la identificación de los clones con dicha actividad por la aparición de un color amarillo, que es el color de los productos de la reacción empleando como sustrato catecol o un derivado catecólico. En este trabajo se utilizaron como sustrato en estos rastreos dos compuestos distintos, con la intención de

ampliar el rango de clones a detectar porque pudieran reconocer uno u otro de los sustratos. Se trataba por un lado de catecol, compuesto con un solo anillo aromático, y por otro de 2,3-dihidroxibifenilo, con dos anillos. Los resultados indicaron que para un mismo número de clones analizados se identificaron más clones distintos con actividad dioxigenasa extradiólica con 2,3-dihidroxibifenilo que con catecol (44 frente a 17). Por otro lado, los ensayos de actividad realizados con la totalidad de los clones identificados empleando 5 sustratos distintos mostraron que un número elevado de clones presentaba más actividad con 2 de estos sustratos (4-metilcatecol y sobre todo 3-metilcatecol) que con catecol o 2,3-dihidroxibifenilo, lo que sugiere que quizás la mejor opción para la realización de estos rastreos sea el empleo de una mezcla de sustratos catecólicos.

La búsqueda de oxigenasas hidroxilantes mediante la detección de colonias de color azul en medio LB por la producción de índigo a partir de indol se ha utilizado ampliamente por ser un ensayo barato y sencillo, y ha permitido la identificación de distintas enzimas de este tipo. Se ha empleado tanto en el estudio de organismos degradadores como en análisis funcionales metagenómicos. En el estudio de organismos degradadores, se ha descrito que las naftaleno dioxigenasas son capaces de emplear distintos compuestos aromáticos como sustrato, entre los que se encuentran el naftaleno, el bifenilo y el indol (Boyd y col., 2001). Una dioxigenasa similar implicada en la ruta de degradación de tetralina puede también producir índigo a partir de indol (Royo y col, 2005a). A la tolueno dioxigenasa, que dioxigena compuestos monoaromáticos, también se le ha atribuido la capacidad de reconocer el indol como sustrato (Kim y col., 2003). Dentro de las monooxigenasas, se pueden citar como ejemplos una flavina monooxigenasa de *Methylophaga* sp. strain SK1 (Choi y col., 2003), la estireno monooxigenasa de *Pseudomonas putida* (O'Connor y col., 1997) y la dibenzotiofeno monooxigenasa de *Bacillus subtilis* WU-S2B expresada en *E. coli* (Furuya y col., 2004). A pesar de estos datos, los resultados de esta tesis indican que el rastreo de producción de índigo realizado en la metagenoteca ha identificado exclusivamente monooxigenasas, ya que ninguno de los 14 genes identificados como responsables de la actividad en los 18 clones productores de índigo secuenciados codificaba una dioxigenasa. Este sesgo tan fuerte hacia las monooxigenasas en principio pensamos que podría deberse a que las dioxigenasas fueran muy infrecuentes en ese suelo. Sin embargo, como se discute en el apartado 4, sí se detectaron genes codificantes de dioxigenasas hidroxilantes en otros clones de la metagenoteca, concretamente en muchos de los clones que presentaron actividad dioxigenasa extradiólica. Resulta improbable que existiendo este tipo de genes en la metagenoteca, no se haya

detectado ninguno si sus productos pudieran generar índigo. Por tanto, la interpretación más apropiada es que la mayoría de las dioxigenasas no son capaces de generar índigo. De hecho, entre los estudios metagenómicos de los que tengamos conocimiento, hasta la fecha solo se ha atribuido la capacidad de transformar el indol en índigo mediante este tipo de rastreos a monooxigenasas o a oxigenasas putativas, entre las que se encuentran una monooxigenasa putativa (Brady y col., 2001), una indol oxigenasa putativa (Lim y col., 2005), una estireno monooxigenasa (van Hellemond y col., 2007) y una 2-naftoato monooxigenasa (Kimura y col., 2010), pero nunca se han identificado enzimas claramente homólogas a dioxigenasas.

Por lo que se acaba de comentar, si se desean realizar análisis funcionales en busca de dioxigenasas sería probablemente aconsejable diseñar sistemas de rastreo basados en otros sustratos, aunque por lo general se trata de ensayos más complejos que el de producción de índigo. Entre las alternativas posibles, se encuentra el uso de sustratos específicos como 2 p-nitrofeniloctil éter, que da lugar a un producto de color amarillo a pH por encima de 7 (Farinas y col. 2001; Kennedy y col., 2011), o cetonas 2-coumariloxi cíclicas y acíclicas, que dan lugar a un compuesto azul fluorescente (Sicard y col. 2005; Kennedy y col., 2011). Otra opción es acoplar reacciones de peroxidación a las de oxidación dando lugar a compuestos coloreados y/o fluorescentes (Joo y col., 1999).

3. RESISTENCIAS A CARBENICILINA IDENTIFICADAS EN LA METAGENOTECA DE PUNTA DE SAN GARCÍA

En la metagenoteca de Punta de San García se aislaron 6 clones distintos que aportan resistencia al antibiótico carbenicilina. De entre ellos, se identificaron 3 genes codificantes de betalactamasas, enzima clave que proporciona un mecanismo de resistencia perfectamente conocido que consiste en destruir el antibiótico. Se trata de una betalactamasa de clase C y de dos betalactamasas de clase B.

Se han descrito diversos análisis metagenómicos funcionales de resistencias a antibióticos betalactámicos de distintos nichos ecológicos y se ha mostrado que estas resistencias pueden encontrarse en ambientes muy diversos (Pehrsson y col., 2013; Mullany, 2014). Dentro de los análisis de suelos, Donato y col. (2010) construyeron una metagenoteca de un suelo de un manzano de Wisconsin en pCC1FOS, que contenía unas 13 Gb de ADN metagenómico. Empleando 7 antibióticos betalactámicos distintos, identificaron 8 enzimas

homólogas a betalactamasas, 2 de cada uno de los 4 grupos establecidos por Ambler (1980): A, B, C y D, lo que sugiere una distribución homogénea de los distintos tipos de betalactamasas en este suelo. Por su parte, Allen y col. (2009) construyeron 8 metagenotecas con muestras de un suelo de Alaska, empleando distintos vectores: pCF430 (Newman y Fuqua, 1999), pCC1BAC (Epicentre) y pCC1FOS (Epicentre), que contenían en total unas 12 Gb de ADN metagenómico. Utilizando 8 antibióticos betalactámicos identificaron 13 secuencias homólogas a proteínas que conferirían resistencia a este tipo de antibióticos, pertenecientes a las 4 clases de betalactamasas descritas (A, B, C y D), concretamente 2 de la clase A, 8 de la clase B, 3 de la clase C y 1 de la clase D, lo que también mostró que se podían encontrar los 4 tipos de betalactamasas en esos suelos teóricamente prístinos. Estos datos de análisis metagenómicos contrastan en gran medida con el hecho de que la gran mayoría de las betalactamasas de las bacterias resistentes aisladas de suelos sean de tipo C (Forsberg y col., 2012).

En la metagenoteca de Punta de San García se identificaron solo 3 betalactamasas distintas, dos de clase B y una de clase C, pero teniendo en cuenta que esa metagenoteca contenía algo menos de 2 Gb de ADN metagenómico y que en las metagenotecas anteriormente mencionadas se obtuvo aproximadamente un gen codificante de betalactamasa por cada Gb de la metagenoteca, no hay motivo para pensar que el suelo costero contaminado fuera más pobre en microorganismos productores de resistencias a ese tipo de antibióticos. Además, puede que se hubieran identificado más betalactamasas si se hubieran empleado más antibióticos betalactámicos en los rastreos, como en los otros dos trabajos. Por otro lado, el hecho de obtener dos betalactamasas de clase B indica que la distribución de betalactamasas detectadas en nuestra metagenoteca se parece más a la obtenida con otras metagenotecas de suelos que a la distribución sesgada hacia el tipo C obtenida de las bacterias cultivadas.

A pesar de que la producción de betalactamasas es la forma de resistencia más frecuente ante antibióticos betalactámicos (Henriques y col., 2006), hay otros dos mecanismos que confieren tolerancia a este tipo de antibióticos, que son el uso de peptidasas de pared celular insensibles a betalactámicos y la expulsión activa de moléculas de betalactámicos de las células mediante bombas de eflujo (Wilke y col., 2005). En las secuencias de los clones identificados en el rastreo realizado, además de las betalactamasas mencionadas, se encontraron los genes codificantes de los componentes de dos bombas de eflujo, una en el clon ETN1 y otra en los clones TN3 y TN4. Nuestros datos muestran que la bomba de eflujo de ETN1 podría estar contribuyendo a la resistencia a carbenicilina en ese clon, ya que tras

subclonar el gen codificante de la betalactamasa de clase C de ETN1, el crecimiento en presencia de Cb 100 mg/L fue peor que el aportado por el clon completo, lo que indica que probablemente haya otro elemento implicado en dicha resistencia. Por otro lado, en los clones TN3 y TN4 no se identificó ningún gen codificante de betalactamasa, por lo que se pensó que la bomba de eflujo producida por estos clones podría ser la responsable de la tolerancia a Cb. Mediante la subclonación de los 3 genes codificantes de los distintos componentes de la bomba de eflujo se confirmó que esta bomba es capaz de proporcionar resistencia a carbenicilina 100 mg/L, y también a ampicilina 50 mg/L, sin necesidad de ningún otro elemento.

La contribución de las bombas de eflujo a la resistencia a diversos antibióticos es un fenómeno conocido y algunas formas de resistencias a diversos antibióticos como cloranfenicol, trimetoprim o tetraciclina aisladas de metagenotecas han resultado ser transportadores (McGarvey y col., 2012). Por el contrario, muy pocas bombas de eflujo se han relacionado con la resistencia a antibióticos betalactámicos y estas se encuentran asociadas a los otros mecanismos de resistencia (Poole, 2005). Hasta la fecha no tenemos constancia de ninguna otra bomba de eflujo que sea capaz de proporcionar la resistencia a antibióticos betalactámicos por sí misma o que haya sido identificada en un análisis metagenómico funcional, ya que en todos los anteriores siempre se identificaron genes codificantes de betalactamasas como los determinantes de la resistencia y en ningún caso llevaban asociadas bombas de eflujo. Por el contrario, en el análisis realizado en esta tesis se habrían identificado dos bombas de eflujo, una que confiere resistencia a carbenicilina por sí misma y otra que probablemente contribuya a esa resistencia junto con una betalactamasa. Posiblemente, esto sea debido a las diferencias en prestaciones de los vectores utilizados anteriormente, que permitían la clonación y transcripción heteróloga de fragmentos pequeños en el caso de plásmidos o no permitían expresión heteróloga en el caso de pCC1FOS.

Por otra parte, si para la búsqueda de resistencias a carbenicilina se hubieran realizado análisis basados en secuencia en lugar de en función, muy probablemente esas dos bombas de eflujo habrían pasado desapercibidas como responsables o participantes en esta actividad, habiéndose limitado los resultados a la identificación de las betalactamasas.

4. ENZIMAS IMPLICADAS EN RUTAS DE DEGRADACIÓN DE CONTAMINANTES AROMÁTICOS

En la metagenoteca de la pila de refinería de la industria petroquímica se identificaron 36 clones con aparente actividad oxigenasa inicial y 58 con actividad dioxigenasa extradiólica, pero muy probablemente existan más clones para ambas actividades. Un reflejo de esto es el bajo grado de repetición observado en los patrones de restricción de los clones positivos identificados para cada rastreo. Tras descartar los clones exactamente iguales, se siguieron identificando algunos genes iguales en distintos clones. Se trataba por un lado de clones con fragmentos solapantes del genoma del mismo organismo, y por otro de clones con fragmentos parecidos que incluían genes que en ocasiones eran exactamente iguales (probablemente pertenecientes a organismos filogenéticamente cercanos). El hecho de que un mismo genoma se encuentre representado en varios clones es una indicación de la abundancia del organismo de procedencia en la muestra ambiental empleada para la construcción de la metagenoteca frente al resto de organismos degradadores de compuestos aromáticos.

En los clones con actividad oxigenasa inicial, detectados por producir índigo, curiosamente todos los genes identificados como posibles responsables de la actividad codificarían monooxigenasas, en ningún caso dioxigenasas. En principio esto hizo pensar que el suelo del que se tomó la muestra para la construcción de esta metagenoteca era bastante más rico en monooxigenasas que en dioxigenasas. Sin embargo sí que se identificó por secuencia la presencia de genes codificantes de dioxigenasas iniciales en clones con actividad EDO con secuencia asignada. Concretamente en 20 de los 28 clones de los que se conocía secuencia se identificaron mono- o dioxigenasas iniciales, siendo solo dos de ellos capaces de generar índigo, ambos con monooxigenasas. A falta de realizar un análisis exhaustivo de esas secuencias, mediante una primera comparación con las secuencias de las bases de datos, entre estos 20 clones se han identificado 13 secuencias codificantes de dioxigenasas. En estas secuencias, para la mayoría de las dioxigenasas se encuentran genes homólogos de genes codificantes tanto de la subunidad α como de la subunidad β , por lo que en principio no les faltaría ninguna subunidad para ser funcionales. Por otro lado, en varios de esos clones se encuentran genes que podrían codificar los componentes del sistema de transporte de electrones necesarios para la actividad de esas dioxigenasas, y podría haber más, al no conocerse la secuencia completa de todos los clones. Por tanto, aunque a algunas de estas dioxigenasas iniciales les pudiera faltar algún componente necesario para realizar su función, no es probable que este sea el problema por el que no se ha detectado actividad oxigenasa en

ninguno de esos clones. Aproximadamente la mitad de estos genes se encuentran en la misma orientación que los promotores heterólogos del vector, por lo que se esperaría que al menos una parte de ellos se estuviera expresando. Al margen de la posibilidad de que algunos de estos genes no se estuvieran expresando en la bacteria hospedadora, el problema en gran parte de ellos puede estar en que no sean capaces de reconocer el indol como sustrato, y por tanto no den lugar a colonias azules por la producción de índigo, aunque sí pudieran ser capaces de actuar sobre otros compuestos, como se propone en el apartado 2. En un reciente estudio realizado con una metagenoteca construida con cultivos de enriquecimiento de bacterias degradadoras de pireno y fenantreno (Singleton y col. 2012), se identificaron diversos genes codificantes de dioxigenasas iniciales. Se clonaron y expresaron en *E. coli* los genes codificantes de seis de esas dioxigenasas y se comprobó si presentaban actividad frente a nueve hidrocarburos aromáticos policíclicos que contenían de dos a cinco anillos. Las dioxigenasas presentaban distinta actividad frente a distintos sustratos y observaron que solo las tres enzimas que reconocieron naftaleno y fluoreno también pudieron transformar el indol hasta índigo. Estos datos indican que algunas dioxigenasas, principalmente las naftaleno dioxigenasas y otras enzimas relacionadas, sí son capaces de oxidar el indol (Boyd y col., 2001; Kim y col., 2003; Royo y col., 2005a), posiblemente por la similitud estructural entre el indol y el naftaleno, pero que el indol no es un sustrato muy adecuado para identificar dioxigenasas en general.

Respecto a los genes identificados como posibles responsables de la producción de índigo en los clones con esa actividad, se trata de 14 genes que por comparación de sus secuencias con las bases de datos codificarían proteínas homólogas a cuatro citocromos P450, dos estireno monooxigenasas, dos alcanosulfonato monooxigenasas, una dimetilnilina monooxigenasa, dos fenol 2-monooxigenasas, una dibenzotiofeno monooxigenasa acompañada de una dibenzotiofeno-5,5-dióxido monooxigenasa y una 2,4-diclorofenol 6-monooxigena. Uno de los clones con fenol 2-monooxigenasa (UPO41) posee más genes implicados en la degradación de fenol, por lo que puede que su función sea efectivamente utilizar el fenol como fuente de carbono y energía. Puede que el clon UPO55, del que se conocen solo unas 3 kb de secuencia que incluyen los genes de una 2,4-diclorofenol 6-monooxigenasa y de una catecol-2,3-dioxigenasa, también pueda codificar una ruta de degradación de contaminantes aromáticos. Sin embargo, en el entorno del gen codificante de la otra fenol 2-monooxigenasa (clon UPO16) no se identificaron otros genes relacionados, lo que posiblemente indique que el organismo de procedencia de ese ADN sea capaz de

hidroxilar fenol o una molécula similar, aunque no sea capaz de degradarlo. Algo similar ocurre con los clones codificantes de la dimetilnilina monooxigenasa y las dos estireno monooxigenasas, que por secuencia podría decirse que podrían estar implicadas en la degradación de compuestos aromáticos, pero al no encontrarse en sus entornos genómicos otros genes codificantes de enzimas implicadas en rutas de degradación, puede que sean hidroxilasas que tengan una función original distinta a la participación en la degradación de compuestos aromáticos. Los citocromos P450 tienen la capacidad de hidroxilar muy diversas moléculas y además de poder estar implicados en la degradación de compuestos aromáticos, a muchos de ellos se les han atribuido un rango de funciones muy amplio y en muchos casos se desconoce su función (Kelly y Kelly, 2013). Tampoco se han detectado otros genes de rutas de degradación en su contexto genómico. Las alcanosulfonato monooxigenasas actúan sobre un amplio rango de compuestos pero su función es la obtención de azufre a partir de compuestos organosulfonados $R-CH_2-SO_3H$ (Ellis, 2011), por lo que inicialmente no estarían implicadas en la degradación de anillos aromáticos. Aunque en un clon codificante de una de ellas (UPO23) se ha detectado el gen codificante de una dioxigenasa extradiólica, que no mostró actividad en un ensayo funcional con catecol, no parece estar relacionado con la alcano sulfonato monooxigenasa. La dibenzotiofeno monooxigenasa junto con las demás actividades codificadas por los genes vecinos constituye la ruta de desulfuración de DBT, consistente en la utilización del azufre del DBT como fuente de azufre, generando 2-hidroxibifenilo, por lo que estos genes tampoco están implicados en la degradación de anillos aromáticos.

