



UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS

DEPARTAMENTO DE CIENCIAS BIOLÓGICAS

Trabajo de Tesis Doctoral

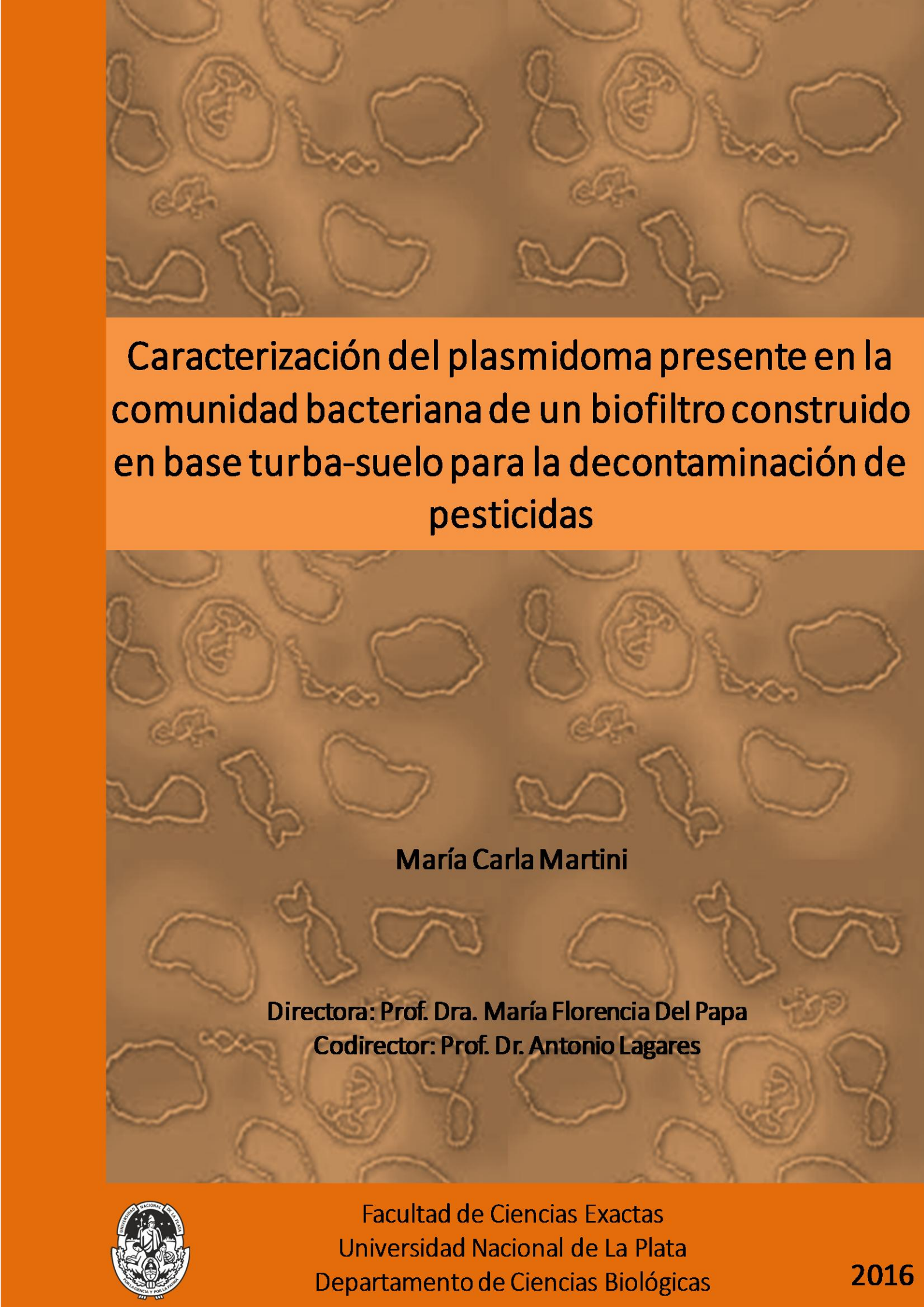
***Caracterización del plasmidoma presente en la
comunidad bacteriana de un biofiltro construido en base
turba-suelo para la decontaminación de pesticidas***

María Carla Martini

Directora: María Florencia Del Papa

Codirector: Antonio Lagares

2016

The background of the entire page is a repeating pattern of microscopic images of plasmids, which are small, circular DNA molecules. The plasmids are shown in various orientations and some are intertwined. The color scheme is primarily orange and brown.

Caracterización del plasmidoma presente en la comunidad bacteriana de un biofiltro construido en base turba-suelo para la decontaminación de pesticidas

María Carla Martini

Directora: Prof. Dra. María Florencia Del Papa

Codirector: Prof. Dr. Antonio Lagares



Facultad de Ciencias Exactas
Universidad Nacional de La Plata
Departamento de Ciencias Biológicas

2016

El presente trabajo de Tesis para optar al grado de Doctor de la Facultad de Ciencias Exactas ha sido realizado en el Instituto de Biotecnología y Biología Molecular de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata, bajo la dirección de la Prof. Dra. María Florencia Del Papa y la codirección del Prof. Dr. Antonio Lagares.

Mi reconocimiento

Al Instituto de Biotecnología y Biología Molecular y a la Facultad de Ciencias Exactas, por haberme brindado el lugar y los equipos necesarios para la realización de este trabajo.

Al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y al Proyecto METAEXPLORE, por haberme otorgado las becas que posibilitaron mi dedicación a la investigación.

Mi agradecimiento

A la Prof. Dra. María Florencia Del Papa, por dirigir mi trabajo con tanta dedicación y compromiso y por iniciarme en la vida científica. Por su estímulo permanente, por haber depositado su confianza en mí, por escucharme y permitirme crecer y sobre todo por su paciencia y su cariño.

Al Prof. Dr. Antonio Lagares, por abrirme las puertas al laboratorio y por codirigir este trabajo con interés y entusiasmo, por su confianza, por su estímulo constante, sus ideas y enseñanzas.

Al Prof. Dr. Mariano Pistorio, por aportar ideas, por estar siempre presente, por su estímulo y sus chistes.

A mis compañeros de laboratorio: Gon, Mauri, Walter, Angie, Piru, Fran, Ile, Jose, Yuli, Flor y Moni, por estar todo este tiempo y poder contar con ellos, por brindarme su amistad, por compartir mates, charlas y hermosos momentos juntos, y sobre todo por hacer divertidos los días en el laboratorio.

A los vecinos R3ses y R3sitos, por compartir lindos momentos, por su alegría, sus chistes y su buena onda.

A los chicos de VA, en especial a Mati, Maia y Emi, por su amistad, por los almuerzos, las charlas y los lindos momentos que compartimos.

A las Bondiolitas, por compartir momentos y los viajes, por las largas y entretenidas charlas y sobre todo por su amistad.

A los integrantes del IBBM, en especial a los becarios, por su compañerismo, por los asados compartidos, los almuerzos, las charlas y las reuniones de todos estos años.

A las chicas del equipo de fútbol del IBBM, por compartir este deporte que nos permitió conocernos y viajar juntas.

A Sil, Pau, Abel, Luciana, Larisa y Ruben, por ayudar y hacer más ameno el trabajo; por su buena predisposición y compañerismo.

A Pity, por preguntar siempre, por preocuparse y por tirar siempre buena onda.

A Gise, por ayudar siempre, por las charlas compartidas y su compañerismo.

A Andrea, por recibirme todas las mañanas con alegría y buena onda.

Al Prof. Dr. Alfred Pühler y al Dr. Andreas Schlüter, por permitirme realizar una estadía en su laboratorio y poder avanzar en la parte informática de mi trabajo. A los integrantes del laboratorio por su buena predisposición y compañerismo.

A los amigos de la facultad y de la vida, por estar siempre.

A mi familia y a Daniel, por acompañarme y aguantarme todos los días.

Índice General

Índice de figuras	11
Índice de tablas	13
Abreviaturas	15
Trabajos científicos publicados durante el transcurso de este trabajo de Tesis doctoral	17
CAPÍTULO I. INTRODUCCIÓN	21
I.1. Los microorganismos en el ambiente	23
I.1.1. Biodiversidad microbiana	23
I.1.2. Estudio de la diversidad bacteriana	23
I.1.2.1. Métodos empleados para el estudio de la diversidad microbiana	24
I.2. Evolución y transferencia de información en los microorganismos	26
I.2.1. Transferencia horizontal de genes	26
I.2.1.1. Mecanismos de transferencia horizontal de genes	28
I.2.1.1.1. Transformación bacteriana	29
I.2.1.1.2. Transducción bacteriana	29
I.2.1.1.3. Conjugación bacteriana	29
I.2.2. Elementos genéticos móviles	31
I.2.2.1. Plásmidos	32
I.2.2.1.1. Clasificación de los plásmidos	32
I.2.2.1.2. Importancia de los plásmidos	35
I.2.2.1.3. Estudio de los plásmidos ambientales: métodos y herramientas	37
I.2.2.1.3.1. Aislamiento endógeno de plásmidos	37
I.2.2.1.3.2. Aislamiento exógeno de plásmidos	38
I.2.2.1.3.3. Nuevos métodos para el aislamiento de plásmidos	38
I.3. Secuenciamiento y abordaje del estudio de genomas microbianos	41
I.3.1. Métodos de secuenciación	41
I.3.2. Secuenciamiento de alto rendimiento y metagenómica: aplicación al estudio de los plásmidos	47
I.3.2.1. Algunos ejemplos de estudios de plásmidos ambientales mediante el uso de secuenciamiento masivo	48
I.4. METAEXPLORE: Consorcio interdisciplinario para el estudio de la biología/ metagenómica de ambientes de interés	50
I.4.1. Selección de la muestra de estudio	51
I.4.2. Sistemas de biopurificación	52
I.4.2.1. Breve descripción del BPS en estudio	53
I.5. Objetivos de esta Tesis	57
I.5.1. Objetivo general	57
I.5.2. Objetivos específicos	57
CAPÍTULO II. MATERIALES Y MÉTODOS	59
II.1. Técnicas microbiológicas	61
II.1.1. Cepas bacterianas utilizadas en este trabajo	61
II.1.2. Plásmidos utilizados en este trabajo	63
II.1.3. Medios de cultivo y conservación de microorganismos	64
II.1.3.1. Medios de cultivo sólidos y líquidos	64
II.1.3.2. Medios de conservación bacterianos	66

II.1.4. Condiciones de cultivo	66
II.1.5. Antibióticos	67
II.1.6. Obtención de los aislamientos a partir de la muestra del biofiltro	67
II.1.7. <i>Screening</i> de aislamientos portadores de plásmidos	67
II.1.8. Determinación de la concentración mínima inhibitoria para antibióticos y metales	68
II.1.9. Evaluación de movilidad plasmídica	68
II.1.9.1. Conjugaciones biparentales	68
II.1.9.2. Conjugaciones triparentales	69
II.1.10. Preparación de células electrocompetentes	69
II.1.11. Transformación bacteriana	69
II.1.12. Recuento de bacterias en medios líquidos	70
II.2. Técnicas moleculares	70
II.2.1. Obtención de ADN genómico para su uso en PCR	70
II.2.2. Preparación de ADN plasmídico	70
II.2.2.1. Obtención a pequeña escala (“miniprep”)	70
II.2.2.2. Obtención de ADN plasmídico a gran escala para secuenciamiento masivo	71
II.2.2.3. Obtención de ADN plasmídico de alto peso molecular para la evaluación de los grupos de incompatibilidad	72
II.2.3. Reacción en cadena de la polimerasa (PCR)	72
II.2.4. Electroforesis de ADN.	73
II.2.4.1. Separación e identificación de fragmentos de restricción o de amplificación por PCR	73
II.2.4.2. Separación e identificación de fragmentos de ADN de bajo peso molecular	73
II.2.4.3. Geles de lisis <i>in situ</i>	73
II.2.5. Purificación de fragmentos de ADN de geles de agarosa	74
II.2.6. Determinación de las concentraciones de ADN en una muestra	75
II.2.7. Secuenciamiento de fragmentos de ADN	75
II.2.8. Síntesis de genes	75
II.2.9. Oligonucleótidos	75
II.2.10. Hibridación por <i>Southern blot</i>	78
II.2.11. Secuenciamiento de los plásmidos de alto peso molecular	78
II.2.12. Tratamientos enzimáticos del ADN	78
II.2.12.1. Digestiones de ADN	78
II.2.12.2. Ligaciones de ADN	79
II.2.13. Desalado de soluciones de ADN	79
II.2.14. Construcción del vector plasmídico para la captura de plásmidos de muestras ambientales	79
II.2.14.1. Introducción de un origen de replicación (<i>oriV</i>) en el minitransposón	79
II.2.14.2. Agregado del sitio de reconocimiento de la endonucleasa <i>Scel</i>	80
II.2.14.3. Remoción del <i>oriT</i> y <i>oriVR6K</i> provenientes del pBAM1	80
II.2.14.4. Introducción de un <i>oriT</i> en el minitransposón	80
II.2.15. Obtención del <i>linker Scel</i>	81
II.2.16. Preparación de proteínas y determinación de su actividad	81
II.2.16.1. Expresión heteróloga de proteínas en <i>E. coli</i>	81
II.2.16.2. Análisis de proteínas mediante geles de poliacrilamida discontinuos en condiciones desnaturizantes (SDS-PAGE)	82
II.2.16.3. Purificación de la lacasa Lac188	82
II.2.16.4. <i>Western blot</i>	82

II.2.16.5. Determinación de la actividad enzimática de la lacasa Lac188	83
II.3. Técnicas bioinformáticas	83
II.3.1. Diseño de <i>primers</i>	83
II.3.2. Determinación de la identidad de fragmentos de ADN purificados y secuenciados	84
II.3.3. Tratamiento de los datos crudos del secuenciamiento de alto rendimiento	84
II.3.4. Búsqueda de genes de resistencia a antibióticos y metales pesados	84
II.3.5. Búsqueda de secuencias de inserción	85
II.3.6. Búsqueda de genes de degradación de compuestos xenobióticos y aromáticos	85
II.3.7. Búsqueda de genes que codifican para lacasas y quitinasas	85
II.3.8. Análisis filogenéticos	85
II.3.8.1. Análisis filogenéticos del gen <i>int11</i>	86
II.3.8.2. Análisis filogenéticos de las relaxasas y del sistema MPF	86

CAPÍTULO III. ESTABLECIMIENTO DE UNA COLECCIÓN DE BACTERIAS PORTADORAS DE PLÁSMIDOS DE ALTO PESO MOLECULAR OBTENIDAS A PARTIR DE UN SISTEMA DE BIOPURIFICACIÓN UTILIZADO PARA LA DECONTAMINACIÓN DE PESTICIDAS

III.1. Introducción	91
III.2. Construcción de una colección de aislamientos portadores de plásmidos de alto peso molecular como material de base para la búsqueda de funciones plasmídicas	93
III.3. La colección de aislamientos como parte del germoplasma presente en el BPS muestreado: Clasificación taxonómica de los aislamientos de la colección bacteriana	94
III.4. Caracterización fenotípica de los aislamientos	95
III.4.1. Evaluación de la tolerancia a antibióticos	97
III.4.2. Evaluación de la tolerancia a metales pesados	97
III.5. Evaluación de los grupos de incompatibilidad plasmídica	99
III.6. Evaluación de la movilidad de los plásmidos de la colección	99
III.7. Discusión	102

CAPÍTULO IV. BÚSQUEDA DE OTROS ELEMENTOS GENÉTICOS MÓVILES EN LOS AISLAMIENTOS REPRESENTATIVOS DE LA COLECCIÓN BACTERIANA DE ESTUDIO

IV.1. Introducción	107
IV.2. Búsqueda y caracterización de integrones	110
IV.2.1. Detección de los genes <i>int11</i> e <i>int12</i> en los aislamientos	110
IV.2.2. Detección del gen <i>sul1</i>	111
IV.2.3. Detección de los genes <i>qacE/qacEΔ1</i>	112
IV.2.4. Detección del <i>orf5</i>	114
IV.2.5. Amplificación y análisis de las regiones variables de los integrones clase 1 presentes en los aislamientos portadores de plásmidos de alto peso molecular seleccionados	114
IV.2.6. Secuenciamiento y análisis de los genes <i>int11</i> identificados en los aislamientos	117
IV.2.7. Localización de los integrones	120
IV.3. Detección de la secuencia de inserción <i>IS1071</i> en los aislamientos	120
IV.4. Discusión	122

CAPÍTULO V. PURIFICACIÓN Y SECUENCIAMIENTO MASIVO DE PLÁSMIDOS PRESENTES EN LA COLECCIÓN BACTERIANA	125
V.1. Introducción	127
V.2. Purificación de plásmidos circulares presentes en la colección de aislamientos	128
V.3. Secuenciamiento del ADN plasmídico en conjunto	129
V.3.1. Información básica del secuenciamiento y ensamblado de la información obtenida	130
V.4. Utilización de la plataforma GenDB para la anotación de las secuencias obtenidas	131
V.4.1. Plataforma GenDB	131
V.4.2. Anotación de las secuencias	132
V.5. Obtención de replicones cerrados	133
V.6. Discusión	137
CAPÍTULO VI. ANÁLISIS BIOINFORMÁTICO DEL SET DE DATOS OBTENIDO A PARTIR DEL SECUENCIAMIENTO MASIVO DEL POOL DE PLÁSMIDOS	141
VI.1. Introducción	143
VI.2. Identificación de genes asociados a funciones plasmídicas	144
VI.2.1. Búsqueda y clasificación de los genes que codifican para proteínas involucradas en la replicación	145
VI.2.2. Búsqueda y clasificación de los genes que codifican para proteínas involucradas en la partición y estabilidad plasmídica	147
VI.2.3. Búsqueda y clasificación de genes que codifican para proteínas involucradas en la movilidad plasmídica	147
VI.2.3.1. Clasificación de las secuencias presentes en los plásmidos de la colección de acuerdo a la relaxasa y al MPF	148
VI.3. Análisis del contenido plasmídico: genes accesorios comúnmente presentes en plásmidos	153
VI.3.1. Presencia de genes asociados con resistencia a antibióticos	154
VI.3.2. Presencia de genes de resistencia a metales	157
VI.3.3. Identificación de genes asociados a la degradación de compuestos xenobióticos e hidrocarburos aromáticos	158
VI.3.4. Presencia de otros MGEs en los plásmidos	159
VI.4. Discusión	163
CAPÍTULO VII. BIOPROSPECCIÓN: BÚSQUEDA, CLONADO, EXPRESIÓN Y CARACTERIZACIÓN DE ENZIMAS CON POTENCIAL INTERÉS INDUSTRIAL	167
VII.1. Introducción	169
VII.2. Búsqueda bioinformática de enzimas de potencial interés industrial	170
VII.2.1. Búsqueda de genes que codifican para quitinasas	171
VII.2.2. Búsqueda de genes que codifican para lacasas	172
VII.2.2.1. Estructura y secuencia de la lacasa Lac188	174
VII.2.2.2. Elección del sistema de expresión	174
VII.2.2.3. Obtención de una construcción portadora del gen de la lacasa Lac188	175
VII.2.2.4. Expresión heteróloga de la enzima Lac188	176
VII.2.2.5. Purificación preliminar de la Lac188	179
VII.2.2.6. Evaluación preliminar de la actividad de Lac188	179
VII.3. Discusión	180

CAPÍTULO VIII. DISEÑO DE UNA ESTRATEGIA Y CONSTRUCCIÓN DE LAS HERRAMIENTAS MOLECULARES NECESARIAS PARA LA CAPTURA DE PLÁSMIDOS	185
VIII.1. Introducción	187
VIII.2. Diseño de una estrategia para la captura de plásmidos: etapas y fundamentos	188
VIII.2.1. Diseño y construcción del vector p-RESC	189
VIII.2.2. Estrategia delineada para la captura de plásmidos	191
VIII.3. Evaluación de la funcionalidad de las herramientas que emplearemos en la estrategia delineada	194
VIII.3.1. Evaluación de la frecuencia de transposición del transposón diseñado	194
VIII.3.2. Evaluación de la actividad de <i>Scel</i> : control de la escisión del p-RESC como consecuencia de la expresión de <i>Scel</i> inducida por la presencia de arabinosa	195
VIII.4. Validación del sistema de captura de plásmidos mediada por el vector p-RESC	197
VIII.4.1. Elección de plásmidos <i>target</i> para la evaluación del sistema	197
VIII.4.2. Prueba de la estrategia diseñada	197
VIII.4.3. Resultados obtenidos	198
VIII.5. Discusión	199
CAPÍTULO IX. DISCUSIÓN GENERAL Y CONCLUSIONES	203
IX.1. Discusión general	205
IX.2. Perspectivas	214
MATERIAL SUPLEMENTARIO	217
BIBLIOGRAFÍA	267

Índice de Figuras

Figura I.1	Mecanismos de transferencia horizontal de genes en bacterias	30
Figura I.2	Mecanismos de transferencia de plásmidos	34
Figura I.3	Esquema general de los métodos más utilizados para la obtención y caracterización de plásmidos en muestras ambientales	39
Figura I.4	Esquema general del secuenciamiento de alto rendimiento utilizando la plataforma Illumina	46
Figura I.5	Imagen del BPS del cual proviene la muestra utilizada en este estudio	54
Figura III.1	Perfiles plasmídicos de los 35 aislamientos representativos por la técnica de lisis <i>in situ</i>	94
Figura III.2	Clasificación taxonómica de los 35 aislamientos representativos	95
Figura IV.1	Estructura típica de un integrón de clase 1	108
Figura IV.2	Detección del gen <i>int1</i> por PCR	111
Figura IV.3	Detección del gen <i>su1</i> por PCR	112
Figura IV.4	Detección del gen <i>qacE</i> (A) y <i>qacEΔ1</i> (B) por PCR	113
Figura IV.5	Amplificación de la región variable de los integrones de clase 1 detectados en la colección de aislamientos utilizando distintos juegos de <i>primers</i>	115
Figura IV.6	Árbol filogenético para el gen <i>int1</i>	119
Figura V.1	Corrida electroforética en gel de agarosa 0,8% de los plásmidos purificados	129
Figura V.2	Esquema de la plataforma Illumina Miseq utilizada para el secuenciamiento de los plásmidos	129
Figura V.3	Interfaz gráfica de la plataforma GenDB	132
Figura V.4	Mapa físico de los 6 replicones cerrados	135
Figura VI.1	Análisis filogenético de las proteínas VirB4 encontradas en los plásmidos de la colección	149
Figura VI.2	Filogenia de las relaxasas identificadas en el set de datos plasmídico	151
Figura VI.3	Filogenia de las relaxasas MOBP encontradas en el set de datos plasmídico	152
Figura VI.4	Genes de resistencia a distintos tipos de antibióticos y acriflavina identificados en los plásmidos	155
Figura VI.5	Clasificación de los genes de resistencia a antibióticos y acriflavina identificados en los plásmidos	156
Figura VI.6	Familias de secuencias de inserción encontradas en el conjunto de secuencias plasmídicas	162
Figura VII.1	Esquema de la estructura de la lacasa Lac188	174
Figura VII.2	Esquema de la construcción pET22b(+):lac188	176
Figura VII.3	Análisis mediante SDS-PAGE de la expresión de Lac188 a distintas temperaturas y tiempos de inducción	177
Figura VII.4	Análisis mediante SDS-PAGE de la expresión de Lac188	177
Figura VII.5	Análisis por <i>Western blot</i> de las fracciones solubles de <i>E. coli</i> BL21(DE3)Star con la construcción pET22b(+):lac188 utilizando un anticuerpo conjugado anti- His-Tag y empleando revelado por quimioluminiscencia	178
Figura VII.6	Actividad enzimática de la Lac ₁₈₈ a diversos pHs empleando distintos sustratos	180
Figura VIII.1	Esquema del vector p-RESC	189
Figura VIII.2	Esquema general de los pasos realizados para la construcción del vector p-RESC	190
Figura VIII.3	Esquema general del procedimiento del sistema utilizado para la captura de plásmidos	193
Figura VIII.4	Esquema de los ensayos realizados a fin de evaluar la inducción por arabinosa y la expresión de Scel	196

Índice de Tablas

Tabla I.1	Algunos ejemplos de estudios de plásmidos de comunidades	42
Tabla I.2	Resumen de los métodos más utilizados para la obtención y caracterización de plásmidos en muestras ambientales	49
Tabla I.3	Pesticidas adicionados al BPS en estudio.	55
Tabla II.1	Cepas utilizadas en este trabajo	61
Tabla II.2	Plásmidos utilizados en este trabajo	63
Tabla II.3	Medios de cultivo utilizados	65
Tabla II.4	Oligonucleótidos utilizados	75
Tabla III.1	Identificación de los aislamientos representativos de la colección bacteriana	96
Tabla III.2	Concentraciones inhibitorias mínimas (MICs) de metales y antibióticos para los aislamientos representativos	98
Tabla III.3	Plásmidos transmisibles y conjugativos detectados en los aislamientos de la colección	101
Tabla IV.1	Identidad a nivel de secuencia nucleotídica (BLASTn) de las regiones variables amplificadas y secuenciadas	116
Tabla IV.2	Identidad a nivel de secuencia aminoacídica (BLASTx) de las regiones variables amplificadas y secuenciadas	116
Tabla IV.3	Resumen de los resultados obtenidos por PCR para los determinantes de integrones de clase 1	121
Tabla V.1	Características principales de los 6 replicones cerrados	133
Tabla VI.1	Genes que contienen un Pfam asociado a replicación plasmídica	146
Tabla VI.2	Lista de posibles genes involucrados en la degradación de xenobióticos e hidrocarburos aromáticos	160
Tabla VII.1	Lacasas y quitinasas detectadas en el conjunto de secuencias de origen plasmídico	172
Tabla VII.2	Actividad lacasa frente a distintos sustratos	179

Abreviaturas

ADN	<u>Á</u> cido <u>D</u> esoxirribonucleico
ADNr	<u>Á</u> cido <u>D</u> esoxirribonucleico <u>r</u> ibosómico
Ap	Ampicilina
ARG	<u>A</u> ntibiotic <u>R</u> esistance <u>G</u> ene, gen de resistencia a antibiótico
ARN	<u>Á</u> cido <u>R</u> ibonucleico
BLAST	<u>B</u> asic <u>L</u> ocal <u>A</u> lignment <u>S</u> earch <u>T</u> ool, herramienta de búsqueda de alineamiento local básico
BPS	<u>B</u> iopurification <u>S</u> ystem, sistema de biopurificación
°C	Grados Celsius
Cb	Carbenicilina
Cm	Cloranfenicol
COG	<u>C</u> lusters of <u>O</u> rtholog <u>G</u> enes, grupos de genes ortólogos
Dex	Medio rico dextrosa ágar
dNTP	deoxinucleótido
DO	<u>D</u> ensidad <u>Ó</u> ptica
EMB	Medio rico eosina-azul de metileno
Er	Eritromicina
g	<u>G</u> ramo
Gm	Gentamicina
GO	<u>G</u> ene <u>O</u> ntology, ontología de genes
GS	Medio mínimo glutamato-sacarosa
h	<u>H</u> ora
HGT	<u>H</u> orizontal <u>G</u> ene <u>T</u> ransfer, transferencia horizontal de genes
ICE	<u>I</u> ntegrative and <u>C</u> onjugative <u>E</u> lement, elemento integrativo y conjugativo
IS	<u>I</u> nsertion <u>S</u> equence, secuencia de inserción
Kb	Kilobases
KEGG	<u>K</u> yoto <u>E</u> ncyclopedia of <u>G</u> enes and <u>G</u> enomes, Enciclopedia de genes y genomas de Kioto
Km	Kanamicina
l	Litro
LB	Medio rico Luria-Bertani
M9	Medio mínimo 9
Mb	Megabases
MCK	Medio rico Mac Conkey
MDA	<u>M</u> ultiple <u>D</u> isplacement <u>A</u> mplification, amplificación por desplazamiento múltiple
MGE	<u>M</u> obile <u>G</u> enetic <u>E</u> lement, elemento genético móvil
MH	Medio rico Mueller-Hinton
MIC	<u>M</u> inimum <u>I</u> nhibitory <u>C</u> oncentration, concentración inhibitoria mínima
min	<u>M</u> inuto
MPF	<u>M</u> ating <u>P</u> air <u>F</u> ormation, formación del par de conjugación
Nm	Neomicina
Nx	Ácido nalidíxico
ORF	<u>O</u> pen <u>R</u> eadin <u>F</u> rame, marco de lectura abierto
oriT	<u>O</u> rigen de <u>T</u> ransferencia
oriV	<u>O</u> rigen de <u>R</u> eplicación
pb	<u>P</u> ares de <u>B</u> ases
PCR	<u>P</u> olymerase <u>C</u> hain <u>R</u> eaction, reacción en cadena de la polimerasa
Rf	Rifampicina

rpm	<u>R</u> evoluciones <u>P</u> or <u>M</u> inuto
Sm	Estreptomicina
Tc	Tetraciclina
Tp	Trimetoprima
TRACA	<i><u>T</u>ransposon <u>A</u>ided <u>C</u>apture</i> , captura mediada por transposón
TY	Medio rico triptona-levadura
UFC	<u>U</u> nidades <u>F</u> ormadoras de <u>C</u> olonias
<i>g</i>	Gravedades

Trabajos científicos publicados durante el transcurso de este trabajo de Tesis doctoral

- Publicaciones en revistas:

- **Relacionadas con este trabajo de Tesis Doctoral**

1. Cultivation-independent screening revealed hot spots of IncP-1, IncP-7 and IncP-9 plasmid occurrence in different environmental habitats

Autores: Simone Dealtry, Guo-Chun Ding, Viola Weichelt, Vincent Dunon, Andreas Schlüter, María Carla Martini, María Florencia Del Papa, Antonio Lagares, Gregory Charles Auton Amos, Elizabeth Margaret Helen Wellington, William Hugo Gaze, Detmer Sipkema, Sara Sjöiling, Dirk Springael, Holger Heuer, Jan Dirk van Elsas, Christopher Thomas, Kornelia Smalla

Año: 2014

Revista: PLOS one

2. Characterization of a collection of plasmid-containing bacteria isolated from an on-farm biopurification system used for pesticide removal

Autores: María Carla Martini, Francisco Javier Albicoro, Eman Nour, Andreas Schlüter, Jan Dirk van Elsas, Dirk Springael, Kornelia Smalla, Mariano Pistorio, Antonio Lagares, María Florencia Del Papa

Año: 2015

Revista: Plasmid

3. Genomics of high molecular weight plasmids isolated from an on-farm biopurification system

Autores: Martini María Carla, Daniel Wibberg, Mauricio Lozano, Gonzalo Torres Tejerizo, Francisco Albicoro, Sebastian Jaenicke, Jan Dirk van Elsas, Alejandro Petroni, María Pilar Garcillán Barcia, Fernando de la Cruz, Andreas Schlüter, Alfred Pühler, Mariano Pistorio, Antonio Lagares, María Florencia Del Papa

Manuscrito enviado

Revista: Scientific Reports

- **Otras (no relacionadas con este trabajo de Tesis Doctoral)**

1. First genomic analysis of the broad-host-range *Rhizobium* sp. LPU83 strain, a member of the low-genetic diversity Oregon-like *Rhizobium* sp. group

Autores: Gonzalo Torres Tejerizo, María Florencia Del Papa, Walter Draghi, Mauricio Lozano, María de los Ángeles Giusti, María Carla Martini, María Eugenia Salas, Ileana Salto, Daniel Wibberg, Rafael Szczepanowski, Andreas Schlüter, Antonio Lagares, Mariano Pistorio

Año: 2011

Revista: Journal of Bacteriology

2. Development of new positive-selection RIVET tools: detection of induced promoters by the excision-based transcriptional activation of an *aacCI* (GmR)-gfp fusion

Autores: Mauricio Lozano, María Eugenia Salas, María de los Ángeles Giusti, Walter Draghi, Gonzalo Torres Tejerizo, María Carla Martini, María Florencia Del Papa, Mariano Pistorio, Antonio Lagares.

Año: 2011

Revista: Journal of Biotechnology

3. Genetic and functional characterization of a yet-unclassified rhizobial Dtr (DNA-transfer-and-replication) region from a ubiquitous plasmid conjugal system present in *Sinorhizobium meliloti*, in *Sinorhizobium medicae*, and in other nonrhizobial Gram-negative bacteria

Autores: María de los Ángeles Giusti, Mariano Pistorio, Mauricio Lozano, Gonzalo Torres Tejerizo, María Eugenia Salas, María Carla Martini, José López, Walter Draghi, María Florencia Del Papa, Daniel Pérez Mendoza, Juan Sanjuán, Antonio Lagares.

Año: 2012

Revista: Plasmid

4. Conjugal transfer of a *Sinorhizobium meliloti* cryptic plasmid evaluated during a field released and in soil microcosms

Autores: María de los Ángeles Giusti, , Mauricio Lozano, Gonzalo Torres Tejerizo, María Carla Martini, María Eugenia Salas, José López, Walter Draghi, María Florencia Del Papa, Mariano Pistorio, Antonio Lagares

Año: 2012

Revista: European Journal of Soil Biology

5. Novel *tnpR*-based transposable promoter traps suitable for RIVET studies in different gram-negative bacteria

Autores: Mauricio Lozano, María Eugenia Salas, María de los Ángeles Giusti, María Carla Martini, José López, Ileana Salto, María Florencia Del Papa, Mariano Pistorio, Antonio Lagares

Año: 2013

Revista: Journal of Microbiol Methods

6. Genome sequence of the acid-tolerant strain *Rhizobium* sp. LPU83

Autores: Daniel Wibberg, Gonzalo Torres Tejerizo, María Florencia Del Papa, María Carla Martini, Alfred Pühler, Antonio Lagares, Andreas Schlüter, Mariano Pistorio

Año: 2014

Revista: Journal of Biotechnology

- Publicaciones en congresos relacionadas con este trabajo de Tesis Doctoral

1. Plasmid isolation and characterization from a biofilter system used for pesticide removal

María Carla Martini *et al.*, póster

Evento: ISME 2012 Gopenhagen: International symposium on microbial ecology, Copenhagen, Dinamarca, 2012

2. Bioinformatic and functional analysis of a plasmid pool from bacterial isolates from an on-farm biopurification system used for pesticide removal

María Carla Martini *et al.*, presentación oral

Evento: X congreso argentino de microbiología general. SAMIGE, Mar del Plata, Argentina, 2014

3. Identification and molecular characterization of class 1 integrons in plasmid-carrying bacterial isolates recovered from a biopurification system used for pesticide removal

Vieta María Paula *et al.*, póster

Evento: XI Congreso Argentino de Microbiología, SAMIGE, Córdoba, Argentina, 2015

CAPÍTULO I

INTRODUCCIÓN GENERAL

I.1. Los microorganismos en el ambiente

I.1.1. Biodiversidad microbiana

En el Planeta Tierra existe una enorme diversidad microbiana que comprende un gran número de organismos procariotas, hongos, protozoos, algas y virus, los cuales conforman la mayor parte de la biomasa terrestre. El conjunto de microorganismos que habitan un determinado ambiente se define como microbiota (Robinson, *et al.*, 2010). La actividad y diversidad de la microbiota condicionan la fertilidad del suelo, la estabilidad y el funcionamiento de ecosistemas naturales y de los agroecosistemas. La diversidad microbiana es esencial en los ciclos de los nutrientes y en la descomposición del material vegetal en todos los ecosistemas, debido a los procesos biológicos como la oxidación, la reducción, la descomposición de materia orgánica y la mineralización, así como en el establecimiento de las relaciones bióticas interespecíficas e intraespecíficas en ambientes tanto terrestres como acuáticos. De esta manera, el funcionamiento de los ecosistemas es gobernado por la dinámica microbiana, por lo que su estudio resulta de gran relevancia para el entendimiento de todos los procesos que se llevan a cabo en el ambiente. Los procariotas (bacterias y arqueas) son componentes esenciales de la microbiota terrestre ya que catalizan transformaciones únicas e indispensables en los ciclos biogeoquímicos de la biósfera, producen componentes importantes de la biósfera terrestre, y representan una gran proporción de la diversidad genética del planeta. Se estima que existen alrededor de $4 - 6 \times 10^{30}$ células procariotas en la biósfera terrestre, localizadas principalmente en la subsuperficie terrestre ($0,25 - 2,5 \times 10^{30}$ células), en la subsuperficie oceánica ($3,5 \times 10^{30}$ células), en el suelo ($2,6 \times 10^{29}$ células) y en aguas oceánicas ($1,2 \times 10^{29}$ células) (Whitman, *et al.*, 1998). En cuanto a la diversidad bacteriana, se ha reportado que una tonelada de suelo puede contener hasta 4×10^6 taxas distintas (Curtis, *et al.*, 2002), lo cual pone en evidencia la gran cantidad y variedad de microorganismos que habitan el Planeta Tierra.

I.1.2. Estudio de la diversidad bacteriana

El estudio de la diversidad microbiana en ambientes naturales es uno de los

retos más desafiantes para los microbiólogos. La principal dificultad que enfrenta la ecología microbiana está dada por la gran diversidad de organismos existentes, pero también por la complejidad de cada ambiente. Las bacterias presentan una amplia diversidad genética y poseen un potencial metabólico altamente flexible. Esto ha dado como resultado una gran capacidad adaptativa posibilitando su supervivencia en ambientes adversos a lo largo de la historia evolutiva (Berdy, 2012). Por este motivo, y dado el importante rol que juegan en distintos entornos, en las últimas décadas se ha puesto un gran interés en el estudio de las comunidades microbianas que habitan ambientes de distintas características. De las miles de especies bacterianas que existen, se cree que sólo entre el 0,1 - 1% pueden ser cultivadas usando técnicas estándar en condiciones de laboratorio (Amann, *et al.*, 1995). El restante 99% de las bacterias, denominadas “no cultivables”, pueden agruparse en dos grandes categorías: **i)** grupos bacterianos sin representantes cultivables (también llamadas “*yet-to-be-cultivated cells*” o “células a ser cultivadas”), las cuales aún no se han podido cultivar debido a la inexistencia de medios de cultivo adecuados, y **ii)** bacterias que pertenecen a grupos que ya han sido cultivados en condiciones de laboratorio, cuyas células están vivas pero no pueden replicar (también llamadas “*non-dividing cells*” o “células que no se dividen”) (Puspita, *et al.*, 2012). Se ha reportado una gran cantidad de estudios estimando las proporciones de células no cultivables en distintos ambientes. En general, se ha encontrado que en ambientes con una baja cantidad de nutrientes (como por ejemplo, desiertos) hay una menor proporción de bacterias cultivables que en aquellos ambientes ricos en nutrientes (como muestras de materia fecal y lodo activado) (Puspita, *et al.*, 2012). Debido a esto, sólo una pequeña porción del *pool* de diversidad de microorganismos ha sido caracterizada hasta la fecha. Sin embargo, el gran avance en el estudio de la diversidad de microorganismos ambientales se dio al empezar a usar técnicas independientes de cultivo basadas en análisis de ácidos nucleicos, como se discutirá en la siguiente sección.

I.1.2.1. Métodos empleados para el estudio de la diversidad microbiana

En las últimas tres décadas, se han desarrollado distintas metodologías que han permitido conocer y estudiar la diversidad microbiana de distintos ambientes. Entre los

primeros métodos utilizados, podemos mencionar el análisis del secuenciamiento del gen que codifica para el ARNr 16S bacteriano (Stahl, *et al.*, 1984), el análisis de huellas digitales o *fingerprints* (estos métodos aprovechan diferentes propiedades de las secuencias amplificadas como ser: su longitud, presencia o ausencia de sitios de restricción y estructura), la electroforesis en gel con gradiente de temperatura (TGGE) y electroforesis en gel con gradiente desnaturalizante (DGGE) (Muyzer, *et al.*, 1993), la hibridización fluorescente *in situ* (FISH) (Giovannoni, *et al.*, 1988) y el polimorfismo de longitud de fragmentos de restricción (RFLP) y su variante polimorfismo de longitud de fragmentos de restricción terminales (T-RFLP) (Moyer, *et al.*, 1994), entre otros. Sin embargo, estos métodos sólo permiten evaluar la diversidad de una comunidad y la identificación y clasificación de los microorganismos, pero no permiten obtener información acerca de las características y funciones que éstos desempeñan.

Más recientemente, con la aparición de métodos de secuenciamiento masivo (descriptos más adelante, en la sección I.3.1), se han realizado numerosos trabajos con el fin de identificar los microorganismos presentes en comunidades bacterianas enteras basados en análisis de ácidos nucleicos. Para Bacteria y Arquea suele utilizarse como marcador el gen de la subunidad ribosomal 16S (Ward, *et al.*, 1990, Zhou, *et al.*, 2002). Woese (1987) fue el primero en utilizar secuencias de genes de ARNr 16S para construir árboles filogenéticos. La elección de este marcador se basa en que es un buen cronómetro evolutivo: está presente en todos los organismos y algunas zonas de la secuencia se encuentran bien conservadas (ideales para diseñar *primers*) y otras son divergentes (Van de Peer, *et al.*, 1997). Si bien se han utilizado también otros marcadores (como *dnaK33*, *amoA*, *nifH* y otros genes del metabolismo de nitrógeno) el gen que codifica para el ARNr 16S sigue siendo el más utilizado para análisis de comunidades microbianas de suelos. Por ejemplo, se ha utilizado en la determinación de la diversidad bacteriana en plantas de biogás utilizadas para la producción de combustible (Heeg, *et al.*, 2014, Stolze, *et al.*, 2015, Vishnivetskaya, *et al.*, 2015), en plantas de tratamientos de efluentes (Whiteley, *et al.*, 2012, Saunders, *et al.*, 2015) y en intestino humano (Drell, *et al.*, 2014, Escobar, *et al.*, 2014), entre otros ambientes. Estas tecnologías permiten determinar además, de una forma relativamente fácil y

poco costosa, las variaciones en la composición de las comunidades microbianas a lo largo de un período de tiempo, en el espacio, luego de la adición de un compuesto o sustrato, o alguna otra perturbación del sistema. Los métodos que emplean el secuenciamiento de alto rendimiento permiten una mayor resolución, calidad y especificidad de la información obtenida que los anteriormente mencionados, aunque algunos factores como la obtención y procesamiento de la muestra o la elección de los *primers* para la PCR pueden influenciar en los resultados (Walker, *et al.*, 2015). Es importante destacar que este tipo de análisis sólo permite conocer la identidad de los microorganismos allí presentes, pero no permite saber qué genes o funciones están presentes en ellos, aunque sí pueden hacerse inferencias en cuanto a las posibles funciones en un ecosistema a partir de la composición y diversidad de comunidades bacterianas (Perez-Valera, *et al.*, 2015).

I.2. Evolución y transferencia de información en los microorganismos

El flujo de material genético en los microorganismos puede darse de dos formas: por transferencia vertical y por transferencia horizontal (o lateral). En la **transferencia vertical**, el material genético pasa de una célula madre a las células hijas durante la división celular. En este mecanismo, el proceso de evolución se va dando por la acumulación de mutaciones, rearrreglos y recombinación del ADN (Goldenfeld & Woese, 2007). La **transferencia horizontal** de genes (HGT, por sus siglas en inglés *Horizontal Gene Transfer*), consiste en la transmisión de material genético entre células bacterianas en forma independiente de la división celular (Lawrence, 2005). La HGT junto con el proceso de recombinación homóloga son los principales mecanismos que operan en la evolución de los genomas bacterianos (Davison, 1999, Arber, 2000).

I.2.1. Transferencia horizontal de genes

Existe abundante evidencia de la transferencia de genes entre diferentes géneros y especies. Mucha de esta información proviene del análisis de las poblaciones bacterianas del suelo y del hallazgo de bacterias diferentes conteniendo, por ejemplo, plásmidos semejantes o, al revés, plásmidos diferentes presentes en bacterias

similares (Vlassak, *et al.*, 1997, Brom, *et al.*, 2002). Esta observación sugiere la existencia de eventos activos de recombinación y transferencia entre cepas de los mismos o diferentes géneros o especies. Se han reportado eventos de transferencia de material genético entre bacterias muy lejanas evolutivamente, e incluso entre distintos dominios: entre bacterias y arqueas, bacterias y eucariotas, arqueas y eucariotas e incluso entre eucariotas (Boto, 2009). Sin embargo, esta transferencia horizontal es más frecuente y efectiva entre microorganismos con una corta distancia evolutiva (Ochman, *et al.*, 2000, Dagan, *et al.*, 2008, Boto, 2009). Esto se debe principalmente a los mecanismos moleculares que operan en la célula al exponerse a material genético foráneo, como por ejemplo los fenómenos de modificación-restricción, que pueden impedir el mantenimiento de ADN exógeno en la célula (Wiedenbeck & Cohan, 2011). Microorganismos estrechamente emparentados suelen tener sistemas de modificación-restricción similares, lo que permite el mantenimiento del ADN adquirido en la célula. Por lo general, el ADN transferido es de longitud relativamente pequeña (con excepción de los plásmidos de alto peso molecular), comprendiendo entre uno y varios genes. Generalmente, fragmentos pequeños de ADN se adaptan más eficientemente al nuevo hospedador cuando la HGT ocurre entre bacterias muy divergentes entre sí (Wiedenbeck & Cohan, 2011).

La información genética necesaria para afrontar los cambios ambientales está concentrada en los genomas bacterianos, tanto en el ADN cromosomal como en elementos genéticos móviles (MGEs, por sus siglas en inglés: ***Mobile Genetic Elements***), los cuales tienen la capacidad de moverse y propagarse entre microorganismos. Estos MGEs (que serán descritos en la sección 1.2.2.) son los responsables de la transferencia lateral de material genético entre microorganismos.

Como hemos mencionado anteriormente, la HGT es uno de los principales mecanismos que contribuyen a la evolución y diversificación de genomas microbianos (de la Cruz & Davies, 2000). De esta afirmación se desprende una pregunta que muchos científicos han tratado de responder: ¿qué proporción de genes en los microorganismos tienen un origen extrínseco? Si bien se sabe que la HGT ha jugado un papel importante desde los orígenes de las divisiones bacterianas (Gogarten, *et al.*,

2002), es difícil establecer cifras exactas con tiempos evolutivos tan largos. Sin embargo, Nakamura y colaboradores (Nakamura, *et al.*, 2004) estudiaron eventos de HGT recientes en 116 genomas bacterianos completos, encontrando que al menos el 14% de los ORFs (*Open Reading Frames*: marcos de lectura abiertos) fueron adquiridos por transferencia lateral. En el mismo trabajo, se reportó también la existencia de un sesgo en el tipo de genes que son preferentemente transferidos. El estudio reveló que la mayoría de estos genes pertenecían a cuatro categorías principales, de acuerdo a la clasificación del Instituto de Investigación Genómica (TIGR): **i**) funciones de plásmidos, fagos y transposones (lo cual es esperable, dado que estos genes intervienen en la HGT entre células), **ii**) envoltura celular (estructuras de superficie, síntesis y degradación de polisacáridos y síntesis de peptidoglicano, entre otras), **iii**) funciones regulatorias (principalmente interacciones de ADN y de proteínas), y **iv**) procesos celulares (transformación de ADN, patogénesis, producción de toxinas y resistencia). Además, encontraron que los genes que menos han sufrido transferencia horizontal, al menos recientemente, son los denominados “informacionales”, es decir, aquellos involucrados en el procesamiento de información celular, como replicación, transcripción y traducción.

El éxito de la HGT depende de dos factores principales: en primer lugar, de la eficiencia del mecanismo de transferencia (conjugación, transducción o transformación) y la posterior recombinación del ADN (si la hay); y en segundo lugar, del mantenimiento en el genoma del material transferido a lo largo de las distintas generaciones (Boto, 2009). La eficiencia y promiscuidad que presenta la transferencia horizontal, hace que este mecanismo sea un importante factor contribuyente a la evolución de los genomas bacterianos.

I.2.1.1. Mecanismos de transferencia horizontal de genes

El flujo horizontal de material genético entre bacterias ocurre a través de al menos tres mecanismos conocidos: transformación, transducción y conjugación, los cuales se describen a continuación y se resumen en la Figura I.1.

I.2.1.1.1. Transformación bacteriana

Puede definirse a la transformación como la incorporación, integración y posterior herencia estable de ADN libre por parte de las bacterias (ver Figura I.1.A). La capacidad de un microorganismo de captar ADN exógeno, se denomina competencia. Algunas especies bacterianas son naturalmente competentes, lo cual se debe a que son capaces de transportar activamente fragmentos de ADN ambiental a través de su envoltura celular hacia su citoplasma (Mell & Redfield, 2014). Esta capacidad requiere de un sistema de captación de ADN específico asociado a la membrana. Si bien este mecanismo de captación de ADN se ha registrado en una gran variedad de ecosistemas naturales (Davison, 1999), la mayoría de las bacterias no son naturalmente competentes (se estima que sólo entre el 1 - 2% de las bacterias lo son (Jonas, *et al.*, 2001)), por lo que este fenómeno está limitado a un cierto grupo de microorganismos.

I.2.1.1.2. Transducción bacteriana

La transducción consiste en la transferencia de información genética desde una célula dadora a otra receptora a través de bacteriófagos. Cuando los fagos empaquetan su material genético, éstos pueden incorporar accidentalmente fragmentos del genoma bacteriano y transferirlos a otras bacterias, las cuales pueden insertar este material en su genoma, adquiriendo nuevos genes (Frost, *et al.*, 2005) (ver Figura I.1.B). Los bacteriófagos juegan un rol crucial en la regulación de la virulencia, diversidad y ecología microbiana. Si bien los fagos son las entidades biológicas más abundantes y diversas de la biósfera (Penades, *et al.*, 2015) y son relativamente estables por las estructuras proteicas que contienen al ARN o ADN (en comparación con el ADN desnudo captado en el proceso de transformación), éstos tienen un rango de huésped restringido, estando incluso a veces limitados a una sola especie bacteriana. Además, las bacterias pueden mutar para volverse resistentes a un determinado fago (Davison, 1999), lo cual hace que la transducción no sea el mecanismo preferido de transferencia lateral entre bacterias.

I.2.1.1.3. Conjugación bacteriana

La conjugación consiste en un intercambio unidireccional de información conte-

nida en plásmidos o elementos integrativos y conjugativos (ICEs, por sus siglas en inglés: *Integrative and Conjugative Elements*) desde una bacteria dadora a otra receptora mediante un contacto físico (Figura I.1.C). El proceso de conjugación requiere de un sistema de secreción tipo IV, que consiste en tres clases de proteínas: las del canal de translocación, las del *pilus* y las ATPasas, que proporcionan la fuerza requerida para el funcionamiento del sistema (Ilangovan, *et al.*, 2015). Estas proteínas, en su conjunto, forman el aparato conjugativo. Si bien la conjugación se describió inicialmente para los plásmidos, hoy se sabe que este mecanismo es utilizado por otros MGEs, como los ICEs mencionados previamente (Johnson & Grossman, 2015).

La adquisición de material genético vía conjugación puede producirse de varias maneras: **i)** por transferencia de un plásmido conjugativo o ICE que posee en su

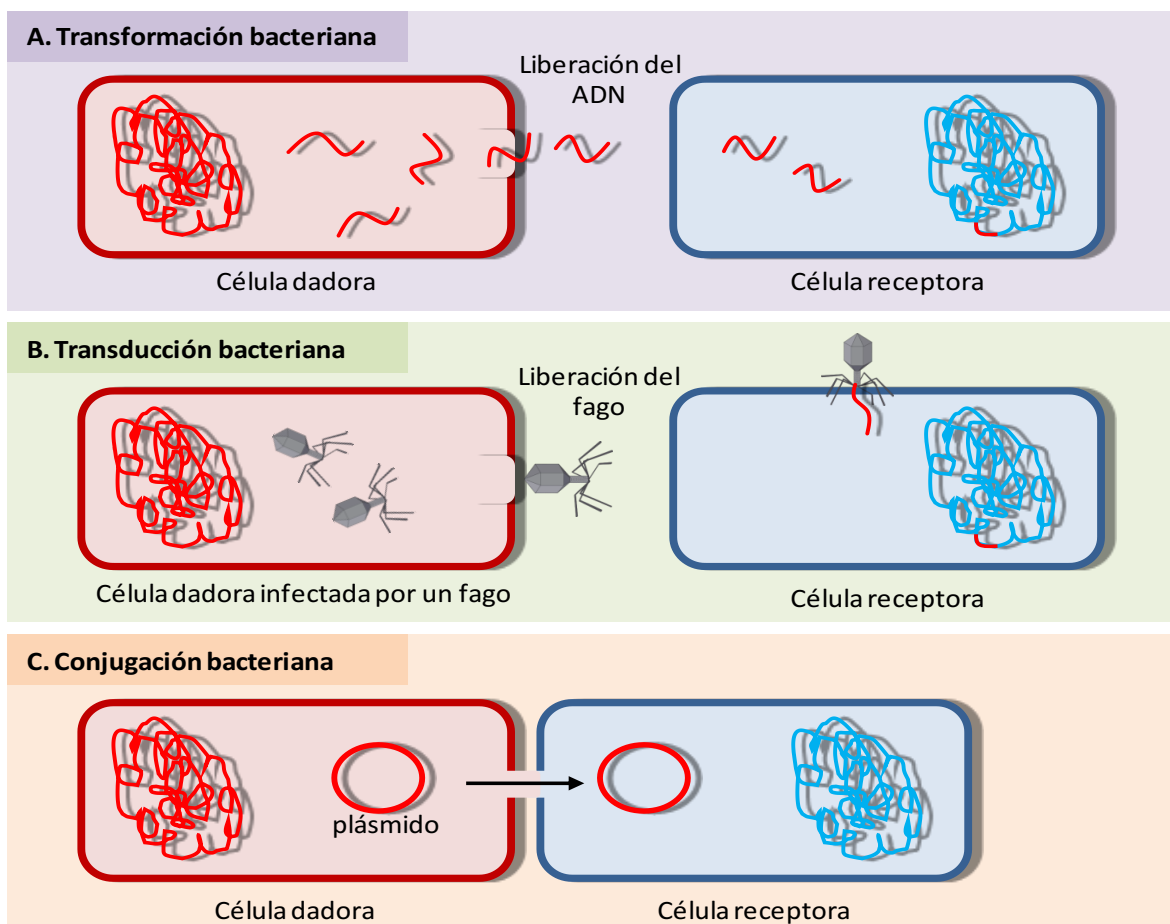


Figura I.1. Mecanismos de transferencia horizontal de genes en bacterias. A) Transformación: una célula libera su ADN desnudo al medio y otra lo capta e integra en su genoma. **B)** Transducción: la transferencia del ADN es mediada por un fago. **C)** Conjugación: el intercambio de información entre dos células ocurre por la transferencia de plásmidos.

secuencia la información completa para su movilización; **ii**) por transferencia de un plásmido movilizable (no-autotransmisible) que contiene el origen de transferencia (*oriT*) y que sólo puede ser movilizado por acción de un plásmido conjugativo o un ICE y; y **iii**) por transferencia de plásmidos por cointegración, donde dos plásmidos diferentes se fusionan para formar un tercero autotransmisible. Luego la resolución del cointegrado puede ocurrir o no una vez incorporado a la célula receptora (ver sección I.2.2.1.1). La fusión de plásmidos generalmente está facilitada por la presencia de transposones o elementos de inserción.

Mientras que la transformación y la transducción dependen de la bacteria receptora para participar en el proceso vía recombinación homóloga, la conjugación depende de funciones específicas que median la transferencia y están presentes en la bacteria dadora, la cual identifica una célula receptora adecuada y transporta el ADN hacia ella en una forma activa que depende de la célula dadora (Frost & Koraimann, 2010). Dado que muchos plásmidos conjugativos exhiben un rango amplio de transferencia y replicación autónoma, se ha postulado a la conjugación plasmídica como el principal mecanismo contribuyente al intercambio horizontal de genes entre bacterias, jugando un rol esencial en la evolución bacteriana (Garcillan-Barcia, *et al.*, 2009, Harrison & Brockhurst, 2012) así como en la respuesta rápida para adaptarse a cambios en las condiciones ambientales (Heuer & Smalla, 2012).

I.2.2. Elementos genéticos móviles

Debido a la capacidad de moverse físicamente entre genomas huéspedes, los MGEs son elementos comunes presentes en las comunidades bacterianas. La adquisición de un MGE puede hacer que un microorganismo gane una nueva función que le permita adaptarse a cambios en el medio ambiente o bien adaptarse a un nuevo nicho. Se han descrito y caracterizado una gran cantidad de MGEs, entre ellos los plásmidos, bacteriófagos, islas genómicas, ICEs, secuencias de inserción, transposones e integrones (Stokes & Gillings, 2011). Si bien estos elementos son denominados móviles por su capacidad de desplazarse dentro de un genoma (movilidad intracelular), no todos ellos tienen la capacidad de moverse intercelularmente de manera autónoma. Por ejemplo, los integrones, secuencias de inserción y transposones tienen

movilidad intracelular gracias a las integrasas o transposasas que codifican, pero sólo pueden moverse de una célula a otra a través de su inserción dentro de otro MGE que sí tenga la capacidad de hacerlo, como los plásmidos conjugativos, fagos e ICEs.

Dentro de los MGEs más eficientes en cuanto a su contenido, promiscuidad y capacidad de moverse y propagarse entre distintas bacterias se encuentran los plásmidos, a quienes prestaremos particular atención en el presente trabajo. En las siguientes secciones de este capítulo se presentan en forma resumida los aspectos más relevantes de la biología, el rol y el estudio de los plásmidos, con el fin de sentar el marco teórico para el desarrollo de este trabajo de Tesis Doctoral.

I.2.2.1. Plásmidos

Los plásmidos son moléculas de ADN doble hebra que se replican de manera independiente del cromosoma bacteriano (Novick, 1987). Si bien pueden ser lineales, los más abundantes son de naturaleza circular. Están presentes en todo tipo de bacterias y puede existir más de un tipo de plásmido dentro de una célula bacteriana, o bien muchas copias de un mismo plásmido. En cuanto a su tamaño, éste puede ser muy variable, rondando entre unos pocos miles y cientos de miles de pares de bases (Smillie, *et al.*, 2010). En general, los plásmidos de mayor tamaño se encuentran en una o unas pocas copias, mientras que los más pequeños pueden estar en cientos de copias por célula. Si bien no son necesariamente esenciales para la supervivencia de los microorganismos, éstos tienen un papel clave en la adaptación y evolución bacteriana.

I.2.2.1.1. Clasificación de los plásmidos

Los plásmidos pueden ser muy diversos en cuanto a tamaño, estructura y contenido genético. Desde su descubrimiento en la década de los '50 (Lederberg, 1952), los plásmidos han sido ampliamente estudiados y, con la aparición de nuevas técnicas microbiológicas y moleculares, se han establecido distintas clasificaciones. Una de ellas, y quizás la más utilizada, es de acuerdo a su movilidad. Como se mencionó previamente, los plásmidos se transmiten verticalmente (segregación a las

células hijas) y horizontalmente (transferencia a otra célula receptora) mediante conjugación bacteriana.

De acuerdo a la movilidad, los plásmidos pueden clasificarse en tres grandes grupos: conjugativos, movilizables y no transmisibles (Smillie, *et al.*, 2010). Los plásmidos **conjugativos** son aquellos que poseen toda la maquinaria necesaria para moverse por sí mismos a otra célula receptora. Esto implica la transferencia del ADN como una partícula nucleoproteica compuesta por una proteína unida a el extremo 5' del ADN monocatenario a partir de un donante unido físicamente al destinatario a través de un poro del llamado aparato de conjugación (Alvarez-Martinez & Christie, 2009). Los componentes importantes de la transferencia conjugativa son: un origen de transferencia (*oriT*, que es el único sitio requerido en *cis* para la transferencia del ADN), una relaxasa encargada de reconocer el *oriT* y hacer un corte en una de las dos hebras de ADN plasmídico para iniciar su transferencia, y todos los genes necesarios para la formación del aparato conjugativo o MPF (por sus siglas en inglés: **Mating Pair Formation**) (Figura 1.2.A). El aparato conjugativo es una estructura compleja, en la que sus proteínas componentes atraviesan la membrana interna y externa de la bacteria donadora formando un canal a través del cual el ADN intermediario es transferido. Se lo considera perteneciente a la familia de Sistemas de Secreción Tipo IV (T4SS), los cuales transportan proteínas monoméricas, multiméricas y complejos ADN-proteína, desde el interior de una bacteria hacia otra célula bacteriana o eucariota (Cascales & Christie, 2003, Lawley, *et al.*, 2003). Los genes para la transferencia y la replicación, la formación del par de acoplamiento y la proteína de acoplamiento, generalmente están organizados en operones (Smillie, *et al.*, 2010).

Los plásmidos **movilizables** sólo contienen el *oriT* y una relaxasa, por lo cual requieren que otro plásmido (o ICE) aporte los genes del MPF para poder ser movilizables a otra célula (Figura 1.2.B). Por último, los plásmidos **no transmisibles** no contienen las funciones mencionadas anteriormente y por lo tanto no pueden moverse, salvo que lo hagan por cointegración o sean captados por la célula por el mecanismo de transformación (Figura 1.2.C).

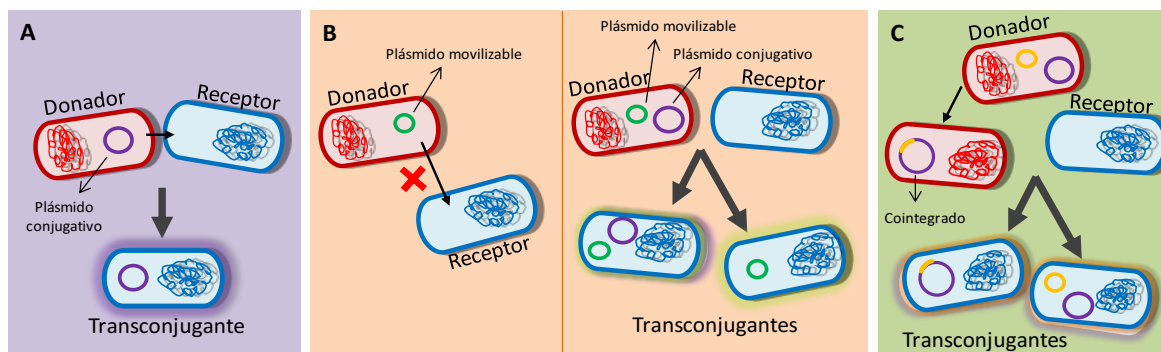


Figura I.2. Mecanismos de transferencia de plásmidos. **A)** Plásmido conjugativo: tiene los genes necesarios para movilizarse por sí mismo. **B)** Plásmido movilizable: necesita de los genes MPF aportados por un plásmido conjugativo (o un ICE) para moverse a otra célula. **C)** Plásmido no movilizable: requiere de su fusión con un plásmido conjugativo para moverse.

Otra clasificación ampliamente utilizada tiene en cuenta la incompatibilidad plasmídica. Este término fue introducido por Novick y colaboradores (Novick, 1987), y hace referencia a la incapacidad de dos plásmidos de coexistir en la misma célula y ser mantenidos de manera estable en ausencia de una selección externa. Este fenómeno de incompatibilidad ocurre cuando dos plásmidos tienen el mismo sistema de replicación y/o estabilidad y partición, y son incapaces de coexistir en la misma célula cuando sus funciones de replicación y partición interfieren. Se han descrito una gran cantidad de grupos de incompatibilidad plasmídica (Dennis, 2005, Fernandez-Lopez, *et al.*, 2006, Phan & Wain, 2008, Loftie-Eaton & Rawlings, 2012, Harmer & Hall, 2015), y su número crece rápidamente debido a la gran diversidad que presentan las proteínas involucradas en replicación, partición y estabilidad. Estos plásmidos pueden ser de amplio o estrecho rango de huésped, es decir, pueden replicar y mantenerse en una amplia variedad de especies bacterianas o sólo replicar en una o unas pocas especies, respectivamente.

Otras clasificaciones agrupan a los plásmidos según sus conformaciones (lineales o circulares), su número de copias (bajo, medio y alto) variando entre una y cientos de copias por célula; según su tamaño (plásmidos pequeños, medianos y megaplásmidos; pueden variar entre poco más de 1 Kb hasta el orden de las Mb), o bien se pueden nombrar según el tipo de genes o funciones que transportan (por

ejemplo, de resistencia a antibióticos o metales pesados y plásmidos catabólicos), según su rango de hospedadores, entre otras.

I.2.2.1.2. Importancia de los plásmidos

Los plásmidos están presentes en una gran cantidad de genomas bacterianos y, si bien no son esenciales para las bacterias, pueden aportarles genes que confieren ventajas adaptativas para establecerse en un determinado nicho, como por ejemplo colonizar un patógeno, establecer una interacción simbiótica con un huésped, degradar un antibiótico o metabolizar una fuente de carbono determinada. Así como portar uno o más plásmidos puede tener un efecto en la bacteria huésped, también pueden transportar genes que produzcan algún efecto en microorganismos vecinos, ya sea en una forma beneficiosa (como por ejemplo a través de la producción de enzimas que degraden antibióticos nocivos para otros microorganismos del mismo nicho) o de manera negativa (como por ejemplo si portan genes que codifiquen para bacteriocinas o para la producción de compuestos de bajo peso molecular con actividad antibiótica que dañen a las células vecinas) (Rankin, *et al.*, 2011). Si bien los plásmidos han resultado de particular interés desde su hallazgo, el conocimiento de los genes que transportan y su rol en la evolución y adaptación bacteriana los ha convertido en elementos ampliamente estudiados en los últimos años.

Desde un punto de vista clínico, los plásmidos pueden constituir un riesgo muy importante para la salud. Esto se debe a que muchos de ellos son portadores de genes de resistencia a antibióticos (ARGs: *Antibiotic Resistance Genes*), lo cual puede dificultar la eliminación de bacterias patogénicas intrahospitalarias que adquieran estos elementos. Además, los plásmidos pueden contener otros MGEs en sus esqueletos tales como transposones, secuencias de inserción e integrones, los cuales pueden acarrear genes de resistencia a compuestos antimicrobianos, compuestos cuaternarios de amonio y metales pesados, entre otros, potenciando aún más la diseminación de estos genes de resistencia. Estos MGEs pueden insertarse en los esqueletos de los plásmidos, ser movilizados a otra célula, y luego insertarse dentro del cromosoma bacteriano o bien continuar en su localización plasmídica. Las infecciones bacterianas constituyen la principal causa de muerte en el mundo y en

parte se debe a la evolución y adquisición de genes de resistencia a antibióticos (Schmieder & Edwards, 2012). A pesar del enorme problema que provoca la resistencia a antibióticos en bacterias patogénicas, poco se sabe de la diversidad, distribución y origen de los genes de resistencia, debido principalmente a la incapacidad de cultivar la mayoría de las bacterias ambientales. Se cree que el suelo es el ecosistema a partir del cual la síntesis de antibióticos ha evolucionado originalmente (D'Costa, *et al.*, 2007). De hecho, hoy en día más del 80% de los antibióticos utilizados en la clínica tienen su origen en bacterias de suelo, ya sea como productos naturales o derivados semisintéticos (Schmieder & Edwards, 2012).

Desde un punto de vista ambiental, los plásmidos resultan de gran relevancia. Han sido descritos numerosos plásmidos que poseen genes involucrados en la degradación de compuestos tóxicos como tolueno (Greated, *et al.*, 2002, Nojiri, *et al.*, 2004), naftaleno (Yen & Serdar, 1988, Li, *et al.*, 2004), xileno (Greated, *et al.*, 2002) e incluso pesticidas y herbicidas como atrazina (Martínez, *et al.*, 2001), 4-clorobenzoato (Layton, *et al.*, 1992), ácido 2,4-diclorofenoxiacético (2,4-D) (Trefault, *et al.*, 2004), entre muchos otros. Estos plásmidos poseen funciones accesorias y participan en la remoción de estos compuestos que resultan dañinos para los seres vivos, por lo cual juegan un rol crucial en la detoxificación ambiental. Se han encontrado plásmidos catabólicos que generalmente son conjugativos y de gran tamaño en una amplia diversidad de ambientes (Segura, *et al.*, 2014). La existencia de plásmidos catabólicos provee una poderosa batería de herramientas biotecnológicas para la remediación de ambientes altamente contaminados. Entre las dos estrategias de biorremediación más desarrolladas, se encuentran el bioaumentación de la cantidad de bacterias que portan un determinado plásmido catabólico de interés y el bioaumentación de los genes catabólicos de interés, esto es, propagar el plásmido de interés a distintos microorganismos, otorgándoles la capacidad de degradar un determinado compuesto.

Además de la importancia clínica y ambiental de los plásmidos, se ha descrito que éstos pueden portar genes que codifican enzimas que poseen actividades de interés industrial. Hoy en día los microorganismos y las enzimas por ellos producidas son utilizados como catalizadores en un vasto repertorio de procesos industriales,

permitiendo eliminar el uso de altas temperaturas, solventes orgánicos o pHs extremos, incrementando la especificidad de la reacción y pureza del producto y disminuyendo el impacto ambiental y/o los costos de producción (Cherry & Fidantsef, 2003). Dada su relevancia, el conocimiento del contenido génico de los plásmidos es importante para el entendimiento de su biología y contribuye al estudio del flujo de información entre distintos microorganismos, sus capacidades adaptativas y evolución en un determinado ambiente.

I.2.2.1.3. Estudio de los plásmidos ambientales: métodos y herramientas

Existe un gran interés en el estudio y conocimiento de los microorganismos que habitan distintos tipos de ambientes. Como ejemplo, podemos mencionar ambientes que poseen un cierto grado de contaminación, los cuales han sido elegidos como sitios blanco para la búsqueda de genes de resistencia a antibióticos (Smalla, *et al.*, 2000), metales pesados (Smit, *et al.*, 1998, Rajkumar, *et al.*, 2012) y degradación de compuestos orgánicos (Choi, *et al.*, 2009, Zhang, *et al.*, 2011), entre otros. Se ha demostrado que muchas de estas funciones están presentes en plásmidos y otros MGEs (Sharma & Thakur, 2009, Tabata, *et al.*, 2011, Anjum, *et al.*, 2012, Shin, *et al.*, 2012). Con el fin de aislar y caracterizar plásmidos de distintos ambientes, se han desarrollado distintas estrategias que permiten su estudio. A continuación se detallan algunos métodos que han sido utilizados con este fin (ver Figura I.3).

I.2.2.1.3.1. Aislamiento endógeno de plásmidos

Este método consiste en un aislamiento directo de la bacteria por plaqueo de la muestra en medio sólido conteniendo algún sustrato de interés, como antibióticos, metales o compuestos orgánicos, seguido de una extracción plasmídica y transformación del ADN plasmídico en una cepa huésped, utilizando estos sustratos para su selección (Hefford, *et al.*, 1997, Guerra, *et al.*, 2002) (ver Figura I.3.A). Este método tiene la desventaja de que requiere marcadores de selección en el plásmido y que su origen de replicación sea activo en la célula receptora, además de estar limitado a plásmidos presentes en microorganismos cultivables. Sin embargo, el aislamiento endógeno tiene la ventaja de que permite conocer la bacteria huésped de

la que proviene el plásmido.