En definitiva, el rastreo de la metagenoteca en busca de clones con actividad productora de índigo ha identificado una serie de enzimas oxigenasas de distintos tipos, algunas de las cuales parecen incluso actuar sobre sustratos lineales o sobre cadenas lineales de sustratos más complejos. Sin embargo, casi ninguna de ellas parece estar implicada en la degradación de contaminantes aromáticos. Estos datos indican que la producción de índigo puede ser un método de rastreo eficaz para la identificación de algunas actividades hidroxilantes que puedan ser de interés biotecnológico. Sin embargo, a diferencia del rastreo de actividades dioxigenasas extradiólicas, como se verá a continuación, no es adecuado para identificar rutas de degradación de anillos aromáticos, ya que la gran mayoría de genes de una metagenoteca cuyos productos pueden generar índigo, en realidad pueden no estar relacionados con las rutas catabólicas de utilización de contaminantes aromáticos, sean estas de la vía *meta* o la *orto*, o al menos no están ligados al resto de los genes degradadores.

Las otras enzimas para las que se realizaron rastreos, las dioxigenasas extradiólicas, forman parte del equipo enzimático necesario para degradar compuestos aromáticos por la vía *meta* (o extradiólica) y utilizarlos como fuentes de carbono y energía, catalizando la apertura extradiólica de un anillo aromático dihidroxilado. Las oxigenasas hidroxilantes (parte de ellas) también están implicadas en estas rutas de degradación, realizando las oxigenaciones iniciales que generan estos anillos aromáticos dihidroxilados. Al ser actividades implicadas en la misma ruta, no es extraño que haya organismos que posean las dos actividades, encontrándose muy a menudo los genes responsables de cada actividad relativamente cercanos en el genoma. De este modo, metagenotecas con clones positivos para una actividad probablemente también contengan clones positivos para la otra, e incluso, si los fragmentos metagenómicos son lo suficientemente grandes, sería posible encontrar frecuentemente genes codificantes de las enzimas responsables de ambas actividades en un mismo clon. Una muestra son los trabajos realizados por Silva y col. (2013) y por Jadeja y col. (2014), en los que en una misma metagenoteca identifican secuencias de genes homólogos a genes codificantes tanto de oxigenasas iniciales como de dioxigenasas extradiólicas.

Cuando se asignó secuencia a parte de los clones con actividad EDO se pudo observar que el 70 % de ellos contenían también genes de oxigenasas hidroxilantes, aunque, como ya se ha comentado, solo dos presentaron actividad frente a indol. Algunas de las oxigenasas hidroxilantes identificadas en estos clones son homólogas a enzimas anotadas como fenol monooxigena, clorobenzoato, benzoato, fenilpropionato, PAH (hidrocarburos poliaromáticos) e ibuprofeno-CoA dioxigenasas. El análisis de las regiones circundantes a los genes codificantes de EDOs mostró también la existencia de genes codificantes para otras actividades de rutas de degradación en algunos de los clones. Entre ellas que se encuentran el resto de enzimas implicadas en la degradación de fenol junto con la fenol 2-monooxigenasa y la catecol 2,3-dioxigenasa: 2-hidroxiacetato semialdehído deshidrogenasa, 2-oxopent-4-enoato hidratasa, 4-oxalocrotonato descarboxilasa, 4-oxalocrotonato tautomerasa y 4-hidroxi-2-oxovalerato aldolasa, encontrándose todas estas enzimas en uno u otro de 4 de los clones. Por otro lado, en otro clon también se encontró el gen codificante de una *cis*-2,3-dihidroxi-2,3-dihidrobifenilo deshidrogenasa y de una 5-oxopent-3-eno-1,2,5-tricarboxilato descarboxilasa, del mismo modo implicadas en rutas de degradación de hidrocarburos aromáticos. Y por último, en uno de los clones, además del gen codificante de una oxigenasa hidroxilante de anillos aromáticos, se encontraron varios genes codificantes de monooxigenasas de hidrocarburos lineales, como metano y propano. Esto pone de manifiesto

la asociación de los genes codificantes de estas actividades degradadoras en los metagenomas y muestra que la identificación de clones con actividad dioxigenasa extradiólica es un procedimiento útil para obtener operones completos o al menos genes con distintas actividades de rutas de degradación de contaminantes por la vía extradiólica.

La secuencia que se asignó a clones concretos con actividad dioxigenasa extradiólica contenía 22 genes codificantes de enzimas de este tipo completos y otro no completo, que se encontraba al final del inserto. El análisis de estas secuencias de EDOs metagenómicas reveló que todas presentaban dos dominios, como la mayoría de las descritas previamente, encontrándose representantes de los tipos I y II descritos por Eltis y Bolin (1996). 18 de las proteínas se ubicaban dentro del tipo I, distribuidas en tres de esas seis familias distintas (I.2, I.3 y I.4), dentro de cada una de las cuales se encontraba al menos una EDO lo suficientemente distinta a las de las subfamilias establecidas como para considerarla miembro de una nueva subfamilia. De este modo se establecerían una nueva subfamilia en la familia I.2 (I.2.H), cuatro en la familia I.3 (I.3.O-R) y dentro de la familia I.4, en la que no se habían establecido aún subfamilias, se ubicaría una de las nuevas secuencias con una identidad con las demás tan baja que podría considerarse perteneciente a una subfamilia distinta. Por otro lado, las otras 4 EDOs más la procedente del gen incompleto fusionado a secuencia de vector pertenecerían al tipo II, grupo que cuenta con muchos menos representantes que el tipo I. El hecho de que de 23 secuencias cinco pertenezcan al tipo II, menos abundante, y que de las otras 18 seis pertenezcan a subfamilias no descritas con anterioridad, indica que la muestra de suelo de la pila de refinería de la industria petroquímica es un ambiente con una gran diversidad de enzimas de este tipo y da una idea de la gran cantidad de enzimas que quedan por conocer. Las secuencias incluidas en este análisis son aquellas asignadas a únicamente 28 de los 58 clones que se detectaron en la metagenoteca con esta actividad y, por tanto, no constituyen la totalidad de las dioxigenasas extradiólicas presentes en esos 58 clones. Posteriormente a la realización de estos análisis se identificaron otras 12 secuencias completas y 3 incompletas de EDOs, cuya incorporación a este estudio en un futuro probablemente permita el establecimiento de otras nuevas subfamilias.

Desde el establecimiento de estas categorías por Eltis y Bolin en 1996 se han ido incorporando nuevas secuencias y definiéndose con algunas de ellas nuevas categorías de EDOs. Las incorporaciones han sido más numerosas dentro de las familias I.2 y I.3. Dos trabajos en los que se han identificado gran cantidad de nuevas EDOs son los realizados por Suenaga y col. (2007) y Brennerova y col. (2009). En ambos casos se realizaron análisis

metagenómicos funcionales con metagenotecas de muestras relacionadas con hidrocarburos aromáticos, lo que vuelve a mostrar la idoneidad de este tipo de muestras para la búsqueda de estas enzimas. Brennerova y col. identificaron 235 clones con actividad dioxigenasa extradiólica, y localizaron 170 genes, mediante amplificación con oligonucleótidos pertenecientes a distintos tipos de EDOs, obteniendo la secuencia de una parte de cada uno de ellos. Todas las enzimas codificadas por esos genes (la región cuya secuencia se conocía) mostraban una identidad de al menos el 55 % con otras previamente descritas, en lo que sin duda habrá influido el modo de identificación de esos genes, quedando el 32 % de los clones sin secuencia conocida, en los que muy probablemente se encontrarían genes con menor grado de identidad con los de las bases de datos. Por otro lado, Suenaga y col. identificaron 91 clones con actividad dioxigenasa extradiólica, y secuenciaron 38, encontrando 43 genes codificantes de EDOs. Más de la mitad de esas EDOs pertenecían a nuevas subfamilias, una dentro de la familia I.1 y tres más en la familia I.2, destacando la nueva subfamilia I.2.G con 20 nuevas secuencias. Ninguna de las secuencias identificadas en esta tesis se ubicaba en ninguna de las subfamilias inauguradas por Suenaga y col. De esta información se concluye que en un análisis metagenómico en el que se obtenga un cierto número de clones con esa actividad, es bastante frecuente la aparición de EDOs con bajo grado de identidad con las enzimas descritas con anterioridad.

Como se puede comprobar en el presente trabajo, una combinación de análisis funcionales con análisis basados en secuencia puede ser ventajosa para la detección de enzimas responsables de actividades relacionadas. En este caso se ha observado que las secuencias de los clones con actividad dioxigenasa extradiólica, además de para identificar el posible gen responsable de dicha actividad, pueden servir para la identificación de un gran número de genes homólogos a genes codificantes de oxigenasas hidroxilantes. Aunque no se ha comprobado la funcionalidad de estas enzimas, es muy probable que muchas de ellas puedan presentar actividad frente a sustratos diferentes al indol.

5. CARACTERIZACIÓN DEL OPERÓN DE DESULFURACIÓN DE DIBENZOTIOFENO

En el clon UPO21, seleccionado como productor de índigo, se identificaron los genes codificantes de los componentes de la ruta de desulfuración de dibenzotiofeno:

dibenzotiofeno-5,5-dióxido monooxigenasa (*dszA*), 2-hidroxi-bifenilo-2-sulfinato desulfinaasa (*dszB*) y dibenzotiofeno monooxigenasa (*dszC*) (Denome y col., 1994; Gray y col., 1996). Estos genes son muy parecidos a los descritos en *Rhodococcus* sp. IGTS8, pero a diferencia de lo que ocurre en la estirpe IGTS8, en UPO21 van precedidos por un gen que hemos denominado *dszE* que codifica una FMN reductasa supuestamente implicada en la transferencia de electrones.

Los 4 posibles genes implicados se subclonaron, de forma que pudieran expresarse desde los promotores heterólogos del vector, confirmándose que eran los responsables de la actividad productora de índigo y que conferían la capacidad de crecer con DBT como única fuente de azufre. Sin embargo, sorprendentemente, no se pudo conseguir la expresión de estos genes ni con la ARN polimerasa de T7 desde el promotor pT7 en la estirpe MPO553, ni con la ARN polimerasa bacteriana desde *psal* con el sistema de antiterminación de N en la estirpe MPO554, y fue necesario aportar además un activador de promotores dependientes de σ^N , codificado por el gen *dszR*, que se subclonó en otro vector. No se ha descrito previamente que reguladores de este tipo estén implicados en la expresión de rutas de desulfuración de DBT, aunque la utilización de DBT como fuente de carbono y energía sí parece estar regulada por este tipo de activadores de la transcripción (Di Gregorio y col., 2004).

Para tratar de aclarar la incapacidad de expresar las dos actividades desde los promotores del vector se construyeron una serie de fusiones a distintos genes del operón, traducionales a *lacZ* y transcripcionales a *gfp*, tratando de analizar la expresión de las fusiones génicas cuando se activaba uno u otro promotor. Los niveles de transcripción desde el propio promotor en presencia del regulador DszR aumentaron respecto a los niveles basales, aunque en varias de las fusiones ese incremento fue mínimo. Por su parte, la inducción desde el promotor *psal* sí se ve reflejada más claramente en un aumento en los niveles de transcripción, aunque de nuevo, estos no son demasiado elevados. Por el contrario, cuando se observan los niveles de expresión de la fusión traducional a *lacZ*, sí que se observa un clarísimo incremento de los niveles de expresión debido al activador DszR, pero cuando la transcripción se realiza desde el promotor *psal*, la traducción del transcrito resultante es prácticamente inexistente.

Estos datos, que concuerdan perfectamente con el hecho de que el subclón pMPO587 no permita la producción de índigo ni el crecimiento con DBT como única fuente de azufre en ausencia del regulador, podrían indicar que DszR fuese un activador traducional necesario para la traducción de estos genes. Esta interpretación podría estar reforzada por el hecho de

que el efecto de DszR sobre los niveles de fluorescencia (fusión transcripcional) sea muy reducido. Sin embargo, la secuencia de DszR indica claramente que tiene exclusivamente el dominio de activación de este tipo de activadores de la transcripción y un dominio de unión a ADN, que en conjunto ocupan la totalidad del gen, sin observarse otras regiones adicionales que pudieran ser responsables de una supuesta activación traduccional. Por otra parte, una baja inducción de la transcripción medida como fluorescencia ya se observó con el vector original en los ensayos para la validación de los niveles de expresión heteróloga del fósido, que indicaron que para obtener unos buenos niveles de transcripción inducida, tanto desde pT7 como desde *psal*, era necesario aumentar el número de copias del fósido (figura 8). Aunque la falta de inducción obvia podría deberse en parte a la fluorescencia intrínseca de la bacteria, que enmascararía la fluorescencia debida a unos bajos niveles de expresión, no tenemos una explicación racional para este fenómeno pero la realidad es que este análisis de expresión se realizó sin inducir el número de copias con arabinosa porque supusimos que debería reflejar mejor la situación real y probablemente por eso no detectemos bien la activación transcripcional por parte de DszR. En cualquier caso, ensayos de expresión recientes realizados por otro miembro del laboratorio, en este caso en alto número de copias, utilizando fusiones traduccionales y transcripcionales *lacZ* a los mismos genes *dsz*, confirman plenamente la activación de la transcripción por parte de *dszR* y confirman a su vez que la traducibilidad de los transcritos producidos desde *psal* es mínima (resultados no publicados y no mostrados en esta tesis).

Cuando los dos promotores, el propio activado por DszR y *psal*, están activos la traducibilidad de los transcritos es menor que cuando los transcritos provienen exclusivamente desde su propio promotor. Esto, cuya interpretación es que de la mezcla de transcritos solo los producidos desde su propio promotor son traducidos, indica que los transcritos provenientes de *psal* no se traducen aun en presencia de DszR activo. Este dato va en contra de la posibilidad de que se necesite DszR como activador traduccional, porque en ese caso DszR podría activar la traducción de estos genes independientemente del transcrito en el que esté.

Ciertamente estos genes parecen estar regulados por algún mecanismo postranscripcional que desconocemos, que permitiría la traducción de los transcritos propios pero evitaría la expresión que pudiera llegar a partir de promotores localizados aguas arriba de los genes de la ruta de desulfuración de DBT. Independientemente del mecanismo concreto, este sistema permite una regulación muy fidedigna de la expresión del operón, evitando una

expresión ilegítima debida a una transcripción desde el exterior, aislándolo por tanto del contexto genómico en el que se encuentre.

Un posible mecanismo podría ser que se generase una estructura secundaria inhibitoria de la traducción en los transcritos que provienen desde *psal*, aguas arriba del promotor propio, mientras que el transcrito proveniente desde su propio promotor carecería de ella. Si se asume que el promotor principal de este operón sería el localizado dentro del gen *dszE*, que transcribiría el operón *dszABC*, esta posibilidad es perfectamente factible desde un punto de vista mecanístico ya que los 3 genes están solapados y presumiblemente acoplados traduccionalmente. La formación de esta estructura inhibidora de la traducción podría consistir en un ARN transcrito en dirección contraria al operón *dszABC*, cuyo extremo 5' se localizase aguas arriba del extremo 5' del promotor *dszABC* e hibridase con *dszE*. La delección de la región aguas arriba de *dszE* que contenía un posible terminador codificado en sentido inverso no alteró los niveles de expresión, de forma que la estructura inhibidora de la traducción seguiría existiendo. En caso de existir un ARNi de este tipo, el posible terminador delecionado no debe ser el terminador real de este ARN ya que si no es previsible que se hubiese alterado la capacidad de asociación con el transcrito proveniente de *psal*.

Un ARNi de este tipo podría ser a su vez responsable de la bajísima traducción de *dszE*, que se encuentra aguas arriba de los otros genes. No sabemos si esos niveles tan bajos de traducción son suficientes para que la FMN reductasa pueda actuar como transportador de electrones para las oxigenasas de la ruta de desulfuración de DBT o si se aprovecha otra reductasa de *E. coli* para esa función. Las monooxigenasas de la ruta de desulfuración de DBT no parecen ser muy específicas en cuanto a la reductasa que requieren. De hecho, los genes de esta ruta de distintas especies se han subclonado en distintas estirpes, sin que se hayan acompañado del gen codificante de la reductasa propia del sistema, *dszD* o el equivalente en otros organismos distintos, habiendo sido en algunas ocasiones suficiente con alguna reductasa aportada por la bacteria hospedadora (Denome y col., 1994; Gallardo y col., 1997). Pero no todas las reductasas permiten a las oxigenasas trabajar al mismo nivel, como describen Galán y col (2000), que subclonaron el gen *hpaC*, codificante de una flavin:NADH reductasa de *E. coli* W, junto con los demás genes de la ruta de *Rhodococcus* sp. IGTS8 en *P. putida* y demostraron que se mejoraban los niveles de actividad frente a los conseguidos con la reductasa endógena de esa bacteria.

Como se ha comentado, la subclonación de los genes de la ruta de desulfuración de DBT permitió conservar las dos actividades observadas en el clon completo UPO21, que son

la actividad oxigenasa inicial, reflejada en la producción de índigo a partir de indol, y la desulfuración de DBT, reflejada en la capacidad de crecer con DBT como única fuente de azufre, aunque para ello fue necesario el aporte del activador de promotores dependientes de σ^N que permitía la expresión de los genes implicados desde su propio promotor.