I.2.2.1.3.2. Aislamiento exógeno de plásmidos

El aislamiento exógeno consiste en la captura de plásmidos por conjugaciones bi o triparentales usando una bacteria receptora; frecuentemente se han utilizado cepas de *E. coli* o *P. putida* como receptoras (Dahlberg, *et al.*, 1997, Smalla, *et al.*, 2000, Ono, *et al.*, 2007, Malik, *et al.*, 2008) (ver Figura I.3.B). Este método tiene la ventaja de que no requiere el cultivo previo de los microorganismos, por lo cual pueden aislarse plásmidos de bacterias no cultivables. Sin embargo, la captura exógena de plásmidos tiene varias limitaciones: en primer lugar, está restringida a plásmidos conjugativos o movilizables, mientras que aquellos no transmisibles no pueden ser capturados. En segundo lugar, requiere la presencia de marcadores de selección plasmídicos y un origen de replicación funcional en la bacteria receptora.

I.2.2.1.3.3. Nuevos métodos para el aislamiento de plásmidos

Recientemente, se han desarrollado nuevas estrategias basadas en metodologías metagenómicas. Una de las más llamativas por su diseño y eficiencia es la técnica basada en la captura de plásmidos asistida por un transposón (TRACA, por sus siglas en inglés: **TR**ansposon **A**ided **CA**pture) (Jones & Marchesi, 2007)(ver Figura I.3.C). Ésta es una técnica que emplea un transposón que contiene un marcador de selección y un origen de replicación, el cual en una reacción de transposición *in vitro* es insertado en los plásmidos que pretenden capturarse, seguido de una transformación bacteriana y posterior selección en un medio adecuado. Los plásmidos pueden ser purificados y secuenciados por la técnica de *primer walking*, que consiste en realizar el secuenciamiento a partir *primers* diseñados sobre el transposón, de secuencia conocida, e ir “caminando” hacia ambos lados del mismo hasta obtener la secuencia completa del plásmido. Este método permite obtener plásmidos sin necesidad de cultivar previamente los microorganismos y es independiente de las características de los plásmidos, es decir, no requiere que éstos posean un marcador de selección ni un origen de replicación compatibles con la bacteria receptora, ya que estas características son aportadas por el transposón utilizado. La principal desventaja de

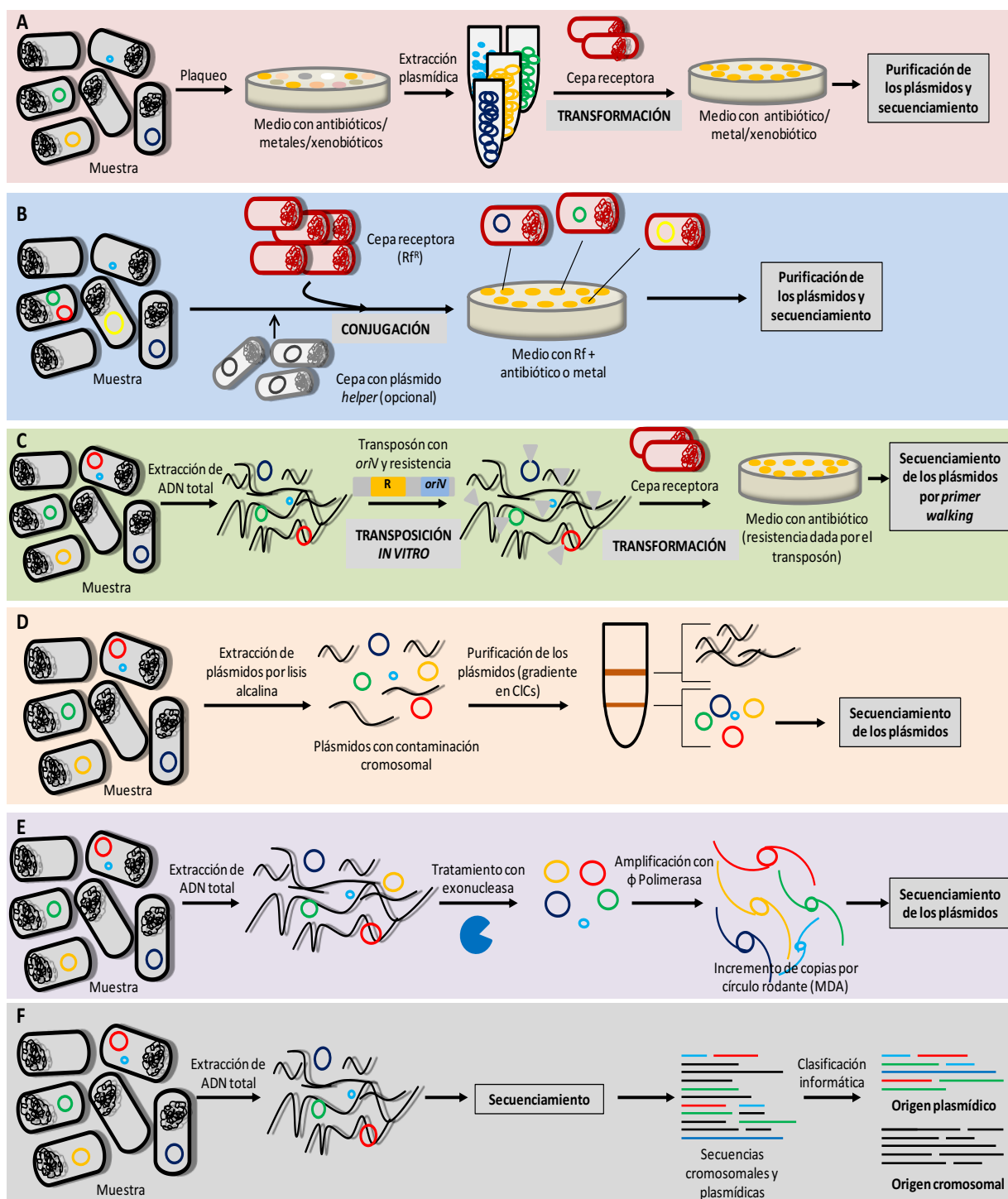


Figura I.3. Esquema general de los métodos más utilizados para la obtención y caracterización de plásmidos en muestras ambientales. Se detallan los aspectos sobresalientes de los siguientes métodos: **A)** Aislamiento endógeno de plásmidos (puede ser por conjugación bi o triparental) **B)** Aislamiento exógeno de plásmidos. **C)** Captura de plásmidos asistida por un transposón (TRACA). **D)** Extracción plasmídica (lisis alcalina u otro método) seguida de gradiente en cloruro de cesio. **E)** Amplificación de plásmidos por desplazamiento múltiple (MDA). **F)** Secuenciación de ADN total y clasificación informática (*in silico*) del ADN de origen plasmídico.

este método reside en el tamaño de los plásmidos que fueron recuperados experimentalmente, limitado principalmente por la dificultad de obtener plásmidos intactos de alto peso molecular y por el proceso de transformación bacteriana. Esta técnica ha permitido la obtención de plásmidos de hasta alrededor de 10 Kpb (Jones & Marchesi, 2007, Warburton, *et al.*, 2011, Zhang, *et al.*, 2011, Dib, *et al.*, 2015). Plásmidos de mayor tamaño, que son los de mayor relevancia ambiental, no han podido ser recuperados por esta técnica.

Otra metodología basada en la metagenómica consiste en el aislamiento del ADN plasmídico total de comunidades bacterianas sin previo cultivo, mediante el uso de técnicas tradicionales (como lisis alcalina o mediante la utilización de kits comerciales de extracción plasmídica) y posterior gradiente en CICs, seguido de secuenciamiento del material obtenido por el uso de tecnologías de alto rendimiento (Sentchilo, *et al.*, 2013) (ver Figura I.3.D). También se han empleado métodos que comprenden el aislamiento de ADN total de una comunidad bacteriana, seguido de la remoción de ADN lineal por el uso de exonucleasas y posterior aumento de la cantidad de ADN remanente (es decir, plásmidos circulares cerrados) mediante la técnica de amplificación por desplazamiento múltiple (MDA, por sus siglas en inglés ***Multiple Displacement Amplification***), permitiendo la recuperación de numerosos plásmidos (Brown Kav, *et al.*, 2013, Jorgensen, *et al.*, 2014, Norman, *et al.*, 2014)(ver Figura I.3.E). En los últimos años, se ha empleado una técnica que consiste en el aislamiento de ADN total de una comunidad seguido de un secuenciamiento de alto rendimiento y posterior clasificación bioinformática de genes o *contigs* (esto es, ensamblado de lecturas originadas por el secuenciamiento) de origen plasmídico (Palenik, *et al.*, 2009, Ma, *et al.*, 2012)(ver Figura I.3.F). En particular, los métodos basados en metagenómica tienen la gran ventaja de que permiten la obtención de un gran número de plásmidos a partir de una comunidad bacteriana, incluso a partir de aquellas bacterias que no pueden cultivarse en el laboratorio, pero su principal desventaja reside en la imposibilidad de recuperar físicamente dichos plásmidos en una célula (ya sea el huésped original u otra), impidiendo tanto su caracterización fenotípica como el conocimiento de la bacteria hospedadora. Si bien los métodos mencionados tienen sus

limitaciones, todos ellos han sido utilizados exitosamente en la obtención y caracterización de plásmidos. En la Tabla I.1 se resumen los métodos más utilizados para la obtención y caracterización de plásmidos en muestras ambientales con sus respectivas ventajas y desventajas.

I.3. Secuenciamiento y abordaje del estudio de genomas microbianos

I.3.1. Métodos de secuenciación

El ADN es la macromolécula donde se almacena la información hereditaria necesaria para el desarrollo y funcionamiento de los organismos vivos. En el año 1975, Frederick Sanger y sus colaboradores desarrollaron un método para determinar el orden de los nucleótidos en una molécula de ADN que revolucionó la biología (Sanger & Coulson, 1975) y posteriormente surgió la técnica de Maxam y Gilbert (Maxam & Gilbert, 1977). Estos dos métodos fueron denominados de **primera generación**. Si bien durante muchos años la técnica de Sanger fue ejecutada en forma manual, resultando muy engorrosa, no tardaron en aparecer los primeros equipos capaces de realizar este proceso en forma automatizada. Durante la década de 1990, se desarrollaron equipos cada vez más poderosos capaces de realizar múltiples reacciones de secuenciación a la vez y generar secuencias de hasta 1000 nucleótidos (también llamados pares de bases, pb) con una muy baja tasa de error. Fue mediante esta metodología, ya sea en forma manual o automatizada, que se obtuvieron las secuencias completas de, primero, algunos pocos genes, más tarde de virus completos y finalmente de genomas de diferentes organismos como bacterias y hasta incluso eucariotas.

El logro más espectacular alcanzado mediante esta técnica ha sido quizás la secuenciación del primer genoma humano, publicado en una versión preliminar en el año 2001 y en forma completa en el año 2003 (Green, *et al.*, 2015).

Si bien el método de Sanger (con algunas modificaciones) es aún ampliamente utilizado, para ciertos objetivos su uso se encuentra limitado principalmente por la longitud de las lecturas que pueden obtenerse, por su rendimiento y su alto costo.

Tabla I.1. Resumen de los métodos más utilizados para la obtención y caracterización de plásmidos en muestras ambientales

Método de aislamiento	Ventajas	Desventajas
Aislamiento endógeno	<ul style="list-style-type: none"> • Conocimiento de la bacteria huésped original • Aplicable a todos los plásmidos de bacterias cultivables (conjugativos, movilizables y no transmisibles) 	<ul style="list-style-type: none"> • Los plásmidos de bajo número de copias o poco abundantes pueden perderse • Plásmidos pequeños (limitado por el proceso de transformación de los plásmidos en la bacteria receptora) • Sólo para plásmidos de bacterias cultivables • Requiere de marcadores de selección codificados por los plásmidos • Requiere que los plásmidos tengan un origen de replicación compatible con la bacteria receptora
Aislamiento exógeno	<ul style="list-style-type: none"> • Independiente de cultivo • Se pueden aislar plásmidos independientemente de la dominancia numérica • Se pueden obtener plásmidos de gran tamaño • Aislamiento selectivo de plásmidos movilizables y conjugativos • Posibilidad de aislar plásmidos lineales si se usa un huésped indicado 	<ul style="list-style-type: none"> • Desconocimiento de la bacteria huésped original • Requiere de marcadores de selección y movilización/conjugación en los plásmidos (no permite obtener plásmidos no transmisibles)
TRACA	<ul style="list-style-type: none"> • Aplicable a todos los plásmidos bacterianos circulares (conjugativos, movilizables y no transmisibles) • Independiente de la dominancia numérica de los plásmidos • Independiente de las características del plásmido (marcador de selección, origen de replicación, funciones de movilización/ conjugación) • Independiente de cultivo • Caracterización de la secuencia de plásmido facilitada por la presencia de la secuencia conocida del transposón 	<ul style="list-style-type: none"> • El paso de transformación bacteriana limita el tamaño de los plásmidos a capturar • El transposón puede inactivar genes de interés • No permite capturar plásmidos lineales • Los transposones disponibles y los huéspedes bacterianos pueden limitar el rango de plásmidos y la caracterización funcional • Desconocimiento de la bacteria huésped original

Tabla I.1 (Continuación)

Método de aislamiento	Ventajas	Desventajas
Extracción plasmídica seguida de gradiente en CICs	<ul style="list-style-type: none"> • Aplicable a todos los plásmidos bacterianos circulares • Independiente de las características del plásmido (marcador de selección, origen de replicación, funciones de movilización/ conjugación) • Independiente de cultivo • Se pueden obtener plásmidos de gran tamaño 	<ul style="list-style-type: none"> • Desconocimiento de la bacteria huésped original • No permite capturar plásmidos lineales si se usan técnicas de aislamiento plasmídico tradicionales (gradiente en CICs, lisis alcalina)
MDA	<ul style="list-style-type: none"> • Aplicable a todos los plásmidos bacterianos circulares • Independiente de las características del plásmido (marcador de selección, origen de replicación, funciones de movilización/ conjugación) • Independiente de cultivo • Se pueden obtener plásmidos de gran tamaño (Norman <i>et al</i> 2014) 	<ul style="list-style-type: none"> • Desconocimiento de la bacteria huésped original • No permite capturar plásmidos lineales
Clasificación bioinformática	<ul style="list-style-type: none"> • Aplicable a todos los plásmidos bacterianos • Independiente de las características del plásmido (marcador de selección, origen de replicación, funciones de movilización/ conjugación) • Se pueden obtener plásmidos de gran tamaño • Independiente de cultivo 	<ul style="list-style-type: none"> • No se conoce la bacteria huésped original de los plásmidos • Pueden obtenerse secuencias parciales de los plásmidos, por la baja abundancia de los mismos en la muestra de ADN total de la comunidad • Puede perderse información, ya que la clasificación depende de la similitud de las secuencias con plásmidos previamente identificados • El ensamblado de los plásmidos puede ser muy difícil o imposible

La necesidad de acceder a una gran cantidad de secuencias y genomas impulsó el desarrollo de tecnologías que permiten la secuenciación, procesamiento y análisis de datos de manera fácil y ágil. En este sentido, en el año 2005 comenzaron a aparecer las primeras plataformas de secuenciamiento de alto rendimiento, también llamadas de **segunda generación**. La llegada de los sistemas de secuenciación por síntesis o

pirosecuenciación de alto rendimiento dieron solución a la mayoría de los problemas arriba mencionados para el análisis de la diversidad microbiana. Las tecnologías de secuenciación de segunda generación se volvieron un método simple, rápido y barato para el análisis de comunidades microbianas en comparación con los laboriosos métodos basados en bibliotecas de clones.

La primera tecnología de alto rendimiento en aparecer fue la plataforma **454 Life-Sciences** de Roche (Metzker, 2005) que emplea un método de pirosecuenciamiento. En esta metodología, se fragmentan bibliotecas de ADN de entre 400 y 800 pb, se ligan a adaptadores y se desnaturalizan las hebras, las cuales son capturadas por perlas, en donde se produce una amplificación por PCR en emulsión. Luego, las perlas se transfieren a placas en donde se agregan los dNTPs (de a uno a la vez, separados por lavados intermedios). Cada vez que se libera un pirofosfato al incorporarse un dNTP, éste impulsa una reacción que convierte luciferina en oxiluciferina, generando una señal lumínica que permite determinar qué nucleótido se ha incorporado. Por ejemplo, con esta tecnología pueden obtenerse hasta 1,3 millones de secuencias de un largo promedio de 700-800 pb en cada corrida de secuenciación (Loman, *et al.*, 2012), comparado con las 96 secuencias de una placa completa en el sistema de Sanger. Por otro lado, el método permite el etiquetado del ADN generando la posibilidad de secuenciar múltiples muestras a la vez y poder diferenciarlas.

Paralelamente, al mismo tiempo que surgían los pirosecuenciadores, se desarrollaron otras dos tecnologías capaces de generar volúmenes de información aún mayores a los del 454: el Solexa, comprado y comercializado luego por Illumina, y el Solid, comercializado por Life Sciences. La plataforma Genome Analyzer de **Illumina** fue liberada al mercado en el año 2006. Ésta utiliza un secuenciamiento por síntesis, luego de un primer paso de amplificación de la muestra de ADN. En esta técnica, primero se fragmenta el ADN de la muestra y se ligan adaptadores en los extremos del ADN fragmentado. Luego, el ADN se selecciona por tamaño, se desnaturalizan las hebras y se adhieren a una celda, en donde se produce una amplificación por PCR en puente para formar *clusters* con cientos de copias de cada fragmento de ADN. El ADN amplificado se convierte en simple hebra y es luego secuenciado. La técnica de

secuenciación emplea fluoróforos terminadores reversibles y permite generar un total de 25 millones – 6 billones de secuencias de un largo de 2x 150 – 2x 300 pb, dependiendo de la serie de Illumina utilizada (<http://www.illumina.com/systems/sequencing>) (ver Figura I.4). Posteriormente, se han desarrollado una gran cantidad de plataformas de secuenciamiento, como Ion Torrent y SOLiD, entre otras, así como variantes de las técnicas previamente mencionadas (Morey, *et al.*, 2013). El Solid, comercializado a partir del 2007, se basó en la tecnología de ligación de oligonucleótidos y permite generar un promedio de 3 mil millones de pares de bases en secuencias de 75 + 35 pb (Loman, *et al.*, 2012).

Todas estas plataformas difieren entre sí en ciertos detalles pero en general siguen pasos similares: preparación de la muestra, amplificación clonal y posterior secuenciamiento masivo en paralelo en rondas cíclicas (Reuter, *et al.*, 2015). Los principales factores a tener en cuenta a la hora de la elección de una plataforma de secuenciamiento son: la longitud de las lecturas generadas, el rendimiento, la precisión y el número de veces que cada lectura es secuenciada independientemente, además del costo por base. La estrategia específica empleada por cada plataforma determina su calidad, cantidad y sesgo de los datos de secuencia resultante y la utilidad de la plataforma para aplicaciones particulares.

La principal limitante de las tecnologías de segunda generación es que la longitud de las lecturas no es tan larga como con las técnicas anteriores, principalmente debido a la disminución progresiva de la eficiencia en el transcurso de la reacción de secuenciación. Las lecturas cortas son un problema a la hora de realizar el alineamiento y ensamblado de las secuencias. Por otro lado, el uso de PCR para la amplificación del ADN puede potencialmente introducir un sesgo (Morey, *et al.*, 2013).

La ambición por abaratar aún más los costos de secuenciación y aumentar la fiabilidad de las secuencias resultantes ha llevado a los denominados secuenciadores de **tercera generación**, basados en la secuenciación de una única molécula de ADN (*single molecule real time sequencing*). El primer secuenciador de este tipo lo comercializa Helicos BioSciences y se basa en la secuenciación a tiempo real de miles

de millones de pequeñas moléculas únicas de ADN adheridas a una superficie sólida. Permite generar de forma fiable fragmentos de entre 25 y 45 bases. Dada la pequeñez de las lecturas generadas, esta tecnología está recomendada para la resecuenciación de genomas y no para la secuenciación *de novo*. En un paso más adelante se sitúa la tecnología desarrollada por la empresa Pacific Biosciences que consigue lecturas (*reads*) de hasta 20.000 pb de longitud con un tamaño medio de entre 4.200 y 8.500 pb, que resuelve los problemas asociados a la segunda generación de secuenciadores (esto es: regiones homopoliméricas, repeticiones en tándem, lecturas cortas). Se trata de un enfoque completamente diferente, ya que lo que se ancla a una superficie sólida (nanoporo) es la enzima ADN polimerasa. Otra tecnología, encuadrada en los secuenciadores de tercera generación, es la desarrollada por ZS Genetics, que utiliza la microscopía electrónica y permite leer la secuencia del ADN directamente sobre una

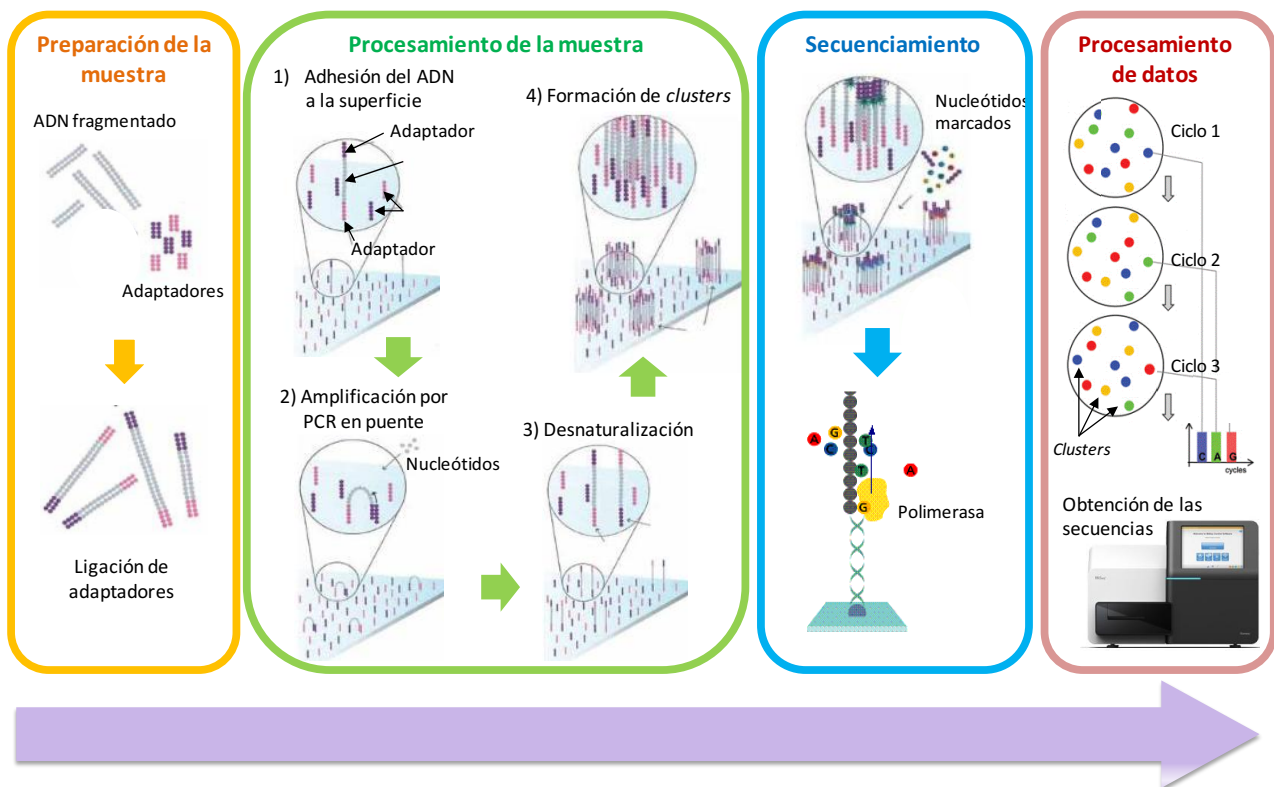


Figura 1.4. Esquema general del secuenciamiento de alto rendimiento utilizando la plataforma Illumina. Consiste en 4 pasos sucesivos que incluyen la preparación de la muestra (fragmentación y ligación de adaptadores), procesamiento (aumento de la cantidad de ADN por PCR en puente), secuenciamiento (mediante el uso de una polimerasa y nucleótidos marcados diferencialmente), y procesamiento de la información obtenida (lectura y posterior ensamblado).

imagen electrónica. La lectura de la secuencia requiere de la replicación previa de una hebra molde de ADN para poder marcarla con bases modificadas con iodo, bromo o triclorometilo antes de analizarlas.

Estos métodos tienen la capacidad de secuenciar moléculas únicas de ADN sin necesidad de realizar una amplificación clonal antes de la secuenciación, lo que evita la introducción de errores que se puede dar por la reacción de PCR y requiere menos manipulación de la muestra en comparación con los métodos de segunda generación. Estas tecnologías por lo general también implican la secuenciación por síntesis química, pero las técnicas de detección se basan principalmente en el reconocimiento físico de las bases de ADN en una cadena de ADN no modificado, en lugar de en la detección de incorporación química. En estas nuevas tecnologías, la reacción de secuenciación no se detiene por lavado y escaneo luego de la incorporación de cada base, por lo cual el tiempo y el consumo de reactivo disminuyen.

I.3.2. Secuenciamiento de alto rendimiento y metagenómica: aplicación al estudio de los plásmidos

El advenimiento de las nuevas tecnologías de secuenciación masiva ha permitido secuenciar grandes cantidades de material genético, abarcando desde microorganismos individuales a comunidades enteras. De la mano de estas técnicas de alto rendimiento, junto con la rápida evolución de las herramientas bioinformáticas, ha surgido la metagenómica. El término **metagenómica** fue mencionado por primera vez por Handelsman y colaboradores en 1998 (Handelsman, *et al.*, 1998), haciendo referencia al acceso de genomas de organismos de suelo que no podían ser cultivados, por aislamiento de su ADN y clonado en microorganismos cultivables. Hoy en día, este término se ha expandido y esta metodología permite estudiar un conjunto de organismos mediante la secuenciación de todos sus genomas a la vez. Suele ser utilizada para la caracterización de comunidades microbianas en muestras ambientales a partir de ADN total obtenido de las mismas. Este tipo de herramienta es una alternativa a otras metodologías empleadas en microbiología para analizar microorganismos (Chistoserdova, 2014) y puede ser utilizada para conocer la identidad

taxonómica de las especies así como su capacidad metabólica. Gracias a los relativamente bajos costos del secuenciamiento de alto rendimiento y el rápido procesamiento y análisis de la información obtenida, se han podido estudiar numerosas comunidades microbianas. Entre los ambientes explorados mediante el uso de distintas aproximaciones metagenómicas, podemos mencionar comunidades microbianas de suelo (Delmont, *et al.*, 2012, Frisli, *et al.*, 2013, Howe, *et al.*, 2014), el estudio de la biodiversidad microbiana oceánica (Wilkins, *et al.*, 2013, Wu, *et al.*, 2013, Kodzius & Gojobori, 2015, Kopf, *et al.*, 2015), metagenoma de ambientes contaminados como plantas de tratamiento efluentes (Ju, *et al.*, 2014, Yang, *et al.*, 2014), ambientes extremos (Cowan, *et al.*, 2015), microbiota de distintos alimentos (Kergourlay, *et al.*, 2015), microbiota del intestino humano (Jones & Marchesi, 2007, Ericsson & Franklin, 2015, Wang, *et al.*, 2015) y rumen de vaca (Ross, *et al.*, 2013), e incluso se han realizado estudios metagenómicos de la microbiota asociada a diversos tejidos y tipos de cáncer (Banerjee, *et al.*, 2015) e inflamaciones intestinales (Serban, 2015). En definitiva, la metagenómica y los métodos de secuenciación masiva de alto rendimiento han permitido conocer y caracterizar comunidades microbianas complejas de manera muy eficiente.

Con la aparición de la metagenómica, los científicos no tardaron en pensar en la potencial aplicación de estas metodologías al estudio de elementos puntuales, como por ejemplo los virus y los plásmidos. Li y colaboradores (Li, *et al.*, 2012) implementaron el término **metamovilómica** para referirse al “metagenoma de los elementos genéticos circulares dentro de una comunidad”. En los últimos años se han realizado numerosos trabajos destinados a estudiar los plásmidos presentes en comunidades bacterianas enteras.

I.3.2.1. Algunos ejemplos de estudios de plásmidos ambientales mediante el uso de secuenciamiento masivo

El estudio de los MGEs mediante el uso de herramientas metagenómicas ha tenido un crecimiento exponencial en la última década. Las metodologías mencionadas en la sección I.3.1 han permitido obtener exitosamente el *pool* de plásmidos de

Tabla I.2. Algunos ejemplos de estudios de plásmidos de comunidades

Método	Muestra/ microorganismos	Método de secuenciamiento	Número de plásmidos ensamblados	Tamaño de los plásmidos aislados	Fenotipos conferidos	Referencia
TRACA	Placa dental humana	<i>Primer walking</i>	32	< 8 Kb	Replicación, movilización, toxina/ antitoxina, integrasa	(Warburton, <i>et al.</i> , 2011)
TRACA	Planta de tratamiento de efluentes	<i>Primer walking</i>	2	~ 3 Kb	Resistencia a antibióticos, transposones	(Zhang, <i>et al.</i> , 2011)
TRACA	Intestino humano	<i>Primer walking</i>	4	3-10 Kb	Replicación, movilización, toxina/ antitoxina	(Jones, <i>et al.</i> , 2010)
Lisis alcalina y gradiente en CICs	Planta de tratamiento de efluentes	Sanger y Roche- 454	40	2-60 Kb	Replicación, transposición, resistencia a antibióticos y metales pesados	(Sentchilo, <i>et al.</i> , 2013)
Lisis alcalina y gradiente en CICs (sólo cultivables)	Planta de tratamiento de efluentes	Roche-454	-	-	Replicación, movilización, transposones, resistencia a antibióticos	(Szczepanowski , <i>et al.</i> , 2008)
Kit de extracción de plásmidos de alto peso molecular	Aislamientos de <i>E.</i> <i>coli</i> de infecciones de tracto urinario	Roche-454	22	1,5-150 Kb	Resistencia a antibióticos	(Brolund, <i>et al.</i> , 2013)
MDA	Intestino de rata	Illumina Hyseq y Roche-454	>160	< 10 Kb	Replicación, movilidad, toxina/ antitoxina, integrasa, transposasa	(Jorgensen, <i>et al.</i> , 2014)

distintas comunidades microbianas permitiendo conocer las funciones que transportan estos elementos móviles, lo cual es importante para comprender la biología de los mismos y la dinámica y flujo de la información genética en tales poblaciones. El estudio de los plásmidos a nivel metagenómico permitió la incorporación del término **plasmidoma**, el cual puede definirse como “todo el ADN plasmídico de un ambiente determinado” (Walker, 2012). En la tabla I.2 se muestran algunos ejemplos de estudios metagenómicos utilizados en el estudio de plásmidos de distintos ambientes. Si bien el plasmidoma hace referencia a todos los plásmidos de una comunidad, implicando así los plásmidos provenientes tanto de microorganismos cultivables como de los no cultivables, algunos de estos trabajos están limitados sólo al estudio de estos últimos (Szczepanowski, *et al.*, 2008, Bleicher, *et al.*, 2013, Brolund, *et al.*, 2013, Song, *et al.*, 2013). Aunque estos estudios pueden resultar incompletos o no representativos de una comunidad entera (y de hecho, lo son) todas las técnicas conocidas y empleadas, incluso aquellas independientes de cultivo, tienen sus limitaciones, impidiendo la obtención de la totalidad de los plásmidos de un ambiente.

I.4. METAEXPLORE: Consorcio interdisciplinario para el estudio de la biología/metagenómica de ambientes de interés

La metagenómica es un área que ha sufrido un crecimiento vertiginoso en el mundo desde el año 2004. Sin embargo, este crecimiento no fue parejo en todos los países del mundo. Esto se debe, en principio, a los elevados costos que tienen los equipos y los reactivos necesarios para secuenciar la cantidad de ADN que este tipo de estudios demanda y por otro lado, debido a la falta de recursos humanos formados para encarar este tipo de trabajos.

Este trabajo de Tesis Doctoral está enmarcado en el proyecto METAEXPLORE: *Metagenomics for bioexploration - tools and application*, del cual nuestro grupo de trabajo forma parte (junto a otros 18 grupos de investigación de distintos países de Europa). Este consorcio interdisciplinario tiene como objetivo principal el desarrollo y la aplicación de técnicas moleculares avanzadas para el clonado, secuenciamiento y

análisis de metagenomas de comunidades microbianas de distintos ambientes acuáticos y terrestres, haciendo hincapié en el moviloma de dichas comunidades, con el fin de buscar y caracterizar actividades enzimáticas, preferentemente relacionadas en la biodegradación de compuestos xenobióticos y recalcitrantes con potencial interés biotecnológico.

El trabajo se organizó en grupos en función de las áreas de especialidad de las personas que formaban parte del proyecto para poder encarar cada objetivo desde diferentes ángulos y en forma integral. Las secciones siguientes detallan el proceso que llevó a la consolidación de nuestra actividad dentro del proyecto.

I.4.1. Selección de la muestra de estudio

La etapa inicial del estudio tuvo como objetivo monitorear diversos hábitats a fin de analizar la abundancia de plásmidos para seleccionar aquéllos más prometedores para el análisis de plásmidos potencialmente diversos. En este sentido, se utilizó un enfoque independiente de cultivo en el que se empleó ADN total de cada comunidad. Se partió de la hipótesis que la frecuencia de ocurrencia de genes que codifican las actividades enzimáticas con potencial biotecnológico se incrementaría en el acervo genético de MGEs diversos y abundantes. En particular, ha sido descrito que los plásmidos que pertenecen a los grupos de incompatibilidad IncP-1, IncP-7 e IncP-9 suelen llevar genes responsables de la degradación de compuestos xenobióticos (artificiales) y de contaminantes de naturaleza orgánica, cumpliendo un rol esencial en la adaptación de las comunidades bacterianas a nuevos compuestos tóxicos liberados en el medio ambiente (Heuer & Smalla, 2012). Por lo tanto, se preseleccionaron diversos ambientes naturales o enriquecidos/contaminados, los cuales fueron analizados con un enfoque independiente de cultivo (por PCR seguido de *Southern Blot*) en relación a la prevalencia de plásmidos pertenecientes a los grupos IncP-1, IncP-7 e IncP-9.

Ninguna de las muestras utilizadas en el presente trabajo consistió en especies en peligro o protegidas. Se analizaron: esponjas marinas (*Halichondria panicea*) de Wageningen, Holanda, y se muestrearon las esponjas marinas *Corticium candelabro*

Petrosia ficiformis en España, los sedimentos de un río y el suelo originarios de Reino Unido. En Landsort (Suecia) se tomaron muestras de un sitio nacional de monitoreo ambiental (BY31) en conjunto con el programa de vigilancia del Mar Báltico. Las muestras de Asko fueron tomadas en el Centro de Investigaciones Marinas de Estocolmo. El sedimento y los suelos de Argentina se obtuvieron de lugares públicos contaminados con petróleo o con residuos industriales (entre ellos Río de La Plata y arroyo El Gato, Prov. de Buenos Aires). Las muestras de los sistemas biopurificación se obtuvieron de tierras privadas con el permiso de los agricultores locales en Kortrijk, Leefdaal, Lierde y Koksijde en Bélgica (Dealtry, *et al.*, 2014). Cada vez que fue necesario se obtuvieron los permisos legales de las autoridades competentes en cada país.

Tanto las señales de hibridación más fuertes para los distintos grupos de incompatibilidad plasmídica como los valores mayores en la cuantificación del ARNr 16S fueron observados empleando el ADN total de la comunidad del biofiltro de Kortrijk, indicando una alta abundancia de poblaciones bacterianas que transportan plásmidos de los diversos grupos de incompatibilidad analizados.