Aunque ambas actividades se presentan tanto con el clon UPO21 como con el subclón con los genes de la ruta más el subclón con el activador, las condiciones necesarias para que se den estas actividades no son las mismas en los dos contextos (tabla 11). Por una parte, en el caso de UPO21 el crecimiento con DBT como única fuente de azufre se puede conseguir incrementando el número de copias del fósido (con arabinosa) o induciendo la expresión heteróloga del regulador desde *psal* (con salicilato) mientras que el color azul debido al índigo solo es detectable en las colonias cuando se suman esas dos condiciones (con arabinosa y salicilato). Por otra parte, en el caso de los subclones para crecer a expensas de DBT como fuente de azufre se requiere inducir la expresión del activador (con IPTG), mientras que se detectan colonias azules siempre que el fósido con los genes de la ruta esté en alto número de copias (con arabinosa).

Una posible explicación a estos distintos comportamientos sería que el clon UPO21 produjera una proteína que inhibiera la función del activador codificado por *dszR*. La estructura de DszR permite suponer que este puede ser el caso, ya que solo tiene el dominio activador y el de unión a ADN pero carece del dominio de regulación amino-terminal. Hay pocos ejemplos de activadores de promotores dependientes de σ^N que carezcan del dominio regulador, solo PspF, HrpR y HrpS (Bush y Dixon, 2012), y en estos casos su función está controlada por las proteínas PspA y HrpV, que se unen a los activadores respectivos e inhiben su función. Esta proteína inhibidora estaría codificada por un gen orientado en sentido contrario a *dszR*, y por tanto en sentido contrario a los promotores heterólogos del vector. De este modo, para detectar producción de índigo en medio LB, además de aumentar el número de copias del fósido con arabinosa hay que añadir también salicilato, de manera que se obtenga mucho más DszR que su inhibidor, con lo que se podría activar el promotor incluso en ausencia de la molécula inductora (supuestamente, DBT). Sin embargo, para el crecimiento con DBT como única fuente de azufre, el medio lleva el supuesto inductor con lo que DszR podría activar sin necesidad de que su producción fuera tan elevada (siendo suficiente con añadir solo arabinosa o solo salicilato). Por otro lado, en el subclón con los genes de la ruta no se encontraría el gen codificante de la proteína inhibidora, por lo que DszR no vería inhibida su función. Esta interpretación está avalada por el análisis de expresión de las fusiones

génicas que indicaron que la expresión activada por DszR podía darse de forma aparentemente constitutiva, en un medio con sulfato como fuente de azufre, sin necesidad de añadir DBT. Por eso, simplemente aumentando el número de copias con arabinosa ya se detecta color azul por producción de índigo. Por otra parte, para crecer a expensas de DBT es necesario aumentar la cantidad de DszR con IPTG, que sería el equivalente al aporte de salicilato en el con UPO21 para inducir la expresión del DszR.

Como candidatos a codificar la proteína inhibidora, existen varios genes transcritos en dirección contraria a *dszR* que no parecen formar operón con otros genes y que codifican proteínas de función desconocida. Cualquiera de ellos podría codificar la proteína inhibidora, pero llama la atención el gen codificante de una hidroxiacil glutatión hidrolasa putativa y no caracterizada. Este gen se encuentra aguas abajo de *dszR*, con el que solapa 32 pb, en sentido contrario.

Los sistemas de transcripción desarrollados en esta tesis han permitido la identificación de un gran número de genes que no eran capaces de expresarse por sí mismos, lo que muestra claramente las ventajas de forzar la transcripción del ADN metagenómico. Sin embargo, hay que tener en cuenta que, aunque la transcripción es la etapa de la expresión génica más frecuentemente regulada, también puede existir regulación a otros niveles, que podría impedir la expresión de los genes implicados aunque se empleen los sistemas de transcripción heteróloga descritos. Un ejemplo lo constituyen los genes de la ruta de desulfuración de dibenzotiofeno, cuya expresión ha sido posible solo aportando las condiciones necesarias para que se produzca transcripción desde el promotor propio y que está sujeta a algún tipo de regulación postranscripcional que impide la expresión cuando los transcritos provienen desde promotores ubicados aguas arriba del promotor propio del operón. En este caso, la transcripción heteróloga del activador de la expresión de este operón ha permitido la identificación de estos genes.

CONCLUSIONES

1. Los sistemas de expresión heteróloga desarrollados permiten la identificación de un mayor número de clones distintos con una actividad de interés que el sistema original sin expresión heteróloga.
2. El sistema basado en la antiterminación de la transcripción mediada por la proteína N permite identificar más clones que el sistema basado en la ARN polimerasa de T7.
3. Ambos sistemas permiten la expresión de genes ubicados al menos a unas 20 kb de los promotores del vector en los clones de la metagenoteca.
4. Una bomba de eflujo puede proporcionar resistencia a carbenicilina por sí sola, en ausencia de betalactamasas, si se expresa suficientemente.
5. La identificación de nuevas dioxigenasas extradiólicas en la pila de compostaje de la industria petroquímica ha permitido establecer seis nuevas subfamilias, incrementando sustancialmente la diversidad conocida de este tipo de enzimas.
6. La búsqueda de clones con actividad dioxigenasa extradiólica en metagenotecas con insertos de cierto tamaño es eficaz para identificar otras enzimas de rutas de degradación de contaminantes aromáticos de actividad no detectable, ya que con frecuencia se encuentran asociadas a ellas.
7. La búsqueda de clones con actividad oxigenasa hidroxilante mediante el rastreo de producción de índigo permite la identificación de oxigenasas de distinta naturaleza, muchas de las cuales no están relacionadas con rutas de degradación de contaminantes aromáticos.
8. El rastreo de oxigenasas hidroxilantes consistente en la detección de colonias de color azul por la producción de índigo presenta un sesgo a favor de las monooxigenasas.
9. La expresión de los genes de la ruta de desulfuración de dibenzotiofeno identificados en esta tesis está regulada por DszR, un activador de promotores dependientes de σ^N que carece de un dominio regulador y cuya actividad podría estar controlada por una proteína inhibidora codificada por el mismo clon.
10. Los genes implicados en la ruta de desulfuración de dibenzotiofeno tienen un mecanismo de regulación postranscripcional que impide la traducción de estos genes desde transcritos que comiencen aguas arriba del propio promotor del operón, por lo que no se han conseguido expresar heterológamente.

ANEXO

Secuencia de proteínas posibles responsables de actividades de interés identificadas en esta tesis. Se indica la proteína con la que presentan un mayor valor de identidad y, entre paréntesis, el clon en el que se encuentra cada proteína.

>Betalactamasa clase C - [pequeño péptido conservado](#) (ETN1)

MSKGWFLIPLLLLNLNAQAEDWPGEQWSSSPPTGPAIEALEQYAFVPRDDATRQGI RTDALLVIQDGR L VYERY
AGPTDAHTPHLTWSISKSLLATVFGVAYGEKRFALDDPAARYYPPLKRHPGMTLADLFHWASGLDWQEDYEYAPL
NSSVVAMLYTRGSHDMAAF TAGHDSFSPPGQVFRYSSGDSNLLAAALKTI VGPASYADY PWTALFDPLGIRHATW
ETDASGTFVGSYAYLTARDLARVGLLMQRDGRWGRQLI PKDWIAFSREPFTRFQAGQDEAVPGGHWWLNRAAD
RTIRPWPDPDPDTFAALGHWGQALYI IPSARLVI VRYGDDRDGYSRHNELLKRMAAFAGTVQP - [MMRKVSLTLM](#)
[AVLLVALFGWVWQERVALRAFPGVISAYTAKEYCSCRYVMQQPAEYCRGYVKQSVPTNDFLDTPEAKRITVSGLG](#)
[RSHSARWMGERQGCRLPEP](#)

>Betalactamasa clase B 1 (TN2)

MRKILAF AAGICV VAGT I SSPAAWAEQPAAWSEPT E PFKI VDDI YVVGTKGLASYLIVSDGEA ILLDGTLEENVG
HIEQNIQALGFSVEDVKI I INTHAHYDHAAGIAGLKRDTGAQVMAMEQERSALENGRHEGDNIYGATRFPAVKVD
SVLHDGDTISVGSVTMTATLTPGHTKGCTTWSMSAKDGAATRQVVFPCSLTVAGNKL VGNAGHPAIVEDYEKSFE
IVKSMKADIVLPAHPEFGDLFDRKAKRDAGDADAFVDPDLLRKLAE TSRAAFEKDLKAAQK

>Betalactamasa clase B 2 (N5 y N6)

MGSRALRIITVIGCAA VLLHGPAANGQANEAWTRPFPF K LIGNI YWVGSYDLSTYLVTTPQGHILINTGVGDT
AQR IKAGVEQLGFKMADV K I LTATHGHWDHVAGLAALKKMTGAALVASEPKELLES GGKSDFRFGNVESSWFEP
VKVDRTFADGGT LSLGGTTLTTHLHPGHTRGATSFTTEVRENGKTYRVVIANLGSINPGVTVTGMPAYPGIERDY
ARTFAAQRAMTIDVFLASHASQFGMHEKYKPGDPYTPDRFVDPKGF LAAVTRLEKLYLDQVAKERA AK

>Bomba de eflujo RND 1 (3 componentes, separados por guiones) (ETN1)

MQFKPAVTALVTAIALASLLSGCKKEEAAPPPPPQVGVVTLKTQPFTLTSELPGRTSAFRVAEVRPQVNGI I LK
RLFKEGADV KAGQQLYQIDPAVYEATLKSAEANLRSTKS IADRYQQLVNEQAVSRQEYDTAVANRLESEANLQTA
QINVRYTKVYAPISGRIGRSSVTEGALVSNQQT DALATIQQLDPIYVDVTQSSVELLQLRRELESGRLQKAGDNA
AMVKLTLEDGSQYPHEGKLEFSEVSVDQTTG SVTLRAVFPNPDHTLLPGMFVHAQLQAGVNSAAI LAPQQGVTRD
LKGMP TAMVVGADNKVELRQLKASRTVGNQWLI EDGLKAGDR LITEGLQYVRPGVEVKATEATNVGAKNPAPAQA
ADKAAGGKGE-MSKFFIDRPIFAWVIALVIMLVGALSILKLPINQYPSIAPPAIAIQVTPGASAQTVQD TVVQV
IEQQLNGIDNLRVVSSESNSDGSMTITATFEQGTSSDTAQVQVQNKLNLATPLLQEVQQQGI RVTKAVKNFLMV
IGVVS RDGSM TKDDLANYIVSNLQDPISR TAGVGD FQVFGAQYAMRIWLDPAKLN FNLT PVDVKNAIAAQNVQV
SSGQLGGLPALPGQQLNATIIGKTRLQTAEQFKAILLKVNP DGSQVRVGDVADVGLGGENYSVSAQFNGAPASGL
AVKLANGANALDTAKALRKTIDDLKPFPPQGV E VVFPYD TTPV VSES I KGVVETLVEAVVLVFLVMFLFLQNFRA
TIITMTVPVVLGTFGILAAAGFSINTLTMFGMVLAIGLLVDDAI VVVENVERVMSEEGLSPK EATK KSMGQIQ
GALVGIALVLSAVLLPMAFFSGSTGVIYKQFSITIVSAMALSVLVALIFTPALCATMLKPIPKGEHGT PKRGFFG
WFNRNFDRGVR SYERGVGNMLRHKAPYLLAYVIIVVGM IWLFTRIPTAFLPEEDQGVLF AQVQTPAGSSAERTQA
VVDKMREFLLRPSKDGEGDAVASVFTVTG FNFAGRGQSSGMAFIMLRPWEERNADNSVFKLAGRAQQHFFTFRD
AMVF AFAPPVLELGNATGFDVFLQDRAGIGHDKLMEARNQFLGMASQSKILSQRPNGLNDEPQYQLEIDDEKA
SALGITLSDINNTLSIALGSSYVNDFIDRGRVKKVYVQGP GARMSPEDLKKWYVRNSAGTMVPFTAF AKGEWVY
GSPKLARYNGVEAMEILGAPAPGYSTGEAMAEVEAIAKKLPAGVGISWTGLSYEERLSGSQAPALYALSLLMVFL
CLAALYESWSIPIAVMLVVPLGIIGALMATS LRGLSNDVYFQVGLLTTIGLAAKNAI L IVEFAKELHEQGRTLME
AAIEACRMRLRPIIMTSLAFV LGVVPLAISTGAGSGSQAHAIGTGVI GGMLTATILAI FWVPLFFVTVSSMGRKN
IDQDHTPETSKEAGQ-MSKSLLSLTIAAVVLSGCSLIPDYQRPEAPVAAQYPQGPAYEPANAPGQAAAEQGWKQF
FHDPALQQLIQVALENRDLRVAALNIDAYAAQYRIQRADLFP AVSATGSGSRQVRPARASQTGEAAISSYSAT
LGISAYELDLFGRVRSLS EQALQS YFATEEARRSTQISLVANVANAYLTWQADKELLKLTQETLGTYEQSLKLT S
RSAE VGVASALDLSQARTAVENARVQLARYTRQVAQDENSLTLLLTGLPANLRTQPLSDDLSEVPAGLPSDLL
QRRPDILQAERNLLAANANIGAARAAFFPSISLTANAGTSLPDSLGLFKGGSGTWT FAPQINLPIFNAGSLRASL
DYAKIQKDINVAQYEKSIQTAFQEVSDGLAARQTYNEQLQAQTD FVAANQDYR LAERRYRIGVDSNLTFLDAQR
QLFSAQQSLITDRLAQLTSEVNLYKALGGGWAETGKNEPVAEKAPETKLF

>Bomba de eflujo RND 2 (3 componentes, separados por guiones) (TN3 y TN4)
MWRSAAVLTLTASALTLAACGKQDAPQAGKQVTVVTLTTPVSLTTELPGRTSAFRVAEVRPQVNGIVQKRLF
TEGGEVKAGQQLYQIDPALYQASLDSQKAALARAQAQKTAALLAERYKPLVATRAVSQQTYDNAVASRDQAAAD
VLSAKAALDTARINLVYTKVLSPIDGIIGRSSVTEGALVTANQGTALASVQQIDPIYVDVTQSSVQLLRLQDALA
SGQLKKADGEQAALVTLTLEDGSQYKQTKLQFSEVTVDQGTGSITLRAVFPNPDRLLPGMFVRARLADGVAAD
GLLVPPQRGVTRNQRGLPTALVNAKNQVELRQLKTDRAIGDKWLVTDGLAAGDKVIVEGLQMRVPGVEVVATEAG
ANAQQPQAANAAPAKQ-MAKFFIDRPVFAWVIAIVLMMAGALSILQLPVSQYPNIAPPAIGIAVTPGASAQTVQ
DTVVQVIEQQMNGLDGLQYISSESNSDGSMSITLTFKQGTNPDTAQVQVQNKLALAQPLLQEVQOQGIKRVTKAT
KNFLIVAGFVSTDGTMTKDDLADYVASYVDPIISRTQGVGDFQLFGSQYAMRIWLDPAKLVNYGLTTVDVVTAIK
EQNVQVSSGQLGGLPSVRGQQLNATIIGPSRLEKPEDFGRILLKVNADGSQVRLGDVSRIELGGQTYAIDSFYNG
KPASGLAIKLAPGANALDTAQAVRDTINNLKPYFPPGMDVVYPYDTPFVSLSIHEVFKTLVEAIIILVFLVMYLF
LQNFRAIIPITLAVPVVLLGTFGVLAAGFYSINTLTMFGMVLAIGLLVDDAIVVVENVERVMAEEGLTPKQATR
SMGQITGALIGIAMVLAAVFI PMAFFGGSTGVIYRQFSITIVSSMVLSVLVAIVFTPALCATLLKPIPKGHHGSK
KGFVGFNRSFERSNGYANTVARGLNRTKRLMVVYLALVIAMGWLFTRIPTAFLPAEDQGI LFAQIQTPAGATA
EATKAVIDDATKYLLTEEKDAVTSVFAVNGFNFGGRQNASILFIKLRDWEERGDASLKAGAVAARANAHFGKTE
RRAQLFVPPPSVMELGNVTGDFQLMDRAGVGHEKLLAARNQLLGEAAKSPVLVGVPRNGIEDAPQYQLDIDRE
KARALGVAUSDINSTLATAWSSYVNDFIDRGRVKKVFAQGEASSRMLPEDLNKWYVRNAGDMVPFSAFSKATW
SFGPQKLNRYNGVPSYNIQQAAPGYSSGAAMEEMEKLAAKLPVGVGFWEFTGLSFEERLSGSQAPALYAIISLIV
FLCLAALYESWSIPTAVMLVPLGIGVLLATLARGLSNDVYFQVGLLTTVGLAAKNAIILIVEFAKEHFESGASL
TESAIHAARQLRPIIMTSLAFILGVTPLAISSGAGSGSQNAIGTGVIGGMLTGTFLAIFVPAFFVIMLRVFKV
KRMSENQDKHDT SANGSQEVS AEGQP-MKLQMRLLSVSLAAALAGCSLAPTYERPDAPVSANYPTGPAYKADGS
PLAQGMSTADVWRDFFSDPLLQQLIELSLANNRDLRVAALNVEAARAQYRIQRADLFPSVGVAGQESAQRTPAD
LSPSGRATTSRSYQVGAAMS AWELDLFGIRISLSDQALESYLALDETRTATQLALIAEVANAYLTLRADQELLGL
TRDTLKSQEDSFKLTQQSYDQGLSTALDLSQAEVSLRTAQRNLSQYTRQAAQDRNALTLVVGQPMSP EIVAALDQ
AVKLLDDGMLPTTL PAGLPD LARRPDIRAAEHQLKGANANIGAARAAFFPTISLTGQAGTASASLGGLFEGGSG
AWSFVPQITVPIFAGGSLQANL DLAKVQKNIQVAQYEKSIQTGFREVADALAGRGTLDQIQAQRLVDANQRAY
DVSDQFRQGIIDYLSVLDSQRSLYTSQQALVDTRLARLSNLVTLTKVVLGGGWERTVTAQTTPAAGS