En base a los resultados obtenidos, y debido a la abundancia y diversidad de elementos genéticos móviles, principalmente plásmidos, se escogió un sistema de biopurificación (BPS, por sus siglas en inglés: *BioPurification System*) ubicado en Kortrijk, Bélgica, que es el sistema en el que nos centraremos para la realización de los objetivos planteados en el presente trabajo.

I.4.2. Sistemas de biopurificación

La contaminación ambiental es un problema creciente que genera preocupación a nivel mundial. La gran variedad de compuestos creados por el hombre (antropogénicos) que son empleados con distintos fines es una de las principales causas de tal polución. En particular, los pesticidas utilizados en campos y granjas para el control de plagas son uno de los factores más importantes que contribuyen a ello, y la contaminación de fuentes naturales de agua por el desecho puntual de pesticidas constituye un serio problema (Carter, 2000). Para abordar esta problemática, se han desarrollado distintos tipos de estrategias; por un lado, se ha intentado mejorar las

prácticas de manejo durante y después del uso de los pesticidas; por otro lado, se han implementado sistemas avanzados de decontaminación utilizando métodos físicos y químicos (Karanasios, *et al.*, 2012). Sin embargo, la mayoría de estas técnicas presentan altos costos o requerimientos tecnológicos difíciles de emplear en granjas, y muchas veces resultan ineficientes. Una estrategia alternativa ha sido la implementación de sistemas de biopurificación o biofiltros, los cuales colectan y decontaminan las aguas residuales provenientes del lavado de equipos de pulverización, derrames o fugas durante la carga, mezcla o vaciado de pesticidas (Torstensson & Castillo, 1997).

Los BPSs son excavaciones o contenedores que contienen una matriz biológicamente activa, llamada biomezcla, en los cuales la remoción de pesticidas ocurre por absorción y biodegradación (Sniegowski & Springael, 2015). Son de composición variable, pero generalmente contienen una mezcla de suelo, materiales lignocelulíticos y otros sustratos orgánicos mezclados en distintas proporciones (Castillo Mdel, *et al.*, 2008). Esta matriz retiene los compuestos tóxicos y potencia su degradación por acción de los microorganismos allí presentes. Las grandes ventajas de esta tecnología residen en los bajos requerimientos de labor y tiempo, bajos costos y la adaptabilidad a distintos climas. El primer BPS fue introducido por Torstensson y Castillo (Torstensson & Castillo, 1997), y desde allí se ha implementado este sistema en distintas partes del mundo. Hoy se sabe que en más de 25 países esta técnica ha sido empleada eficientemente, tanto de manera experimental como a gran escala en campos agrícolas (Karanasios, *et al.*, 2012).

La comunidad bacteriana es el factor clave en el control de la capacidad de depuración, y el conocimiento de las dinámicas microbianas dentro de estos sistemas permite su optimización. Si bien se han realizado numerosos estudios sobre estos sistemas, poco se conoce aún de la composición microbiana, su dinámica y metabolismo.

I.4.2.1. Breve descripción del BPS en estudio

Como se mencionó en la sección I.4.1, el BPS utilizado en este trabajo se

seleccionó en base a un previo *screening* en el que se analizó la presencia de ciertos elementos genéticos móviles. Dealtry y colaboradores (Dealtry, *et al.*, 2014, Dealtry, *et al.*, 2014) demostraron la presencia de plásmidos de distintos grupos de incompatibilidad, integrones de clase 1 y clase 2 y algunos genes asociados a ellos, como resistencia a sulfonamidas, estreptomicina y a compuestos cuaternarios de amonio. Adicionalmente, Dunon y colaboradores (Dunon, *et al.*, 2013) detectaron la presencia de la secuencia de inserción *IS1071*, la cual se encuentra frecuentemente asociada a genes involucrados en la degradación de compuestos orgánicos xenobióticos.

El BPS de estudio consiste en un contenedor de dimensiones de 20 m de largo, 1,2 m de ancho y 1 m de alto (ver Figura I.5). Está compuesto principalmente por una mezcla de turba, suelo y paja y otros

sustratos como estiércol de vaca, virutas de coco y compost, entre otras cosas. En este biofiltro son desechados más de 30 pesticidas producto de las prácticas agrícolas, como el lavado de maquinaria utilizada en el esparcimiento de los pesticidas, derrames y desecho directo de estos compuestos. La tabla I.3 detalla la lista de pesticidas que los empleados agrícolas declaran haber adicionado al BPS seleccionado.

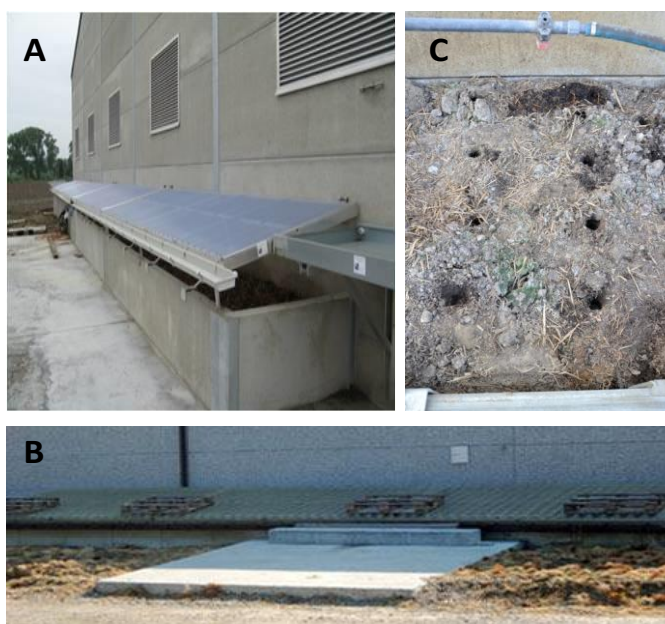


Figura I.5. Imagen del BPS del cual proviene la muestra utilizada en este estudio. A y B) Vistas exteriores del contenedor. C) Vista de la superficie del contenedor. Se observa la mezcla de suelo, turba y otros componentes.

Tabla I.3. Pesticidas adicionados al BPS en estudio

Nombre del Producto	Compuesto activo	Nombre del Producto	Compuesto activo
Aramo	Tepraloxidim	Starane	Fluoroxipir
Artist	Flufenacet Metribuzin	Stomp	Pendimetalina
Aspect T	Flufenacet Terbutilazina	CIPC	Clorprofam
Asulox	Asulam	Comet	Piraclostrobin
Atlantis	Yodosulfurón-metil-sodio Mefenpir-dietil Metsulfurón-metil	Defi	Prosulfocarb
Bacara	Diflufenican Flurtamone	Dianal	Fenmedifam
Beetix	Metamitrón	Diquanet	Dicuat
Bofix	Clopiralid Fluroxipir Ácido 2-Metil-4-clorofenoxiacético(MCPA)	Dual Gold	S-Metaloclor
Butizyl	Ácido 4-(4-Cloro-o-toliloxi)butírico (MCPB)	Equip	Foramsulfurón
Centium	Clomazón	Ethomat	Etofumesato
Challenge	Aclonifen	Frontier	Dimetenamida-P
Cyocel	Clormecuat	Goltix	Metamitrón
		Linuron	Linurón
		Metrigon	Clopiralid
		MCPA	MCPA
		Mikado	Sulcotriona
		Mitron	Metamitrón
		Opus	Epoconazol
		Pyramin	Cloridazón
		Primstar	Florasulam
		Primus	Florasulam
		Round-Up	Glifosato
		Samson	Nicosulfurón

Nota: los compuestos activos resaltados en negrita son aquellos cuya presencia fue detectada en el BPS por

I.5. Objetivos de este trabajo de Tesis

Los objetivos del presente trabajo han sido delineados para avanzar en el conocimiento de base de los plásmidos presentes en aislamientos recuperados de un biofiltro enriquecido con pesticidas, para conocer la relación de los mismos con los presentes en otras bacterias y otros ambientes, y para obtener nueva evidencia molecular que ayude a comprender de modo más acabado el peso de la transferencia horizontal de plásmidos en la dinámica de transferencia de ADN y en la adaptación/tolerancia de las bacterias presentes en dicho ambiente. Asimismo esperamos contribuir en el desarrollo de herramientas que permitan avanzar en la captura de plásmidos y estudio.

I.5.1. Objetivo general

Caracterización molecular y funcional de plásmidos aislados de un biofiltro utilizado para la decontaminación de pesticidas.

I.5.2. Objetivos específicos

En el marco del objetivo general expuesto, en la presente Tesis Doctoral se establecen los siguientes objetivos específicos:

- Obtención de una colección de bacterias portadoras de plásmidos de alto peso molecular. Caracterización de dichos aislamientos.
- Purificación y secuenciamiento del conjunto de plásmidos de la colección bacteriana. Análisis bioinformático del contenido plasmídico.
- Búsqueda, identificación, clonado, expresión y caracterización de actividades con potencial interés industrial.
- Diseño, construcción y validación de una nueva herramienta para la captura de plásmidos de muestras ambientales y cultivos puros.

CAPÍTULO II

MATERIALES Y MÉTODOS

II.1. Técnicas microbiológicas

II.1.1. Cepas bacterianas utilizadas en este trabajo

Las cepas bacterianas utilizadas en este trabajo y sus características más relevantes se muestran en la Tabla II.1.

Tabla II.1. Cepas utilizadas en este trabajo

Cepa	Características relevantes	Referencias
<i>Escherichia coli</i> CV601	<i>thr, leu, thi</i> , Rf ^R , GFP	(Smalla, <i>et al.</i> , 2006)
<i>Escherichia coli</i> CSH26	F- <i>ara</i> Δ(<i>lac-pro thi</i>), Rf ^R	(Miller, 1972)
<i>Escherichia coli</i> HB101	<i>pro, leu, thi, lacY, endA, recA, hsdR, hsdM</i> , Rf ^R , Sm ^R	(Boyer & Roulland-Dussoix, 1969)
<i>Escherichia coli</i> HB101 pRK600	<i>tra</i> ⁺ , Cm ^R , Rf ^R	(Finan, <i>et al.</i> , 1986)
<i>Escherichia coli</i> HB101 pRK2013	<i>tra</i> ⁺ , Km ^R , Rf ^R	(Figurski & Helinski, 1979)
<i>Ensifer meliloti</i> MVII-1	Sm ^R	(Hynes, <i>et al.</i>), 1986
<i>Escherichia coli</i> DH5α	<i>recA1, endA1, 80d, lacZ, dm15</i>	Bethesda Laboratory
<i>Escherichia coli</i> S17-1	294 <i>recA</i> , RP4 integrado cromosomalmente, Tc ^R	(Simon, <i>et al.</i> , 1983)
<i>Escherichia coli</i> BL21(DE3)Star	F- <i>ompT hsdSB (rB-mB-) gal dcm rne131</i> (DE3)	Invitrogen
<i>Agrobacterium tumefaciens</i> UBAPF2	Curada del plásmido pAtC58, Rf ^R	(Hynes, <i>et al.</i> , 1985)
<i>Proteus mirabilis</i> PR9	Aislamiento clínico portador de un integrón de clase 1 y clase 2	(Arduino, <i>et al.</i> , 2012)
<i>Escherichia coli</i> pLT06	Cepa portadora del plásmido pLT06, Cm ^R	(Thurlow, <i>et al.</i> , 2009)
<i>Enterobacter</i> sp. pAci39	Cepa portadora del plásmido pAci39, Ag ^R	Salto y Pistorio, comunicación personal
BF02 <i>Ochrobactrum</i> sp.	Aislamiento portador de plásmido/s de alto peso molecular obtenido del BPS de Kortrijk, Bélgica *	(Martini, <i>et al.</i> , 2015)
BF13 <i>Ochrobactrum</i> sp.	<i>Idem</i> *	Martini, <i>et al.</i> , 2015
BF14 <i>Cellulosimicrobium cellulans</i>	<i>Idem</i> *	Martini, <i>et al.</i> , 2015
BF15 <i>Ochrobactrum</i> sp.	<i>Idem</i> *	Martini, <i>et al.</i> , 2015

Tabla II.1. (continuación)

Cepa	Características relevantes	Referencias
BF19	<i>Variovorax</i> sp.	<i>Idem</i> *
BF21	<i>Paenibacillus</i> sp.	<i>Idem</i> *
BF22	<i>Paenibacillus tundra</i>	<i>Idem</i> *
BF25	<i>Pseudomonas putida</i>	<i>Idem</i> *
BF27	<i>Pseudomonas</i> sp.	<i>Idem</i> *
BF28	<i>Paenibacillus xylanexedens</i>	<i>Idem</i> *
BF30	<i>Bordetella</i> sp.	<i>Idem</i> *
BF31	<i>Cellulosimicrobium</i> sp.	<i>Idem</i> *
BF33	<i>Pseudomonas putida</i>	<i>Idem</i> *
BF36	<i>Microbacterium</i> sp.	<i>Idem</i> *
BF37	<i>Serratia</i> sp.	<i>Idem</i> *
BF42	<i>Alcaligenes</i> sp.	<i>Idem</i> *
BF43	<i>Serratia</i> sp.	<i>Idem</i> *
BF44	<i>Microbacterium oxydans</i>	<i>Idem</i> *
BF46	<i>Serratia</i> sp.	<i>Idem</i> *
BF48	<i>Microbacterium</i> sp.	<i>Idem</i> *
BF52	<i>Brevibacterium</i> sp.	<i>Idem</i> *
BF53	<i>Agromyces cerinus</i>	<i>Idem</i> *
BF55	<i>Pseudomonas putida</i>	<i>Idem</i> *
BF56	<i>Acinetobacter</i> sp.	<i>Idem</i> *
BF58	<i>Pseudomonas</i> sp.	<i>Idem</i> *
BF59	<i>Sphingobacterium</i> sp.	<i>Idem</i> *
BF60	<i>Sphingobacterium</i> sp.	<i>Idem</i> *
BF61	<i>Pseudomonas</i> sp.	<i>Idem</i> *
BF62	<i>Pseudomonas</i> sp.	<i>Idem</i> *
BF67	<i>Pseudomonas</i> sp.	<i>Idem</i> *
BF68	<i>Pseudomonas</i> sp.	<i>Idem</i> *
BF70	<i>Bacillus pumilus</i>	<i>Idem</i> *
BF71	<i>Bacillus</i> sp.	<i>Idem</i> *
BF73	<i>Bacillus megaterium</i>	<i>Idem</i> *
BF75	<i>Pseudomonas</i> sp.	<i>Idem</i> *

II.1.2. Plásmidos utilizados en este trabajo

Los plásmidos utilizados en este trabajo y sus características más importantes se muestran en la Tabla II.2.

Tabla II.2. Plásmidos utilizados en este trabajo

Plásmido	Características relevantes	Referencias
pBAM1	<i>tnpA</i> , <i>oriR6K</i> , <i>oriT</i> , Ap ^R , Km ^R	(Martinez-Garcia, <i>et al.</i> , 2011)
pET22b(+)	Vector de expresión, Ap ^R	Novagen
pET22b(+>:: <i>lac188</i>	Construcción que contiene el gen de la lacasa <i>lac188</i> , Ap ^R	Este trabajo
pK18mob	Vector de clonado, <i>mob</i> , Km ^R	(Schafer, <i>et al.</i> , 1994)
pG18mob	Vector de clonado, <i>mob</i> , Gm ^R	(Kirchner & Tauch, 2003)
pUC19mob	Vector de clonado de alto número de copias, <i>mob</i> , Ap ^R	New England Biolabs
pSM10	<i>mob</i> , Gm ^R	(Selbitschka, <i>et al.</i> , 1995)
pUC57	Vector de clonado, Ap ^R	GenScript
pLT06	Origen de replicación termosensible, <i>P-pheS</i> , <i>cat</i> , <i>lacZ</i> , Cm ^R	(Thurlow, <i>et al.</i> , 2009)
pAci39	Ag ^R	Salto y Pistorio, comunicación personal
RN3	Plásmido utilizado como control positivo para grupo de incompatibilidad IncN	(Gotz, <i>et al.</i> , 1996)
R388	Plásmido utilizado como control positivo para grupo de incompatibilidad IncW	(Llosa, <i>et al.</i> , 1991)
RSF1010	Plásmido utilizado como control positivo para grupo de incompatibilidad IncQ	(Nagahari & Sakaguchi, 1978)
pKJK5	Plásmido utilizado como control positivo para grupo de incompatibilidad IncP-1	(Bahl, <i>et al.</i> , 2007)
pCAR1	Plásmido utilizado como control positivo para grupo de incompatibilidad IncP-7	Izmalkova, <i>et al.</i> , 2005
pM3	Plásmido utilizado como control positivo para grupo de incompatibilidad IncP-9 α	(Dealtry, <i>et al.</i> , 2014) b
pBS2	Plásmido utilizado como control positivo para grupo de incompatibilidad IncP-9 β	Dealtry, <i>et al.</i> , 2014 b
pBS265	Plásmido utilizado como control positivo para grupo de incompatibilidad IncP-9 γ	Dealtry, <i>et al.</i> , 2014 b

Tabla II.2 (continuación)

Plásmido	Características relevantes	Referencias
pSN11	Plásmido utilizado como control positivo para grupo de incompatibilidad IncP-9 δ	Dealtry , <i>et al.</i> , 2014 b
pMG18	Plásmido utilizado como control positivo para grupo de incompatibilidad IncP-9 ϵ	Dealtry , <i>et al.</i> , 2014 b
pNL60	Plásmido utilizado como control positivo para grupo de incompatibilidad IncP-9 ζ	Dealtry , <i>et al.</i> , 2014 b
pNL15	Plásmido utilizado como control positivo para grupo de incompatibilidad IncP-9 η	Dealtry , <i>et al.</i> , 2014 b
pSVS15	Plásmido utilizado como control positivo para grupo de incompatibilidad IncP-9 θ	Dealtry , <i>et al.</i> , 2014 b
pNL22	Plásmido utilizado como control positivo para grupo de incompatibilidad IncP-9 ι	Dealtry , <i>et al.</i> , 2014 b
pBAM1-oriV	Km ^R , Ap ^R . Construcción derivada del pBAM1 con el fragmento que contiene el <i>oriV</i> del pSM10	Este trabajo
pBAM1-oriV-Scel	Km ^R , Ap ^R . Construcción derivada del pBAM1 que contiene el <i>oriV</i> del pSM10 y el sitio de reconocimiento de la endonucleasa <i>Scel</i> en el sitio <i>Swal</i>	Este trabajo
pBAM1-oriV-Scel Sin oriR6K/oriT	Km ^R , Ap ^R . Construcción derivada del pBAM1-oriV- <i>Scel</i> ; se le eliminaron el <i>oriVR6K</i> y <i>oriT</i> originales del pBAM1	Este trabajo
p-RESC	Km ^R , Ap ^R . Construcción derivada del pBAM1-oriV- <i>Scel</i> , sin <i>oriR6K/oriT</i> con el <i>oriT</i> del pUC19mob. Utilizada para la captura de plásmidos.	Este trabajo

II.1.3. Medios de cultivo y conservación de microorganismos

II.1.3.1. Medios de cultivo sólidos y líquidos

Los medios de cultivo utilizados en este trabajo, así como su composición detallada, se indican en la Tabla II.3. Todos los medios ricos fueron preparados y luego esterilizados en autoclave a 120 °C por 20 minutos a 1 atm de sobrepresión. Los medios mínimos fueron preparados esterilizando previamente las soluciones *stock* de los micronutrientes (1000 X) por filtración y luego agregando las cantidades adecuadas

a una solución de ágar-agua previamente esterilizada en autoclave. La composición de los medios líquidos es similar a la composición de los sólidos, excepto que a los líquidos no se les adicionó 1,5% de ágar. Todos los medios fueron llevados a volumen con agua destilada.

Tabla II.3. Composición de los medios de cultivo utilizados

Medio	Composición (por litro)	Referencia
Medios de cultivo complejos		
LB (Luria-Bertani)	Glucosa	1 g
	NaCl	5 g
	Triptona	10 g
	Agar	15 g
		(Sambrook, <i>et al.</i> , 1989)
TY (Tryptone-Yeast)	CaCl ₂ .2H ₂ O	0,7 g
	Triptona	5 g
	Extracto de levadura	3 g
	Agar	15 g
		(Beringer, 1974)
EMB (Eosin Methylene Blue)	Peptona	10 g
	Lactosa	5 g
	Sacarosa	5 g
	Fosfato dipotásico	2 g
	Eosina	0,4 g
	Azul de metileno	0,065 g
	Agar	13,5 g
		(Levine, 1918)
Dex (Dextrose-agar)	Digerido enzimático de caseína	10 g
	Dextrosa	40 g
	Agar	15 g
		Sigma-Aldrich
MCK (Mac Conkey)	Peptona	17 g
	Pluripeptona	3 g
	Lactosa	10 g
	Mezcla de sales biliares	1,5 g
	NaCl	5 g
	Rojo neutro	0,03 g
	Cristal violeta	0,001 g
	Agar	13,5 g
		(Macconkey, 1905)
MH (Mueller-Hinton)	Infusión de carne	300 g
	Peptona ácida de caseína	17,5 g
	Almidón	1,5 g
	Agar	15 g
		Oxoid Ltd.

Tabla .3 (continuación)

Medios de cultivo mínimos

GS (Glutamato- sacarosa)	Sacarosa	29,2 mM	
	Glutamato sódico	2,67 mM	
	K ₂ HPO ₄ ·3H ₂ O	0,15 mM	
	KH ₂ PO ₄	0,15 mM	
	Na ₂ SO ₄	0,7 mM	
	MgSO ₄ ·7H ₂ O	1 mM	
	CaCl ₂ ·2H ₂ O	1 mM	
	H ₃ BO ₃	0,048 mM	
	MnSO ₄ ·H ₂ O	0,01 mM	
	ZnSO ₄ ·7H ₂ O	0,01 mM	(del Papa, <i>et al.</i> , 1999)
	CuSO ₄ ·5H ₂ O	0,048 mM	
	CoCl ₂ ·6H ₂ O	0,0005 mM	
	Na ₂ MoO ₄ ·2H ₂ O	0,001 mM	
	FeCl ₃ ·3H ₂ O	0,001 mM	
	Tiamina HCl	0,00295 mM	
	Ácido Pantoténico	0,0042 mM	
	Biotina	0,000082 mM	
PIPES	20 mM		
agar	15 g		
M9	Na ₂ HPO ₄	5,8 g	
	KH ₂ PO ₄	3 g	
	NaCl	0,5 g	(Miller, 1972)
	NH ₄ Cl	1 g	
	Glucosa	10 g	
	agar	15 g	

II.1.3.2. Medios de conservación bacterianos

Para conservar a largo plazo las bacterias, éstas se cultivaron previamente hasta fase logarítmica tardía en LB o el medio adecuado y se suplementaron con glicerol al 50% (v/v) y 25% (v/v) para su conservación a -20 °C y -80 °C respectivamente. Para uso rutinario, alícuotas de estas reservas se hicieron crecer en medio sólido con el/los antibióticos correspondientes.

II.1.4. Condiciones de cultivo

Los cultivos líquidos se incubaron en todos los casos en agitación a 180 rpm. Las cepas de *Escherichia coli* fueron incubadas a 37 °C, tanto en medio líquido como en

sólido. Todos los aislamientos obtenidos en este trabajo y las demás cepas utilizadas fueron incubados a 28 °C.

II.1.5. Antibióticos

Los antibióticos, cuando se requirieron, fueron adicionados a los medios de cultivo a partir de soluciones concentradas de los mismos preparados en agua destilada salvo para los casos de tetraciclina, eritromicina, cloranfenicol, cicloheximida y rifampicina, que en los cuatro primeros se utilizó etanol y en el último metanol como solvente. Las soluciones preparadas en agua se esterilizaron por filtración empleando membranas de nitrocelulosa de 0,2 µm de tamaño de poro. Las concentraciones utilizadas (excepto que específicamente se indique otra) fueron las siguientes: 120 µg/ml neomicina (Nm), 400 µg/ml estreptomina (Sm), 200 µg/ml ampicilina (Ap), 10 µg/ml tetraciclina (Tc), 25 µg/ml kanamicina (Km), 50 µg/ml gentamicina (Gm), 20 µg/ml trimetoprima (Tp), 200 µg/ml eritromicina (Er), 20 µg/ml cloranfenicol (Cm), 5 µg/ml ácido nalidíxico (Nx), 50 carbenicilina (Cb), 100 µg/ml rifampicina (Rf), 25 µg/ml zeocina (Zeo) o 0,5 µg/ml Ciprofloxacina (Cp).

II.1.6. Obtención de los aislamientos a partir de la muestra del biofiltro

El procesamiento de la muestra se llevó a cabo como se indica a continuación: 5 g de muestra proveniente del biofiltro (fecha de muestreo: Marzo de 2011) fueron resuspendidos en 50 ml de solución fisiológica estéril, incubados en agitación a 180 rpm a 28 °C por 1 h y luego se dejó decantar. Alícuotas del sobrenadante fueron plaqueadas en los 8 medios de cultivo sólidos indicados en la tabla II.3 (sección II.1.3.1), conteniendo distintos antibióticos (en las concentraciones indicadas en la sección II.1.5.). Además, todos los medios fueron suplementados con cicloheximida en una concentración de 200 µg/ml, para evitar el crecimiento de hongos. Las placas se incubaron a 28 °C durante 48 h o tiempos superiores.

II.1.7. Screening de aislamientos portadores de plásmidos

Para llevar a cabo este *screening*, se partió de cultivos líquidos de las colonias obtenidas por plaqueo en los distintos medios de cultivo sólidos mencionados anteriormente. Los cultivos fueron agitados a 180 rpm por 24 - 48 h a 28 °C y analizados

por la técnica de lisis *in situ* de Eckhardt modificada, como se indica en la sección II.2.4.3.

II.1.8. Determinación de la concentración mínima inhibitoria para antibióticos y metales

La determinación de la concentración mínima de antibióticos/metales que inhiben el crecimiento bacteriano (MIC: *minimum inhibitory concentration*) para los aislamientos, se realizó como se indica a continuación: se crecieron los cultivos en medio LB líquido a 28 °C a 180 rpm hasta alcanzar una $DO_{600\text{ nm}} = 0,5$. Luego se realizaron diluciones seriadas (1:10) y 10 μl conteniendo 10^6 bacterias/ml fueron inoculados en forma de gota en placas de Petri conteniendo medio LB suplementado con distintos antibióticos y metales, en distintas concentraciones. Las soluciones de metales fueron esterilizadas por filtración y adicionadas en concentraciones que variaron en el rango de 2,5 μM - 20 mM. Los metales ensayados fueron los siguientes: Ni^{+2} , Cd^{+2} , Cu^{+2} , Mn^{+2} , Co^{+2} , Zn^{+2} , y Hg^{+2} ($\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$, CdCl_2 , $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ y HgCl_2). Los antibióticos evaluados, así como el rango de concentraciones, se indican a continuación: Ap (6,25–1.600 $\mu\text{g/ml}$), Cm (0,625–160 $\mu\text{g/ml}$), Km (1,56–400 $\mu\text{g/ml}$), Nx (1,25–40 $\mu\text{g/ml}$), Nm (3,75–480 $\mu\text{g/ml}$), Rif (3,13–100 $\mu\text{g/ml}$), Sm (3,13–1.600 $\mu\text{g/ml}$), Gm (0,75–50 $\mu\text{g/ml}$), Tp (2,5–320 $\mu\text{g/ml}$), Cp (0,016–0,5 $\mu\text{g/ml}$) y Tc (0,313–40 $\mu\text{g/ml}$). Las placas fueron incubadas a 28 °C por 48 h.

II.1.9. Evaluación de movilidad plasmídica

II.1.9.1. Conjugaciones biparentales

Las conjugaciones fueron realizadas empleando la técnica de Simon y colaboradores (1983) con modificaciones menores según se indica a continuación. En un microtubo de polipropileno de 1,5 ml se mezclaron 0,7 ml de un cultivo de la cepa donadora ($DO_{600\text{ nm}} = 0,4 - 0,5$) con 0,7 ml de un cultivo de igual densidad óptica de la cepa receptora. La mezcla fue luego centrifugada a 640 *g* durante 8 minutos y el sedimento fue resuspendido suavemente en aproximadamente 50 μl de medio líquido LB. La suspensión fue colocada en una caja de Petri con medio LB (sin antibióticos) e incubada a 28 °C durante 24 hs. La mezcla de bacterias fue finalmente resuspendida en

medio LB líquido y sembrada en medio LB sólido suplementado con distintos antibióticos o metales, separadamente, para seleccionar los transconjugantes. En todas las conjugaciones se realizaron controles correspondientes a la bacteria dadora y la receptora.

II.1.9.2. Conjugaciones triparentales

Las conjugaciones triparentales fueron llevadas a cabo de manera similar a las biparentales, con la diferencia de que se utilizó 0,5 ml de los cultivos bacterianos de la cepa dadora y de la receptora y se agregó además 0,5 ml de cultivo en fase logarítmica de una cepa *helper* (de ayuda). Según el caso, se utilizó *E. coli* HB101 conteniendo el plásmido pRK600 o pRK2013, los cuales aportaron en *trans* los genes necesarios para la conjugación de plásmidos movilizables (no-conjugativos).

II.1.10. Preparación de células electrocompetentes

Se siguió la técnica de Tung y Chow (1995) con pequeñas modificaciones. Se inocularon 500 ml de medio LB sin NaCl con 1 ml de un cultivo de la cepa de *E. coli* de interés que crecida desde el día anterior. El cultivo fue incubado a 37 °C y agitado hasta que la $DO_{600\text{ nm}}$ llegó a 0,8 unidades. Las células fueron centrifugadas a 4.000 *g* por 15 minutos a 4 °C y se lavaron 2 veces con 200 ml de glicerol 10% a 4 °C. Luego, las células se resuspendieron suavemente en 0,5 ml de glicerol al 10% y se fraccionaron en alícuotas de 100 µl, las cuales fueron almacenadas a -80 °C.

II.1.11. Transformación bacteriana

Las electrotransformaciones fueron llevadas a cabo usando un equipo Gene Pulser (Bio -Rad) y en cubetas de 0,2 ml bajo las condiciones recomendadas por el fabricante (25 µF, 200Ω, 2,5 kV). Luego de la electrotransformación se adicionó inmediatamente a las células 1 ml de medio LB líquido, y la mezcla se incubó 1 h a 37 °C para permitir la expresión de la información contenida en el ADN plasmídico incorporado. Finalmente, las células fueron plaqueadas en medio LB sólido selectivo y se incubaron a 37 °C.

II.1.12. Recuento de bacterias en medios líquidos

El recuento del número de unidades formadoras de colonias (UFC) en un medio líquido se determinó por el método de recuento de gotas en placa (Herigstad *et al.*,2001). Para ello, se realizaron sucesivas diluciones del cultivo cuyo número de UFC se quiso determinar, en solución fisiológica estéril. Para cada dilución, se sembraron 6 gotas de 10 μ l cada una en medio LB sólido (conteniendo o no antibiótico, según el caso). Luego se incubó durante 24 h a 28 ó 37 °C (dependiendo del microorganismo) y pasado este tiempo se contaron las colonias en cada gota. En aquellas diluciones en las que se visualizaron entre 1 y 20 colonias, se realizó un promedio de las 6 gotas y se refirió al número de UFC/ml del cultivo original.

II.2. Técnicas moleculares

II.2.1. Obtención de ADN genómico para su uso en PCR

Para cada aislamiento, se partió de una colonia aislada y se resuspendió en 300 μ l de una solución de NaCl 1 M. Luego se incubó a 25 °C por 15 minutos, vorteadando en distintas oportunidades. Posteriormente se centrifugó 3 minutos a 17.000 *g* y los *pellets* se lavaron con 300 μ l de agua destilada estéril. Luego se resuspendieron los *pellets* en 150 μ l de resina Chelex (BioRad) al 6% p/v y se los incubó por 20 minutos a 56 °C y luego durante 8 minutos a 99 °C. Finalmente, se centrifugó a 17.000 *g* 3 minutos y se utilizó el sobrenadante como molde para las reacciones de PCR.

II.2.2. Preparación de ADN plasmídico

II.2.2.1. Obtención a pequeña escala (“miniprep”)

Los plásmidos de *E. coli* fueron preparados por la técnica de lisis alcalina descrita por Sambrook y colaboradores (1989), con la siguiente modificación: luego de realizar la lisis alcalina y la neutralización con acetato de potasio 3M (pH=5,4) como se describe en dicho protocolo, se recogió el sobrenadante y se le agregó la cantidad apropiada de la suspensión de sílica (1,33 g/ml SiO₂). Se estima una capacidad de unión de 200 ng de ADN/ μ l de suspensión. Se centrifugó 10 segundos a máxima velocidad y

se lavó tres veces con 500 µl de solución de lavado (NaCl 50 mM, Tris 10 mM pH=7,5, EDTA 2.5 µM, etanol 50%). Luego se hizo una centrifugación final a máxima velocidad y el *pellet* se secó al aire. La elución se realizó agregando el volumen deseado de TE o agua, incubando 5 minutos a 60 °C; luego se centrifugó y se recuperó el sobrenadante.

II.2.2.2. Obtención de ADN plasmídico a gran escala para secuenciamiento masivo

La preparación de plásmidos de alto peso molecular a gran escala se realizó utilizando dos metodologías distintas: utilizando el kit Large Construction (QIAGEN) o bien empleando la técnica descrita por Juanin *et al* (Jouanin, *et al.*, 1981) seguido de un gradiente en CsCl. En este último caso se partieron de 50 ml de cultivo bacteriano en LB líquido crecidos por 24 h a 28 °C en agitación, y se recogieron las bacterias centrifugando el cultivo a 5.000 *g* durante 5 minutos. Las células se trataron con Sarcosyl 0,1% en TE (Tris 50 mM, EDTA 20 mM, pH=8,0), se centrifugaron y se lavaron con el mismo tampón dos veces.

Las células resuspendidas en TE se lisaron con SDS a una concentración final de 1% y 5 mg/ml de pronasa a 37 °C durante 2 h. El lisado se llevó a pH=12,45 adicionando NaOH 3N de a poco y agitando suavemente. Una vez alcanzado el pH se dejó en reposo durante 20 - 30 minutos, para luego neutralizar con Tris 2 M, pH=7 hasta pH=8,5. La preparación se incubó en hielo 15 minutos y se añadió NaCl 5 M hasta una concentración final de 1 M. Después de 1 h a 0 – 4 °C se sedimentaron los restos de proteínas y ADN cromosómico centrifugando 30 minutos a 6.500 *g*. El sobrenadante se precipitó añadiendo polietilenglycol (PEG 6000) al 20% hasta una concentración final de 10% dejándolo durante toda la noche a 0-4 °C. El ADN plasmídico se recogió centrifugando a 2.500 *g* 10 minutos y el precipitado se resuspendió en TE. En este punto, la muestra se chequeó mediante un gel de agarosa al 0,8% p/v. A la solución resultante se le añadió 1 g de CsCl sólido por cada 1 ml de solución. Posteriormente se adicionó bromuro de etidio disuelto en TE (10 mg/ml) hasta llegar a una concentración final de 400 µg/ml. Se centrifugó 30 minutos para eliminar el exceso de PEG y la suspensión se transfirió a tubos apropiados (BECKMAN) para someterlos nuevamente a una centrifugación durante 72 h a 100.000 *g* a 20 °C.

La banda inferior, correspondiente al ADN plasmídico, fue recogida usando una aguja perforando por debajo de la banda. Se utilizó luz UV de 360 nm para mejorar la observación. La eliminación del bromuro de etidio se consiguió agregando un volumen igual de isopropanol saturado en agua. Se mezcló y se extrajo la fase superior. El proceso se repitió 4 veces hasta desaparición de todo el color rosado de la solución acuosa. A ésta se le agregaron 2 volúmenes de agua y se precipitó el ADN con 6 volúmenes de etanol a -20 °C durante una noche. Pasado este tiempo, se centrifugó a 15.000 g 20 minutos; se lavó el precipitado 2 veces con etanol al 70% y se disolvió en TE conservándose a -20 °C. El ADN plasmídico de los 35 aislamientos fue purificado por separado y luego se mezclaron alícuotas de cada uno de ellos para realizar el secuenciamiento de alto rendimiento.