>UPOEDO1 (UPO64)

MAQIVGGFIMPHDPLITGNPEIADQGOVAAVEAFAFARVVERVKALKADTAIIVIGDDHFTMFGPHCLPQYLIGIGD
VEGPEENWLRIERYTVPNNALAEHIMNFGFDRGFDWAVAKSLVVDHGTMIPIHKAVTPAGMRTIPIYTAAGVTP
LLRTKRAKELGQMIGAAVAAYPGNDRVVVYGTGGISHWVGTAEMGRVNEFDRLVLGMVERGDVDGLVNLSDIYV
LEHGGNGALEIRNWI VAMAAMPAGVSGELFCYEPMPPIWITGIGIAELKLA

>UPOEDO2 (UPO64)

MASIVGGFCVPHDPAITAFPHVANQRQAKNIMAGFDRLARRVAELGADTAIIVIGDDHFALFGPHCLPSFLIGIGD
VEGPEENWMHIDRYVPVNNVPLAEHIMHYGFDHGFDWSVAKSLVLDHGTMI PVHLAVTPNPGVRTIPIYTAAGVTP
PLLRMKRALALGRMIGAAVAAPGNDRVVVMGCGGISHRVGTADMGRVNEPFDRMILDMVVRGDVEAMTELDDAY
VLREGGNGAFEIRNWI VAMAAMPFRKGEVICYEPVPEWITGLGLAELKVAA

>UPOEDO3 (UPO49 y UPO50)

LPEKRRDAESSIAGIPKKAETERPRARSFYDKHADHKEQNMGMVRIGHVSLQVMDIDASVRHYVDVVGMDVTHRD
SDGTTYLKCWDEWDKYSLILTPSNAARLNHIA YKVENDADLDVLGKRIADAGVEVTKSPAGALPFCGRSISFLLP
SQQRMYLFAQKDFLKGKDVGTLPDPWPDSKHGIGAHWLDHCLILGEMNPEKGINRVQDSSDFLQNVLD FHM SERA
MVGPNHSM AASFLFRSSKPHDIALVAAPTNGFHHISFFLDGWSIDLHAGDILSKKRVKIELLPSRHGITRGETI
YFFDPSGIRNETFAGMGYQAYPDMPIWITWKEEDLARGIFYTAREFVESFSAAYTPL

>UPOEDO4 (UPO57)

MDLEEAI VHYRDNLGLDLVSQEDEFAYL RAYDEFDRHSVVLREAQEPGLDVLGFKVIVSGAALTDFAKRLNEMGVK
TDDIDAGDHPGVGRRVRFVPTGHVIDLYAEMELSDNGPMTKNPDIWREEPRGMRI SRFDHCAVLGVDIAASAEI
FEKALDFVVTERRAIDPKGNPVAIFLTCSNKAHDIAFLASPVDAKLHHSFLVESWNDVGHACDLIAQRNISLDIG
PTRHGITRGQTVYFFDPSGNRNECFAGGYQHYPDNVREWTVDQIGKAIFYEKEVHDFRFTSVMT

>UPOEDO5 (UPO52, UPO55, UPO58 y UPO72)

MREAIMGITSRGVHSIDHYALYVPSLQQARHFFEHFGLDVQGSEAWLDISAADGHRWARILPGPGKSLAWISFNC
 FEDDFATLREQVERSGALLEEGAGGGFWFRDPDGNLIQVRVGEKTMPDSRPTSPQVVRDLRGVGSRDVAVCQVRPR
 RLSHVLFSPDVPRAVAFYEQALGLKLSDRSQDIIAFTHARHGCDHHLVAFKSGAKGFHHAADVPTIDDVGNNG
 AEQMAAAGYVEGWGTGRHCLGSNYFHYVRDPWGSFSEYSTEMDFIEAGSEWPAGDFPAESSLYLWGPPVPDYFIL
 NTEPGG

>UPOEDO6 (UPO67)

MLANDQQDIFMPRQPALFVSHGSPMLAFEDVPARSFLLGLAPQLAEPRAVLVI SAHHEEQPTVTTSPAPGTIHD
 FYGFPQALYRLTYPAPGAPALAEERVAALLSERGHVDVARDPERGLDHGAWVPLLLINPYADI PVVQLSISPHRDA
 WHLALGEALRPLRDEGVLIILASGSI THNLRDLTRTEAGATPAAWASEFTTEWVHDRVGTHTRTAELVRYRELAPHAV
 HAHTDEHFLPFVVALGAAHGDEPAERLHHSFTY GALGMDVYAFGKLG

>UPOEDO7 (UPO37, UPO60 y UPO89)

MSVSGSLGYVGFVSDLAAWDAFFTDFLGLMRGQNGEAA RYRMDERAFRFHLSK GKADDISFIGLEVETHEGMRA
 LREKLDRAKIKYSAGSDNLKAERGVRDLIVLNDPAGLQVELYV GATSLTSAPFSSPAGVSGFLTGAQGLGHVLA
 AERIDEVRDFYTRALGFRLSDTIGMALGPDFTLLEFYHCNPRHHTLALAPMPAPKRLH HFMVEMATLDDVGFAM
 DRVERLGVKQTTTLGKHSNDQMVSYAATPSGFDVEVGTAGIAVHDEVVRVTHHEVTSMWGHRCH

>UPOEDO8 (UPO78)

MREAIMGITSRGVHSIDHYALYVPSLQQARHFFEHFGLDVQGSEAWLDISAADGHRWARILPGPGKSLAWISFNC
 FEDDFATLREQVERSGALLEEGAGGGFWFRDPDGNLIQVRVGEKTMPDSRPTSPQVVRDLRGVGSRDVAVCQVRPR
 RLSHVLFSPDVPRAVAFYEQALGLKLSDRSQDIIAFTHARHGCDHHLVAFKSSAKGFHHAADVPTIDDVGNNG
 AEQMAAAGYVEGWGTGRHCLGSNYFHYVRDPWGSFSEYSTEMDFIKAGSEWPAGDFPAESSLYLWGPPVPDYFIL
 NTEPGG

>UPOEDO9 (UPO90)

MSLHTLSYVGFNATDLDLDDWRQYATGVLGMQVVDQTADTLRLRMDEKHHRF A IHKSATPGFAYS GWDVATRAELDR
 LTTEL TALGYPAVEGTEAERTFRKVVDLRFVTDPGGNRVELFYGQLSGCPFTPARQFGGFKTGQLGLGHIVFMTR
 NMDAMMAFYGVGLSVSDYIFIKPVQSRACFMHLNARHHSFALLGGPQDVFH HIMI EANSLDDVGMANELALRQG
 RTVTMSLGKHTNDEVVSFYTMTPSGF ELEYGCGRTIDNEEDWQVVEYDDISFWGHLGPLRAPV

>UPOEDO10 (UPO42)

MVSVTELGYIGIGVSNMDTWREYAGKVMGLEVVDDGEADR FYLRMDYQHHRITVHNSGSDLDYTGWRVAGPEEF
 EQMLRQLSAAGVKFQRGSEAECKERHVLDLVKFLDPGDNPT E IYHGPRVDYHKPFHPGRGMHGRFMTAQEGLGHL
 IMRQTDVMKAYRFYTDVLMGRGSI EYHLAVPQIGITAKPVFLHCNDRDHS LAF LGAPAAKRINHLMLAVDNIDDV
 GMT HDIVRDLKIPVSV ALGKHANDQMFTFYSATPSDWLLEYGGVGR TASHQSEY YEDMWGHDNETQGYGLDLRL
 RYDWSGLLKK

>UPOEDO11 (UPO36)

MAMTGILRPGHVQLRVMELGPALH H YKEMVGLIETARDAEGRVYLKAWDEHDLFSIVLREADRPGMDFYGFKVL D
 TKALEHYAKRLTDYGVAIERI PAGELKGTGERIRFETPSGHLIELYAEKQHVGNGLPDVNP DVWPDGLKGMHPT R
 FDHCLLNGDSIPQTAKLFQEVLD FDLTERILGPDGSMVGAFLACSDKAHDVAFIHEDSKGR LHHSFYLESWDEI
 RHAADIMSKNYIPVEYGPGRHGLTRGLTIYFDPGSGNRNETFHGGYIRYPDHPVITWTHEETPKGIFYYGHKMVE
 SFVTANT

>UPOEDO12 (UPO35)

MRIGHVSLQVMDMDASVRHYVDVGM DVTHRSDGTTYLKCWDEWDKYS LIILTPSNAARLNHIAYKVENDADLDV
 LGKRIADAGVEVTKSAAGALPFCGRSISFLLPSEQRMYLFAQKDFL GKDVGT LNPDWPWDSKHGIGAHWLDHCLI
 LGEMNPEKGINRVQDSSDFLQNVLD FHM SERAMVGNH S IMAASFLFRSSKPHDIALVASPTNGFHHSIFFLDGW
 SDILHAGDILSKKRVKIE LLPSRHGITRGETIYFFDPSGIRNETFAGMGYQAYPDMPVITWKEEDLARGIFYTAR
 EFVESFSAAATPV

>UPOEDO13 (UPO46)

MGVMRIGHANIRVMDMAAAVRHYEKVLGLKSTHSDAQGNVYLKWCDEWDKYSLVLTTPSETAGIAHVAYKVEHDDAD
LDALKKRIEDYGIKTEFLPAEAMPFIFGRMLKFNLPSGHEMRLYAQKDCVGTDVGSNNPDPWPDDIRGAGAHWLDH
CLLMCELNPEKGINKVAENTRFMMQALDFYMSEQIVVGPDNISIQLATWLFRSSTPHDLAFVGGPKQGLHHIAFFL
DSWNDILKAADVMAKNKVRIDVPPTRHGVTRGETIYFYDPSGNNRNETFAGLGYMAQPDRPVTTWDESHLGRGIFY
HTGEVNPAPFTEVYT

>UPOEDO14 (UPO47)

MRQGGLLTAARALRRFHPAQEDSPMAMTGILRPGHVQLRVMELGPALHHYKEMIGLIETGRDAEGRVYLKAWDEH
DLFSLVLRREADRPGMDFYGFKVRDKQALEHYAERLADYGVAIERVPAGELKGTGERVRFETPSGHLIELYAEKQH
VGNGLPDVNPDPVWDGLKGMHPTRFDHCLLNGDSIPETAKLFQEVLDLFDLSEILAPDGSVMGAFSLSCSDKAHDV
AFIHEDSKGRLLHVSFYLESWDEIRHAADIMSKNYIPVEYGPGRHGLTRGLTIYFFDPSGNNRNETFHGGYIRYPD
NPVITWTHEETPKGIFYYGHKMVESFVTANT

>UPOEDO15 (UPO43)

MAMTGILRPGHVQLRVMELGPALHHYKEMVGLIETARDAEGRVYLKAWDEQDLFSIVLRESDRPGMDFYGFKVRD
KAALEHYAKRLTEYGAVERIAAGELKGTGERVRFETPSGHLIELYAEKQHFNGNGLPDVNPDPVWDGLKGMHPSR
FDHVLNNGDSIPETAKLFQEALDFDLSEIRIVGPDGSVMGAFACSTKAHDVAFIHEESKGRLLHVSFHLLESWDEI
RHAADIMSKNYIPVEYGPGRHGLTRGLTIYFFDPSGNNRNETFHGGYIRYPDMPVITWTHEEVKGIIFYYGHKMVE
SFVTANT

>UPOEDO16 (UPO38)

MGVMRIGHVSLRVMMAAAVKHYENVLGMKRTMEDRHGNVYLKWCDEWDKYSLQLKPADRAGMEHVAYKVEHEAD
LDALQKRIAAAGIATQMLPEGALPSTGRMLQFTLPSGHEMRLYAMKECVGTAVGSMNPDWPDDLKGAGAKWLDH
VALVCELNPPAGVNRVADNVRFFKDCDFYLSEQIVVGGPGAIQAAAFLLFRSTKPHDIAFLPGPQPGIHHASFFL
DSWHDVLKAADVMAKNRTRKIDLPTRHGI TRGETIYFFDPSGNNRNETFAGLGYLAQPDRPVNTWTTEENMWRGIIY
HTGEAMPSFTEVYT

>UPOEDO17 (UPO45)

MAEVVAGIAASHVPMAMQPQFELAEQGQRNRVAVAGMSEARKIVEQARPDVIVIFSTDHDFDRCFYNNLPPFLVGV
GEEAEGPIAEWMHLPKVKLKIASDLAQFIVREGLDNGVDFALSQELPLDHAEEVPLSFITPQWNIPIVPIVNVAF
APPMPSLKRCWQVGAFIRETIERWPLKRKVGVTGGISHWVGPPETGRINAEFDRWFMQCLVEGKAAEVIIEKYK
DAEELEKTAGNGGQEI RDWLAVAGAMP SRLKARVLAYEPIQWATGTGVMGWA

>UPOEDO18 (UPO45)

MASIAQLGYLGLGVRDNEWERFASNVLGLQPNGRDPDGSFLRMDEYHHRFIVHPNGNDDLAYIGWEVTTEQAM
AAMAEQLRQAGVEVRPGTAAEADARRVAGLIKADPSGLPSEIFYGPLVTFDKPFQSPRPISGFKTEGQGLGHFV
VAVDDFDRSLAFYRDVLGMRISDFVNLS PAPNMQLKVAFFHCNPRHHTLAFWAMPQSPKRLHHFMLQLNSLDDVG
STYYLCQNQQTPITMNLGKHTNDHMVSFYLRTPSGFAVEYGWGAREVDDATWQVQVHTAGSIWGHQGS

>UPOEDO19 (UPO41)

LEEVMALRGLLRVGEVAIRVMDMKLRRHYGDRMGLHEVMEADGRVYKAWDEHDHHSIVLRQADSPGLDYFA
FKVYDDKTLTDLEPKIRAFGLEVKHIAAGVYPKSGRRLEFMLPSGHRMQLYAHKEQVGNTLGYLNPVGVVDEHVI
RGFNINRLDHVLLGGTNIDDNVRLFTEVFEDLSEKLIDHESKASLAVFLSGSTKPHDVAFVLQPEPGRFHVVSF
LLESADHVIKAADRIGKYRIPVDVGPNRHGVTRGATIYFFDPSGNNRNEVFAEGYVHYPDTQTLVWDTTELGAATF
HQDFTPRESFLTFLT

>UPOEDO20 (UPO33)

LPEKRRDAESSIAGIPKKAETERPRARSFYDRHANHKEQNMGMVIRIGHVSLQVMDIDASVRHYVDVVGMDVTHRD
SDGTTYLKCWDEWDKYSLILTPSNAARLNHIAKYVENDADLDVLGKRIADAGVEVTKSPAGALPFCGRSISFLLP
SEQRMYLFAQKDFLKGKDVGTLPNDPWPDSKHGIGAHWLDHCLILGEMNPEKGINRVQDSSDFLQNVLD FHM SERA
MVGPNH SIMAASFLFRSSKPHDIALVAAPTNGFHHISFFLDGWSDILHAGDILSKKRVKIELLPSRHGITRGETI
YFFDPSGIRNETFAGMGYQAYPDMPVITWKEEDLARGIFYTAREFVESFSAAYTPL

>UPOEDO21 (UPO61)

MQVSSFGYLGSLVSDLSRWKSFADFLGLGVASEDADTIKFRNDDRAWRIAHRGDADDIAYAGFEVANANALAA
IRDKLQEGGVAIEEATADLRKQRGVKDLVVCRDPLDGLRVEIFYGPLEKFDTPPFRSGCVVSGFVTGDQGLGHIVLM
TPDLPRMRGFYSIGILGFKLSDTIDIPISGELTISLEFYHCNRRHHTVALGPTMGPGGKLLHHFMLQVASLDDVGF
ALDRVTRYGVPLRNTLGRHTNDHMVSFYAATPSGFVEVEFGTGARTVDDDEDWSVVSHEKISTWGHVPVA

>UPOEDO22 (UPO51) (el gen no estaba completo, acaba con **secuencia de vector**)

MLANDQQDIFMPRQPALFVSHGSPMLAFEDVPARSFLLGLAPQLAEPRAVLVI SAHHEEQPTVTTSPAPGTIHD
FYGFPPQALYRLTYPAPGAPALAEVAALLSERGHVDVARDPERGLDHGAWVPLLLINPYADI PVVQLSISPHRDA
WHLALGEALRPLRDEGVLIILASGSI THNLRDLTRTEAGAAPAAWASEFTTEWVHDRVGTHTRTAELVRYRELAPHAV
HAHPTDEHFLPFFVALGAAHGDEPAERLHHSFTY GALGMDGGIDNYNGPKVASVSGSSRVDLQACKLRGRRYR

>UPOEDO23 (UPO68 y UPO87)

MIKVTDIAYVRFRAPDLDEMEKFLIEFGMARARSGTHLHMRGTDAHPFVHETEKGEVGLGFFAESIADLEK
LAGEHGLPVEDNPAPGGGKIVRFTDPDGFVVDVAGQSAPQKLPVTRKLIHNDGDARVRAGQPLRVAPGPSQVQR
LGHAVINVTDFRLSEAWYKSRFGLVTSDEIELAPGAPLGAFMRCDRGRTHVDHHTLFLVGTGQPGFNHAAFEVAH
FDDL MAGHNHLEKAGRRHEWGVGRHILGSQIFDYWRDPWGHTVEHWTDGDLFNNETPPEIRGLGDLMGNQWGPEA
PPSMGS

>Estireno monooxigenasa 1 (UPO08 y UPO11)

MKSIAIIGAGPVGLHFGFALLKAGIGRVTLFSDRTPPEAYGAGKLANTAFFAYVQEKEKSLGLNFWDVTPRGMG
LHLEVHLPTGATVATIDAAFRDGVTAQGVDIRIRIPRWMQEFVHRGGDLRIENIDAERLDAIASEYDATFVAVGK
GAMNGLFERDVERSEFDKPARHMGVLLLEGESITGRGAWPEVGFPHLRHTIFGVGDVFSIPMYGKPGVPCRGMA
FEAIPGSPFDQFADAQSGEALLEIGLNLMRKFLPGDVPEFSQASLCEPGSWITGALTPTIRKPVGHTLSGYPVMG
LGDVVMLTDLPLAGQGANNIRMAHRVAEELIDLGPRIADADWMKQTFEDYWDYGYPIEFTNRLKDLGPSLMV
ILQAAGSDQRVADAFATAFTDPKTMVPWLRSVEDAKRFVENARQDSSSEAGIRQVV

>Estireno monooxigenasa 2 (UPO12, UPO13, UPO14 y UPO25)

MKSIAIIGAGPVGLHFGFALLKAGSGKVTLFSDRTPPEAYGEGKLSNTAFAFFAYVQOKERELGLNFWDVTPRGMG
IHLEAHLPTGATVASITAAFRHGVTAQGVDIRIRIPRWMQEFVHRGGDLRIENIDEARLDAIAREHDVTFVAVGK
GAMSKLFERDAERSEFDRPARHMGVLLLEGESITGQDAWPEVGFPHLRHTIFGVGDVFSIPMFGKPGVPCRGMA
FEAIPGSPFDQFAGAQSGEELLQVGLDLMRKFLPGDVPEFSRATLCEPGSWIAGALTPTIRKAVGYTRSGHPVMG
LGDVVMLTDLPLAGQGANNIRMVHRVAEELMELGTQPAGADWMEQAFEDYWNDYGYPIEFTNRLKDPGPPLMV
ILQAAGSDQRVADAFATAFTDPKTMVPWLRSVEDAKRFVENARQDSSSEAGIRQVV

>Alcanosulfonato monooxigenasa 1 (UPO09 y UPO19)

MVKIFTVDSPFENNVPSPALARQARKWEAAEVIDYVLCGYSSLWPHNLVTSPLLFAESSRVKLI VAHRPGVMHP
AAAARCFATMDVMSGARLAINFVTGSSDKDVHREGDYADKATRYERAGEYVEVMKRCWSEPKSFDYQGEFYRAEA
VRQLLRPLKNHVPVFMGGSDAAIEFGARHADTYMLWGEPLAGTRERVERVLAAA AKHGRKPEFSLSLRFLGDT
DEAAWAAAHAEDTILAAQGSNRFARSSAGDTSVGRERQLSLTNADVHDDCFWVSLVKKLLGGFANSALVGT PDR
VMATLKQYHALGIGAFLEFTGGADGLWEPSLKEFLVRVKQELQ

>Alcanosulfonato monooxigenasa 2 (UPO23)

MVKIFTVDSPFENNVPSPALARQARKWEAAEVIDYVLCGYSSLWPHNLVTSPLLFAESSRVKLI VAHRPGVMHP
AAAARCFATMDVMSGARLAINFVTGSSDKDVHREGDYADKATRYERAGEYVEVMKRCWSEPKSFDYQGEFYRAEA
VRQLLRPLKNHVPVFMGGSDAAIEFGARHADTYMLWGEPLAGTRERIERVLA AA EKHGRKPEFSLSLRFLGDT
DEAAWAAAHAEDTILAAQGSNRFARSSAGDTSVGRERQLSLTNADVHDDCFWVSLVKKLLGGFANSALVGT PDR
VMATLKQYHALGIGAFLEFTGGADGLWEPSLKDFLVRVKQELQ

>Dimetilanilina monooxigenasa (UPO26)

MTRVAVIGAGPSGLAQLRAFQSARKKGAKIPEVVCFEKQEDWGGLWNYTWRTGLDEHGEPVHGSMYRYLWSNGPK
ECLEFADYTFEEHF GKPIASYPPRAVLWDYIKGRVQKAGVRKWRVFCPTVPRMVTYSKTKKFTVTAHDRKKDRMY
AEEFDYVVVASGHFSTPNVPIFEGISTFNGRVMHSHDFRDALEFKGKDVLI VGRSYSAEDIGSQCWKYGAKSVTT
TYRSKPMGFKWPENFKERPLLQKLVNKTAFKDGTTKEVDAILCTGYLHHFPFLTDDLRKLTANRLWPLGLYKG
VVWEENPRLFYIGMQDQFYTFNMFDAQAWYARDVIMGRIKLP SKEKMIKHSKAWRKREEKLEDAEQMIWYQGDYV
KELLQDIDYPKFDIEATNKTFMEWEHKAEDIMGFRNNAYRSLMTGTMSPAHHTPWVDAMDDSMESYLGDKGRGS
RREAAE

>2-4-diclorofenol 6-monooxigenasa (UPO55) (no se conoce el final de la
secuencia)

MKLQNTSEIIEVPVLI VGGGPSGLTASLLLSRYGIRNLLINKYRWTANS PRAHITNQTMEVFRDAGVESQVVA
ASPHELMANNVWATSFSGQEFGRLLTWGNAIERKSDYELASPSAMCNI PQHLMEPILANEALRAGS (...)

>Fenol 2-monooxigenasa 1 (componentes P0-P5, separados por guiones) (UPO16)

MMTKDDAALI QHRYVRVERREDGFVEFRFSIGYPEYYAELVLSEKDFEVFCKENAVMFLPEGESARTAAEPNGL
TWSLRDVAKRVRP-MTVEIKAVEIKPLRNTFTGVAELIGGDKPASRYQEATLGAQSRVNFHYRPTWEPEFELFD
PRRTRVRMRDWDLRDPRQLYSSWTIQRAKEQETQEANFDFVESRLLHLNLPQEVKDRAIAVLLPLRHVAWGAN
MNSEL CARGYSEMI IAPAIMHAMDNLGVAQYLTRLGLLLGD AELLDAKAQWMDDPAWQPLRHLVEDTFVLRDF
MEIFTAQNLVMDGLLYPLVYDRFVDDRLTSMGGSAIAMLTAFMPNWFSSQCGKWDATLKTVVAESEENKAI VSAW
TQQWRQRLDALAPIATKAFGADGDGMLAGVAEEFDARARKLGI AV-MSKVFIIMQANE EARPIIEA ILRDNPTA
AVQES PGLVRIEAEQGLTVRRSTVEEVTGQPFDIQQINVLNLSISGHIDEEDDYLTVAWKN-MDMKVDKKMGLK
ERYAYMTRGLGWEPYQPADAVFPYDKYEGIKIHDWDKWDVDFRLTMDAYWKYQGEKDRKLYAVIDAFAQNNHGL
NITDARYLSAIKLFWQGVSPLENYVVQGFARTGRHFRGEAPRVACQM QAVDEL RHYQTQAHSISNYNKYYNGFHN
QMYWFD RVWYLSVPKSYGDDALTAGPF EYMAI SFSFEYVLTNLLFVPMFMSGAAYNGDMSTVTFGFS AQSDSRH
MTLGI EAIKFLLEQDPANVPIVQRWVDKWFWRGARLLSLVGMMDYMLPKRVMSWKEAWEIYAEENGGALFRDLA
RYGIRPPKGWADACEAKEHITHQVMLLFYSYGHVTAFTSLPTPDEL DWLSAKYPNTFDKYRPRFEHLREMKER
GERFYQEALPTLCATCQIPLGFTEPGDHTRIGHRET VHKGEKYVFCSEHCKEIFEYEPEKYVQAWLPVNQIYQGN
CFPPDVPDGPADFNPLTEVLKFFKFNLGRDNGDFDGS EDQRNFATWRGQATKN-MAVTALGVYDFPSKDR AENFH
GNILVNFCWDRHVLFAAPFCFAVSPSMTLREAMTTLVAKAIRAH PDAGEIDWD AVRWQKDDQPWQPDV DKT LAEN
HVGHKDYIVFNT PGLMGLGGAGL-MSHSLTIEPLGETIEVEEGQTI L DAALRAGIYI PNACGHGLCATCKVQVLD
GEVEHGQASPFALMDFERDEGKTLACCATLLDDVTIEAEIEEDADAERI PVRDFAGVVNRLVDLTPTIKGVFITL
DEPIHFQAGQYIQFVIPGKHIERAFSLANPPSTGREIELNVRI VAGGEGTSFIHQQLREGDRVAIKGPYGRFYVR
QSARKPLLFMAGGSGLSSPRSMILDLLAKGWDLPI TLVYGQRTRAE LYHDEF LALAEKHPNFQYIAALSDEPEN
SGWQGFGRGF AHQAATVVFPEGFAGHHAYLCGPPVMIDACIT TLMQGR LFEDDIYFERFISAADAQQVRSPLFRKI

>Fenol 2-monooxigenasa 2 (componentes P0-P5, separados por guiones) (UPO41)

MQAHNARSLLDGVDTKIRFVRRNRVNDHGYVEFSFGIGSPDLMV ELVLSERAYAEFCCKTNQCIELTPAQSLAVDL
EQAKWRYGKAGITE-MSVDIRTNSVKPIRNTFAHLARRFGDKPASRYQEIAYDVQSEVNFHYKPLWDPGEIYDK
RRTAITMADWYALKDPRQFYYSYTTTRAKQQDALDRQLDFAEKRDLLRQLPEEVRAQIVFALLPLRHYEWGANM
NNCFISAYGYGTAITQAAIMATTDRLGMAQHISRIGLLVDGNSGESLTRGRALWIDDPAWQGLRRRIERL FVTRD
WFEVMVAQN LALDGLVY PFFQ QFDKHIAASHGTGLSTLTDYLLRWHEETAKVWDATVKTAA AENEQNRAAIVGW
FLAWREAGIDALKPIAERVLGDAEAALGAVAAQLDARA AKLGLKV-MSNIQPVFIALQANNDTLP IIEAICSDN
PQAQVNEY PAMVKINCPGRMVI RRSSVEERMGRDFDLRELQINLISLAGEVDESEDEFVLFWKA-MAADGDLKPI
KKLSLKENYALLTRGLGW EPTYEKKEEIFRTIDFEGIKIHDWEKWEDPFR L TMDAYWKYQGEKERKLYA I IDAFQ
QNNGQFNVT DARYVQALKFLN GIVGLE YASHRGFNILGREFPVGVPRIAAQMQAIDELRHATTQIHAI SHYNR
FEGMADFPHQFDRVWYLSIPKSF FEDAVSSGPF EIVAISFAFEYVLTNLAFMPWMSGAAFNGDMSTV SFGFSAQ
SDEARHMTLGI EVIKFLLEQDPDNVPLVQKWIDKWFWRGYRL LALLATMMDYMLPKRVMSWSEAWTIYYEENGA
LFADLARYGIRPPKYADVTI KEKQRV SHEMWSIFYSVRNAVGF PVWLPSAEEMDFSEKYPDTFDRYRPRFEHW
RELEEKGTPFTNMTLPQLCQVCQWPC IYTEPDDPTEFVHRETEYQGETFHFCS DGCKDIFCNEPEKYVHAWLPVH
QIYQGNCFPEGADPTAPDNPLVAALQYMGIKPGVDGGLMSQHVDASRWEQWTSR PANLHEGAITQAVRKT T-MP
VKAIKPGYHGEFRDRVENFRGQQLLNICWEKHTMFGWPMCI PVPPEMPFGALLEQVLPGIYGAHPDFARIK WNEV
QWSSAKGPFPTDPSKSMVEHGFKHK TQIRFRTPGLDGHNGAG-VSYQLTIEPLGQTI EVADDQ TMLDAALRAGI W

LPHACCHGLCATCKITVKEGDI EHG DASP FALMDFERDEGKALACCATLQSDAVIEADIDEDPDAIVI PVRDFHG
 TVTRIDALTPTIKGVHVQLDEAIEFQAGQYINLAIPLPGQAVTRAFSLASQPSAGRDIELNIRAVPGGVGTGFVH
 KDLAVGDRLLTSGPYGRFFVRQSAGVPMIFMAGGSSLSPRSMILDLLSESGSDLPITLVYQVRVREELYHDEFLL
 ALAEKHPNFRYPALSSSEASDSGWTGFRGFVHEAALAHFNGDFRGHKAYLCGPPMMIDACISALMKGRLFERDIY
 HEKFISAADAQQNRSPLFRNI

Ruta de desulfuración de dibenzotiofeno (UPO21):

>FMN reductasa, DszE (UPO21)

MAKLVLINGSPSAQSRTGRLLDHFATELERDGTVDHLALRTPAVALLGADFAAPDLIAARERVEAAAAIIVAT
 PVYKGAYTGLLKAFIDVLPQRALVDKSVLP IASGGTLAHYGVIDHALKPVLAALGATFTLPGLFVLDTDLKSAEA
 GGGFSAEILLRLESARLALGRLHGLTIRPATEADLPLRQYGS

>Dibenzotiofeno-5,5-dióxido monooxigenasa, DszA (UPO21)

MSTPRQLHLAGFFSAGNVTHAHGAWRHARASNGFLT KAYYQHIARTLERGKFDLLFLPDGLAVEDGYGDNLETGV
 RYGGQGAVALEPGAVIATMAAVTEHLGLGATVSTTYPPYHVARLFATLDQLSAGRVSWNVVTSLNHAEARNFGF
 EQHLEHDARYDRADEFLLTAVRKLWNSWDADALKLDKAAGLFADPAKVRYVDHRGEWLNVRGPLQVPRSPQGEPI
 LQAGLSDRGKRFAGRWAEAVFSVAPNLAFMQANYRDIKAQVKAAGRDPAQTKIFTAVMPVLGETEAIARERLEYL
 NALVHPEVGLSTLSSHTGINLAAYPLDAMPSEVLRDLKSRNVPTPLHMFAMAMQAHDLTGEMGRRYGTNVAFVP
 QWVGTGEQIADQLEAHFKAEAADGFIISPAFLPGAYDEFVDQVVPILQARGLFRTEYEGHTLRDHLGLRQPQLG
 DDAWQTQPNNQQAIA

>2-hidroxibifenilo-2-sulfinato desulfinasa, DszB (UPO21)

MANPAEPTSHRSLSQTQAGPDTLTYSNCPVPNALLTALSSGILRQRDIALDILSGQQGAVHFYDQPAYTRFGGEI
 PPLVSEGLRAPGRTRLLGITPVLRGGYVVRADSPIRTPADLAGRRVGVSAATRI LRRELGNYLELDPWRQTLV
 ALGTWEARALLHTLAAGGLTVDDVDLVR IESPGVDVPPERLQAASSLKGADLFPDVGQQAAI LEQDRAEALFAW
 LPWAAELESRIGARPVLDLGLDARNAYASVWTVSSALVESRPGVVQRLVDAAVDAGLWARQHGEVDTALHAANLG
 VSEAAIRKGFADFNQHLVPRLDVDALALLERTQQFLLDHKLLLEPVPLDAWAAPFLNNSLQRQSARKIA

>Dibenzotiofeno monooxigenasa, DszC (UPO21)

MSQSAEHLTRVRASNVEIARQLADKWRATAVERDRAGGSATAERQDLRDSGLLSLLVPTEYGGWADWPPTTLEV
 IREIAKVDGSLGHLFGYHTGSAPMIELFGSAEQKERLYTQLAQNDWWTGNASSENNSHVLDWKSATPTGDDGYL
 LNGTKHFCSGAKGSNLLFVFGVIEEESPLKGAIAAAIPTDREGVHINDDWNAI GMRQTDSGSTEFHNAKVYPNE
 VLGAPNAFILA FMGAKRGS LWTMPVQLIFSTVYLGIAHGALEAAREYTRTQTRPWT PAGVKATDDPYVVRTYGE
 FAIALQGADAAAAREAGRLLQQVWEKGDALTPQERGELMVKVS GVKALSTKVALDITSRI FEVIGSRGTHPRFGFD
 RFWRNVRTHSLHDLPLAYKIVEVGNHALNGRYVPVPGFTS

>UPOCYTP1 (UPO10 y UPO24)

MNATATAAAAARCPYASTTKPSPDDVPLDQIDVADADLIRSGAVFGYFDRLRREDPIHYCTTNELYGPYWSVTRFK
 DIVEVETKPHIFSSSEAGGITIEDNLLGLGTPMFIAMDPPKHDEQRMSVSPAVSPETLAQLEAGIRQRTSQVLDL
 PLNQTFDWVDRVSIELTTQMLAILFGFPWEHRRMLTRWSEVTTASPEGSNVVESPEARMRELEFCRGYLTALWH
 QRADAEPQFNLSMMAHAPATRDLP SRPQEF LGNMLLLI IGVGNDTTRNTMSGALLALNRNPEEYRKL CARPELVA
 SFVPEAIRWQTPLSHMRRTALCDTMLGGKQIRKGDKLV MWYLSGNRDPQAIERPDEFIVDRKNARHLSFGFGI
 RCLGNRLAELQLRVLWEEVLKRWSKPMQIEVVGEVQTDTSIMFRGLRHLPVRINA

>UPOCYTP2 (UPO28 y UPO30)

MNTTATAAAVARRPDVSATMPLPDNVPLDQIDVADNELFRSGAVFDYFDRLRREDPIHYCTTNKLYGPYWSVTRFK
 DIVEIETKPHIFSSSEAGGITIEDGILGLGTPMFI GMDPPKHDEQRMAASPAVSPENLRLTRLEAGVRQRTSQVLDGL
 PLNQTFDWVDRVSIELTTQMLAILFGFPWEHRRLLTRWSEVGSASPEGSNVVESQEARMRELDRCRGYLTALWH
 QRADAEPQPDLLSMAHAPATRDLP SRPEEFLGNMLLLI IGVGNDTTRNTMSGVSVLALNRNPEEYRKL CARPELVT
 SFVPEAIRWQTPLTHMRRTALCDTTLGGKQIRKGDKLV MWYLSGNRDPQAIERPDEFIVDRKNPRQHLSFGFGI
 RCLGNRLAELQLRVLWEEVLKRWNKPMQLEVVGEPELVDTSIIIFRGRHLPVRINA