II.2.2.3. Obtención de ADN plasmídico de alto peso molecular para la evaluación de los grupos de incompatibilidad

Los aislamientos fueron crecidos en placas de medio R2A (Sigma-Aldrich) por 48 h a 28 °C. Las colonias frescas fueron tratadas con 2 ml de MgSO₄ y centrifugadas a 3.000 g por 3 minutos. Los *pellets* fueron lavados 3 veces con MgSO₄ y luego el ADN plasmídico fue extraído según fue descrito por Smalla y colaboradores (2000). La calidad del ADN plasmídico fue chequeada por electroforesis en gel de agarosa al 0,8% p/v.

II.2.3. Reacción en cadena de la polimerasa (PCR)

Los ensayos de PCR fueron realizados empleando los termocicladores de tubos de pared delgada BioRad ICycler o BioRad Mycycler. Las concentraciones estándar que se utilizaron para la reacción de PCR son las siguientes: *buffer* de reacción 1X (KCl 50 mM, Tris-HCl 20 mM, pH=8,0), dNTPs en concentración 200 µM de cada uno, MgCl₂ 3 mM, 5 µM de cada *primer* y 1 U de Taq Polimerasa Pegasus (Productos Bio-Lógicos). Cuando se requirió amplificar fragmentos con alta fidelidad se utilizó Pfx polimerasa (Invitrogen). Para la amplificación de fragmentos de gran longitud se utilizó la Taq Polimerasa LongAmp (BioLabs). Cuando fue necesario, se variaron las concentraciones de MgCl₂ para aumentar la especificidad de la reacción.

Las condiciones de ciclado empleadas variaron de acuerdo a los *primers* utilizados y al tamaño del producto de amplificación esperado siempre siguiendo el protocolo estándar compuesto por: una etapa de desnaturalización inicial a 94 °C durante 4 minutos; 35 ciclos de: desnaturalización a 94 °C por 30 segundos, *annealing* a la temperatura adecuada por 30 segundos y elongación a 72 °C (tiempo variable, dependiendo de la longitud del fragmento a amplificar; se estimó 1 minuto cada 1.000 pb) y una etapa final de 5 minutos a 72 °C.

II.2.4. Electroforesis de ADN.

II.2.4.1. Separación e identificación de fragmentos de restricción o de amplificación por PCR

La verificación de la calidad de las preparaciones de ADN y la separación de fragmentos de restricción se llevaron a cabo mediante electroforesis horizontal en geles de agarosa de concentración entre 0,8% y 1,5% en *buffer* TBE (Tris 89 mM, EDTA Na₂ 2,5 mM, H₃BO₃ 89 mM, pH=8,2). Este mismo *buffer* se utilizó como solución de corrida. El voltaje de trabajo varió entre 50 y 120 volts/cm según los tamaños a separar. Como solución de siembra se utilizó sacarosa 40% y azul de bromofenol 0,25 % en una relación 5 µl solución de ADN en agua o TE / µl solución de siembra. El agregado de 0,5 µg/ml de bromuro de etidio a los geles de agarosa permitió la visualización de los fragmentos de ADN por transiluminación con luz ultravioleta de $\lambda = 260$ nm. El registro de los resultados se realizó mediante fotografía de los geles con una cámara digital bajo iluminación UV.

La estimación del tamaño de los fragmentos de ADN obtenidos por PCR se realizó utilizando como patrón de peso molecular el marcador comercial Ladder 50 bp o Ladder 100 bp (Productos Bio-Lógicos) o bien el ADN del fago λ digerido con la endonucleasa de restricción *HindIII*.

II.2.4.2. Separación e identificación de fragmentos de ADN de bajo peso molecular

Para la visualización de fragmentos de ADN de longitud < 80 pb se utilizaron geles de poliacrilamida en condiciones no desnaturalizantes (sin la adición de SDS) de

la misma composición que los utilizados para las electroforesis de proteínas (ver Sección II.2.16.2).

II.2.4.3. Geles de lisis *in situ*

Para la visualización de plásmidos de alto peso molecular, se utilizó la técnica de lisis *in situ* (Eckhardt, 1978) modificada como se indica a continuación: a partir de un cultivo crecido 24 h se obtuvo por centrifugación (14.000 *g*, 3 minutos) la masa bacteriana equivalente a 1 ml de cultivo a $DO_{600\text{ nm}} = 0,4$. El *pellet* se lavó con 1 ml de sarcosyl 0,3% en TE frío, mezclando suavemente con una punta de pipeta, y luego de remover el sarcosyl se secó el *pellet* por inversión de los tubos. Posteriormente se agregaron 15 μl de ficoll 20% p/v, manteniendo las muestras en un baño de hielo. Para la siembra de la muestra, se adicionaron 15 μl de *buffer* de lisis (Tris 50 mM, EDTA 25 mM, RNAsa 0,4 mg/ml, lizozima 1mg/ml), y 15 μl de esta mezcla se sembraron en el gel de agarosa.

Para la preparación del gel, se utilizó agarosa 0,8% en TBE 1X. Una vez listo, se colocó en la cuba electroforética con *buffer* TBE 1X de manera tal que el mismo no sobrepase la altura del gel. Luego se sembró 15 μl del colorante xylene cyanol en todas las fosas del gel y se corrió con la polaridad invertida a 110 V (hasta que el colorante azul se salga de las fosas). Luego se agregó TBE 1X hasta cubrir completamente el gel, se sembró la muestra y se corrió por 1 h a 30 V y luego se subió el voltaje a 110 V, y se corrió 1 h más. Finalmente se realizó el revelado en una cuba conteniendo una solución de bromuro de etidio (1 $\mu\text{g/ml}$) y se observó bajo luz UV.

Para estimar el peso molecular de los plásmidos, se utilizó la cepa de referencia *E. meliloti* MVII-I, que contiene plásmidos de peso molecular conocido (Hynes, *et al.*, 1985).

II.2.5. Purificación de fragmentos de ADN de geles de agarosa

Los productos de PCR fueron amplificados y corridos en geles de agarosa (1-2%). Las bandas correspondientes fueron eluidas del gel y purificadas usando el kit comercial Gel Purification Kit (Bioneer) o alternativamente los *primers* y nucleótidos en exceso fueron eliminados mediante precipitación del producto de PCR con una

solución PEG 8000 20% (Rosenthal, *et al.*, 1993) o por tratamiento enzimático con ExoSAP siguiendo las instrucciones del proveedor. Esta alternativa provee una forma sencilla de limpiar los productos de PCR antes de la secuenciación. La exonucleasa I elimina los cebadores en exceso, mientras que la fosfatasa alcalina de camarón elimina fosfatos 5' terminales de los dNTPs restantes.

II.2.6. Determinación de las concentraciones de ADN en una muestra

La concentración y calidad de las preparaciones de ADN se analizó por medidas de la absorbancia a 260 y 280 nm. Las medidas y cálculos fueron realizados en un equipo Nanodrop ND1000 (Nanodrop Technologies) según las especificaciones del fabricante. En el caso particular de la muestra correspondiente al *pool* de ADN plasmídico, la concentración de ADN se estimó utilizando el kit Quant-iT™ PicoGreen® dsDNA (Invitrogen).

II.2.7. Secuenciamiento de fragmentos de ADN

Para secuenciar fragmentos de ADN, las muestras previamente purificadas se enviaron al servicio de INTA-Castelar Argentina, en donde se secuenciaron utilizando la metodología de Sanger.

II.2.8. Síntesis de genes

Cuando se requirió sintetizar un gen químicamente, la síntesis fue llevada a cabo por la empresa GenScript (USA).

II.2.9. Oligonucleótidos

Los oligonucleótidos generados o utilizados en el presente trabajo fueron sintetizados en Genbiotech SRL. En la siguiente tabla se muestran sus secuencias.

Tabla II.4. Oligonucleótidos utilizados

Nombre	Oligonucleótidos: Secuencia 5' → 3'	Referencia
<i>ADNr 16S</i>		
27f	GAGTTTGATCCTGGCTCAG	(Weidner, <i>et al.</i> , 1996)
1385r	CGGTGTGTRCAAGGCC	(Weidner, <i>et al.</i> , 1996)

Tabla II.4 (continuación)

Nombre	Oligonucleótidos: Secuencia 5' → 3'	Referencia
IS1071		
IS-F	GCTTGGTCACTTCTGGGTCTTC	(Providenti, <i>et al.</i> , 2006)
IS-R	CTATGCCCGTCTATCGTTACCC	Providenti <i>et al.</i> 2006
Grupos de incompatibilidad plasmídica		
IncP-1 (<i>trfA</i>)		
trfA 733 f (α,β,ε)	TTCACSTTCTACGAGMTKGCCAGGAC	(Bahl, <i>et al.</i> , 2009)
trfA 1013 r	GWCAGCTTGCGGTA CTCTCCC	Bahl <i>et al.</i> , 2009
trfA g-F (γ)	TTCACSTTTTACGAGCTTGCAGCGAC	Bahl <i>et al.</i> , 2009
trfA g-R	GTCAGCTCGCGGTA CTCTCCCA	Bahl <i>et al.</i> , 2009
trfAg-208f (γ-l.)	TTCACSTTCTACGAACTGTGTAAT	Bahl <i>et al.</i> , 2009
trfAg-208r	GTCAAGGCCCGATACTTCTCCCA	Bahl <i>et al.</i> , 2009
trfA d-F (δ)	TTCACSTTCTACGAGCTTGCACAGAC	Bahl <i>et al.</i> , 2009
trfA d-r	GACAGCTCGCGGTA CTTTCCCA	Bahl <i>et al.</i> , 2009
trfA z-f (ζ)	TTCACSTTCTACGAAATCTGCAAAGAC	Bahl <i>et al.</i> , 2009
trfA z-r	GATAGCTTCGGATACTTTCCCA	Bahl <i>et al.</i> , 2009
IncP-9 (<i>oriV-rep</i>)		
IncP-9 ori 69f	GAGGGTTTGGAGATCAT(AT)AGA	Dealtry <i>et al.</i> , 2014 b
IncP-9 rep 679r	GGTCTGTATCCAGTT(AG)TGCTT	Dealtry <i>et al.</i> , 2014 b
IncP-7 (<i>rep</i>)		
P7repA	CCCTATCTCACGATGCTGTA	(Izmalkova, <i>et al.</i> , 2005)
P7repB	GCACAAACGGTCGTCAG	Izmalkova <i>et al.</i> , 2005
IncQ (<i>oriV</i>)		
IncQ-oriV-1	CTCCCGTACTAACTGTCACG	(Gotz, <i>et al.</i> , 1996)
IncQ-oriV-2	ATCGACCGAGACAGGCCCTGC	Götz <i>et al.</i> , 1996
IncN (<i>repA</i>)		
IncN-rep-1	AGTTCACCACCTACTCGCTCCG	Götz <i>et al.</i> , 1996
IncN-rep-2	CAAGTTCTTCTGTTGGGATTCCG	Götz <i>et al.</i> , 1996
IncW (<i>oriV</i>)		
IncW-oriV-1	GACCCGGAAAACCAAAAATA	Götz <i>et al.</i> , 1996
IncW-oriV-2	GTGAGGGTGAGGGTGCTATC	Götz <i>et al.</i> , 1996
Integrones		
Región conservada		
Orf513F	ATGGTTTCATGCGGGTT	(Arduino, <i>et al.</i> , 2003)
Orf513R	CTGAGGGTGTGAGCGAG	(Orman, <i>et al.</i> , 2002)

Tabla II.4 (continuación)

Nombre	Oligonucleótidos: Secuencia 5' → 3'	Referencia
Sul1 lower	TTTGAAGGTTTCGACAGC	(Barbolla, <i>et al.</i> , 2004)
Sul1 upper	GACGGTGTTCGGCATTCT	Barbolla <i>et al.</i> , 2004
qacEDelta1A	CATCGTGGAAGTCGCT	Quiroga <i>et al.</i> , 2013
qacEDelta1B	GCGATAACAAGAAAAAGCC	Quiroga <i>et al.</i> , 2013
QaceR	GCGAAGTAATCGCAACATCC	Arduino <i>et al.</i> , 2003
QaceF	AGCCCCATACCTACAAAGCC	Arduino <i>et al.</i> , 2003
int1FP	GGGTCAAGGATCTGGATTTCCG	Mazel <i>et al.</i> , 2000
int1FR	ACATGCGTGTAATCATCGTCC	Mazel <i>et al.</i> , 2000
int1F	CGAGGCATAGACTGTAC	(Kristiansson, <i>et al.</i> , 2011)
int1R	TTCGAATGTCGTAACCGC	Quiroga <i>et al.</i> , 2007
Inti2F	GCAAATGAAGTGCAACGC	Orman <i>et al.</i> , 2006
Inti2R	ACACGCTTGCTAACGATG	Orman <i>et al.</i> , 2006
Región variable		
Sulpro3	GCCTGACGATGCGTGGA	(Levesque, <i>et al.</i> , 1995)
TniR	TTCAGCCGCATAAATGGAG	(Post, <i>et al.</i> , 2007)
3'CsNew	AAGCAGACTTGACCTGATAG	(Quiroga, <i>et al.</i> , 2013)
HS287	GGAATCCTCSGCTKGARCGAMTTGTTAGVC	Stokes <i>et al.</i> , 2001
HS286	GGGATCCGCSGCTKANCTCVRRCGTTAGSC	(Stokes & Gillings, 2011)
AJH60	CRRSKTCGGCTTGAAYGARTTG	(Gillings, <i>et al.</i> , 2005)
Is26F	GCTGGCTGAACGCGGAG	(Traglia, <i>et al.</i> , 2012)
Is26R	ATACCTTTGATGGTGCC	Traglia <i>et al.</i> , 2012
Lacasas		
Tag2_80-NdeI	GGGCCATATGAATACCCTTTCACGACG	Este trabajo
Tag2_80-XhoI	ATATTTAGCGGCCGCTGCGCTGACCGCGAATTCGGTCATC	Este trabajo
Construcción y chequeo del vector p-RESC		
pBAM-NcoI-Fw	GCCCATGGATTGTATCGGTTTACCAATG	Este trabajo
pBAM-NcoI-Rv	ATCCATGGACACTTAACGGCTGACATGG	Este trabajo
TnpA_1	AGCCATCCAGTTTGCTTTGC	Este trabajo
TnpA_2	CGAACGTTTTGTGGTGCGT	Este trabajo
tnpA_3'A	CCACGCTTTGTGAAATTCTTCA	Este trabajo
tnpA_3'G	CCACGCTTTGTGAAATTCTTCC	Este trabajo
tnpA_R	GGAAGCGACCACCTTTTCGTA	Este trabajo
Linker IScel H-SFw	CGCTAGGGATAACAGGGTAATATAGAAGCTTCCCGGGCG	Este trabajo
Linker IScel H-SRv	CGCCCGGAAGCTTCTATATTACCCTGTTATCCCTAGCG	Este trabajo

II.2.10. Hibridación por *Southern blot*

El ADN plasmídico fue digerido con las enzimas de restricción *Pst*I y *Bst*Z17I o *Not*I (Fermentas GmbH, St. Leon Rot, Germany). Los fragmentos digeridos fueron separados por electroforesis en gel de agarosa al 1% p/v y luego se realizó el *blotting*. Las hibridaciones para los plásmidos digeridos con *Pst*I y *Bst*Z17I fueron usadas para las sondas de IncP-1 mezcladas (subgrupos α , β , γ , δ y ϵ), IncP-7, IncP-9 o IncQ (Tabla II.4). La sonda de IncP-1 es una mezcla de sondas marcadas con dioxigenina, obtenida por amplificaciones por PCR a partir de distintos plásmidos de los subgrupos. Las sondas IncW e IncN fueron utilizadas en la hibridación por *Southern blot* para los plásmidos digeridos con *Not*I. Las hibridaciones fueron llevadas a cabo tal como se describió en (Binh, *et al.*, 2008).

II.2.11. Secuenciamiento de los plásmidos de alto peso molecular

El secuenciamiento del *pool* de plásmidos obtenidos a partir de los 35 aislamientos representativo del BPS se llevó a cabo en el CeBiTec (Center for Biotechnology), Bielefeld, Alemania. El protocolo utilizado fue el siguiente: cantidades aproximadamente equimolares de los plásmidos previamente purificados fueron mezclados para su secuenciamiento en conjunto. La concentración de ADN en la muestra final fue determinada mediante el uso de Bioanalyzer (Agilent technologies), que permitió determinar (además de la cantidad), la calidad y tamaño de los fragmentos de ADN. 50 ng de la muestra fueron utilizados para preparar las bibliotecas utilizando el protocolo Nextera Libraries (Illumina) de acuerdo con las instrucciones del proveedor. Finalmente se llevó a cabo el secuenciamiento de la muestra utilizando la tecnología MiSeq de Illumina.

II.2.12. Tratamientos enzimáticos del ADN

II.2.12.1. Digestiones de ADN

Las digestiones con endonucleasas de restricción se realizaron en las condiciones óptimas para cada enzima, siguiendo las indicaciones de reacción (temperatura y *buffer*) propuestas por los proveedores comerciales (Fermentas y Promega) y las recomendaciones descritas por Sambrook y colaboradores (1989). Las

reacciones se llevaron a cabo habitualmente con 0,5 - 5 µg de ADN, el volumen adecuado de *buffer* suministrado por el proveedor y 0,5 - 1 U de la enzima a utilizar en un volumen final de 20 - 25 µl. Las mezclas de reacción fueron incubadas a la temperatura óptima de las enzimas utilizadas por un período de tiempo mayor que 5 h.

II.2.12.2. Ligaciones de ADN

Las ligaciones de fragmentos de ADN fueron llevadas a cabo utilizando la enzima ligasa del fago T4 (Promega), según las indicaciones del comerciante. Las reacciones de ligación se incubaron a 16 °C por al menos 4 h. Para el clonado de fragmentos de ADN en vectores de clonado/expresión, se utilizó una relación molar de 3:1 de inserto:vector.

II.2.13. Desalado de soluciones de ADN

Para desalar soluciones de ADN se realizó microdiálisis utilizando las membranas Millipore VSWP P02500 de tamaño de poro de 0,025 µm. Las soluciones conteniendo el ADN a desalar se colocaron durante 20 minutos a temperatura ambiente sobre las membranas, las cuales se ubicaron sobre una placa de Petri conteniendo 30 ml de agua bidestilada estéril.

II.2.14. Construcción del vector plasmídico para la captura de plásmidos de muestras ambientales

El vector p-RESC se obtuvo mediante una serie de pasos de clonado sucesivos mediante los cuales se modificó el vector pBAM1 (ver mapa plasmídico, Figura VII.2), como se indica a continuación:

II.2.14.1. Introducción de un origen de replicación (*oriV*) en el minitransposón

El plásmido pBAM1 se digirió con la endonucleasa *Xba*I para su linearización y clonado del *oriV* proveniente del plásmido pSM10, el cual fue previamente digerido con las enzimas *Nhe*I y *Xba*I. Ambas digestiones se mezclaron y se realizó la ligación. Luego, el producto de ligación se transformó en *E. coli* DH5α, seleccionando los clones recombinantes por su capacidad de crecer en medio con Km, Ap y su capacidad de

replicar en esta bacteria por el *oriV* introducido. La construcción se chequeó por miniprep y digestión con las enzimas *EcoRI* y *HindIII*.

II.2.14.2. Agregado del sitio de reconocimiento de la endonucleasa *SceI*

La construcción obtenida en el punto anterior se digirió con la enzima de restricción *SwaI* a fin de linearizar el vector. Luego de la linearización se clonó el *linker* *SceI*, obtenido previamente como se indica en la sección II.2.15. La ligación se transformó en *E. coli* DH5 α y se seleccionaron los transformantes en medio con Ap y Km. La construcción se chequeó por análisis del perfil de restricción cuando se digirió con la enzima *HindIII*.

II.2.14.3. Remoción del *oriT* y *oriVR6K* provenientes del pBAM1

Para eliminar el *oriT* (origen de transferencia plasmídica) y el *oriVR6K* (origen de replicación plasmídico) que se encuentran por fuera del minitransposón, se realizó una PCR con *primers* inversos (pBAM-*NcoI*-Fw/ pBAM-*NcoI*-Rv, que contienen un sitio *NcoI*) para amplificar el plásmido completo, excluyendo la región deseada. El producto de PCR se religó y transformó en *E. coli* DH5 α , seleccionándose los transformantes resistentes a Ap y Km. La construcción se chequeó por miniprep y digestión con *NcoI*, que lineariza al vector.

II.2.14.4. Introducción de un *oriT* en el minitransposón

Para el clonado de un nuevo *oriT* dentro del transposón, se digirió la construcción obtenida previamente con la enzima de restricción *KpnI*, y luego se ligó con el *oriT* del plásmido pUC19mob (previamente digerido con *KpnI*). El producto de ligación se transformó en *E. coli* DH5 α y luego de 1 h en incubación en medio LB líquido (como se indica en la sección II.1.11) 500 μ l de este cultivo se utilizaron para realizar una conjugación triparental (como se indica en la sección II.1.9.2), utilizando la cepa de *E. coli* HB101 portadora del plásmido pRK600 como *helper* y la cepa HB101 como receptora. Los transconjugantes se seleccionaron en medio LB conteniendo los antibióticos Rf, Km y Ap. Las colonias que crecieron en este medio corresponden a la cepa de *E. coli* HB101 que contiene la construcción con el *oriT* clonado, el cual le permitió al plásmido moverse desde la cepa transformada de *E. coli* DH5 α hacia la cepa

E. coli HB101 receptora. La construcción resultante se chequeó en una primera instancia por digestión con la enzima *KpnI*, y luego se secuenció completamente y se la llamó p-RESC.

II.2.15. Obtención del *linker* Scel

La obtención del *linker* de ADN de doble hebra se obtuvo como se indica a continuación: se mezclaron los oligonucleótidos de secuencia complementaria en una cantidad de 25 μ M de cada uno, junto con NaCl 2,5 mM y DMSO (dimetilsulfóxido) al 10%. La mezcla se hirvió durante 5 minutos y luego se dejó enfriar lentamente hasta temperatura ambiente para permitir la hibridación de las hebras de ADN complementario. La correcta formación del *linker* de 39 pb se chequeó en un gel nativo de poliacrilamida (PAGE 15%). Para su posterior clonado, se digirió con la endonucleasa *SmaI* (cuyo sitio de reconocimiento se adicionó a ambos lados del *linker*), para generar ambos extremos fosforilados.

II.2.16. Preparación de proteínas y determinación de su actividad

II.2.16.1. Expresión heteróloga de proteínas en *E. coli*

La cepa de *E. coli* BL21(DE3)Star portadora de la construcción pET22b(+):*lac188* fue crecida en medio LB líquido suplementado con Ap (100 μ g/ml) en erlenmeyers de 2 l, a 37 °C con agitación continua de 200 rpm. Cuando el cultivo alcanzó una $DO_{600\text{ nm}} = 0,6$ se adicionó IPTG (isopropil- β -D-1-tiogalactopiranosido) a una concentración final de 0,4 mM. Al mismo tiempo, con el fin de optimizar la expresión: se ensayaron diversas condiciones de inducción. Por ejemplo, se ensayó con y sin el agregado de $CuSO_4$ 0,25 mM y se compararon varias temperaturas de inducción (37 °C, 25 °C o 16 °C). Luego, las células fueron colectadas a distintos tiempos, centrifugadas a 7.500 *g* durante 10 minutos a 4 °C. Los *pellets* se resuspendieron en PBS (*buffer* fosfato salino) 10X conteniendo DNAsa 10 μ g/ml, PMSF (fluoruro de fenilmetilsulfonilo) 0,19 mg/ml y pepstatina A 0,7 μ g/ml; luego las células fueron lisadas por sonicación (6 ciclos de 30 segundos cada uno, en hielo), seguido de una centrifugación a 34.000 *g* por 60 minutos a 4 °C. Finalmente, las fracciones proteicas fueron analizadas mediante SDS-PAGE (electroforesis en gel de poliacrilamida con sodio dodecil sulfonato).

II.2.16.2. Análisis de proteínas mediante geles de poliacrilamida discontinuos en condiciones desnaturizantes (SDS-PAGE)

Las electroforesis en geles de poliacrilamida con SDS y β -mercaptoetanol, se llevaron a cabo según el procedimiento descrito por Laemmli (1970). El gel de apilamiento se preparó con 5% de poliacrilamida en Tris-HCl a pH=6,8, mientras que los geles de separación se prepararon con 12,5 a 15% de poliacrilamida en Tris-HCl a pH=8,8. Las corridas se llevaron a cabo durante 30 a 60 minutos en equipos Miniprotean II (BioRad) a 100 - 120 volts. Las proteínas en el gel fueron visualizadas por tinción con Azul Coomassie (metanol 50% v/v, ácido acético 5% v/v, Coomassie Brilliant Blue R2500, 1% p/v). Los geles se incubaron con la solución colorante a temperatura ambiente durante 12 horas aproximadamente. Luego fueron desteñidos mediante una serie de lavados sucesivos con solución decolorante (metanol 50% v/v, ácido acético 5% v/v). El registro digital se realizó sobre un transiluminador con luz blanca.

II.2.16.3. Purificación de la lacasa Lac188

La lacasa fusionada en su extremo C-terminal al *tag* de 6 histidinas fue purificada de las fracciones solubles por columnas de afinidad HiTrap™ Chelating HP (GE Healthcare Life Sciences) equilibradas con *buffer* fosfato 20 mM (pH 6,7), conteniendo imidazol 20 mM y Na₂SO₄ 300 mM. La proteína recombinante fue eluida en el mismo *buffer* conteniendo imidazol 250 mM. La pureza de la proteína fue chequeada mediante SDS-PAGE. Para la estimación de la masa molecular, se utilizaron marcadores de peso molecular (GE Healthcare Sciences, Little Chalfont). La concentración proteica fue estimada mediante el método de Bradford (1976) empleando seroalbúmina bovina (BSA) como proteína estándar o por análisis densitométrico de las bandas del gel de SDS-PAGE utilizando el *software* Quantity One (Bio Rad, Hercules, CA, USA).

II.2.16.4. Western blot

Los geles conteniendo la proteína Lac188 fueron transferidos a una membrana de PVDF (difluoruro de polivinilideno). La proteína de interés fue detectada por quimioluminiscencia empleando el anticuerpo monoclonal anti His-Tag conjugado HRP

(Novagen Inc., Madison, USA) y el kit ECL Western Blotting Detection System (GE Healthcare Sciences, Little Chalfont, UK) siguiendo las indicaciones del proveedor.

II.2.16.5. Determinación de la actividad enzimática de la lacasa Lac188

La actividad lacasa fue medida a 25 °C en una mezcla de reacción de 1,1 ml conteniendo 500 µl de acetato de sodio 100 mM (concentración final 45 mM, pH 5,0), 50 µl de ABTS (ácido 2,2'-azino-bis(3-etilbenzotiazolina-6-sulfónico)) concentración final de 4,5 mM, 450 µl de agua destilada y 100 µl de la muestra proteica. La oxidación del sustrato fue determinada espectrofotométricamente mediante la medida de la absorbancia a 420 nm y 25 °C durante 5 minutos. La oxidación se calculó como la pendiente de la recta ajustada a las primeras medidas que exhibieron una dependencia lineal de la absorbancia con el tiempo. Como control negativo, se corrió en paralelo la misma mezcla en ausencia de enzima. Una unidad enzimática (U) fue definida como la cantidad de lacasa Lac188 requerida para oxidar 1 µmol de ABTS por minuto.

La capacidad de la lacasa Lac188 de oxidar distintos compuestos fue chequeada gracias a una colaboración con el grupo del Dr. F. Marinelli (Italia) para los siguientes sustratos a las longitudes de onda indicadas: 100 mM de 2,6-dimetoxifenol (DMP) (468 nm), 100 mM de pirocatecol (450 nm), 100 mM de pirogalol (450 nm), 8 mM de ácido vanílico (316 nm), 10 mM de $K_4Fe(CN)_6$ (405 nm), 2 mM de tirosina (475 nm), 5 mM de L-3,4-dihidroxifenilalanina (L-DOPA) (475 nm), 100 mM de guayacol (468 nm), 10 mM de ácido sirínico (300 nm), 4 mM de ácido ferúlico (287 nm), 1mM de siringaldazina (525 nm) ó 1 mM siringaldehído (320 nm). El pH óptimo para la actividad sobre distintos sustratos fue evaluado a 25 °C en *buffer* de pH 2,0–8,0.

II.3. Técnicas bioinformáticas

II.3.1. Diseño de *primers*

Para el diseño de *primers*, se utilizó el servidor web Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) y se variaron los parámetros según las necesidades.

II.3.2. Determinación de la identidad de fragmentos de ADN purificados y secuenciados

La identidad de los fragmentos amplificados por PCR se determinó a partir de una búsqueda de similitud de secuencia a nivel ADN (BLASTn: *Basic Local Alignment Search Tool*-nucleotid) o a nivel proteína (BLASTp: *Basic Local Alignment Search Tool*-protein o BLASTx, el cual traduce la secuencia de ADN a los seis marcos de lectura posibles y luego realiza una búsqueda de similitud a partir de las seis secuencias peptídicas generadas). Esta búsqueda se realizó en el servidor web de NCBI, *National Center for Biotechnology Information* (www.ncbi.nlm.nih.gov/BLAST/), el cual contiene una base de datos de secuencias biológicas. En las búsquedas, se utilizaron los parámetros estandarizados del servidor.

II.3.3. Tratamiento de los datos crudos del secuenciamiento de alto rendimiento

Las lecturas obtenidas a partir del secuenciamiento fueron ensambladas en *contigs* utilizando el *software* Newbler 2.6. Para la predicción de regiones como ORFs, se utilizó el *software* GLIMMER (Delcher, *et al.*, 1999). Los *contigs* fueron luego importados a la plataforma de anotación GenDB versión 2.2 (Meyer, *et al.*, 2003), la cual contiene una interfaz fácil de utilizar y permite una visualización simple y análisis de la información. Esta plataforma también permite refinar los datos manualmente.

Para el análisis de las lecturas crudas (esto es, sin ensamblar) se utilizó la plataforma MGX (CeBiTec, Alemania). Las anotaciones Pfam fueron creadas realizando una búsqueda BLASTx versus la base de datos EggNOG (release 3.0), usando un e-value de corte de $1e^{-5}$. La familia Pfam del mejor *hit* fue el utilizado para anotar la secuencia.

II.3.4. Búsqueda de genes de resistencia a antibióticos y metales pesados

Para el análisis de los genes de resistencia a antibióticos (ARGs), aquellos genes automáticamente anotados como genes de resistencia en GenDB fueron analizados individualmente (para su clasificación a nivel de familia) utilizando la base de datos ARDB (*Antibiotic Resistance Data Base*), la cual contiene 23.137 genes de resistencia pertenecientes a 380 tipos que confieren resistencia a 249 antibióticos (<http://arbd.cbcb.umd.edu/>) (Liu & Pop, 2009). En el caso de los genes de resistencia a

metales, los mismos fueron corroborados utilizando el servidor BLASTp de NCBI, mencionado anteriormente.

II.3.5. Búsqueda de secuencias de inserción

Para el análisis de secuencias de inserción, se trabajó con la base de datos de ISfinder (Siguiet, *et al.*, 2006), de la cual se extrajeron todas las secuencias depositadas para generar una base de datos contra la cual se realizó un BLASTn de los *contigs* del set de datos plasmídicos obtenidos en este trabajo.

II.3.6. Búsqueda de genes de degradación de compuestos xenobióticos y aromáticos

El análisis de genes posiblemente implicados en rutas metabólicas de degradación de compuestos xenobióticos/aromáticos cíclicos se realizó a partir de la búsqueda de dichos genes en la plataforma GenDB. La plataforma permite la visualización y contextualización de las enzimas codificadas por estos genes en una interfaz gráfica de KEGG (Kyoto Encyclopedia of Genes and Genomes), que muestra su ubicación y actividad enzimática en una determinada ruta.

II.3.7. Búsqueda de genes que codifican para lacasas y quitinasas

La identificación *in silico* de lacasas y quitinasas de potencial interés industrial se llevó a cabo sobre las secuencias de origen plasmídico en la plataforma GenDB, utilizando dos herramientas distintas: **i)** la búsqueda a partir de la anotación automática de genes y **ii)** la búsqueda a partir de los números EC (*Enzyme Commission numbers*), que clasifican a las enzimas de acuerdo a las reacciones químicas que catalizan. Para la última estrategia, se utilizaron los números de identificación EC 3.2.2.14 y EC 1.10.3.2 para quitinasas y lacasas, respectivamente. Aquellos ORFs candidatos fueron comparados contra la base de datos de GeneBank para corroborar su identidad.

En el caso de las lacasas, adicionalmente a las dos herramientas mencionadas, se realizó en colaboración con el grupo del Dr. I. Mandic-Mulec (Eslovenia) una comparación de todas las secuencias ensambladas de origen plasmídico contra una base de datos del tipo HMM (*Hidden Markov Models*) generada por Ausec y

colaboradores (2011), la cual fue construida a partir una recopilación de genes que codifican para lacasas de dos y tres dominios identificados en más de 1000 genomas y metagenomas bacterianos disponibles en distintas bases de datos públicas.

II.3.8. Análisis filogenéticos

II.3.8.1. Análisis filogenéticos del gen *intl1*

Las secuencias correspondientes a los genes *intl1* identificados en este trabajo (*intl1_BF02*, *intl1_BF19*, *intl1_BF25*, *intl1_BF27*, *intl1_BF44* e *intl1_BF48*) fueron comparadas filogenéticamente con otros alelos *intl1* previamente identificados a partir de ambientes clínicos y no clínicos, recopilados por Nardelli y colaboradores (2012) y extraídos de GenBank. Las secuencias fueron alineadas con la herramienta CLUSTALW utilizando el software MEGA 5.05 con los parámetros *default*. La filogenia fue realizada utilizando el método de Neighbour Joining. Se estimó el mejor modelo para la reconstrucción filogenética, siendo éste el *Nakamura-3 parameter model*. Las posiciones con *gaps* fueron removidas para el análisis.

II.3.8.2. Análisis filogenéticos de las relaxasas y del sistema MPF

Las relaxasas del set de datos plasmídico BJP_1037, BJP_380 AP_265, BJP_6246, BJP_4415, BJP_1723, pMC3_7, BJP_1498, AP_669, BJP_441, BJP_847, BJP_634, BJP_2740, AP_394, AP_113, BJP_237, BJP_387, BJP_198, y pMC6_1 fueron usadas para realizar una búsqueda BLASTp. Para cada familia MOB, las proteínas fueron cortadas de manera de analizar sólo el dominio del extremo N-terminal de las relaxasas (300 amino ácidos para las familias MOB_F, MOB_Q y MOB_V y 500 amino ácidos para MOB_P) y luego fueron alineadas con MUSCLE (Edgar, 2004). Gracias a una colaboración con el Dr. M. P. Garcillán y F. de la Cruz (España), la reconstrucción filogenética fue llevada a cabo por Maximum Likelihood (ML), usando RAxML versión 7.2.7 (Stamatakis, 2006). Se ejecutaron 20 árboles ML usando el modelo JTTGAMMA. Los árboles *bootstrap* fueron inferidos usando el criterio *autoFC stop* para obtener los valores para cada nodo del mejor árbol ML. Las familias y subfamilias MOB fueron definidas por Garcillán-Barcia *et al.* (Garcillan-Barcia, *et al.*, 2009).