>UPOCYTP3 (UPO28 y UPO30)

MNATATAVPSAHKSAAYAI PLEDIDVSNPELYRSRSTMWGYFERLRREDPVHYCAKSMFGPYWSVTKYKDIIEV
ETQPGIFSSDSRRGGVTIMDTNTAVQLPMFIQMDPPQHDEHRKAVSPIVAPENLARLESIRKRTGEALDSLPGV
EEFNWVEKVSINLTTQMLATLFDLPWEDRTKLTYSWDATLALPGEGLVDSEEHLQIMGECAYMTRLWNERVNA
EPRADLISMAHDERTRNMTPEEFLGNVLLLI VGGNDTTRNSMTGGLLALNENPEQYAKLCANPALVTSMVPEII
RWQTPLSHMRRTATQDTVIGGKNIRKGDKVVMMWYVSGNRDEEAIENPNAFIIDRARPRQHLSFGFGIHRVGNRL
AEMQLRVLWEEILKRWSKPGQLKVIGEPVVRVLSPFRLRGYASMRVRIDA

>UPOCYTP4 (UPO28)

MNATDTATAVPSAHKSAAYAI PLEDIDVSNPELYRSNTMWGYFERLRREDPVHYCAKSMFGPYWSVTKYKDIMQV
ETQPRIFSSDSRRGGITIMDSNAAASLPMFIAMDPPKHDEQREVVSPIVAPENLAKLEQLIRQRTAASLDELPIG
QTFNWVELVSINLTTKMLATLFDLPWEERSKLTYSWDVTTALPGAGLIDSEEERMAKLKECAAYMTRLWNERVNA
EPRADLISMAHDERTRNMTPEEFLGNVLLLI VGGNDTTRNSMTGGLLALNENPEQYAKLCANPALVTSMVPEII
RWQTPLSHMRRTATQDTVIGGKNIRKGDKVVMMWYVSGNRDEEAIENPNAFIIDRARPRQHLSFGFGIHRVGNRL
AEMQLRVLWEEILKRWSKPGQLKVIGEPVVRVLSPFVRGYESMRVRIDA

>UPOCYTP5 (UPO61)

MSETLVAARDKSIADAYAI PLEKIDVSNPELFRANAIWPYFERLRREDPVHYKDYSGFWSVTKYKDIMHVD
NHGIYSSEATLGGVALRDQEEGFFLPMFIMMDPPKHDAQRKVVSPIVAPGNLAKLESTIRERAGTILDSLPI
FNWVDRVSIELTTQMLATLFDLPWEERRKLTRWSDVATAGTAFGDEEAEKARRNELRDCAAYFTELWNQ
PGNDLITMLAQGEATKNMGPMYELGNILLI VGGNDTTRNSITGGLLALNRNPDQYKLRDNPALVESMVPEI
RWQTPLSHMRRTALQDTELGGKQIKKGDKVVMMWYVSGNRDEEAIENANSFIIDRKRPRQHLSFGFGIHRVGNRL
EMQLRVVWEEILKRWPKPIEVIGEPERVFSNLINGYSSMPVRI PG

>UPOCYTP6 (UPO87)

MTDRTIDNAIVNPKTYAHVDFHKLFTQLRKEEPVRWTEPDGFRPFWTVSKHADIMEVERQNDKFLNDPRLTLQS
IELEEEVKKFTGGQSKLIRSLVMDNPDHRNYRGLTQAWFMPPNLKAI SARVEALAEKYIDRLEAKGGECDFVSD
VAVWYPLRVIMTVLGVPEEDEPIMLKLQELFGSTDPDMKRDPATETVNTVTEFFNYFTAMTEDRRKNARDDVAS
VIANGTIDGQPLRHLEAISYYIIVATAGHDTTSSSTAAGLLALMQNPAEFAKLANPEGLLGGGAIDEMIRWTT
KHFFRTAAVDYELRGQIKAGDNLLMCYWSANRDEDAFDDPFAFRIERSPNKHLAFGYGAHLCLGQHLAKMEIRA
LYKELLARLDHIELAGDPAWVEASFVSGLKRLPVRYSMKRKAA

>UPOCYTP7 (UPO37 y UPO89)

MTENPQNVQRDYRRKNAETAAIWKIDVNRDPYEIPIEELDPAHDDLFAANKALPYFERLRKEDPVHLSEGGPYGR
YWSITKYDDIMHVDTHHQLFSSDIRNGGIRLGGQRLEGEPPDPLTYLPMFIMEDQPKHDEQRKAVQPMFT
PQSLANLEPLIRERAGLILDNLPRGETFNWVREVAVELTGRTLATLFDVPQEDRHKLHWSDTV
ERLGDPEYFETPEEGFKELWSCWEYFDAVWKERLSRKEPGSDLISMLAHSEATRNMPPNEYLG
NMILLI VGGNDTTRNSITGGVLALNQNPDQYRKL IENPGI VPMVSEIIRWQSPVAHMCRTA
LEDTEIRGKKIRKWDKIAMWYVSGNRDEEKIARANEFLIDRE GARHLSFGFGIHRVGNRL
GEMQLRILWEEILKRFKKVEVVGEPKYLRNSNFI RGITELPVVVH GK

BIBLIOGRAFÍA

- Aakvik, T., Degnes, K. F., Dahlsrud, R., Schmidt, F., Dam, R., Yu, L. y col. (2009) A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. *FEMS Microbiol Lett* 296: 149-158.
- Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A. y Handelsman, J. (2009) Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J* 3: 243-251.
- Allen, H. K., Donato, J., Wang, H. H., Cloud-Hansen, K. A., Davies, J. y Handelsman, J. (2010) Call of the wild: antibiotic resistance genes in natural environments. *Nat Rev Microbiol* 8: 251-259.
- Altenbuchner, J., Viell, P. y Pelletier, I. (1992) Positive selection vectors based on palindromic DNA sequences. *Methods Enzymol* 216: 457-466.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. y Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. y Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402.
- Ambler, R. P. (1980) The structure of beta-lactamases. *Philos Trans R Soc Lond B Biol Sci* 289: 321-331.
- Anderson, A. S., Clark, D. J., Gibbons, P. H. y Sigmund, J. M. (2002) The detection of diverse aminoglycoside phosphotransferases within natural populations of actinomycetes. *J Ind Microbiol Biotechnol* 29: 60-69.
- Andújar Pulido, M. E. (2001) Caracterización molecular y bioquímica del gen *thnC* e identificación de nuevos genes relacionados con la ruta de degradación de tetralina en *Sphingomonas* sp. estirpe TFA. Tesis Doctoral. Universidad de Sevilla.
- Andujar, E., Hernaez, M. J., Kaschabek, S. R., Reineke, W. y Santero, E. (2000) Identification of an extradiol dioxygenase involved in tetralin biodegradation: gene sequence analysis and purification and characterization of the gene product. *J Bacteriol* 182: 789-795.
- Angelov, A., Mientus, M., Liebl, S. y Liebl, W. (2009) A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. *Syst Appl Microbiol* 32: 177-185.
- Asturias, J. A., Eltis, L. D., Prucha, M. y Timmis, K. N. (1994) Analysis of three 2,3-dihydroxybiphenyl 1,2-dioxygenases found in *Rhodococcus globerulus* P6. Identification of a new family of extradiol dioxygenases. *J Biol Chem* 269: 7807-7815.
- Barrios, H., Valderrama, B. y Morett, E. (1999) Compilation and analysis of sigma(54)-dependent promoter sequences. *Nucleic Acids Res* 27: 4305-4313.
- Bartossek, R., Nicol, G. W., Lanzen, A., Klenk, H. P. y Schleper, C. (2010) Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context. *Environ Microbiol* 12: 1075-1088.

- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S. y col. (2004) The Pfam protein families database. *Nucleic Acids Res* 32: D138-41.
- Batie, C. J., Ballou, D. P. y Correll, C. C. (1991) Phthalate dioxygenase reductase and related flavin-iron-sulfur containing electron transferases. En *Chemistry and biochemistry of flavoenzymes*, Vol III. (Ed: Muller, F.). Boca Raton, FL, CRC Press, pp. 544-554.
- Bayer, S., Birkemeyer, C. y Ballschmiter, M. (2011) A nitrilase from a metagenomic library acts regioselectively on aliphatic dinitriles. *Appl Microbiol Biotechnol* 89: 91-98.
- Bhat, M. K. (2000) Cellulases and related enzymes in biotechnology. *Biotechnol Adv* 18: 355-383.
- Biver, S., Steels, S., Portetelle, D. y Vandenberg, M. (2013) *Bacillus subtilis* as a tool for screening soil metagenomic libraries for antimicrobial activities. *J Microbiol Biotechnol* 23: 850-855.
- Boyd, D. R., Sharma, N. D. y Allen, C. C. (2001) Aromatic dioxygenases: molecular biocatalysis and applications. *Curr Opin Biotechnol* 12: 564-573.
- Brady, S. F., Chao, C. J., Handelsman, J. y Clardy, J. (2001) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org Lett* 3: 1981-1984.
- Brennerova, M. V., Josefiova, J., Brenner, V., Pieper, D. H. y Junca, H. (2009) Metagenomics reveals diversity and abundance of meta-cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-sparging bioremediation. *Environ Microbiol* 11: 2216-2227.
- Bush, M. y Dixon, R. (2012) The role of bacterial enhancer binding proteins as specialized activators of sigma54-dependent transcription. *Microbiol Mol Biol Rev* 76: 497-529.
- Camacho, E. M. y Casadesus, J. (2005) Regulation of traJ transcription in the *Salmonella* virulence plasmid by strand-specific DNA adenine hemimethylation. *Mol Microbiol* 57: 1700-1718.
- Canada, K. A., Iwashita, S., Shim, H. y Wood, T. K. (2002) Directed evolution of toluene ortho-monooxygenase for enhanced 1-naphthol synthesis and chlorinated ethene degradation. *J Bacteriol* 184: 344-349.
- Cebolla, A., Sousa, C. y de Lorenzo, V. (2001) Rational design of a bacterial transcriptional cascade for amplifying gene expression capacity. *Nucleic Acids Res* 29: 759-766.
- Chai, W. y Stewart, V. (1999) RNA sequence requirements for NasR-mediated, nitrate-responsive transcription antitermination of the *Klebsiella oxytoca* M5al *nasF* operon leader. *J Mol Biol* 292: 203-216.
- Cherepanov, P. P. y Wackernagel, W. (1995) Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of F1p-catalyzed excision of the antibiotic-resistance determinant. *Gene* 158: 9-14.

- Choi, H. S., Kim, J. K., Cho, E. H., Kim, Y. C., Kim, J. I. y Kim, S. W. (2003) A novel flavin-containing monooxygenase from *Methylophaga* sp strain SK1 and its indigo synthesis in *Escherichia coli*. *Biochem Biophys Res Commun* 306: 930-936.
- Chung, C. T., Niemela, S. L. y Miller, R. H. (1989) One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A* 86: 2172-2175.
- Clarke, L. y Carbon, J. (1976) A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome. *Cell* 9: 91-99.
- Coco, W. M., Levinson, W. E., Crist, M. J., Hektor, H. J., Darzins, A., Pienkos, P. T. y col. (2001) DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotechnol* 19: 354-359.
- Collins, J. y Hohn, B. (1978) Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. *Proc Natl Acad Sci U S A* 75: 4242-4246.
- Connon, S. A. y Giovannoni, S. J. (2002) High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl Environ Microbiol* 68: 3878-3885.
- Cormack, B. P., Valdivia, R. H. y Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 173: 33-38.
- Courtois, S., Frostegard, A., Goransson, P., Depret, G., Jeannin, P. y Simonet, P. (2001) Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environ Microbiol* 3: 431-439.
- Courtois, S., Cappellano, C. M., Ball, M., Francou, F. X., Normand, P., Helynck, G. y col. (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol* 69: 49-55.
- Craig, J. W., Chang, F. Y. y Brady, S. F. (2009) Natural products from environmental DNA hosted in *Ralstonia metallidurans*. *ACS Chem Biol* 4: 23-28.
- Craig, J. W., Chang, F. Y., Kim, J. H., Obiajulu, S. C. y Brady, S. F. (2010) Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Appl Environ Microbiol* 76: 1633-1641.
- Daniel, R. (2005) The metagenomics of soil. *Nat Rev Microbiol* 3: 470-478.
- Daniel, R. (2004) The soil metagenome--a rich resource for the discovery of novel natural products. *Curr Opin Biotechnol* 15: 199-204.
- Datsenko, K. A. y Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 97: 6640-6645.

- Degtyarenko, K. N. y Archakov, A. I. (1993) Molecular evolution of P450 superfamily and P450-containing monooxygenase systems. *FEBS Lett* 332: 1-8.
- Demaneche, S., David, M. M., Navarro, E., Simonet, P. y Vogel, T. M. (2009a) Evaluation of functional gene enrichment in a soil metagenomic clone library. *J Microbiol Methods* 76: 105-107.
- Demaneche, S., Philippot, L., David, M. M., Navarro, E., Vogel, T. M. y Simonet, P. (2009b) Characterization of denitrification gene clusters of soil bacteria via a metagenomic approach. *Appl Environ Microbiol* 75: 534-537.
- Demaneche, S., Sanguin, H., Pote, J., Navarro, E., Bernillon, D., Mavingui, P. y col. (2008) Antibiotic-resistant soil bacteria in transgenic plant fields. *Proc Natl Acad Sci U S A* 105: 3957-3962.
- Denome, S. A., Oldfield, C., Nash, L. J. y Young, K. D. (1994) Characterization of the desulfurization genes from *Rhodococcus* sp. strain IGTS8. *J Bacteriol* 176: 6707-6716.
- Desai, C. y Madamwar, D. (2007) Extraction of inhibitor-free metagenomic DNA from polluted sediments, compatible with molecular diversity analysis using adsorption and ion-exchange treatments. *Bioresource Technol* 98: 761-768.
- Di Gregorio, S., Zocca, C., Sidler, S., Toffanin, A., Lizzari, D. y Vallini, G. (2004) Identification of two new sets of genes for dibenzothiophene transformation in *Burkholderia* sp. DBT1. *Biodegradation* 15: 111-123.
- Diaz, E., Jimenez, J. I. y Nogales, J. (2013) Aerobic degradation of aromatic compounds. *Curr Opin Biotechnol* 24: 431-442.
- Donato, J. J., Moe, L. A., Converse, B. J., Smart, K. D., Berklein, F. C., McManus, P. S. y Handelsman, J. (2010) Metagenomic analysis of apple orchard soil reveals antibiotic resistance genes encoding predicted bifunctional proteins. *Appl Environ Microbiol* 76: 4396-4401.
- Ekkers, D. M., Cretoiu, M. S., Kielak, A. M. y Elsas, J. D. (2012) The great screen anomaly-- a new frontier in product discovery through functional metagenomics. *Appl Microbiol Biotechnol* 93: 1005-1020.
- Ellis, H. R. (2011) Mechanism for sulfur acquisition by the alkanesulfonate monooxygenase system. *Bioorg Chem* 39: 178-184.
- Eltis, L. D. y Bolin, J. T. (1996) Evolutionary relationships among extradiol dioxygenases. *J Bacteriol* 178: 5930-5937.
- Ensley, B. D., Ratzkin, B. J., Osslund, T. D., Simon, M. J., Wackett, L. P. y Gibson, D. T. (1983) Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. *Science* 222: 167-169.
- Entcheva, P., Liebl, W., Johann, A., Hartsch, T. y Streit, W. R. (2001) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl Environ Microbiol* 67: 89-99.

- Eschenfeldt, W. H., Stols, L., Rosenbaum, H., Khambatta, Z. S., Quate-Randall, E., Wu, S. y col. (2001) DNA from uncultured organisms as a source of 2,5-diketo-D-gluconic acid reductases. *Appl Environ Microbiol* 67: 4206-4214.
- Farinas, E. T., Schwaneberg, U., Glieder, A. y Arnold, F. H. (2001) Directed evolution of a cytochrome P450 monooxygenase for alkane oxidation. *Adv Synth Catal* 343: 601-606.
- Ferrer, M., Beloqui, A., Timmis, K. N. y Golyshin, P. N. (2009a) Metagenomics for mining new genetic resources of microbial communities. *J Mol Microbiol Biotechnol* 16: 109-123.
- Ferrer, M., Beloqui, A., Vieites, J. M., Guazzaroni, M. E., Berger, I. y Aharoni, A. (2009b) Interplay of metagenomics and in vitro compartmentalization. *Microb Biotechnol* 2: 31-39.
- Ferrer, M., Beloqui, A., Zumarraga, M., Alcalde, M. y Golyshin, P. N. (2008) Microbes and enzymes: recent trends and new directions to expand protein space. En: *Protein Engineering Handbook* (Eds: Stefan Lutz, Uwe T. Bornscheuer; Wiley-VCH Verlag GMBH & Co), 1: 233-264.
- Ferrer, M., Golyshina, O., Beloqui, A. y Golyshin, P. N. (2007) Mining enzymes from extreme environments. *Curr Opin Microbiol* 10: 207-214.
- Ferrer, M., Golyshina, O. V., Chernikova, T. N., Khachane, A. N., Reyes-Duarte, D., Santos, V. A. y col. (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol* 7: 1996-2010.
- Fetzner, S. (2012) Ring-cleaving dioxygenases with a cupin fold. *Appl Environ Microbiol* 78: 2505-2514.
- Figurski, D. H. y Helinski, D. R. (1979) Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci U S A* 76: 1648-1652.
- Fishman, A., Tao, Y., Rui, L. y Wood, T. K. (2005) Controlling the regiospecific oxidation of aromatics via active site engineering of toluene para-monooxygenase of *Ralstonia pickettii* PKO1. *J Biol Chem* 280: 506-514.
- Forsberg, K. J., Reyes, A., Wang, B., Selleck, E. M., Sommer, M. O. y Dantas, G. (2012) The shared antibiotic resistome of soil bacteria and human pathogens. *Science* 337: 1107-1111.
- Fortin, N., Beaumier, D., Lee, K. y Greer, C. W. (2004) Soil washing improves the recovery of total community DNA from polluted and high organic content sediments. *J Microbiol Methods* 56: 181-191.
- Fozo, E. M., Hemm, M. R. y Storz, G. (2008) Small toxic proteins and the antisense RNAs that repress them. *Microbiol Mol Biol Rev* 72: 579-89, Table of Contents.
- Frostegard, A., Courtois, S., Rami se, V., Clerc, S., Bernillon, D., Le Gall, F. y col. (1999) Quantification of bias related to the extraction of DNA directly from soils. *Appl Environ Microbiol* 65: 5409-5420.

- Furuya, T., Takahashi, S., Ishii, Y., Kino, K. y Kirimura, K. (2004) Cloning of a gene encoding flavin reductase coupling with dibenzothiophene monooxygenase through coexpression screening using indigo production as selective indication. *Biochem Biophys Res Commun* 313: 570-575.
- Gabor, E., Liebeton, K., Niehaus, F., Eck, J. y Lorenz, P. (2007) Updating the metagenomics toolbox. *Biotechnol J* 2: 201-206.
- Gabor, E. M., de Vries, E. J. y Janssen, D. B. (2004a) Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environ Microbiol* 6: 948-958.
- Gabor, E. M., Alkema, W. B. y Janssen, D. B. (2004b) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* 6: 879-886.
- Gabor, E. M., de Vries, E. J. y Janssen, D. B. (2003) Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. *FEMS Microbiol Ecol* 44: 153-163.
- Galan, B., Diaz, E. y Garcia, J. L. (2000) Enhancing desulphurization by engineering a flavin reductase-encoding gene cassette in recombinant biocatalysts. *Environ Microbiol* 2: 687-694.
- Gallardo, M. E., Ferrandez, A., De Lorenzo, V., Garcia, J. L. y Diaz, E. (1997) Designing recombinant *Pseudomonas* strains to enhance biodesulfurization. *J Bacteriol* 179: 7156-7160.
- Garcia-Gonzalez, V., Govantes, F., Porrua, O. y Santero, E. (2005) Regulation of the *Pseudomonas* sp. strain ADP cyanuric acid degradation operon. *J Bacteriol* 187: 155-167.
- Gibson, D. T., Resnick, S. M., Lee, K., Brand, J. M., Torok, D. S., Wackett, L. P. y col. (1995) Desaturation, dioxygenation, and monooxygenation reactions catalyzed by naphthalene dioxygenase from *Pseudomonas* sp. strain 9816-4. *J Bacteriol* 177: 2615-2621.
- Gong, X., Gruniniger, R. J., Forster, R. J., Teather, R. M. y McAllister, T. A. (2013) Biochemical analysis of a highly specific, pH stable xylanase gene identified from a bovine rumen-derived metagenomic library. *Appl Microbiol Biotechnol* 97: 2423-2431.
- Gonzalez-Pastor, J. E. y Mirete, S. (2010) Novel metal resistance genes from microorganisms: a functional metagenomic approach. *Methods Mol Biol* 668: 273-285.
- Gotoh, O. (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* 267: 83-90.
- Govantes, F. y Santero, E. (1996) Transcription termination within the regulatory *nifLA* operon of *Klebsiella pneumoniae*. *Mol Gen Genet* 250: 447-454.

- Gray, K. A., Pogrebinsky, O. S., Mrachko, G. T., Xi, L., Monticello, D. J. y Squires, C. H. (1996) Molecular mechanisms of biocatalytic desulfurization of fossil fuels. *Nat Biotechnol* 14: 1705-1709.
- Greenblatt, J., Mah, T. F., Legault, P., Mogridge, J., Li, J. y Kay, L. E. (1998) Structure and mechanism in transcriptional antitermination by the bacteriophage lambda N protein. *Cold Spring Harb Symp Quant Biol* 63: 327-336.
- Guan, C., Ju, J., Borlee, B. R., Williamson, L. L., Shen, B., Raffa, K. F. y Handelsman, J. (2007) Signal mimics derived from a metagenomic analysis of the gypsy moth gut microbiota. *Appl Environ Microbiol* 73: 3669-3676.
- Hain, T., Otten, S., von Both, U., Chatterjee, S. S., Technow, U., Billion, A. y col. (2008) Novel bacterial artificial chromosome vector pUvBBAC for use in studies of the functional genomics of *Listeria* spp. *Appl Environ Microbiol* 74: 1892-1901.
- Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95-98.
- Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166: 557-580.
- Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68: 669-685.
- Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. y Goodman, R. M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5: R245-9.
- Hannemann, F., Bichet, A., Ewen, K. M. y Bernhardt, R. (2007) Cytochrome P450 systems--biological variations of electron transport chains. *Biochim Biophys Acta* 1770: 330-344.
- Harayama, S., Kok, M. y Neidle, E. L. (1992) Functional and evolutionary relationships among diverse oxygenases. *Annu Rev Microbiol* 46: 565-601.
- Hardeman, F. y Sjoling, S. (2007) Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiol Ecol* 59: 524-534.
- Harpel, M. R. y Lipscomb, J. D. (1990) Gentisate 1,2-dioxygenase from *Pseudomonas*. Purification, characterization, and comparison of the enzymes from *Pseudomonas testosteroni* and *Pseudomonas acidovorans*. *J Biol Chem* 265: 6301-6311.
- Hart, S., Koch, K. R. y Woods, D. R. (1992) Identification of indigo-related pigments produced by *Escherichia coli* containing a cloned *Rhodococcus* gene. *J Gen Microbiol* 138: 211-216.
- Hatta, T., Mukerjee-Dhar, G., Damborsky, J., Kiyohara, H. y Kimbara, K. (2003) Characterization of a novel thermostable Mn(II)-dependent 2,3-dihydroxybiphenyl 1,2-dioxygenase from a polychlorinated biphenyl- and naphthalene-degrading *Bacillus* sp. JF8. *J Biol Chem* 278: 21483-21492.

- Healy, F. G., Ray, R. M., Aldrich, H. C., Wilkie, A. C., Ingram, L. O. y Shanmugam, K. T. (1995) Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl Microbiol Biotechnol* 43: 667-674.
- Heiss, G., Stolz, A., Kuhm, A. E., Muller, C., Klein, J., Altenbuchner, J. y Knackmuss, H.-J. (1995). Characterization of a 2,3-dihydroxybiphenyl dioxygenase from the naphthalenesulfonate-degrading bacterium strain BN6. *J Bacteriol* 177: 5865-5871.
- Henriques, I., Moura, A., Alves, A., Saavedra, M. J. y Correia, A. (2006) Analysing diversity among beta-lactamase encoding genes in aquatic environments. *FEMS Microbiol Ecol* 56: 418-429.
- Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. y Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77: 51-59.
- Holloway, B. W. (1979) Plasmids that mobilize bacterial chromosome. *Plasmid* 2: 1-19.
- Hu, P., Janga, S. C., Babu, M., Diaz-Mejia, J. J., Butland, G., Yang, W. y col. (2009) Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol* 7: e96.
- Ilmberger, N., Meske, D., Juergensen, J., Schulte, M., Barthen, P., Rabausch, U. y col. (2012) Metagenomic cellulases highly tolerant towards the presence of ionic liquids--linking thermostability and halotolerance. *Appl Microbiol Biotechnol* 95: 135-146.
- Inoue, H., Nojima, H. y Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96: 23-28.
- Jacobsen, C. S. y Rasmussen, O. F. (1992) Development and application of a new method to extract bacterial DNA from soil based on separation of bacteria from soil with cation-exchange resin. *Appl Environ Microbiol* 58: 2458-2462.
- Jadeja, N. B., More, R. P., Purohit, H. J. y Kapley, A. (2014) Metagenomic analysis of oxygenases from activated sludge. *Bioresour Technol* 165: 250-256.
- Jeng, S. T., Gardner, J. F. y Gumport, R. I. (1990) Transcription termination by bacteriophage T7 RNA polymerase at rho-independent terminators. *J Biol Chem* 265: 3823-3830.
- Jimenez, J. I., Canales, A., Jimenez-Barbero, J., Ginalska, K., Rychlewski, L., Garcia, J. L. y Diaz, E. (2008) Deciphering the genetic determinants for aerobic nicotinic acid degradation: the *nic* cluster from *Pseudomonas putida* KT2440. *Proc Natl Acad Sci U S A* 105: 11329-11334.
- Joo, H., Arisawa, A., Lin, Z. y Arnold, F. H. (1999) A high-throughput digital imaging screen for the discovery and directed evolution of oxygenases. *Chem Biol* 6: 699-706.
- Kaeberlein, T., Lewis, K. y Epstein, S. S. (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* 296: 1127-1129.

- Kakirde, K. S., Wild, J., Godiska, R., Mead, D. A., Wiggins, A. G., Goodman, R. M. y col. (2011) Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. *Gene* 475: 57-62.
- Kasai, Y., Shindo, K., Harayama, S. y Misawa, N. (2003) Molecular characterization and substrate preference of a polycyclic aromatic hydrocarbon dioxygenase from *Cycloclasticus* sp. strain A5. *Appl Environ Microbiol* 69: 6688-6697.
- Kelly, S. L. y Kelly, D. E. (2013) Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? *Philos Trans R Soc Lond B Biol Sci* 368: 20120476.
- Kennedy, J., O'Leary, N. D., Kiran, G. S., Morrissey, J. P., O'Gara, F., Selvin, J. y Dobson, A. D. (2011) Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems. *J Appl Microbiol* 111: 787-799.
- Kim, J. Y., Lee, K., Kim, Y., Kim, C. K. y Lee, K. (2003) Production of dyestuffs from indole derivatives by naphthalene dioxygenase and toluene dioxygenase. *Lett Appl Microbiol* 36: 343-348.
- Kim, S. J., Lee, C. M., Kim, M. Y., Yeo, Y. S., Yoon, S. H., Kang, H. C. y Koo, B. S. (2007) Screening and characterization of an enzyme with beta-glucosidase activity from environmental DNA. *J Microbiol Biotechnol* 17: 905-912.
- Kim, S. J., Lee, C. M., Han, B. R., Kim, M. Y., Yeo, Y. S., Yoon, S. H. y col. (2008) Characterization of a gene encoding cellulase from uncultured soil bacteria. *FEMS Microbiol Lett* 282: 44-51.
- Kim, U. J., Shizuya, H., de Jong, P. J., Birren, B. y Simon, M. I. (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res* 20: 1083-1085.
- Kimura, N., Sakai, K. y Nakamura, K. (2010) Isolation and characterization of a 4-nitrotoluene-oxidizing enzyme from activated sludge by a metagenomic approach. *Microbes Environ* 25: 133-139.
- Knietsch, A., Waschowitz, T., Bowien, S., Henne, A. y Daniel, R. (2003a) Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on *Escherichia coli*. *J Mol Microbiol Biotechnol* 5: 46-56.
- Knietsch, A., Waschowitz, T., Bowien, S., Henne, A. y Daniel, R. (2003b) Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl Environ Microbiol* 69: 1408-1416.
- Knietsch, A., Bowien, S., Whited, G., Gottschalk, G. y Daniel, R. (2003c) Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Appl Environ Microbiol* 69: 3048-3060.

- Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., 2nd y Peterson, K. M. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166: 175-176.
- Kulakov, L. A., Delcroix, V. A., Larkin, M. J., Ksenzenko, V. N. y Kulakova, A. N. (1998) Cloning of new *Rhodococcus* extradiol dioxygenase genes and study of their distribution in different *Rhodococcus* strains. *Microbiology* 144 (Pt 4): 955-963.
- Kweon, O., Kim, S. J., Baek, S., Chae, J. C., Adjei, M. D., Baek, D. H. y col. (2008) A new classification system for bacterial Rieske non-heme iron aromatic ring-hydroxylating oxygenases. *BMC Biochem* 9: 11-2091-9-11.
- Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685.
- Lakay, F. M., Botha, A. y Prior, B. A. (2007) Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. *J Appl Microbiol* 102: 265-273.
- Lammle, K., Zipper, H., Breuer, M., Hauer, B., Buta, C., Brunner, H. y Rupp, S. (2007) Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J Biotechnol* 127: 575-592.
- Langer, M., Gabor, E. M., Liebeton, K., Meurer, G., Niehaus, F., Schulze, R. y col. (2006) Metagenomics: an inexhaustible access to nature's diversity. *Biotechnol J* 1: 815-821.
- Lee, S. W., Won, K., Lim, H. K., Kim, J. C., Choi, G. J. y Cho, K. Y. (2004) Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Appl Microbiol Biotechnol* 65: 720-726.
- Li, X. Z., Ma, D., Livermore, D. M. y Nikaido, H. (1994) Role of efflux pump(s) in intrinsic resistance of *Pseudomonas aeruginosa*: active efflux as a contributing factor to beta-lactam resistance. *Antimicrob Agents Chemother* 38: 1742-1752.
- Lim, H. K., Chung, E. J., Kim, J. C., Choi, G. J., Jang, K. S., Chung, Y. R. y col. (2005) Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Appl Environ Microbiol* 71: 7768-7777.
- Lindahl, V. y Bakken, L. R. (1995) Evaluation of methods for extraction of bacteria from soil. *FEMS Microbiol Ecol* 16: 135-142.
- Livermore, D. M. (1996) Are all beta-lactams created equal? *Scand J Infect Dis Suppl* 101: 33-43.
- Lombard, N., Prestat, E., van Elsas, J. D. y Simonet, P. (2011) Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. *FEMS Microbiol Ecol* 78: 31-49.
- Lopez-Sanchez, A., Floriano, B., Andujar, E., Hernaez, M. J. y Santero, E. (2010) Tetralin-induced and ThnR-regulated aldehyde dehydrogenase and beta-oxidation genes in *Sphingomonas macrogolitabida* strain TFA. *Appl Environ Microbiol* 76: 110-118.

- Luna, G. M., Dell'Anno, A. y Danovaro, R. (2006) DNA extraction procedure: a critical issue for bacterial diversity assessment in marine sediments. *Environ Microbiol* 8: 308-320.
- Lynd, L. R., Weimer, P. J., van Zyl, W. H. y Pretorius, I. S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66: 506-77, table of contents.
- Martinez, A. y Osburne, M. S. (2013) Preparation of fosmid libraries and functional metagenomic analysis of microbial community DNA. *Methods Enzymol* 531: 123-142.
- Martinez, A., Kolvek, S. J., Yip, C. L., Hopke, J., Brown, K. A., MacNeil, I. A. y Osburne, M. S. (2004) Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl Environ Microbiol* 70: 2452-2463.
- Martinez-Perez, O., Moreno-Ruiz, E., Floriano, B. y Santero, E. (2004) Regulation of tetralin biodegradation and identification of genes essential for expression of *thn* operons. *J Bacteriol* 186: 6101-6109.
- Martin-Laurent, F., Philippot, L., Hallet, S., Chaussod, R., Germon, J. C., Soulas, G. y Catroux, G. (2001) DNA extraction from soils: old bias for new microbial diversity analysis methods. *Appl Environ Microbiol* 67: 2354-2359.
- Mason, J. R. y Cammack, R. (1992) The electron-transport proteins of hydroxylating bacterial dioxygenases. *Annu Rev Microbiol* 46: 277-305.
- McGarvey, K. M., Queitsch, K. y Fields, S. (2012) Wide variation in antibiotic resistance proteins identified by functional metagenomic screening of a soil DNA library. *Appl Environ Microbiol* 78: 1708-1714.
- Medina, C., Camacho, E. M., Flores, A., Mesa-Pereira, B. y Santero, E. (2011) Improved expression systems for regulated expression in Salmonella infecting eukaryotic cells. *PLoS One* 6: e23055.
- Miettinen-Oinonen, A. y Suominen, P. (2002) Enhanced production of *Trichoderma reesei* endoglucanases and use of the new cellulase preparations in producing the stonewashed effect on denim fabric. *Appl Environ Microbiol* 68: 3956-3964.
- Miller, J. H., (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Miller, D. N., Bryant, J. E., Madsen, E. L. y Ghiorse, W. C. (1999) Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl Environ Microbiol* 65: 4715-4724.
- Monticello, D. J. y Finnerty, W. R. (1985) Microbial desulfurization of fossil fuels. *Annu Rev Microbiol* 39: 371-389.
- Morimoto, S. y Fujii, T. (2009) A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking. *Appl Microbiol Biotechnol* 83: 389-396.