El análisis de las proteínas del tipo VirB4 del set de datos plasmídico incluyó 13

proteínas pertenecientes al grupo Pfam CagE-TrbE-VirB y representativas de los 8 sistemas MPF definidos en http://conjdb.web.pasteur.fr/conjdb/_design/conjdb/index.html (Guglielmini, *et al.*, 2014) (GenBank Acc. No. BAA78009.1: TraU_R64 y AAN87691.1: TraU_pCTX-M3 para MPF_I; BAB78290.1: Alr7206_pCC7120alpha para MPF_C; WP_014326569.1: Tfc16_ICEHin10810 para MPF_G; AAB60017.1: Orf16_Tn916 para MPF_{FA}; YP_002790923.1: TrsE_pGO1, YP_195783.1: PrgJ_pCF10, YP_001086885.1: CD418_CTn2 y CAD47019.1: Gbs1360_ICESaNEM316 para MPF_{FATA}; AAC44180.1: TraC_F para MPF_F; NP_809006.1: TraG_CTnDOT para MPF_B; AAO43554.1: TrbE_p42a, YP_009077466.1: TrwK_R388, CAC82751.1: TraE_pIPO2T, NP_065362.1: TraE_R721; CAJ85691.1: TrbE_RP4; NP_053384.1: VirB4_pTi for MPF_T). Todas las proteínas fueron alineadas con MUSCLE y la reconstrucción filogenética de Maximum-Likelihood fue realizada con FastTree 2 (Price, *et al.*, 2010).

CAPÍTULO III

ESTABLECIMIENTO DE UNA COLECCIÓN DE BACTERIAS PORTADORAS DE PLÁSMIDOS DE ALTO PESO MOLECULAR OBTENIDAS A PARTIR DE UN SISTEMA DE BIOPURIFICACIÓN UTILIZADO PARA LA DECONTAMINACIÓN DE PESTICIDAS

III.1. Introducción

En la primera parte de este trabajo nos hemos enfocado en la construcción de una colección de aislamientos portadores de plásmidos de alto peso molecular presentes en un biofiltro utilizado para la descontaminación de diversos pesticidas. Como se mencionó previamente en la introducción de esta Tesis, la elección del sistema de trabajo aquí utilizado surge de un *screening* previo independiente de cultivo en el que se empleó ADN total de distintas comunidades microbianas y se analizó la presencia, abundancia y diversidad de MGEs presentes en los distintos ambientes preseleccionados. Luego de este *screening*, decidimos trabajar con un BPS en el cual se ha detectado la mayor abundancia de plásmidos de distintos grupos de incompatibilidad plasmídica, integrones de clase 1 y 2, y la secuencia de inserción IS1071. Es importante destacar que estos marcadores se encuentran en la microbiota total del biofiltro, es decir, tanto en los microorganismos cultivables como los no cultivables que conforman la comunidad allí presente.

Para comenzar a trabajar con este sistema, tuvimos en cuenta tres grandes aspectos que merecen ser mencionados y explicados antes de comenzar:

- En primer lugar, nos limitamos al estudio de aquellas bacterias portadoras de plásmidos, principalmente los de alto peso molecular. Esto se debe a que estamos particularmente interesados en el estudio del *pool* de genes móviles de la comunidad, que puede contener funciones asociadas a la adaptación bacteriana. Los plásmidos de alto peso molecular son particularmente interesantes debido a varios aspectos: **i)** generalmente son movilizables o conjugativos, por lo cual juegan un rol clave en la adaptación y evolución bacteriana, y pueden diseminarse fácilmente en una comunidad formando parte del genoma de diversas bacterias; **ii)** plásmidos pequeños generalmente contienen sólo los genes que le permiten su replicación y algunos otros pocos, mientras que los plásmidos de mayor tamaño pueden contener una gran cantidad de genes en sus esqueletos; y **iii)** se ha encontrado, entre otras funciones, genes catabólicos involucrados en la degradación de compuestos recalcitrantes y xenobióticos presentes en plásmidos de alto peso molecular, lo cual resulta

interesante en este trabajo, dado que el ambiente de estudio se encuentra contaminado con este tipo de compuestos.

- En segundo lugar, decidimos limitarnos sólo a los microorganismos cultivables. Si bien es cierto que éstos representan sólo una pequeña proporción de la comunidad entera presente en un determinado ambiente, la obtención de plásmidos de alto peso molecular sin previo cultivo de los microorganismos es muy ineficiente; todos los métodos de aislamiento de plásmidos de comunidades bacterianas enteras independientes de cultivo hasta hoy desarrollados (mencionados previamente en la sección I.2.2.1.3), tienen la gran desventaja de que sólo permiten obtener plásmidos relativamente pequeños.

- Por último, en caso que los aislamientos presentes en el biofiltro posean plásmidos transmisibles o conjugativos, es esperable que éstos estén presentes tanto en bacterias cultivables como no cultivables y que la probabilidad de encontrarlos aumente con el número de aislamientos que se analicen. En general, es escasa la información sobre abundancia, distribución y compatibilidad de plásmidos en estos ambientes de importancia ecológica y agrícola, existiendo sólo en unos pocos casos información genético-molecular de los sistemas de transferencia en ambientes agrícolas. Esperamos que la caracterización de los plásmidos presentes y el estudio de sus sistemas de movilización permitan dimensionar con más precisión la significación de los eventos de transferencia horizontal en la generación de variabilidad y adaptabilidad de las bacterias en este tipo de ambientes.

Comenzamos con la selección de bacterias del BPS enriquecido en pesticidas. Esta estrategia ha sido concebida en la posible existencia de una selección natural adaptativa y positiva para las bacterias que puedan recuperarse de nichos altamente contaminados con pesticidas. Luego, con el objetivo de determinar algunos rasgos fenotípicos de los aislamientos, avanzaremos en la evaluación de su tolerancia a antibióticos y metales pesados. Dado que seleccionaremos aislamientos portadores de plásmidos, podría ocurrir que estos elementos posean genes de resistencia asociados, ya que los plásmidos juegan un rol clave en la diseminación de diversos genes de

resistencia, exista (como ocurre en hospitales o en granjas con desechos animales) o no presión de selección.

En este marco, por las razones expuestas en los párrafos precedentes, el objetivo de esta parte del trabajo se concentró en la búsqueda sistemática y en la caracterización de una colección de aislamientos bacterianos, cultivables y portadores de plásmidos de alto peso molecular, presentes en un BPS enriquecido en pesticidas que siendo manejable y práctica desde el punto de vista experimental, tenga una buena representatividad de la diversidad presente en los aislamientos analizados.

III.2. Construcción de una colección de aislamientos portadores de plásmidos de alto peso molecular como material de base para la búsqueda de funciones plasmídicas

Para la obtención de la colección de aislamientos bacterianos se partió de una muestra del BPS, la cual fue procesada como se indica detalladamente en la sección II.1.6 de Materiales y Métodos. El plaqueo de la suspensión obtenida a partir de la muestra en los 8 medios de cultivo empleados, así como en medio LB conteniendo 12 antibióticos diferentes generó alrededor de 10.000 colonias en los distintos medios, principalmente en los medios ricos, ya sea conteniendo o no antibióticos.

De las miles de colonias obtenidas por plaqueo en los distintos medios de cultivo, se eligieron 1.400 al azar. Para avanzar en su caracterización, cada aislamiento fue examinado en su perfil plasmídico por la técnica de lisis *in situ* en geles de agarosa (Capítulo II, sección II.2.4.3). Esta técnica nos permitió visualizar fácilmente el contenido genómico de las bacterias, ver si contienen plásmidos de alto peso molecular y, en caso de que los tuvieran, estimar su tamaño. Entre los 1.400 aislamientos analizados, encontramos que 75 poseían plásmidos de alto peso molecular. Si bien en la mayoría de los aislamientos se observó la presencia de un solo plásmido, en otros se detectaron dos, tres e incluso cuatro plásmidos de distinta migración electroforética y tamaño. De estos 75 aislamientos, muchos presentaron un perfil plasmídico similar, además de poseer ciertas características fenotípicas similares (como morfología y coloración de colonia). En base a sus perfiles plasmídicos, estos

aislamientos se clasificaron en 35 grupos, y de allí en adelante se siguió trabajando con esta colección bacteriana que se generó tomando un aislamiento representativo de cada uno de estos 35 grupos. En la figura III.1 se muestran los perfiles plasmídicos de los 35 aislamientos representativos de la colección bacteriana. La colección entera comprende al menos 53 plásmidos que tienen un tamaño que varía entre 30 y 1.500 Kb aproximadamente, siendo más de la mitad de ellos mayores a 100 Kb.

Es importante en este punto realizar la observación de que algunos plásmidos de bajo peso molecular y/o muy bajo número de copias no siempre son visibles con esta técnica, dadas las condiciones utilizadas como los tiempos largos de corrida electroforética. Por estas razones decimos que la colección bacteriana comprende al menos 53 plásmidos, pero no descartamos la posibilidad de que haya más.

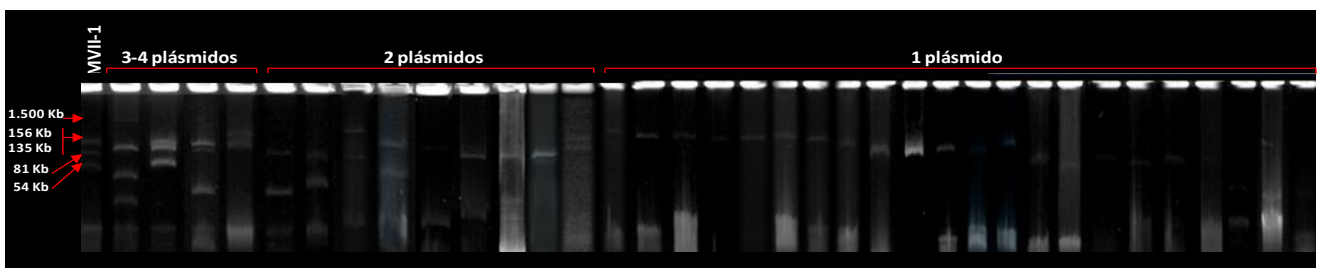


Figura III.1. Perfiles plasmídicos de los 35 aislamientos representativos por la técnica de lisis *in situ*. El tamaño de los plásmidos fue estimado por comparación con los plásmidos de la cepa MVII-1 de *E. meliloti*, la cual contiene 5 plásmidos de peso molecular conocido (Hynes *et al.*, 1986).

III.3. La colección de aislamientos como parte del germoplasma presente en el BPS muestreado: clasificación taxonómica de los aislamientos de la colección bacteriana

Una vez establecida la colección de 35 aislamientos, se procedió a determinar la identidad de los mismos a partir del secuenciamiento del fragmento del ADNr 16S. Para ello se secuenciaron los productos de amplificación (aproximadamente 1.200 pb) obtenidos con la combinación de *primers* 27f/1385r (Materiales y Métodos, Tabla II.4). Una vez obtenida la secuencia de este fragmento para los 35 aislamientos, se evaluó su identidad por comparación con las presentes en la base de datos GenBank del NCBI (National Center for Biotechnology Information).

En la tabla III.1 se detallan, junto a otros datos de interés, los géneros y/o especies y el número de acceso del mejor *hit* para cada aislamiento. Los resultados obtenidos indican que de los 35 aislamientos analizados 22 corresponden a bacterias Gram-negativas y 13 a Gram-positivas. En la figura III.2 se muestra el número de aislamientos correspondientes a diferentes clases y géneros. Como puede observarse, la clase más representada es la de gammaproteobacterias con un 40% del total, dominada por el género *Pseudomonas*, seguida por la clase Actinobacteria (20%), Bacilli (17%), Betaproteobacteria (9%), Alfaproteobacteria (8%) y Sphingobacteria (6%). Estos resultados muestran una amplia diversidad bacteriana en la colección de aislamientos portadores de plásmidos de alto peso molecular recuperados del BPS en estudio.

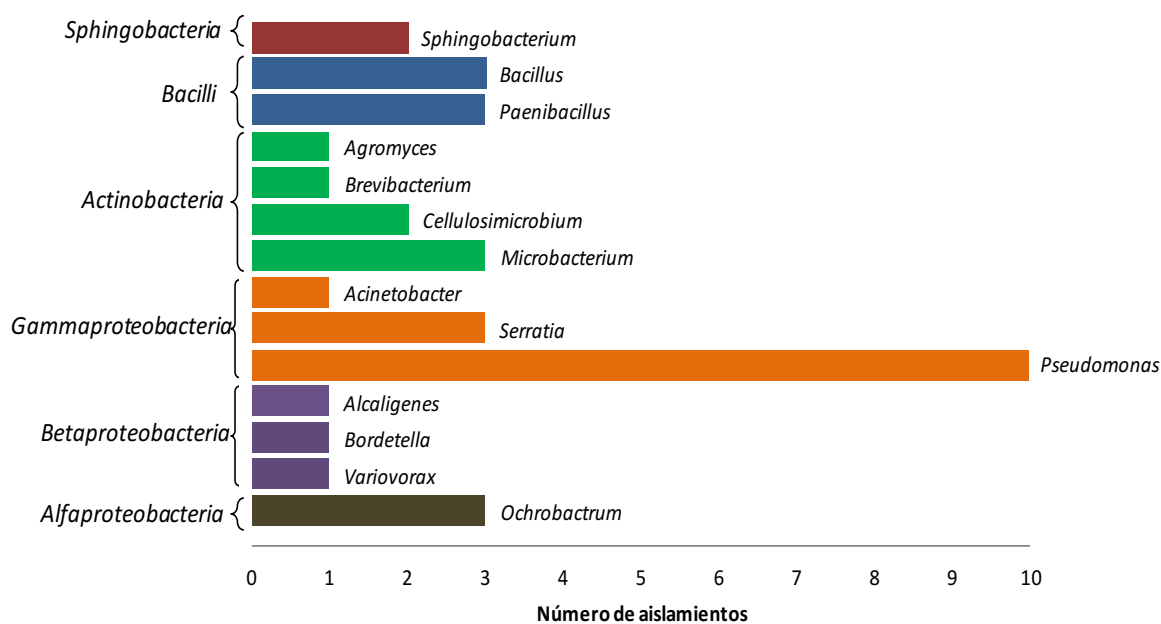


Figura III.2. Clasificación taxonómica de los 35 aislamientos representativos.

III.4. Caracterización fenotípica de los aislamientos

Con el objetivo de determinar algunos rasgos fenotípicos de los aislamientos, comenzamos con la evaluación de su tolerancia a antibióticos y metales pesados.

Tabla III.1. Identificación de los aislamientos representativos de la colección bacteriana

Aislamiento	Medio de aislamiento	Número de plásmidos ^a	Número de acceso del mejor hit	Porcentaje de identidad	Género/ Especie	Coloración de Gram
BF02	LB Km	4	AM231060.1	99%	<i>Ochrobactrum</i> sp.	negativa
BF13	LB Km	1	AM231060.1	98%	<i>Ochrobactrum</i> sp.	negativa
BF14	LB Gm	1	KC660139.1	92%	<i>Cellulosimicrobium cellulans</i>	positiva
BF15	LB Km	3	DQ987618.1	96%	<i>Ochrobactrum</i> sp.	negativa
BF19	LB Sm	2	JN863504.1	99%	<i>Variovorax</i> sp.	negativa
BF21	LB Sm	1	JX912456.1	96%	<i>Paenibacillus</i> sp.	positiva
BF22	LB Sm	3	JX469406.1	99%	<i>Paenibacillus tundra</i>	positiva
BF25	LB Sm	1	JQ086574.1	99%	<i>Pseudomonas putida</i>	negativa
BF27	LB Sm	1	AB679945.1	98%	<i>Pseudomonas</i> sp.	negativa
BF28	LB Sm	1	KC355299.1	99%	<i>Paenibacillus xylanexedens</i>	positiva
BF30	LB	1	GQ263693.1	98%	<i>Bordetella</i> sp.	negativa
BF31	LB Gm	1	JQ782927.1	98%	<i>Cellulosimicrobium</i> sp.	positiva
BF33	LB	1	KC688876.1	78%	<i>Pseudomonas putida</i>	negativa
BF36	LB	2	JX514852.1	99%	<i>Microbacterium</i> sp.	positiva
BF37	LB	1	AB749367.1	97%	<i>Serratia</i> sp.	negativa
BF42	LB	2	HM352330.1	99%	<i>Alcaligenes</i> sp.	negativa
BF43	LB	1	KC337094.1	99%	<i>Serratia</i> sp.	negativa
BF44	LB	1	KC355284.1	94%	<i>Microbacterium oxydans</i>	positiva
BF46	LB	2	KC583444.1	98%	<i>Serratia</i> sp.	negativa
BF48	LB	1	GU300597.1	99%	<i>Microbacterium</i> sp.	positiva
BF52	TY	2	DQ347560.1	99%	<i>Brevibacterium</i> sp.	positiva
BF53	TY	1	FM998720.1	100%	<i>Agromyces cerinus</i>	positiva
BF55	TY	1	JX514404.1	96%	<i>Pseudomonas putida</i>	negativa
BF56	TY	1	KC211306.1	99%	<i>Acinetobacter</i> sp.	negativa
BF58	TY	3	JF208064.1	96%	<i>Pseudomonas</i> sp.	negativa
BF59	TY	2	KC311627.1	94%	<i>Sphingobacterium</i> sp.	negativa
BF60	LB	1	AJ786800.1	81%	<i>Sphingobacterium</i> sp.	negativa
BF61	EMB	1	HF678976.1	97%	<i>Pseudomonas</i> sp.	negativa
BF62	EMB	1	HF678986.1	99%	<i>Pseudomonas</i> sp.	negativa
BF67	YEM	1	JX501692.1	99%	<i>Pseudomonas</i> sp.	negativa
BF68	Dex	1	JX501692.1	99%	<i>Pseudomonas</i> sp.	negativa
BF70	Dex	1	JN995577.1	99%	<i>Bacillus pumilus</i>	positiva
BF71	Dex	3	GU377063.1	98%	<i>Bacillus</i> sp.	positiva
BF73	Dex	2	KC236718.1	91%	<i>Bacillus megaterium</i>	positiva
BF75	Mck	2	JX501692.1	94%	<i>Pseudomonas</i> sp.	negativa

(*) Número de plásmidos visualizados con la técnica de lisis *in situ*

III.4.1. Evaluación de la tolerancia a antibióticos

Para determinar la tolerancia de los aislamientos frente a distintos antibióticos, se determinó la concentración mínima inhibitoria (MIC, por sus siglas en inglés: *Minimum Inhibitory Concentration*), esto es, la mínima concentración del antibiótico que impide el crecimiento bacteriano. La MIC para cada antibiótico analizado fue determinada como se indica en Materiales y Métodos, sección II.1.8. Los resultados obtenidos se indican en la tabla III.2. Las MICs más altas registradas entre los aislamientos representativos para los distintos antibióticos fueron: 320 µg/ml para Tp, 1600 µg/ml para Ap, 40 µg/ml para Nx, 100 µg/ml para Rf, 0,5 µg/ml para Cp, 1600 µg/ml para Sm, 480 µg/ml para Nm, 40 µg/ml para Tc, 400 µg/ml para Km, 200 µg/ml para Gm y 160 µg/ml para Cm. Como puede observarse en la Tabla III.2, las MICs para cada antibiótico fueron muy variables para cada aislamiento, incluso dentro de un mismo género o especie. Tal es el caso de los aislamientos de *Pseudomonas putida* BF25, BF33 y BF55, los cuales presentaron distinta capacidad de crecimiento en medio suplementado con cloranfenicol, estreptomicina y ampicilina. Esta observación sugiere que algún plásmido u otro MGE podría ser el responsable de tales diferencias. Otro punto a resaltar es que la mayoría de los aislamientos mostraron tolerancia a más de un antibiótico, lo que indica la presencia de fenotipos de multiresistencia en distintos géneros bacterianos.

III.4.2. Evaluación de la tolerancia a metales pesados

La MIC para cada metal analizado fue determinada como se indica en Materiales y Métodos, sección II.1.8. Los resultados obtenidos se muestran en la tabla III.2. Las máximas MICs registradas para los distintos metales fueron: 40 mM para Mn^{+2} , 2 mM para Co^{+2} , 40 mM para Zn^{+2} , 2 mM para Cd^{+2} y 0.2 µg/ml para Hg^{+2} . Para Cu^{+2} y Ni^{+2} , la MIC fue 2 mM para todos los aislamientos de la colección analizados, por ello no se indica en la tabla. Al igual que lo ocurrido para los antibióticos, se detectaron marcadas diferencias de tolerancia para la mayoría de los metales (Hg^{+2} , Mn^{+2} , Zn^{+2} , Cd^{+2} , y Co^{+2}), incluso dentro de un mismo género bacteriano, sugiriendo una posible localización de estos genes en algún MGE.

Tabla III.2. Concentraciones mínimas inhibitorias (MICs) de metales y antibióticos para el crecimiento de los aislamientos representativos del biofiltro

Aislamiento	Tp ^a	Ap ^a	Nx ^a	Rf ^a	Cp ^a	Sm ^a	Nm ^a	Tc ^a	Km ^a	Gm ^a	Cm ^a	MnCl ₂ . 4H ₂ O ^b	CoCl ₂ .6 H ₂ O ^b	ZnSO. 7H ₂ O ^b	CdCl ₂ ^b	HgCl ₂ ^b
BF02 <i>Ochrobactrum</i> sp.	320	1600	nd	25	0,0624	1600	60	5	400	6	80	2	0,2	2	0,2	2
BF13 <i>Ochrobactrum</i> sp.	320	800	2,5	25	0,0624	1600	60	5	400	6	80	2	0,2	2	0,2	2
BF14 <i>Cellulosimicrobium cellulans</i>	nd	200	40	25	0,25	6,25	7,5	28	400	6	40	2	0,2	2	0,02	2
BF15 <i>Ochrobactrum</i> sp.	320	1600	5	25	0,25	1600	30	5	200	3	160	2	0,2	2	0,2	2
BF19 <i>Variovorax</i> sp.	40	800	5	25	0,0624	1600	nd	nd	nd	6	2,5	0,2	2	0,2	2	0,2
BF21 <i>Paenibacillus</i> sp.	5	nd	nd	6,25	0,016	1600	nd	nd	3,12	Nd	10	20	0,2	2	0,2	20
BF22 <i>Paenibacillus tundra</i>	160	nd	2,5	6,5	0,016	1600	30	nd	6	Nd	10	2	0,2	2	0,02	2
BF25 <i>Pseudomonas putida</i>	320	800	5	50	nd	1600	15	5	3,12	3	160	2	0,2	2	0,2	2
BF27 <i>Pseudomonas</i> sp.	320	800	10	50	0,0312	1600	15	5	3,12	3	40	2	0,2	40	2	2
BF28 <i>Paenibacillus xylanexedens</i>	10	nd	nd	nd	nd	25	nd	20	3,12	Nd	10	2	0,2	2	0,02	2
BF30 <i>Bordetella</i> sp.	20	25	40	25	0,25	25	60	0,625	400	3	2,5	2	0,2	2	0,2	2
BF31 <i>Cellulosimicrobium</i> sp.	40	25	40	25	0,25	12,5	7,5	2,5	200	50	10	2	0,2	0,2	0,02	2
BF33 <i>Pseudomonas putida</i>	320	400	20	25	0,016	25	15	0,625	3,12	1,5	80	20	0,2	2	0,2	20
BF36 <i>Microbacterium</i> sp.	10	nd	40	25	0,125	12,5	nd	5	25	3	1,25	2	0,2	0,2	2	2
BF37 <i>Serratia</i> sp.	40	25	1,25	25	nd	25	30	5	6	3	10	20	0,2	2	2	20
BF42 <i>Alcaligenes</i> sp.	160	nd	2,5	6,25	0,0312	400	nd	2,5	nd	1,5	2,5	2	0,2	0,2	0,2	2
BF43 <i>Serratia</i> sp.	20	25	2,5	50	0,016	12,5	7,5	5	6	3	10	20	0,2	2	2	20
BF44 <i>Microbacterium oxydans</i>	nd	12,5	40	nd	0,25	25	7,46	40	3,12	3	10	2	0,2	2	0,2	2
BF46 <i>Serratia</i> sp.	5	400	2,5	100	0,0312	12,5	30	40	3,12	3	10	40	2	2	0,2	40
BF48 <i>Microbacterium</i> sp.	320	nd	40	nd	0,125	25	nd	5	3,12	3	1,25	2	0,2	2	0,02	2
BF52 <i>Brevibacterium</i> sp.	320	nd	40	nd	0,5	25	7,5	nd	3,12	6	1,25	2	0,2	2	0,02	2
BF53 <i>Agromyces cerinus</i>	nd	nd	20	6,25	0,25	12,5	nd	0,625	3,12	1,5	1,25	0,2	0,2	0,2	0,01	0,2
BF55 <i>Pseudomonas putida</i>	320	25	5	50	0,016	25	7,5	2,5	3,12	1,5	80	2	0,2	2	0,02	2
BF56 <i>Acinetobacter</i> sp.	160	nd	2,5	6,25	0,016	12,5	7,5	0,625	6	1,5	10	2	0,2	0,2	0,02	2
BF58 <i>Pseudomonas</i> sp.	320	1600	20	50	0,016	25	7,5	nd	3,12	1,5	160	2	0,2	2	0,2	2
BF59 <i>Sphingobacterium</i> sp.	80	800	20	25	0,25	800	480	20	400	200	10	2	0,2	2	0,2	2
BF60 <i>Sphingobacterium</i> sp.	320	800	5	nd	0,25	800	480	20	400	200	10	2	0,2	2	0,2	2
BF61 <i>Pseudomonas</i> sp.	320	1600	5	50	0,0624	25	15	5	3,12	1,5	80	2	0,2	2	0,2	2
BF62 <i>Pseudomonas</i> sp.	320	1600	5	50	0,0312	12,5	15	5	3,12	1,5	80	2	0,2	2	0,2	2
BF67 <i>Pseudomonas</i> sp.	40	25	5	25	nd	25	15	0,625	6	3	10	2	0,2	0,2	0,2	2
BF68 <i>Pseudomonas</i> sp.	40	25	nd	nd	nd	25	15	0,625	6	3	10	2	0,2	0,2	0,2	2
BF70 <i>Bacillus pumilus</i>	nd	nd	2,5	nd	0,016	12,5	nd	nd	3,12	Nd	10	20	0,2	2	0,02	20
BF71 <i>Bacillus</i> sp.	10	nd	nd	nd	nd	6,25	nd	nd	nd	3	1,25	20	0,2	0,02	0,01	20
BF73 <i>Bacillus megaterium</i>	10	nd	5	nd	0,016	6,5	nd	nd	6	3	1,25	2	0,2	0,2	0,01	2
BF75 <i>Pseudomonas</i> sp.	40	25	40	6,5	0,0312	25	7,5	0,625	25	3	10	2	0,2	2	0,02	2

^a Los números representan la concentración más baja (µg/ml) del antimicrobiano que inhibe completamente el crecimiento bacteriano en medio LB.

^b Los números representan la concentración más baja (mM) del metal que inhibe completamente el crecimiento bacteriano en medio LB.

nd: no determinado.

III.5. Evaluación de los grupos de incompatibilidad plasmídica

Originalmente los plásmidos de bacterias fueron clasificados basándose en su grupo de incompatibilidad, característica que está asociada al mecanismo por el cual los plásmidos controlan el inicio de la replicación y su herencia estable. Los plásmidos pertenecientes a un mismo grupo de incompatibilidad comparten estos mecanismos y como consecuencia no pueden ser coheredados en forma estable. En pocas palabras, dos plásmidos son incapaces de coexistir en la misma célula si sus funciones de replicación y partición interfieren. Esto es causado por dos características del mantenimiento de los plásmidos bacterianos: la selección al azar de las moléculas de plásmido que se van a replicar y la distribución al azar en la partición. Este tipo de clasificación es útil desde un punto de vista práctico, ya que en una cepa sólo pueden mantenerse plásmidos compatibles (de distinto grupo de incompatibilidad). Por ello con el objetivo de establecer qué tipos de plásmidos se encuentran presentes en nuestra colección, se estudió la presencia de los grupos de incompatibilidad plasmídica en todos los aislamientos Gram negativos seleccionados. En particular nos centramos en el análisis de grupos de incompatibilidad de plásmidos de amplio rango de huésped conjugativos IncP (subgrupos IncP-1, IncP-7 e IncP-9), IncW e IncN y movilizable IncQ empleando la técnica de *Southern blot* (Materiales Y Métodos, sección II.2.10). Estos ensayos revelaron la presencia de un plásmido del subgrupo IncP-1 en el aislamiento *Alcaligenes* sp. BF42 y un plásmido IncP-7 en el aislamiento *Pseudomonas putida* BF25 (figura no mostrada). Los aislamientos restantes no mostraron señales de hibridación positivas para ninguno de los grupos analizados, indicando la presencia de otros tipos de plásmidos en la colección de aislamientos seleccionados del biofiltro.

III.6. Evaluación de la movilidad de los plásmidos de la colección

Un aspecto importante de los plásmidos que merece ser analizado es su movilidad. Como se mencionó en la introducción de esta Tesis (Capítulo I, sección I.2.2.1.1), los plásmidos pueden clasificarse, de acuerdo a su movilidad, en conjugativos, movilizables y no transferibles. Este fenotipo puede determinarse por

ensayos de conjugación bacteriana entre los aislamientos portadores de plásmidos y cepas bacterianas de referencia, como *E. coli* o *P. putida*. En nuestro sistema de estudio, la mayor limitante para evaluar la movilidad plasmídica es el desconocimiento del contenido de los plásmidos (de los cuales en principio no conocemos más que su tamaño aproximado) y consecuentemente la falta de marcadores requeridos para la selección de transconjugantes. Sin embargo, dado que en los aislamientos seleccionados se encontraron algunos fenotipos de tolerancia a antibióticos y metales pesados, podría esperarse que algunos de los genes que confieren dichas resistencias posean localización plasmídica. Bajo esta hipótesis, en una primera etapa realizaron ensayos de conjugación triparentales entre los 35 aislamientos de la colección y cepas de *E. coli* (HB101 o CV601). Los transconjugantes se seleccionaron en medios conteniendo los 11 antibióticos y 7 metales (separadamente) que se emplearon para determinar la MIC de los aislamientos. Para las conjugaciones triparentales, se utilizaron cepas portadoras de plásmidos *helper* (esto es, plásmidos conjugativos que aportan los genes MPF necesarios para movilizar plásmidos de tipo movilizables). Las conjugaciones triparentales generaron transconjugantes para 6 aislamientos distintos en 7 condiciones de selección diferentes (ver Tabla III.3), indicando la presencia de plásmidos movilizables en estos aislamientos de la colección.

En una etapa posterior se realizaron conjugaciones biparentales con el propósito de determinar si estos 6 aislamientos, que generaron transconjugantes en conjugaciones triparentales, poseían plásmidos automovilizables (conjugativos). Se realizaron ensayos utilizando *E. coli* HB101 o CV601 y cada uno de los 6 aislamientos como cepas receptoras y dadoras, respectivamente. Los ensayos fueron repetidos varias veces empleando una relación [dadora:receptora] de aproximadamente $[10^8:10^8]$. Como resultado de las conjugaciones biparentales, se obtuvieron transconjugantes cuando se utilizó el aislamiento *Alcaligenes* sp. BF42 como dador y Hg^{+2} como agente de selección, indicado que este aislamiento posee un plásmido conjugativo portador de un gen u operón que confiere la resistencia a Hg^{+2} .

En el resto de las conjugaciones tri y biparentales no se detectaron transconjugantes. Esto sugiere que dichos plásmidos no serían movilizables/conjugativos o bien que se transferirían a una frecuencia muy baja (menor que 10^{-8} /célula donadora). Sin embargo, estos resultados deben considerarse con cautela. Por un lado la ausencia de marcadores de selección conocidos para el seguimiento de la transferencia de muchos plásmidos puede hacer que las frecuencias de conjugación sean a veces difíciles de detectar y cuantificar. Además, la expresión de genes del sistema de transferencia conjugativa puede estar altamente regulada y requerir de inductores que no siempre están presentes en las condiciones experimentales de laboratorio (More, *et al.*, 1996, Li, *et al.*, 1998). Asimismo, la transferencia conjugativa puede depender del *background* genómico de la cepa dadora y requiere que los plásmidos sean replicativos en las cepas de *E. coli* utilizadas como receptoras.

Como puede observarse en la Tabla III.3, el aislamiento de *Alcaligenes* BF42 posee dos plásmidos: uno conjugativo (de resistencia a Hg^{+2}) y otro movilizable (de resistencia a tetraciclina). Hemos comprobado que las tolerancias al antibiótico y al metal fueron transferidas independientemente. Otro punto interesante, es la presencia de plásmidos movilizables en tres aislamientos Gram-positivos: *Cellulosimicrobium cellulans* BF14, *Microbacterium* sp. BF36 y *Bacillus* sp. BF71, los

Tabla III.3. Plásmidos transmisibles y conjugativos detectados en los aislamientos de la colección

Aislamiento dador	Conjugación triparental	Conjugación biparental	Resistencia
<i>Cellulosimicrobium cellulans</i> BF14	+	-	estreptomicina, gentamicina
<i>Microbacterium</i> sp. BF36	+	-	estreptomicina
<i>Alcaligenes</i> sp. BF42	+	-	tetraciclina
<i>Alcaligenes</i> sp. BF42	+	+	Hg^{+2}
<i>Pseudomonas</i> sp. BF61	+	-	ácido nalidíxico
<i>Bacillus</i> sp. BF71	+	-	Zn^{+2}
<i>Pseudomonas</i> sp. BF75	+	-	Zn^{+2}

cuales confieren resistencia a gentamicina/estreptomicina, estreptomicina y Zn^{+2} , respectivamente, indicando la capacidad de tales plásmidos de replicar tanto en bacterias Gram-positivas como Gram-negativas.

III.7. Discusión

En este Capítulo hemos establecido una colección de 75 aislamientos bacterianos portadores de plásmidos de alto peso molecular, entre los cuales 35 fueron seleccionados como aislamientos representativos. Se realizó una primera caracterización genotípica de los aislamientos en la que se determinó la identidad de los mismos, y una caracterización fenotípica, analizándose su tolerancia a distintos antibióticos y metales pesados. Dado que estas bacterias son portadoras de plásmidos, era esperable encontrar fenotipos de resistencia ya que los plásmidos juegan un rol clave en la diseminación de este tipo de genes, no sólo en aquellos ambientes que presentan una presión de selección (como los hospitales o desechos animales en granjas), sino también en ambientes que no la presentan. Se han reportado plásmidos conteniendo genes de resistencia en ambientes prístinos libres de contaminación antropogénica (Allen, *et al.*, 2009, Berglund, 2015). Si bien los plásmidos son los elementos más importantes en el esparcimiento de estos genes, el hecho de que los aislamientos presenten un fenotipo de tolerancia a un antibiótico o metal no implica que esta resistencia se encuentre codificada en un plásmido.

En cuanto a los plásmidos presentes en los aislamientos seleccionados, se estimó que la colección de aislamientos representativos cuenta con al menos 50 plásmidos de alto peso molecular y mediante la técnica de lisis *in situ* y se determinó que sus tamaños varían entre 30 y 1.500 Kpb aproximadamente. Dos de estos plásmidos pudieron ser asignados a grupos de incompatibilidad plasmídica de amplio rango de huésped, mientras que los restantes no pudieron clasificarse. Esto sugiere la presencia de plásmidos de otros grupos de incompatibilidad plasmídica no ensayados en este trabajo, o bien de grupos que aún no han sido caracterizados. Por otro lado, se ha evaluado la movilidad de los plásmidos, encontrándose al menos 6 replicones movilizables y 1 conjugativo. Tal como mencionamos en la sección III.6, estos

resultados deben analizarse con cuidado dado que la metodología empleada sólo permite visualizar eventos de conjugación siempre y cuando haya un gen de resistencia a alguno de los antibióticos o metales utilizados en el ensayo; probablemente exista un mayor número de plásmidos en la colección que tengan un fenotipo móvil, pero que no contengan en sus esqueletos algún gen de resistencia. Asimismo, la transferencia conjugativa puede estar altamente regulada, depender del *background* genómico de la cepa dadora y además requiere que los plásmidos sean capaces de replicar en la cepa receptora utilizada.

Adicionalmente, encontramos que el aislamiento de *Alcaligenes* BF42 posee al menos dos plásmidos: uno conjugativo (de resistencia a Hg⁺²) y otro movilizable (de resistencia a tetraciclina), y este aislamiento mostró señal positiva en el abordaje utilizado para detectar replicones IncP-1. En particular, los plásmidos IncP-1 son relativamente grandes (50 a 500 kpb) y por lo general contienen otros MGEs asociados, tales como integrones, secuencias de inserción y transposones, llevando genes de resistencia a antibióticos y metales, así como agrupaciones de genes catabólicos (Dennis, 2005). A los plásmidos IncP-7 se los ha asociado con la degradación de contaminantes orgánicos y naturales (Dennis, 2005). La baja proporción de plásmidos similares a IncP dentro de la colección bacteriana sugiere que plásmidos de otros grupos de incompatibilidad podrían ser los responsables de la difusión de información en la comunidad BPS. Por otra parte, es concebible que estos plásmidos codifiquen funciones de replicación y genes de mantenimiento diferentes a los de los plásmidos caracterizados anteriormente de aislamientos clínicos y ambientales.

Para dar continuidad al estudio realizado en el presente capítulo, presentaremos en los capítulos siguientes los resultados referidos a la caracterización genotípica y al análisis *in silico* de los tipos de replicones y genes asociados a movilidad presentes en nuestra colección de aislamientos.

CAPÍTULO IV

**BÚSQUEDA DE OTROS ELEMENTOS GENÉTICOS
MÓVILES EN LOS AISLAMIENTOS REPRESENTA-
TIVOS DE LA COLECCIÓN BACTERIANA DE ESTUDIO**

IV.1. Introducción

Como hemos mencionado previamente, la selección del biofiltro utilizado en este trabajo se hizo en base a un *screening* previo de MGEs en el que se utilizó un enfoque independiente de cultivo que empleó ADN total de diversos ambientes naturales y enriquecidos/contaminados. Dealtry y colaboradores (Dealtry, *et al.*, 2014, Dealtry, *et al.*, 2014) detectaron la presencia de integrones de clase 1 y clase 2, así como genes usualmente ligados a estos MGEs (genes de resistencia a sulfonamidas, estreptomicina y a compuestos cuaternarios de amonio) en ADN total extraído de la comunidad del biofiltro. Los integrones pueden definirse como plataformas de ensamblado que permiten la incorporación de ORFs por recombinación sitio específica, convirtiéndolos en genes funcionales al asegurar su correcta expresión (Mazel, 2006). Si bien estos elementos fueron inicialmente descritos en aislamientos clínicos, posteriormente se los ha encontrado en otros ambientes, como aguas y suelos, y hoy se sabe que están presentes en aproximadamente el 10% de los genomas bacterianos que han sido secuenciados (Domingues, *et al.*, 2012). Los integrones pueden clasificarse en dos grandes grupos: **i)** los súperintegrones, que se encuentran localizados en cromosomas bacterianos y sólo se asocian esporádicamente a genes de resistencia; en general, estos integrones son de gran tamaño y contienen más de 20 genes en casete; y **ii)** los integrones móviles, usualmente asociados a plásmidos o transposones, que presentan casetes, en su mayoría de resistencia a antibióticos, agrupados en arreglos relativamente pequeños (Gillings, 2014). A su vez, estos últimos pueden clasificarse en los tipos 1, 2 y 3, de acuerdo a la secuencia aminoacídica de su integrasa. Las integrasas de clase 1 y 2 (*int1* e *int2*, respectivamente) difieren en más del 50% de su secuencia, aunque los integrones de clase 1 y 2 pueden incorporar el mismo tipo de genes (Deng, *et al.*, 2015). Estos integrones pueden moverse fácilmente por transferencia horizontal, lo cual genera un flujo dinámico de información en poblaciones bacterianas, dando como resultado la adquisición de nuevas funciones.

Si bien los integrones pueden ser muy diversos entre sí, todos ellos poseen una estructura mínima que incluye tres elementos generales: un gen que codifica para una

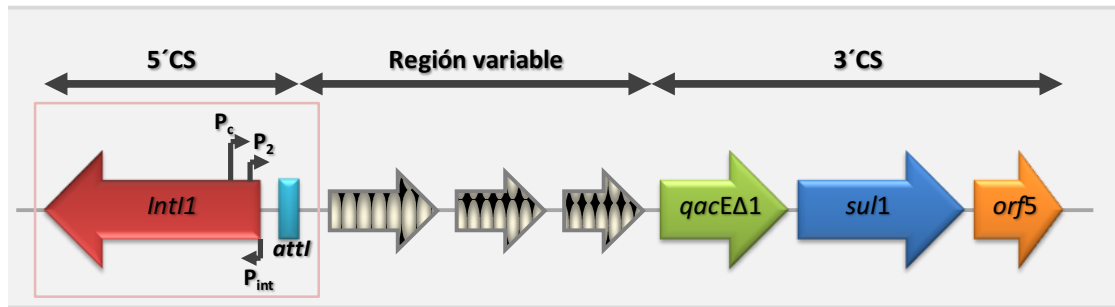


Figura IV.1. Estructura típica de un integrón de clase 1. Se observa en rojo el gen que codifica para la integrasa (*int11*), con su promotor (P_{int}) y los dos promotores ubicados dentro de este gen con orientación contraria (P_c y P_2), que permiten la expresión de los genes presentes en la región variable (gris). En verde, azul y naranja se indican los genes *qacEΔ1*, *sul1* y *orf5*, respectivamente, que forman parte del fragmento 3' de secuencia conservada (3'CS). El recuadro rosa contiene los elementos mínimos presentes en un integrón.

integrasa (*int1*), un sitio adyacente de recombinación (*attI*) y al menos un promotor, aunque a veces dos (P_c , P_2), orientados para la expresión de los genes capturados. El gen *int1* contiene su propio promotor y en la mayoría de los casos está en orientación opuesta respecto de los genes codificados en los casetes. La proteína para la que codifica este gen, la integrasa, pertenece a la familia de las tirosina-recombinasas y cataliza la recombinación entre secuencias específicas, necesarias para la integración y escisión de genes en casete. En la Figura IV.1 se muestra la estructura general simplificada de un integrón de clase 1 típico, y se resaltan los elementos mínimos, común a todos ellos. En la misma figura, pueden visualizarse tres zonas principales: **i)** el fragmento 5' conservado (5'CS), el cual contiene el gen de la integrasa *int11*, el sitio de recombinación (*attI*) y uno o dos promotores que controlan la expresión de los casetes génicos incorporados en la región variable, **ii)** el fragmento 3' conservado (3'CS) que está menos definido, aunque generalmente incluye los genes *qacEΔ1* (que confiere resistencia a compuestos cuaternarios de amonio), *sul1* (de resistencia a sulfonamidas) y el *orf5* (de función desconocida); y **iii)** una región variable, comprendida entre ambos segmentos conservados, que suele contener un número variable de casetes o bien no poseer ninguno. En total, se han descrito alrededor de 130 casetes de genes de resistencia a antibióticos diferentes en integrones clase 1, junto con algunos otros casetes génicos de función desconocida (Gillings, 2014). Sin

embargo, lo más habitual es la presencia de dos o tres casetes (Di Conza & Gutkind, 2010).

Un punto que es importante resaltar respecto a los integrones es que, si bien son considerados MGEs por su capacidad de adquirir casetes génicos y propagarse por los genomas, éstos no tienen la capacidad de hacerlo por sí mismos, ya que no poseen funciones asociadas a la automovilidad. Por este motivo se asocian frecuentemente a otros MGEs, como los plásmidos, los cuales promueven su transferencia entre bacterias. En este Capítulo nos concentraremos en la búsqueda y caracterización de los integrones móviles de clase 1 y 2, cuya presencia ha sido demostrada previamente en la muestra del biofiltro enriquecido en pesticidas que seleccionamos (Dealtry, *et al.*, 2014, Dealtry, *et al.*, 2014) y poco se conoce acerca de la estructura de estos elementos en ambientes no clínicos. Recientemente Gillings y colaboradores (Gillings, *et al.*, 2015) revisaron la relación entre los impactos antropogénicos y la abundancia de *int11* y delinearon un enfoque mediante el cual *int11* podría servir como un indicador de contaminación antropogénica.

A partir de una muestra de ADN total del mismo biofiltro, Dunon y colaboradores (Dunon, *et al.*, 2013) detectaron además la presencia de la secuencia de inserción *IS1071*. Su estudio también sugiere que las bacterias portadoras de *IS1071* proliferan como respuesta a la exposición a pesticidas y estos autores han hecho hincapié en el papel ecológico de estos MGEs. En este sentido, ha sido reportado que esta secuencia de inserción puede contener genes u operones enteros asociados a la degradación de compuestos xenobióticos, como ácido 2,4-diclorofenoxiacético (Vedler, *et al.*, 2000), atrazina (Martinez, *et al.*, 2001), y clorobenzoato (Nakatsu, *et al.*, 1991), entre otros, por lo que su identificación suele usarse como un posible indicador de la presencia de este tipo de genes.

Por las razones antes expuestas, en este capítulo realizamos un *screening*, utilizando metodologías moleculares clásicas, para la búsqueda e identificación de la secuencia de inserción *IS1071* e integrones clase 1 y 2, así como de otros genes que frecuentemente se encuentran formando parte de la estructura típica de estas

plataformas de ensamblado (*qacE/qacEΔ1*, *sul1* y *orf5*). Finalmente indagaremos la localización genómica de los MGEs (integrones e *IS1071*) que detectemos.

IV.2. Búsqueda y caracterización de integrones

Esta sección tiene como objetivo la búsqueda, identificación y caracterización de integrones de clase 1 y 2 en los aislamientos de la colección previamente establecida. La evaluación de la presencia y caracterización de estas plataformas genéticas en cepas o comunidades bacterianas, es generalmente abordada por una búsqueda preliminar del gen de la integrasa *intl1* e *intl2* para integrones de clase 1 y clase 2, respectivamente, el cual es indicativo de la presencia de integrones. Si bien estos genes presentan variaciones de secuencia entre distintos microorganismos, así como versiones truncadas, se han podido diseñar exitosamente diversos pares de *primers* a partir de regiones altamente conservadas que permiten la detección eficiente de estos genes (Hoyle, *et al.*, 2006).

Nuestro estudio fue abordado en dos etapas: la primera consistió en un *screening* por PCR para detectar la presencia de integrones clase 1 y 2. Luego, en una etapa posterior, para aquellos aislamientos con amplificación positiva para *intl* se determinó la composición de la región variable.

IV.2.1. Detección de los genes *intl1* e *intl2* en los aislamientos

Para comenzar con la búsqueda de integrones en los aislamientos, en una primera instancia se realizó un *screening* por PCR del gen *intl1*, el cual codifica para la integrasa y se ubica en el segmento 5' conservado de la estructura génica de los integrones de clase 1. Esta integrasa es imprescindible para la incorporación y escisión de casetes génicos que forman parte de los integrones. El gen *intl1* presenta un tamaño aproximado de 1.000 pb, y para su identificación se utilizó un par de *primers* (Int1FP/int1RP) que amplifican un fragmento interno de este gen de 483 pb (Materiales y Métodos, Tabla II.4). Los resultados obtenidos se muestran en la Figura IV.2. Como puede observarse, 6 de los 35 aislamientos analizados presentan un

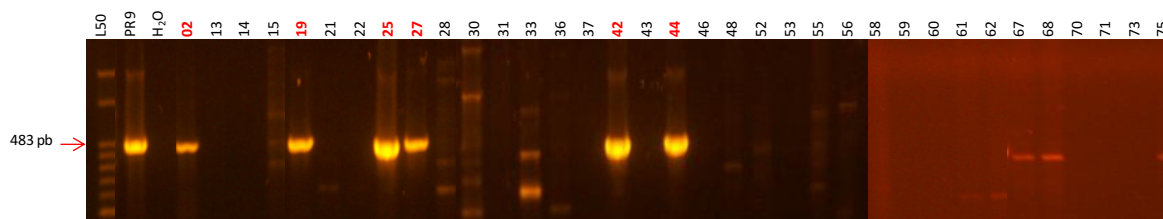


Figura IV.2. Detección del gen *int1* por PCR. Se muestran los productos de amplificación por PCR corridos en gel de agarosa 1,5%. En cada calle se indica el número de aislamiento. PR9 corresponde a la cepa de *P. mirabilis* usada como control positivo y H₂O corresponde al control negativo. L50: marcador de peso molecular. En rojo se indican los aislamientos que dieron un amplicón del tamaño esperado.

amplicón del tamaño esperado, sugiriendo la presencia de integrones clase 1 en dichos aislamientos. En segundo lugar, se realizó el *screening* de la integrasa *intI2* perteneciente a integrones de clase 2 utilizando los *primers* Inti2F/Inti2R descritos en Materiales y Métodos (Tabla II.4), que amplifican un fragmento de 466 pb. Esta PCR no arrojó resultados positivos, es decir, no hemos podido detectar este gen (o bien no está presente) en ninguno de los aislamientos analizados, sugiriendo la ausencia de integrones de clase 2. Por este motivo, de aquí en adelante continuaremos con la identificación de los genes que forman parte de la estructura conservada de los integrones de clase 1.

IV.2.2. Detección del gen *su1*

El gen *su1* codifica para una proteína que proporciona resistencia a sulfonamidas (antimicrobianos sintéticos). Este gen se encuentra usualmente ubicado en el segmento 3'CS de los integrones clase 1, el cual está menos definido que el 5'CS. Con el fin de detectar la presencia del gen *su1* en los aislamientos, se realizó un *screening* por PCR de dicho gen utilizando el juego de *primers* Sul1Lower/Sul1Upper (Materiales y Métodos, Tabla II.4) que amplifican un fragmento de 580 pb. Los resultados mostrados en la Figura IV.3 evidencian la presencia de productos de amplificación del tamaño esperado en 8 aislamientos de la colección. En el caso del aislamiento BF30, se observa una banda de mayor tamaño al del control positivo pero de intensidad fuerte. Este resultado podría indicar la presencia de una variante del gen

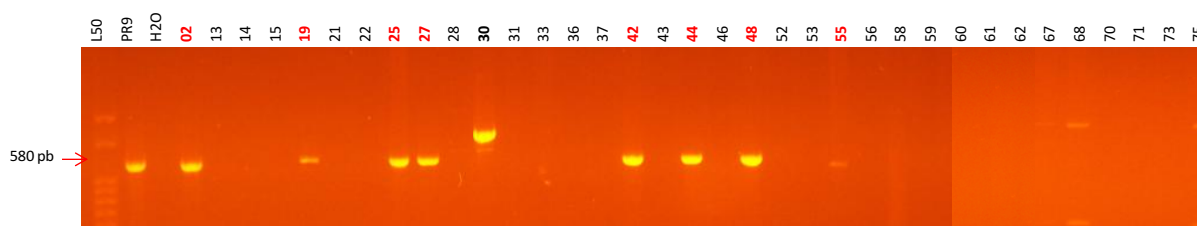


Figura IV.3. Detección del gen *sul1* por PCR. Se muestran los productos de amplificación por PCR corridos en gel de agarosa 1,5%. En cada calle se indica el número de aislamiento. En rojo se marcan los aislamientos que dieron positivo para el gen *sul1*. PR9 es la cepa de *P. mirabilis* utilizada como control positivo y H₂O el control negativo. L50: marcador de peso molecular.

sul1 en este aislamiento, o bien podría deberse a la amplificación de un producto inespecífico. En este sentido, y considerando resultados posteriores (el aislamiento BF30 no dio amplificación positiva para ningún otro gen asociado a integrones clase 1 analizado), se decidió no secuenciar dicho amplicón ni avanzar más en este punto.

Si se tiene en cuenta que sólo 6 de los 8 aislamientos positivos para *sul1* mostraron amplificación positiva también para el gen *int11*, se puede deducir que al menos 2 de estos aislamientos contienen el gen *sul1* en ausencia de la integrasa del integrón de clase 1. Si bien el gen *sul1* se encuentra frecuentemente asociado a integrones clase 1, se ha reportado su presencia en otros MGEs como plásmidos (Phuong Hoa, *et al.*, 2008, Wu, *et al.*, 2010), y tampoco puede descartarse su localización cromosomal independiente de la presencia de este tipo de elementos. Por otro lado, hay autores que han descrito la ausencia de este gen en hasta un 10% de los integrones clase 1 estudiados (Leverstein-Van Hall, *et al.*, 2002). Finalmente, cabe mencionar que la presencia de *sul1* en un aislamiento *int11* positivo no implica que dicho gen se encuentre necesariamente formando parte del integrón.

IV.2.3. Detección de los genes *qacE/qacEΔ1*

El gen *qacE* codifica para una proteína de eflujo que media la resistencia a compuestos cuaternarios de amonio (que son utilizados como antisépticos) y a colorantes intercalantes, como el bromuro de etidio. Este gen se encuentra frecuentemente en su versión truncada *qacEΔ1*, que es un derivado semifuncional de

dicho gen (Paulsen, *et al.*, 1993) y forman parte del segmento 3'CS de los integrones de clase 1.

Con motivo de determinar si los genes *qacE* y *qacEΔ1* se encuentran presentes en los aislamientos de nuestra colección bacteriana, se realizó su análisis por PCR utilizando dos juegos de *primers*: QaceF/QaceR y QacEΔelta1A/ QacEΔelta1B (Materiales y Métodos, Tabla II.4). Como puede observarse en la figura IV.4.A, 12 de los 35 aislamientos en estudio presentan un amplicón del tamaño esperado (240 pb) para el gen *qacE*.

La amplificación por PCR para el gen *qacEΔ1* generó varios productos inespecíficos (figura IV.4.B), incluso cuando se aumentó la temperatura de *annealing* hasta 64°C. Sin embargo, 5 aislamientos mostraron una banda de tamaño similar a la observada en el control positivo. Para corroborar que estos productos corresponden al gen *qacEΔ1* y descartar la posibilidad de que sean productos inespecíficos de amplificación, se procedió a la elución de estas bandas del gel de agarosa y purificación para su posterior secuenciamiento. Las secuencias obtenidas se analizaron en la base de datos de NCBI y su identificación se realizó utilizando BLASTx. Los 5 amplicones

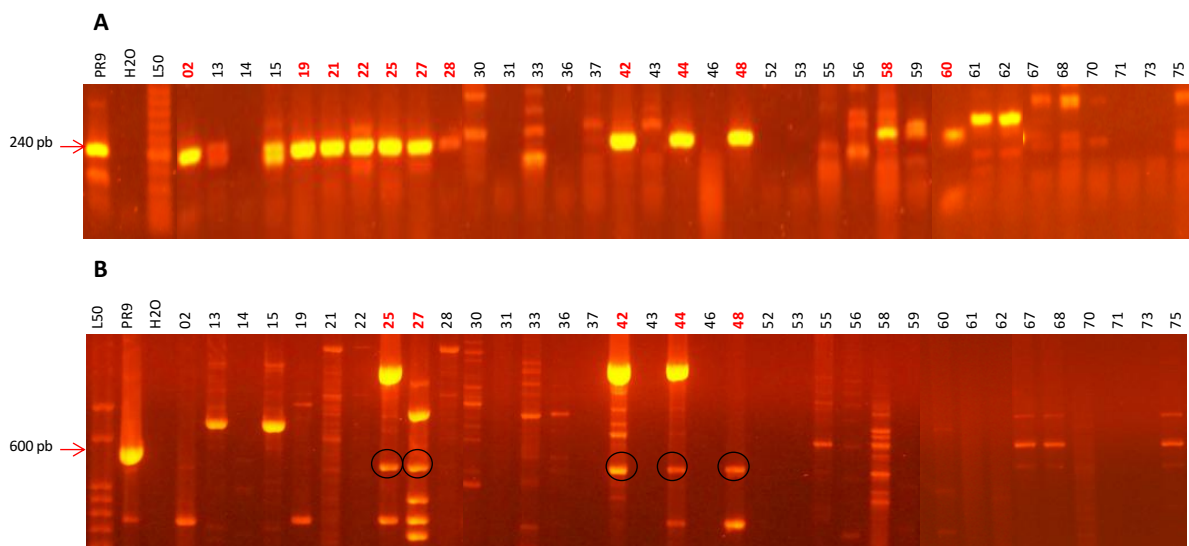


Figura IV.4. Detección del gen *qacE* (A) y *qacEΔ1* (B) por PCR. Se muestran los productos de amplificación por PCR corridos en gel de agarosa 1,5%. En cada calle se indica el número de aislamiento. En rojo se marcan los aislamientos que dieron positivo para el gen *qacE*. PR9 es el control positivo y H₂O el control negativo. L50: marcador de peso molecular. Con círculos se indican las bandas del gel que fueron purificadas y secuenciadas.

analizados mostraron un alto porcentaje de identidad con el gen *qacE* Δ 1.

Los resultados obtenidos en esta sección sugieren que un alto porcentaje de los aislamientos de la colección poseen alguna versión del gen *qacE*, lo cual indica su presencia en otro lugar del genoma bacteriano. Estos resultados son esperables, dado que ambas versiones del gen *qacE* no son exclusivas de los integrones de clase 1. Por el contrario, se los encuentra muy frecuentemente presentes en cromosomas, plásmidos u otros MGEs en bacterias (Kazama, *et al.*, 1998, Gillings, *et al.*, 2009).

IV.2.4. Detección del *orf5*

El *orf5* corresponde a un marco de lectura abierto de función aún desconocida. Este gen se ha encontrado frecuentemente acompañando a los genes *sul1* y *qacE/qacE* Δ 1 en el segmento 3'CS. La presencia de este ORF se evaluó por PCR para los 35 aislamientos en estudio, utilizando el juego de *primers* Orf513F/Orf513R (Materiales y Métodos, Tabla II.4), el cual amplifica un fragmento de 474 pb.

En ningún caso se detectó amplificación del fragmento correspondiente, por lo cual se puede decir que ninguno de los aislamientos portaría este gen o bien que podría existir dentro de nuestra colección de aislamientos variantes de dicho marco de lectura abierto de función desconocida que no pueden ser detectadas con el par de cebadores empleados.

En la tabla IV.3, al final de este capítulo, se muestra un resumen de los resultados obtenidos para la identificación de todos los genes asociados a integrones de clase 1.

IV.2.5. Amplificación y análisis de las regiones variables de los integrones clase 1 presentes en los aislamientos

Como se mencionó previamente, los integrones poseen una zona variable en la cual se pueden encontrar casetes génicos que carecen de promotor propio y que pueden ser correctamente expresados gracias a la presencia del promotor ubicado en el gen *intI1* corriente arriba de la zona variable y adecuadamente orientado. A fin de investigar qué genes podrían estar formando parte de las zonas variables de los

integrones detectados en los aislamientos analizados, se procedió a amplificar dichas regiones para su posterior secuenciamiento y su análisis *in silico*. Para ello se utilizaron *primers* previamente diseñados que alinean con secuencias conservadas de las regiones 5'CS y 3'CS, flanqueando esta región. La amplificación de las regiones variables con distintos juegos de *primers* se realizó para los 6 aislamientos en los que previamente se había detectado amplificación positiva para el gen de la integrasa (*intI1*) presente en integrones de clase 1. Se utilizó un único *primer* directo, *sulpro3*, que se encuentra en el segmento 5'CS, en combinación con 7 *primers* reversos: 3'CSNew, HS287, AJH60, TniR y HS286, IS26F e IS26R (Materiales y Métodos, Tabla II.4), complementarios al segmento 3'CS del integrón.

De las 7 combinaciones de *primers* ensayadas, 3 de ellas (*sulpro3*/3'CSnew, *sulpro3*/HS287 y *sulpro3*/AJH60) permitieron amplificar la región variable en 3 aislamientos distintos. Como puede verse en la Figura IV.5, el juego *sulpro3*/3'CSnew amplificó un fragmento de alrededor de 1.200 pb en los aislamientos de *Ochrobactrum* sp. BF02, *Pseudomonas putida* BF25 y *Pseudomonas* sp. BF27; el par *sulpro3*/HS287 amplificó un fragmento de la misma longitud en los aislamientos BF25 y BF27, mientras que el juego *sulpro3*/AJH60 permitió amplificar la zona variable en el aislamiento BF25, también de 1.200 pb. Las 4 combinaciones de *primers* restantes (3'CSNew/TniR, 3'CSNew/HS286, 3'CSNew/IS26F y 3'CSNew/IS26R) arrojaron resultados negativos con todos los moldes de ADN ensayados.

Estos resultados indican que al menos 3 de los aislamientos analizados poseen una zona variable de aproximadamente 1.200 pb. El resto de los aislamientos, para los que no se obtuvo amplificación, probablemente contengan un segmento 3'CS cuya

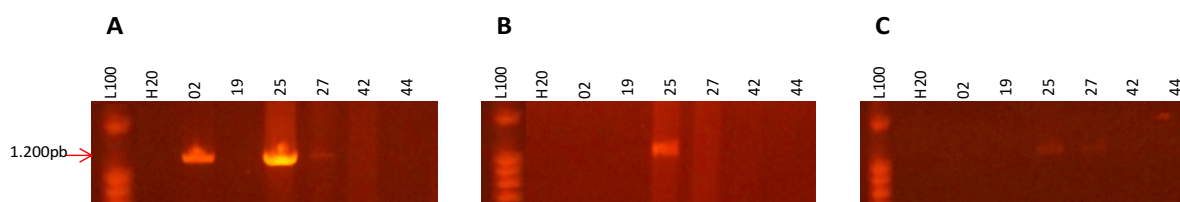


Figura IV.5. Amplificación de la región variable de los integrones de clase 1 detectados en la colección de aislamientos utilizando distintos juegos de *primers*: A) *sulpro3*/ 3'CS new. B) *sulpro3* /HS287. C) *sulpro3*/ AJH60. Los productos de PCR se corrieron en un gel de agarosa al 0,8%.

Tabla IV.1. Identidad a nivel de secuencia nucleotídica (BLASTn) de las regiones variables amplificadas y secuenciadas

Aislamiento	Descripción	Cobertura	E value	Identidad	N° acceso
BF02	<i>Salmonella enterica</i> subespecie entérica serovar <i>Typhimurium</i> , plásmido incHI2	98%	0.0	92%	LN794248.1
BF25	<i>Klebsiella pneumoniae</i> cepa 34618, genoma completo	98%	0.0	99%	CP010392.1
BF27	<i>Klebsiella pneumoniae</i> cepa 34618, genoma completa	100%	0.0	99%	CP010392.1

Tabla IV.2. Identidad a nivel de secuencia aminoacídica (BLASTx) de las regiones variables amplificadas y secuenciadas

Aislamiento	Descripción	Organismo	Cobertura	E value	Identidad	N° acceso
BF02	aminoglicósido nucleotidiltransferasa ANT3	<i>Raoultella planticola</i>	85%	0.0	91%	WP_022635574.1
BF25	aminoglicósido nucleotidiltransferasa ANT3	<i>Escherichia coli</i>	84%	0.0	100%	WP_001297038.1
BF27	aminoglicósido nucleotidiltransferasa ANT3	<i>Escherichia coli</i>	85%	0.0	100%	WP_001297038.1

secuencia nucleotídica difiera en algún grado de la descrita. Dado el origen ambiental de los aislamientos analizados, no es sorprendente que el extremo 3'CS de estos integrones difiera de los previamente depositados en bases de datos, que son mayoritariamente de origen clínico.

Los amplicones de la zona variable generados por PCR en los tres aislamientos fueron secuenciados utilizando los productos generados con la combinación de *primers* sulpro3/3'CSNew. Se determinó la identidad de las secuencias de las regiones variables amplificadas, realizando una búsqueda de similitud a nivel ADN (BLASTn) y a nivel proteína (BLASTx). Esta búsqueda se realizó en la base de datos de NCBI. Los resultados se muestran en las Tablas IV.1 y IV.2.

En los tres casos, el fragmento amplificado de aproximadamente 1.200 pb contiene un gen que codifica para una aminoglicósido nucleotidil-transferasa (*aadA*), la cual confiere resistencia a estreptomicina/ espectinomicina (Gestal 2005). A nivel de secuencia nucleotídica, puede verse que en los aislamientos BF25 y BF27 esta secuencia es similar y posee un 99% de identidad con la misma secuencia presente en el genoma de *Klebsiella pneumoniae*, mientras que la secuencia del aislamiento BF02 posee una menor identidad con la misma secuencia presente en el plásmido IncHI2 hallado en un aislamiento de *Salmonella typhimurium*.

Se han reportado numerosos casos en los que este gen *aadA* se encuentra integrado en la región variable en integrones de clase 1 (Michael, *et al.*, 2005, Binh, *et al.*, 2009, Lopes, *et al.*, 2014, Shahcheraghi, *et al.*, 2014). Si bien los aislamientos BF02, BF25 y BF27 presentan este gen en su zona variable, en el aislamiento BF02 el porcentaje de identidad con el gen *aadA* más cercano depositado en la base de datos NCBI es sólo de 91%, por lo cual se podría decir que este integrón posee una nueva variante de la familia *aadA*, de la cual hasta la fecha se han descrito al menos 26 proteínas distintas (depositadas en la base de datos de NCBI: <http://www.ncbi.nlm.nih.gov/>). En el caso de los aislamientos BF25 y BF27, ambos genes poseen un 100% de identidad con el gen *aadA* más cercano depositado en esta base de datos.

IV.2.6. Secuenciamiento y análisis de los genes *intI1* identificados en los aislamientos

Para realizar un análisis genético de las secuencias de las 6 integrasas identificadas en los aislamientos, se procedió a realizar un análisis filogenético de las secuencias obtenidas del gen *intI1* amplificado por PCR utilizando los cebadores *intI1F/intI1R* (Tabla II.4 de Materiales y Métodos). En el caso particular del aislamiento BF19 no logramos amplificar (en ninguna de las diversas condiciones ensayadas) el gen entero, por lo cual se amplificó y secuenció un fragmento de 486 pb obtenido con el juego de *primers* *intI1FP/intI1RP* que previamente hemos utilizado para realizar el

screening de las integrasa asociada a los integrones clase 1 (sección 2.1 de este capítulo).

Las 6 secuencias de obtenidas fueron comparadas filogenéticamente utilizando el método de *Neighbour Joining*, como se indica en la sección II.3.8.1 de Materiales y Métodos. Los genes *intl1* obtenidos fueron comparados con otros previamente reportados por Nardelli y colaboradores (2012) quienes identificaron alelos de este gen en ambientes no clínicos con distintos grados de contaminación antropogénica y los compararon con otros recopilados de distintos ambientes (clínicos y no clínicos) depositados en GenBank. En la figura IV.6 se muestra el árbol filogenético obtenido, y se indica el origen de cada alelo. Como puede observarse, las seis integrasas pertenecientes a los integrones identificados en los aislamientos de nuestra colección son similares entre sí, agrupándose en un mismo clado. El alineamiento de las seis secuencias *intl1* obtenidas utilizando CLUSTALW mostró diferencias sólo en 4 posiciones puntuales (a excepción de la integrasa BF19, que carece del extremo 5' del gen, aunque el fragmento secuenciado es similar a las 5 integrasas restantes; ver Material Suplementario, S1). La comparación de las secuencias *intl1* con aquéllas depositadas en la base de datos de GenBank mediante BLASTx reveló que las 5 integrasas que pudieron ser completamente secuenciadas poseen un 99% de identidad a nivel aminoacídico con una integrasa de *Escherichia coli* BIDMC 82 y un 99% de identidad a nivel nucleotídico con la integrasa presente en el plásmido pH11 de *Klebsiella pneumoniae* (en los aislamientos BF02, BF42, BF44) o con la integrasa presente en el plásmido pHK0653 de *Salmonella enterica* (para los aislamientos BF25 y BF27). En cuanto al fragmento de la integrasa obtenido a partir del aislamiento BF19, éste resultó 100% idéntico al gen previamente descrito en *Acinetobacter baumannii* ABNIH6, que también está truncada. Esta observación nos permite suponer que podría tratarse de un pseudogén que posee el extremo 5' (amino terminal de la proteína) truncado, lo cual explica la imposibilidad de amplificar el gen completo en este aislamiento.

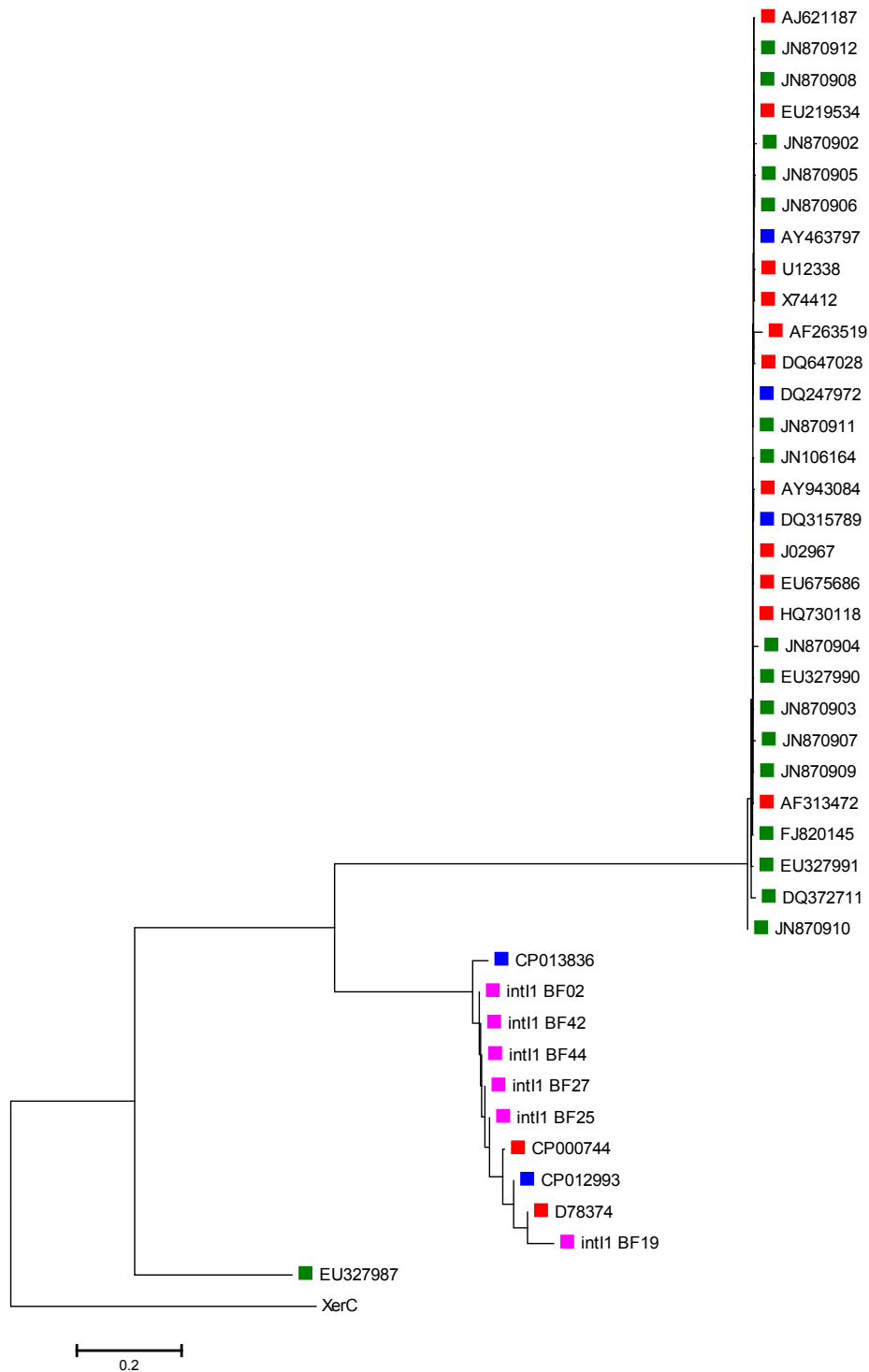


Figura IV.6. Árbol filogenético para el gen *int11*. En la figura se comparan las seis secuencias del gen *int11* obtenidas en este trabajo (rosa) con otros alelos previamente recopilados de distintas fuentes por Nardelli y colaboradores (2012) y depositados en GenBank: en rojo se marcan los aislamientos de origen clínico, en verde aquéllos de origen no clínico, y en azul aquéllos que han sido encontrados tanto en ambientes clínicos como no clínicos. El árbol filogenético fue obtenido con el programa MEGA 5.0 utilizando parámetros preestablecidos, a partir de un alineamiento utilizando CLUSTALW e inferido por el método de *Neighbour Joining*.