- Mullany, P. (2014) Functional metagenomics for the investigation of antibiotic resistance. *Virulence* 5: 443-447.
- Nacke, H., Engelhaupt, M., Brady, S., Fischer, C., Tautzt, J. y Daniel, R. (2012) Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. *Biotechnol Lett* 34: 663-675.
- Nam, J. W., Nojiri, H., Yoshida, T., Habe, H., Yamane, H. y Omori, T. (2001) New classification system for oxygenase components involved in ring-hydroxylating oxygenations. *Biosci Biotechnol Biochem* 65: 254-263.
- Nebert, D. W., Adesnik, M., Coon, M. J., Estabrook, R. W., Gonzalez, F. J., Guengerich, F. P. y col. (1987) The P450 gene superfamily: recommended nomenclature. *DNA* 6: 1-11.
- Nebert, D. W., Nelson, D. R., Adesnik, M., Coon, M. J., Estabrook, R. W., Gonzalez, F. J. y col. (1989) The P450 superfamily: updated listing of all genes and recommended nomenclature for the chromosomal loci. *DNA* 8: 1-13.
- Nebert, D. W., Nelson, D. R., Coon, M. J., Estabrook, R. W., Feyereisen, R., Fujii-Kuriyama, Y. y col. (1991) The P450 superfamily: update on new sequences, gene mapping, and recommended nomenclature. *DNA Cell Biol* 10: 1-14.
- Nelson, D. R., Kamataki, T., Waxman, D. J., Guengerich, F. P., Estabrook, R. W., Feyereisen, R. y col. (1993) The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol* 12: 1-51.
- Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J. y col. (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 6: 1-42.
- Neveu, J., Regeard, C. y DuBow, M. S. (2011) Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts. *Appl Microbiol Biotechnol* 91: 635-644.
- Newman, J. R. y Fuqua, C. (1999) Broad-host-range expression vectors that carry the L-arabinose-inducible *Escherichia coli* araBAD promoter and the araC regulator. *Gene* 227: 197-203.
- Nguyen, N. H., Maruset, L., Uengwetwanit, T., Mhuantong, W., Harnpicharnchai, P., Champreda, V. y col. (2012) Identification and characterization of a cellulase-encoding gene from the buffalo rumen metagenomic library. *Biosci Biotechnol Biochem* 76: 1075-1084.
- Nies, D. H. (2003) Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiol Rev* 27: 313-339.
- Nikaido, H. (2011) Structure and mechanism of RND-type multidrug efflux pumps. *Adv Enzymol Relat Areas Mol Biol* 77: 1-60.

- Nodwell, J. R. y Greenblatt, J. (1991) The nut site of bacteriophage lambda is made of RNA and is bound by transcription antitermination factors on the surface of RNA polymerase. *Genes Dev* 5: 2141-2151.
- Nogales, J., Canales, A., Jimenez-Barbero, J., Garcia, J. L. y Diaz, E. (2005) Molecular characterization of the gallate dioxygenase from *Pseudomonas putida* KT2440. The prototype of a new subgroup of extradiol dioxygenases. *J Biol Chem* 280: 35382-35390.
- Nordlund, I., Powlowski, J. y Shingler, V. (1990) Complete nucleotide sequence and polypeptide analysis of multicomponent phenol hydroxylase from *Pseudomonas* sp. strain CF600. *J Bacteriol* 172: 6826-6833.
- O'Connor, K. E., Dobson, A. D. y Hartmans, S. (1997) Indigo formation by microorganisms expressing styrene monooxygenase activity. *Appl Environ Microbiol* 63: 4287-4291.
- Ogram, A., Sayler, G. S. y Barkay, T. (1987) The extraction and purification of microbial DNA from sediments. *J Microbiol Methods* 7: 57-66.
- Ono, A., Miyazaki, R., Sota, M., Ohtsubo, Y., Nagata, Y. y Tsuda, M. (2007) Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Appl Microbiol Biotechnol* 74: 501-510.
- Owen, J. G., Robins, K. J., Parachin, N. S. y Ackerley, D. F. (2012) A functional screen for recovery of 4'-phosphopantetheinyl transferase and associated natural product biosynthesis genes from metagenome libraries. *Environ Microbiol* 14: 1198-1209.
- Owen, J. G., Reddy, B. V., Ternei, M. A., Charlop-Powers, Z., Calle, P. Y., Kim, J. H. y Brady, S. F. (2013) Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proc Natl Acad Sci U S A* 110: 11797-11802.
- Pace, N. R. (2009) Mapping the tree of life: progress and prospects. *Microbiol Mol Biol Rev* 73: 565-576.
- Pace, N. R., Stahl, D. A., Lane, D. J. y Olsen, G. J. (1985). Analyzing natural microbial populations by rRNA sequences. *ASM News* 51: 4-12.
- Palzkill, T. (2013) Metallo-beta-lactamase structure and function. *Ann N Y Acad Sci* 1277: 91-104.
- Pehrsson, E. C., Forsberg, K. J., Gibson, M. K., Ahmadi, S. y Dantas, G. (2013) Novel resistance functions uncovered using functional metagenomic investigations of resistance reservoirs. *Front Microbiol* 4: 145.
- Peng, Q., Zhang, X., Shang, M., Wang, X., Wang, G., Li, B. y col. (2011) A novel esterase gene cloned from a metagenomic library from neritic sediments of the South China Sea. *Microb Cell Fact* 10: 95-2859-10-95.
- Peng, R. H., Xiong, A. S., Xue, Y., Fu, X. Y., Gao, F., Zhao, W. y col. (2010) A profile of ring-hydroxylating oxygenases that degrade aromatic pollutants. *Rev Environ Contam Toxicol* 206: 65-94.

- Peng, R. H., Xiong, A. S., Xue, Y., Fu, X. Y., Gao, F., Zhao, W. y col. (2008) Microbial biodegradation of polyaromatic hydrocarbons. *FEMS Microbiol Rev* 32: 927-955.
- Poole, K. (2005) Efflux-mediated antimicrobial resistance. *J Antimicrob Chemother* 56: 20-51.
- Precigou, S., Goulas, P. y Duran, R. (2001) Rapid and specific identification of nitrile hydratase (NHase)-encoding genes in soil samples by polymerase chain reaction. *FEMS Microbiol Lett* 204: 155-161.
- Prodromou, C. y Pearl, L. H. (1992) Recursive PCR: a novel technique for total gene synthesis. *Protein Eng* 5: 827-829.
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C. y col. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464: 59-65.
- Rees, H. C., Grant, S., Jones, B., Grant, W. D. y Heaphy, S. (2003) Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles* 7: 415-421.
- Rhee, J. K., Ahn, D. G., Kim, Y. G. y Oh, J. W. (2005) New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Appl Environ Microbiol* 71: 817-825.
- Rossello-Mora, R. y Amann, R. (2001) The species concept for prokaryotes. *FEMS Microbiol Rev* 25: 39-67.
- Royo, J. L., Moreno-Ruiz, E., Cebolla, A. y Santero, E. (2005a) Stable long-term indigo production by overexpression of dioxygenase genes using a chromosomal integrated cascade expression circuit. *J Biotechnol* 116: 113-124.
- Royo, J. L., Manyani, H., Cebolla, A. y Santero, E. (2005b) A new generation of vectors with increased induction ratios by overimposing a second regulatory level by attenuation. *Nucleic Acids Res* 33: e169.
- Royo, J. L., Becker, P. D., Camacho, E. M., Cebolla, A., Link, C., Santero, E. y Guzman, C. A. (2007) In vivo gene regulation in *Salmonella* spp. by a salicylate-dependent control circuit. *Nat Methods* 4: 937-942.
- Rubin, E. M. (2008) Genomics of cellulosic biofuels. *Nature* 454: 841-845.
- Sabehi, G. y Beja, O. (2013) Preparation of BAC libraries from marine microbial populations. *Methods Enzymol* 531: 111-122.
- Sakai, M., Masai, E., Asami, H., Sugiyama, K., Kimbara, K. y Fukuda, M. (2002) Diversity of 2,3-dihydroxybiphenyl dioxygenase genes in a strong PCB degrader, *Rhodococcus* sp. strain RHA1. *J Biosci Bioeng* 93: 421-427.
- Sala-Trepat, J. M., Murray, K. y Williams, P. A. (1972) The metabolic divergence in the meta cleavage of catechols by *Pseudomonas putida* NCIB 10015. Physiological significance and evolutionary implications. *Eur J Biochem* 28: 347-356.

- Sambrook, J. y Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Sauvage, E., Kerff, F., Terrak, M., Ayala, J. A. y Charlier, P. (2008) The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol Rev* 32: 234-258.
- Schell, M. A. (1985) Transcriptional control of the nah and sal hydrocarbon-degradation operons by the *nahR* gene product. *Gene* 36: 301-309.
- Schell, M. A. y Wender, P. E. (1986) Identification of the *nahR* gene product and nucleotide sequences required for its activation of the *sal* operon. *J Bacteriol* 166: 9-14.
- Schloss, P. D. y Handelsman, J. (2006) Toward a census of bacteria in soil. *PLoS Comput Biol* 2: e92.
- Schmieder, R. y Edwards, R. (2012) Insights into antibiotic resistance through metagenomic approaches. *Future Microbiol* 7: 73-89.
- Shingler, V., Powlowski, J. y Marklund, U. (1992) Nucleotide sequence and functional analysis of the complete phenol/3,4-dimethylphenol catabolic pathway of *Pseudomonas* sp. strain CF600. *J Bacteriol* 174: 711-724.
- Shingler, V., Franklin, F. C., Tsuda, M., Holroyd, D. y Bagdasarian, M. (1989) Molecular analysis of a plasmid-encoded phenol hydroxylase from *Pseudomonas* CF600. *J Gen Microbiol* 135: 1083-1092.
- Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y. y Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci U S A* 89: 8794-8797.
- Sicard, R., Chen, L. S., Marsaioli, A. J. y Reymond, J. L. (2005) A fluorescence-based assay for Baeyer–Villiger monooxygenases, hydroxylases and lactonases. *Adv Synth Catal* 347: 1041-1050.
- Silva, C. C., Hayden, H., Sawbridge, T., Mele, P., De Paula, S. O., Silva, L. C. y col. (2013) Identification of genes and pathways related to phenol degradation in metagenomic libraries from petroleum refinery wastewater. *PLoS One* 8: e61811.
- Simon, C. y Daniel, R. (2011) Metagenomic analyses: past and future trends. *Appl Environ Microbiol* 77: 1153-1161.
- Simon, C. y Daniel, R. (2010) Construction of small-insert and large-insert metagenomic libraries. *Methods Mol Biol* 668: 39-50.
- Simon, C., Herath, J., Rockstroh, S. y Daniel, R. (2009) Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl Environ Microbiol* 75: 2964-2968.
- Simons, R. W., Hoopes, B. C., McClure, W. R. y Kleckner, N. (1983) Three promoters near the termini of IS10: pIN, pOUT, and pIII. *Cell* 34: 673-682.

- Simu, K. y Hagstrom, A. (2004) Oligotrophic bacterioplankton with a novel single-cell life strategy. *Appl Environ Microbiol* 70: 2445-2451.
- Singleton, D. R., Hu, J. y Aitken, M. D. (2012) Heterologous expression of polycyclic aromatic hydrocarbon ring-hydroxylating dioxygenase genes from a novel pyrene-degrading betaproteobacterium. *Appl Environ Microbiol* 78: 3552-3559.
- Sommer, M. O., Dantas, G. y Church, G. M. (2009) Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* 325: 1128-1131.
- Sosio, M., Giusino, F., Cappellano, C., Bossi, E., Puglia, A. M. y Donadio, S. (2000) Artificial chromosomes for antibiotic-producing actinomycetes. *Nat Biotechnol* 18: 343-345.
- Spence, E. L., Kawamukai, M., Sanvoisin, J., Braven, H. y Bugg, T. D. (1996) Catechol dioxygenases from *Escherichia coli* (MhpB) and *Alcaligenes eutrophus* (MpcI): sequence analysis and biochemical properties of a third family of extradiol dioxygenases. *J Bacteriol* 178: 5249-5256.
- Stach, J. E., Bathe, S., Clapp, J. P. y Burns, R. G. (2001) PCR-SSCP comparison of 16S rDNA sequence diversity in soil DNA obtained using different isolation and purification methods. *FEMS Microbiol Ecol* 36: 139-151.
- Steffan, R. J., Goksoyr, J., Bej, A. K. y Atlas, R. M. (1988) Recovery of DNA from soils and sediments. *Appl Environ Microbiol* 54: 2908-2915.
- Steiner, R. A., Janssen, H. J., Roversi, P., Oakley, A. J. y Fetzner, S. (2010) Structural basis for cofactor-independent dioxygenation of N-heteroaromatic compounds at the alpha/beta-hydrolase fold. *Proc Natl Acad Sci U S A* 107: 657-662.
- Stephen, D., Jones, C. y Schofield, J. P. (1990) A rapid method for isolating high quality plasmid DNA suitable for DNA sequencing. *Nucleic Acids Res* 18: 7463-7464.
- Streit, W. R. y Schmitz, R. A. (2004) Metagenomics--the key to the uncultured microbes. *Curr Opin Microbiol* 7: 492-498.
- Studier, F. W. y Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189: 113-130.
- Suenaga, H., Ohnuki, T. y Miyazaki, K. (2007) Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environ Microbiol* 9: 2289-2297.
- Tabor, S. y Richardson, C. C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci U S A* 82: 1074-1078.
- Taguchi, K., Motoyama, M. y Kudo, T. (2004) Multiplicity of 2,3-dihydroxybiphenyl dioxygenase genes in the Gram-positive polychlorinated biphenyl degrading bacterium *Rhodococcus rhodochrous* K37. *Biosci Biotechnol Biochem* 68: 787-795.

- Tao, Y., Fishman, A., Bentley, W. E. y Wood, T. K. (2004) Oxidation of benzene to phenol, catechol, and 1,2,3-trihydroxybenzene by toluene 4-monooxygenase of *Pseudomonas mendocina* KR1 and toluene 3-monooxygenase of *Ralstonia pickettii* PKO1. *Appl Environ Microbiol* 70: 3814-3820.
- Taupp, M., Lee, S., Hawley, A., Yang, J. y Hallam, S. J. (2009) Large insert environmental genomic library production. *J Vis Exp* (31). pii: 1387. doi: 10.3791/1387.
- Teather, R. M. y Wood, P. J. (1982) Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl Environ Microbiol* 43: 777-780.
- Teeling, H. y Glockner, F. O. (2012) Current opportunities and challenges in microbial metagenome analysis--a bioinformatic perspective. *Brief Bioinform* 13: 728-742.
- Thomas, C. M., Stalker, D. M. y Helinski, D. R. (1981) Replication and incompatibility properties of segments of the origin region of replication of the broad host range plasmid RK2. *Mol Gen Genet* 181: 1-7.
- Thompson, J. D., Higgins, D. G. y Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.
- Tomas-Gallardo, L., Santero, E., Camafeita, E., Calvo, E., Schlomann, M. y Floriano, B. (2009) Molecular and biochemical characterization of the tetralin degradation pathway in *Rhodococcus* sp. strain TFB. *Microb Biotechnol* 2: 262-273.
- Troeschel, S. C., Thies, S., Link, O., Real, C. I., Knops, K., Wilhelm, S. y col. (2012) Novel broad host range shuttle vectors for expression in *Escherichia coli*, *Bacillus subtilis* and *Pseudomonas putida*. *J Biotechnol* 161: 71-79.
- Tyson, G. W. y Banfield, J. F. (2005) Cultivating the uncultivated: a community genomics perspective. *Trends Microbiol* 13: 411-415.
- Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M. y col. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428: 37-43.
- Uchiyama, T. y Miyazaki, K. (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* 20: 616-622.
- Uchiyama, T., Abe, T., Ikemura, T. y Watanabe, K. (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol* 23: 88-93.
- van Hellemond, E. W., Janssen, D. B. y Fraaije, M. W. (2007) Discovery of a novel styrene monooxygenase originating from the metagenome. *Appl Environ Microbiol* 73: 5832-5839.

- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A. y col. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304: 66-74.
- Vieites, J. M., Guazzaroni, M. E., Beloqui, A., Golyshin, P. N. y Ferrer, M. (2009) Metagenomics approaches in systems microbiology. *FEMS Microbiol Rev* 33: 236-255.
- Wackett, L. P. y Gibson, D. T. (1983) Rapid method for detection and quantitation of hydroxylated aromatic intermediates produced by microorganisms. *Appl Environ Microbiol* 45: 1144-1147.
- Wang, F., Li, F., Chen, G. y Liu, W. (2009) Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. *Microbiol Res* 164: 650-657.
- Wang, G. Y., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J. y col. (2000) Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett* 2: 2401-2404.
- Warren, R. L., Freeman, J. D., Levesque, R. C., Smailus, D. E., Flibotte, S. y Holt, R. A. (2008) Transcription of foreign DNA in *Escherichia coli*. *Genome Res* 18: 1798-1805.
- Werck-Reichhart, D. y Feyereisen, R. (2000) Cytochromes P450: a success story. *Genome Biol* 1: REVIEWS3003.
- Wery, N., Gerike, U., Sharman, A., Chaudhuri, J. B., Hough, D. W. y Danson, M. J. (2003) Use of a packed-column bioreactor for isolation of diverse protease-producing bacteria from antarctic soil. *Appl Environ Microbiol* 69: 1457-1464.
- Wexler, M., Bond, P. L., Richardson, D. J. y Johnston, A. W. (2005) A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environ Microbiol* 7: 1917-1926.
- Whitman, W. B., Coleman, D. C. y Wiebe, W. J. (1998) Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* 95: 6578-6583.
- Wilke, M. S., Lovering, A. L. y Strynadka, N. C. (2005) Beta-lactam antibiotic resistance: a current structural perspective. *Curr Opin Microbiol* 8: 525-533.
- Williamson, L. L., Borlee, B. R., Schloss, P. D., Guan, C., Allen, H. K. y Handelsman, J. (2005) Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Appl Environ Microbiol* 71: 6335-6344.
- Yen, K. M. y Gunsalus, I. C. (1982) Plasmid gene organization: naphthalene/salicylate oxidation. *Proc Natl Acad Sci U S A* 79: 874-878.
- Yokouchi, H., Fukuoka, Y., Mukoyama, D., Calugay, R., Takeyama, H. y Matsunaga, T. (2006) Whole-metagenome amplification of a microbial community associated with scleractinian coral by multiple displacement amplification using phi29 polymerase. *Environ Microbiol* 8: 1155-1163.
- Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E. J., Short, J. M. y Keller, M. (2002) Cultivating the uncultured. *Proc Natl Acad Sci U S A* 99: 15681-15686.

Zhou, J., Bruns, M. A. y Tiedje, J. M. (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62: 316-322.

<http://umbbd.ethz.ch/index.html>, University of Minnesota. Biocatalysis / Biodegradation Database.