IV.6.7. Localización de los integrones

Para avanzar con el estudio del rol en la biología e importancia evolutiva de los integrones en los aislamientos recuperados del biofiltro, nos propusimos analizar si los integrones clase 1 detectados se encontraban dentro de los esqueletos plasmídicos o si poseían una localización cromosomal. Para ello, se realizó un gel de lisis *in situ* (Materiales y Métodos, sección II.2.4.3) para los 6 aislamientos positivos para *intl1* y se eluyeron las bandas correspondientes a los plásmidos de distinta movilidad electroforética y al ADN cromosomal. Luego, se realizó la amplificación por PCR para el gen *intl1*, utilizando el ADN extraído de cada banda, con el mismo juego de *primers* que se utilizó en el *screening* de este gen. Los resultados obtenidos permitieron determinar la localización plasmídica de los integrones clase 1 en 4 (BF02, BF25, BF27 y BF42) de los 6 aislamientos analizados, mientras que para el aislamiento BF19, que presentaba la integrasa truncada en su extremo N-terminal, se determinó una localización cromosomal. Resta aún dilucidar la localización del integron clase 1 en el aislamiento BF44. En este caso no se pudo determinar su ubicación debido a la proximidad de las bandas plasmídica y cromosómica en el gel de lisis *in situ*, lo que derivó en amplicones positivos cuando se empleó como molde ADN extraído de ambas bandas.

IV.3. Detección de la secuencia de inserción IS1071 en los aislamientos

Curiosamente, las secuencias de inserción IS1071 son a menudo encontradas en plásmidos, principalmente los del grupo IncP-1, en los que frecuentemente flanquean a grupos de genes catabólicos formando transposones compuestos. La identificación de la secuencia de inserción IS1071, se realizó mediante PCR a partir de un set de *primers* (Materiales y Métodos, Tabla II.4) que amplifican el gen de la transposasa, *tnpA*, asociada a este elemento. Como resultado, se encontraron 24 aislamientos positivos para este marcador, lo cual indica una gran abundancia de esta transposasa en la colección bacteriana estudiada. Los resultados obtenidos se muestran en la Tabla IV.3. A fin de corroborar que se tratara de esta secuencia en

Tabla IV.3. Resumen de los resultados obtenidos por PCR para los determinantes de integrones de clase 1 y la secuencia de inserción IS1071

Aislamiento	Integrones clase 1					IS1071
	<i>intl1</i>	<i>sul1</i>	<i>qacE</i>	<i>qacEΔ1</i>	<i>orf513</i>	<i>tnpA</i>
BF02 <i>Ochrobactrum</i> sp.	+	+	+	-	-	+
BF13 <i>Ochrobactrum</i> sp.	-	-	-	-	-	+
BF14 <i>Cellulosimicrobium cellulans</i>	-	-	-	-	-	+
BF15 <i>Ochrobactrum</i> sp.	-	-	-	-	-	+
BF19 <i>Variovorax</i> sp.	+	+	+	-	-	+
BF21 <i>Paenibacillus</i> sp.	-	-	+	-	-	+
BF22 <i>Paenibacillus tundrae</i>	-	-	+	-	-	+
BF25 <i>Pseudomonas putida</i>	+	+	+	+	-	+
BF27 <i>Pseudomonas</i> sp.	+	+	+	+	-	+
BF28 <i>Paenibacillus xylanexedens</i>	-	-	+	-	-	-
BF30 <i>Bordetella</i> sp.	-	?	-	-	-	+
BF31 <i>Cellulosimicrobium</i> sp.	-	+	-	-	-	-
BF33 <i>Pseudomonas putida</i>	-	-	-	-	-	+
BF36 <i>Microbacterium</i> sp.	-	-	-	-	-	+
BF37 <i>Serratia</i> sp.	-	-	-	-	-	+
BF42 <i>Alcaligenes</i> sp.	+	+	+	+	-	+
BF43 <i>Serratia</i> sp.	-	-	-	-	-	+
BF44 <i>Microbacterium oxydans</i>	+	+	+	+	-	+
BF46 <i>Serratia</i> sp.	-	-	-	-	-	-
BF48 <i>Microbacterium</i> sp.	-	+	+	+	-	+
BF52 <i>Brevibacterium</i> sp.	-	-	-	-	-	+
BF53 <i>Agromyces cerinus</i>	-	-	-	-	-	-
BF55 <i>Pseudomonas putida</i>	-	+	-	-	-	-
BF56 <i>Acinetobacter</i> sp.	-	-	-	-	-	-
BF58 <i>Pseudomonas</i> sp.	-	-	+	-	-	-
BF59 <i>Sphingobacterium</i> sp.	-	-	-	-	-	+
BF60 <i>Sphingobacterium</i> sp.	-	-	+	-	-	-
BF61 <i>Pseudomonas</i> sp.	-	-	-	-	-	+
BF62 <i>Pseudomonas</i> sp.	-	-	-	-	-	-
BF67 <i>Pseudomonas</i> sp.	-	-	-	-	-	-
BF68 <i>Pseudomonas</i> sp.	-	-	-	-	-	+
BF70 <i>Bacillus pumilus</i>	-	-	-	-	-	-
BF71 <i>Bacillus</i> sp.	-	-	-	-	-	+
BF73 <i>Bacillus megaterium</i>	-	-	-	-	-	+
BF75 <i>Pseudomonas</i> sp.	-	-	-	-	-	+

particular, cinco productos de PCR elegidos al azar fueron secuenciados. Las secuencias obtenidas se analizaron en la base de datos ISFinder, la cual contiene una gran compilación de secuencias de inserción, y en la base de datos de GenBank. Ambos análisis indicaron que los fragmentos de PCR analizados correspondían a la secuencia del gen *tnpA* asociado a la IS1071.

Dado que se ha sido descrita a esta secuencia de inserción formando parte de esqueletos plasmídicos, particularmente aquéllos pertenecientes al grupo de incompatibilidad IncP (Tan, 1999, Ma, *et al.*, 2007, Szczepanowski, *et al.*, 2008), se procedió a analizar la presencia del gen *tnpA* en los transconjugantes obtenidos en las conjugaciones tri y biparentales detalladas previamente (Capítulo III, sección III.6). Nos propusimos evaluar si en los aislamientos portadores de plásmidos movilizables y conjugativos la IS1071 presenta localización plasmídica. Los resultados de la amplificación por PCR de secuencias específicas del gen *tnpA* en los transconjugantes obtenidos indicaron que en 4 de los 7 plásmidos movilizables (presentes en los aislamientos *Microbacterium* sp. BF36, *Pseudomonas* sp. BF61 y dos en *Alcaligenes* sp. BF42) la IS1071 se encontraba presente en dicho elemento.

IV.4. Discusión

En esta sección se ha analizado la presencia de dos MGEs potencialmente asociados a genes de resistencia a antibióticos y compuestos cuaternarios de amonio (integrones) y genes catabólicos (IS1071).

El trabajo realizado nos ha permitido identificar la presencia de integrones de clase 1 en 6 de los 35 aislamientos de la colección bacteriana. De los 6 aislamientos en los que se detectó la presencia de integrones, 5 corresponden a los géneros *Ochrobactrum*, *Variovorax*, *Pseudomonas* y *Alcaligenes* (todos ellos Gram-negativos) y uno a *Microbacterium* (Gram-positivo). Los integrones de clase 1 son considerados de alta relevancia clínica, dado que en estos ambientes es en donde más se los ha encontrado y estudiado. Esta clase de integrones en particular resulta de gran interés debido a su amplia diseminación entre proteobacterias y su implicancia en la

propagación de genes de resistencia a antibióticos y desinfectantes, lo cual genera un grave problema a la hora de erradicar microorganismos potencialmente patógenos en ambientes hospitalarios. De hecho, se ha reportado en reiteradas ocasiones la presencia de integrones de clase 1 en altos porcentajes (hasta más del 50%) de bacterias aisladas de ambientes clínicos (Yousefi, *et al.*, 2010, Nikokar, *et al.*, 2013, Ahangarkani, *et al.*, 2015). Sin embargo, estos elementos no son exclusivos de ambientes nosocomiales; por el contrario, se han encontrado en otros ambientes como suelos, aguas y sedimentos, aunque en estos casos la frecuencia de aparición es mucho menor que en hospitales. Trabajos previamente publicados indican frecuencias que oscilan entre valores menores al 1% y poco más del 10% cuando se trata de aislamientos no hospitalarios (Rosser & Young, 1999, Gaze, *et al.*, 2005, Stokes, *et al.*, 2006, Rosewarne, *et al.*, 2010, Nardelli, *et al.*, 2012). En nuestro caso, un 17% de los aislamientos de la colección transporta integrones de este tipo, lo cual indica una alta ocurrencia de estos elementos. También fueron identificados en una gran proporción de los aislamientos los genes *qacE* (34%), *qacEΔ1* (14%) y *sul1* (25%). El hecho de que haya un porcentaje mayor de aislamientos conteniendo estos genes indica que los mismos, en muchos casos, serían dispersados en el ambiente de manera independiente de los integrones de clase 1.

Teniendo en cuenta que se ha reportado la presencia de integrones de clase 1 en campos y granjas donde se crían animales y en otros ambientes contaminados con pesticidas y otros compuestos contaminantes (Binh, *et al.*, 2009, Jechalke, *et al.*, 2013, Dealtry, *et al.*, 2014, Lopes, *et al.*, 2014), se ha propuesto recientemente al gen de la integrasa *intl1* como un posible marcador ambiental de contaminación antropogénica (Gillings, *et al.*, 2015). Por lo tanto, el *screening* de este gen (y otros asociados a integrones) resulta una poderosa herramienta no sólo para la identificación de estas plataformas de ensamblado, sino para inferir en el grado de polución.

Finalmente, se detectó la secuencia de inserción *IS1071* en un alto porcentaje de los aislamientos de la colección (68%), y se determinó su localización plasmídica en cuatro casos. La presencia de secuencias de inserción en plásmidos

movilizables/conjugativos en bacterias ambientales es una posible causa de su amplia distribución en bacterias (Soto, *et al.*, 1992). Además de su participación en potenciales eventos de recombinación homóloga, su capacidad de transposición desde plásmidos a cromosomas y viceversa puede dar lugar a posibles intercambios de información genética entre los distintos replicones. Además, la presencia del elemento *IS1071* es indicativa de la posible presencia de genes u operones catabólicos en estos elementos. Sin embargo, deberían realizarse ensayos adicionales si se quisiera establecer una asociación concreta entre las *IS1071* detectadas y posibles genes catabólicos transportados en MGEs presentes en los aislamientos.

CAPÍTULO V

**PURIFICACIÓN Y SECUENCIAMIENTO MASIVO DE
PLÁSMIDOS PRESENTES EN LA COLECCIÓN
BACTERIANA**

V.1. Introducción

En los últimos años se ha reportado una gran cantidad de trabajos enfocados en el secuenciamiento y caracterización de plásmidos de distintos ambientes (Szczepanowski, *et al.*, 2008, Kristiansson, *et al.*, 2011, Zhang, *et al.*, 2011, Bengtsson-Palme, *et al.*, 2014). Esto se debe en parte a la fortaleza y fidelidad que presentan las plataformas de secuenciación, sumado al interés de los científicos de conocer el *pool* de genes móviles que son compartidos por una comunidad de microorganismos. Muchos de estos trabajos se han realizado utilizando la plataforma de Roche 454 o Illumina-MiSeq o HiSeq. Estas plataformas permiten el secuenciamiento y procesamiento de una gran cantidad de lecturas, por lo que pueden secuenciarse grandes cantidades de ADN en una sola corrida. Si bien es importante la elección de la plataforma de estudio, el análisis posterior de la información obtenida es un paso clave en este proceso. Este análisis suele ser abordado implementando dos estrategias generales: la primera, comprende el análisis de las lecturas crudas (o *reads*) generadas por el secuenciamiento. Estas lecturas crudas, si bien presentan una longitud corta (100-300 pb), pueden compararse con distintas bases de datos de dominios o familias de proteínas, como Pfam, para tener una idea global sobre el tipo de funciones que están presentes en los plásmidos y en qué proporción están representadas. Sin embargo, en este tipo de análisis la abundancia de una determinada función puede estar sesgada por la cantidad relativa de los distintos replicones presentes en la muestra. Por ejemplo, si se analiza el conjunto de plásmidos de una comunidad bacteriana sin cultivo previo y se evalúa la presencia de un gen o grupo de genes que se encuentran localizados en un plásmido de bajo número de copias o en un plásmido de un microorganismo que está poco representado en la muestra, éstos estarán poco representados, mientras que los genes presentes en un plásmido de alto número de copias o en un plásmido presente en uno o varios microorganismos abundantes en la muestra, estarán sobrerrepresentados. Otra estrategia para avanzar en el análisis de la información obtenida consiste en el ensamblado de las lecturas y formación de *contigs*, con la posibilidad de cerrar plásmidos enteros ya sea circulares o lineales (Dib, *et al.*, 2015). Esta metodología es más completa que la anterior, ya que pueden

obtenerse genes completos y contar con su contexto génico, determinando incluso la presencia de operones, rutas biosintéticas o catabólicas, secuencias regulatorias, promotores, ARN regulatorios, etc.

En los capítulos anteriores, hemos detallado la obtención y caracterización de los aislamientos de la colección bacteriana obtenida a partir de un biofiltro utilizado para mitigar la contaminación de pesticidas y realizamos un análisis preliminar de los plásmidos de la colección (tamaño, grupos de incompatibilidad plasmídica y movilidad). En este capítulo se presenta la purificación del ADN plasmídico circular y su secuenciamiento, y se discute el análisis preliminar del conjunto de datos obtenidos, a fin de avanzar en el conocimiento del contenido genético del material extracromosomal.

V.2. Purificación de plásmidos circulares presentes en la colección de aislamientos

La obtención preparativa de ADN plasmídico de alta pureza y de buena calidad, esto es, ADN circular cerrado intacto sin daño físico generado por estrés mecánico (por ejemplo fuerzas de cizalla), es un paso clave para su posterior secuenciamiento. La extracción de los plásmidos se realizó individualmente para los 35 aislamientos seleccionados. Para ello, se utilizó un kit de extracción de plásmidos de alto peso molecular o bien la metodología de lisis alcalina seguida de un gradiente en CICs, el cual permite separar el ADN doble hebra circular cerrado (plásmidos) del ADN lineal (cromosomal y plásmidos rotos y lineales). En el último caso, luego de haber realizado la lisis para los aislamientos individuales, las muestras fueron colocadas de a pares en cada tubo de centrifuga. La banda perteneciente a los plásmidos fue eluida y concentrada en un pequeño volumen. La integridad del ADN obtenido se chequeó mediante un gel de agarosa 0,8% (Figura V.1) y la cantidad de cada preparación se evaluó según su absorbancia a 260 nm con un equipo Nanodrop. La concentración de las muestras de ADN presentó una gran variación dependiendo de los aislamientos y el método de purificación utilizado. Cuando las purificaciones fueron llevadas a cabo utilizando un kit de extracción plasmídica, las concentraciones variaron entre 20 – 500

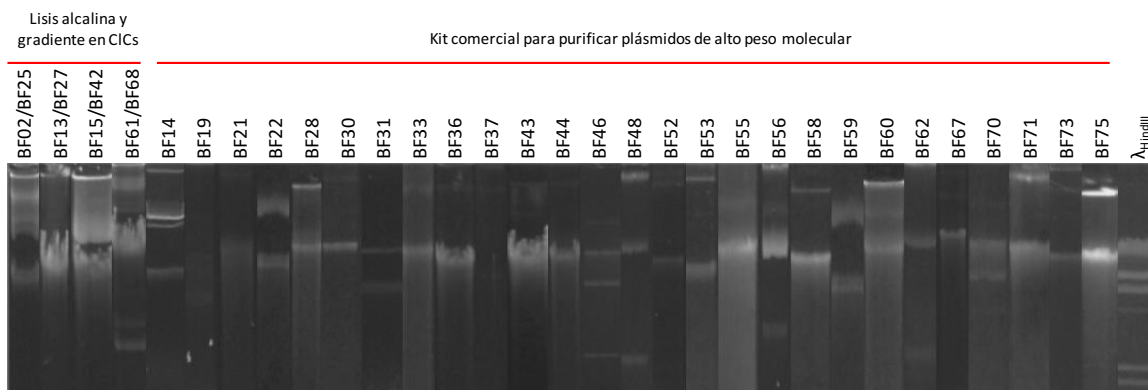


Figura V.1. Corrida electroforética en gel de agarosa 0,8% de los plásmidos purificados. Se indica si los mismos fueron purificados por el método de lisis alcalina seguido de un gradiente en CICs o si fueron extraídos utilizando un kit para la purificación de plásmidos de alto peso molecular (QIAGEN). $\lambda_{HindIII}$ corresponde a ADN del fago lambda digerido con *HindIII*.

ng/ μ l para las distintas muestras, mientras que la purificación por gradiente en CICs generó concentraciones de 100 – 600 ng/ μ l. Una vez determinadas las cantidades para cada preparación, se procedió a unificarlas. Con el objetivo de colocar cantidades equimolares de los plásmidos en la muestra final, se tuvo en cuenta el tamaño aproximado de los plásmidos (estimado en los geles de lisis *in situ*) y la cantidad de cada plásmido en las preparaciones individuales. De esta manera se conformó la muestra que se utilizó para el secuenciamiento, cuya cantidad, calidad e integridad se analizó nuevamente antes del secuenciamiento utilizando un equipo Bioanalyzer (Agilent technologies). Como resultado de la unificación de las muestras, se obtuvo una solución final de 100 μ l con una concentración de 28 ng/ μ l (2,8 μ g de ADN totales).

V.3. Secuenciamiento del ADN plasmídico en conjunto

Para el secuenciamiento de los plásmidos *contigs* decidimos utilizar la plataforma Illumina MiSeq (ver Figura V.2), dado que la misma permite obtener una mayor cantidad de lecturas que otras



Figura V.2. Esquema de la plataforma Illumina MiSeq utilizada para el secuenciamiento de los plásmidos.

tecnologías (por ejemplo, que Ion torrent), cubriendo una mayor cantidad de secuencias, presentando una mayor longitud de lecturas individuales y de mejor calidad (baja tasa de error) de las mismas. Para realizar el secuenciamiento mediante esta tecnología, se partió de 50 ng de la muestra de la muestra de ADN unificada. El ADN fue fragmentado por sonicación y luego se corrió en un gel de agarosa, de donde fueron eluidos y purificados aquellos fragmentos de aproximadamente 500 pb. El secuenciamiento con Illumina MiSeq, permite la lectura de los fragmentos de 500 pb desde ambos extremos, y las dos lecturas generadas pueden unirse para formar una lectura larga, en caso de que el solapamiento de sus extremos sea lo suficientemente bueno.

V.3.1. Información básica del secuenciamiento y ensamblado de la información obtenida

El secuenciamiento generó un total de 2.065.817 de lecturas de una longitud promedio de 250 pb (516 Mb de datos de secuencia). Estas lecturas fueron ensambladas en *contigs* utilizando el servidor Newbler (versión 2.6). En total, se obtuvieron 19.962 *contigs* de una longitud mayor a 500 pb, comprendiendo un total de 20,1 Mb de secuencias no redundantes. El tamaño promedio de los *contigs* fue de 1.513 pb y el *contig* más largo obtenido fue de 76,6 Kb. La predicción automática de genes utilizando Prodigal resultó en un total de 24.069 ORFs en los *contigs* de longitud mayor a 500 pb.

Como generalmente las preparaciones de ADN plasmídico están contaminadas con ADN cromosomal, en una etapa posterior hemos utilizado una estrategia de filtrado *in silico* de aquellos *contigs* que contenían secuencias de origen cromosomal, a fin de eliminar los *contigs* que provienen de secuencias cromosómicas y tener representados sólo aquellos que se presuman de origen plasmídico. Para esto, los *contigs* fueron comparados con los cromosomas de las bacterias huéspedes (previamente identificadas por secuenciamiento del gen ARNr 16S, Capítulo III, sección II.3) disponibles en GenBank. Los *contigs* que presentaron un porcentaje de identidad mayor al 95% con los cromosomas de referencia en más del 90% de cobertura de su

secuencia fueron considerados de origen cromosomal y, por lo tanto, fueron removidos de la colección de datos. Si bien en todos los casos se corroboró que estos *contigs* tuvieran un valor bajo de número de lecturas por longitud de *contig*, no podemos descartar que no fueran posibles transferencias de ADN cromosomal hacia los plásmidos crípticos. Como resultado de este filtrado, se obtuvieron 9.386.079 pb no redundantes consideradas de origen plasmídico, lo que de aquí en adelante llamaremos “set de datos plasmídico”. La sumatoria de todos los *contigs* plasmídicos comprendió un total de 11.839 ORFs.

V.4. Utilización de la plataforma GenDB para la anotación de las secuencias obtenidas

V.4.1. Plataforma GenDB

Antes de comenzar con el análisis *in silico*, realizaremos una breve descripción de la plataforma que utilizaremos de aquí en adelante para abordar el estudio bioinformático. El sistema de anotación GenDB versión 2.0 (Meyer, *et al.*, 2003) permite cargar secuencias ensambladas y realizar distintos tipos de análisis. Esta plataforma posee una interfaz gráfica muy intuitiva y una gran cantidad de herramientas que permiten, entre otras cosas, buscar dentro del set de datos determinados genes por su ID (código otorgado por GenDB), por función o palabra clave; permite explorar las rutas metabólicas (clasificación según KEGG, *Kioto Encyclopedia of Genes and Genomes*) y clasificar los genes según su GO (*Gene Ontology*: ontología génica) o COG (*Clusters of Orthologous Groups*: agrupaciones por grupos de ortólogos), como se muestra en la figura V.3. Además, permite el acceso remoto e ingresar secuencias externas para realizar búsquedas de homología utilizando diversos algoritmos de búsqueda (BLASPp, BLASTx y BLASTn) del NCBI (<http://www.ncbi.nlm.nih.gov>) con los sets de datos importados y depositados en GenDB, empleando parámetros establecidos por defecto o modificándolos. Finalmente, es importante resaltar que esta plataforma permite realizar todo tipo de

The image shows a screenshot of the GenDB web interface. At the top left, there is a navigation menu with options: Search, BLAST Interface, Create region manually, Create new mutation, Import sequences, Update supporting observation, and SNP Detection. A red arrow points from the text 'Permite visualizar los distintos sets de datos cargados' to this menu. In the top center, there is a box labeled 'Gene Ontology' with links for KEGG and COG. A red arrow points from the text 'CDS seleccionada' to a specific green box on the contig map. The contig map itself shows a horizontal line with various green boxes representing genes and arrows indicating their orientation. A red box highlights one of these boxes. To the right of the map is the text 'Esquema de un contig'. Below the map is a 'Region Info' table with the following data:

Region Info	
Name	Biofilter_Just_Plasמידs_18
Start	6837
Stop	5989
Length	849
Frame	-1
Status function	automatically annotated
Status region	status 1
Associated	-
Aliases	
Start/Stop Codon	atg / tag
%GC	56.537
%AC	0.000
#AA	282
Molweight	31641.264
IEP	7.764
Starting 8	atgagcat
Ending 8	ttgcgcta
HTH Motif	n

To the right of the table is an 'Action for region:' menu with options: Annotate region, Show observations, Add Alias, Show region report, Edit region, (Re)Start tools on region, Tool status, Show sequence, Zoom region neighborhood, DNA sequence, and AA sequence. A red box highlights this menu, with a red arrow pointing from the text 'ID de la CDS seleccionada' to the menu title 'Biofilter_Just_Plasמידs_18'. Another red arrow points from the text 'Puede descargarse la secuencia del gen en cuestión, observarse y editar la región.' to the 'Show sequence' option.

Figura V.3. Interfaz gráfica de la plataforma GenDB. Se indican las funciones principales utilizadas para el análisis bioinformático utilizado.

modificaciones sobre las anotaciones automáticas, lo cual resulta de gran utilidad ya que la información puede actualizarse o refinarse de manera manual.

V.4.2. Anotación de las secuencias

Los *contigs* ensamblados se importaron en la plataforma de anotación GenDB para el análisis del contenido genético. Luego, se realizó una predicción de función automática usando una combinación de herramientas bioinformáticas como BLAST e InterPro (Mulder & Apweiler, 2007). En una etapa posterior, la información de secuencia fue refinada manualmente: anotación de genes, asignación de nombres a los genes, números EC (clasificación enzimática), COG y GO. Finalmente, la identidad de las proteínas fue determinada utilizando BLASTp.

El contenido promedio de GC fue de 56,24%. En relación a la asignación de funciones, de los 11.839 ORFs identificados en el set de datos plasmídico, un 64% de ellos pudo ser asignado a algún gen, mientras que el 36% restante de los ORFs predichos no generó coincidencias (*e-value*: 1×10^{10}) con las secuencias depositadas en la base de datos GenBank. Una observación similar fue reportada en otros estudios metagenómicos de plásmidos (Sentchilo, *et al.*, 2013).

V.5. Obtención de replicones cerrados

El ensamblado de replicones puede resultar difícil de lograr si se tiene en cuenta la corta longitud de las lecturas generadas por el secuenciamiento con Illumina y la alta abundancia de secuencias de inserción, transposones y otros elementos repetitivos presentes en plásmidos que dificultan el correcto solapamiento de lecturas. Sin embargo, el ensamblado de las lecturas permitió obtener 6 replicones completos, a los que llamamos pMC1 a pMC6. Los ensamblajes circulares obtenidos nos permitieron analizar en mayor detalle la cobertura de cada uno de los *contigs* que formaban parte de cada uno de ellos por separado. Estos *contigs* generados tuvieron una cobertura homogénea a lo largo de cada ensamblaje, lo esperable para *contigs* que son parte, efectivamente, de un mismo replicón dentro de la mezcla secuenciada más allá de la representatividad que ese plásmido tenga dentro de la misma. Estos replicones

Tabla V.1. Listado y descripción general de los 6 replicones que pudieron ser cerrados luego del secuenciamiento

Nombre	Tipo de replicón	gen rep	ID gen rep (ver anexo S2)	Pfam rep	Tamaño	Número de ORFs	% GC
pMC1	fago*	sí	ORF_18	no detectado	39.636 bp	51	51,94
pMC2	fago	no	-	-	37.598 bp	50	60,9
pMC3	plásmido	sí	ORF_8	no detectado	7.915 bp	16	31,78
pMC4	plásmido	Sí	ORF_8	pfam02486	7.144 bp	12	57,5
pMC5	plásmido	Sí	ORF_6	pfam02486	4.059 bp	6	30,92
pMC6	plásmido	sí	ORF_17	pfam01051	38.140 bp	44	45,69

ID: código de identificación en la plataforma GenDB

(*): Este replicón posee funciones fágicas y una posible proteína de replicación plasmídica

cerrados poseen un tamaño de entre aproximadamente 4 y 40 Kb de longitud. Las principales características de los mismos y sus mapas físicos se muestran en la Tabla V.1 y en la Figura V.4 respectivamente. El análisis de secuencia de estos seis replicones completos mostró una gran cantidad de ORFs que codifican para proteínas hipotéticas a las que aún no ha sido posible asignarles ninguna función, sugiriendo la presencia de funciones no caracterizadas hasta el momento. En la Tabla Suplementaria S2 se muestra la lista y descripción de todas las proteínas hipotéticas codificadas por los ORFs predichos en estos plásmidos. Es importante aclarar que cuando en el análisis BLAST, la primer similitud (el mejor *hit*) correspondió a una proteína hipotética (sin función asignada), al ORF se le asignó la función siguiente más cercana que poseía una anotación funcional, cuando ésta estaba disponible.

Según el análisis *in silico*, 4 de estos elementos circulares cerrados (pMC3, pMC4, pMC5 y pMC6) pudieron ser clasificados como plásmidos de acuerdo a la presencia de proteínas de replicación plasmídica, mientras que los otros dos (pMC1 y pMC2), podrían ser considerados fagos teniendo en cuenta que la mayoría de los ORFs que contienen codifican para proteínas fágicas, aunque el replicón pMC1 contiene un gen que codifica para una proteína hipotética, que posee un bajo porcentaje de identidad con una proteína de replicación plasmídica. Además, este replicón contiene genes que codifican para proteínas del tipo toxina-antitoxina (HicA/HicB) bacterianas. Los plásmidos pMC3 y pMC6 también transportan genes que codifican para proteínas de los sistemas toxina-antitoxina.

Los tres plásmidos más pequeños (pMC3, pMC4 y pMC5) contienen un gen que codifica para una proteína de replicación y la mayoría de los ORFs predichos codifican para proteínas hipotéticas. La mayoría de las proteínas codificadas por el pMC3 mostraron homología con proteínas presentes en *Bacillus cereus*. Este replicón posee un ORF que presenta una alta similitud con el gen de la antitoxina barstar. Esta proteína juega un papel como inhibidor de la barnasa, que es una ribonucleasa que cliva el ARN generando 3'-nucleótidos a través de un intermediario 2,3-cíclico. Ambas

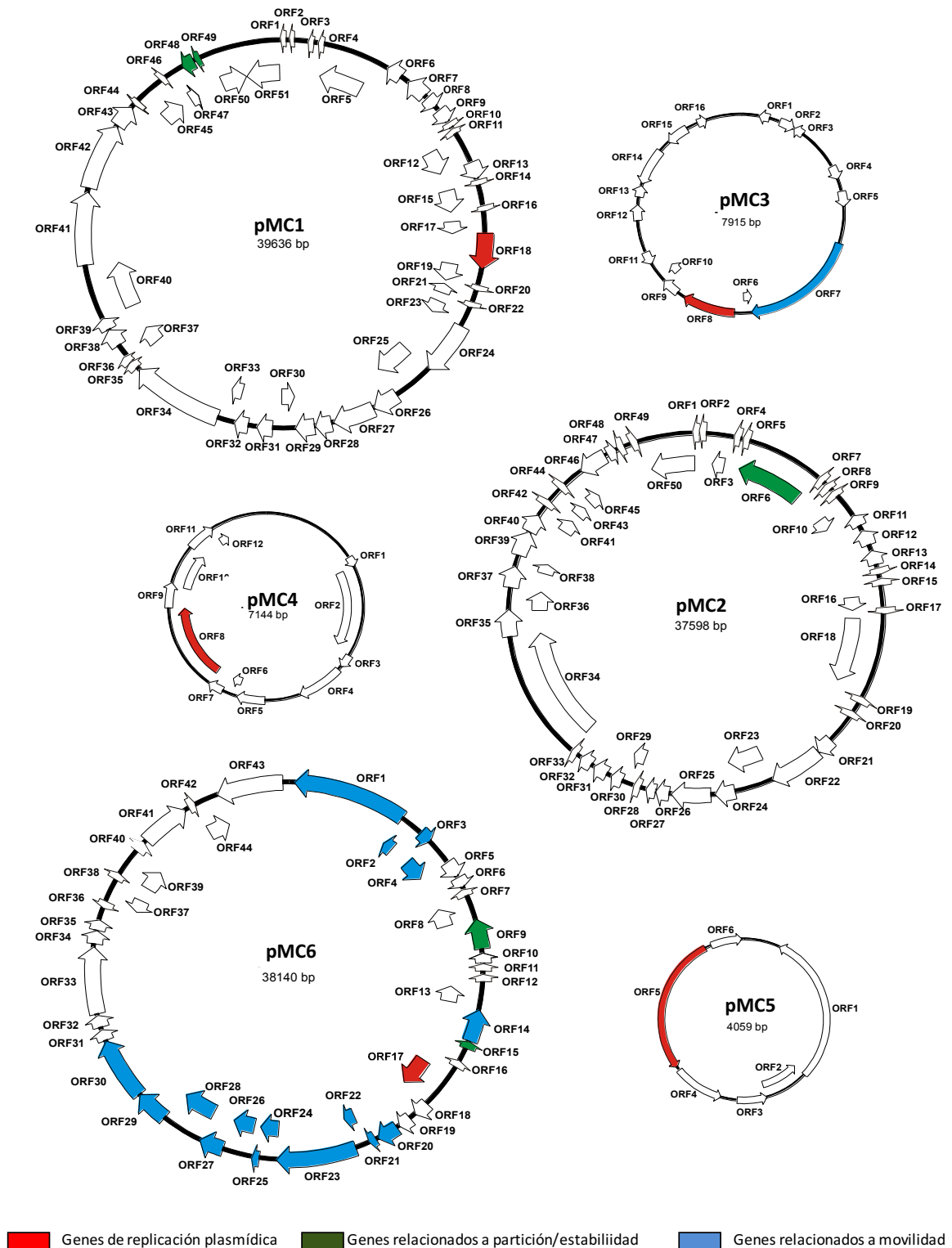


Figura V.4. Mapa físico de los 6 replicones cerrados. Se muestran los ORFs con diversos colores representando la mejor similitud funcional obtenida en búsquedas BLASTp ($e\text{-value} \leq 10^5$). En rojo se indican los genes que están involucrados en la replicación plasmídica; en verde los involucrados en partición y estabilización, en celeste los implicados en movilización y en blanco se muestran aquellos ORFs que no se asignan a estas funciones o bien que (con los parámetros de búsqueda utilizados) no fue posible realizar la predicción bioinformática de la función.

proteínas forman un complejo toxina-antitoxina uno a uno (Hartley, 1989). Sin embargo, el plásmido pMC3 no codifica para ninguna barnasa. La ocurrencia de las ribonucleasas del tipo de la barnasa se encuentra limitada a un cierto número de organismos, lo cual sugiere que esta enzima puede haber evolucionado recientemente (Aravind & Koonin, 2001). Sin embargo, la presencia de la antitoxina barstar podría proteger a la bacteria portadora de pMC3 de efectos perjudiciales ocasionados por la presencia de la toxina barnasa en el genoma de la célula huésped o bien permitir la incorporación de otros plásmidos que codifiquen para esta toxina.

Los plásmidos pMC4 y pMC5, de 7 y 4 Kb respectivamente, sólo poseen (además de los genes que codifican para las proteínas de replicación) unos pocos ORFs, la mayoría de los cuales aún no se le ha podido asignar ninguna función.

El plásmido pMC6, de mayor tamaño, contiene algunos genes *tra* y *mob* que codifican proteínas de movilización y que participan en la conjugación, incluyendo el gen que codifica para la relaxasa, y genes que codifican para el MPF clase T (de acuerdo a Guglielmini y colaboradores (2014)), lo cual sugiere que este plásmido es conjugativo. Este plásmido tiene una composición quimérica; 29 de sus 44 ORFs están relacionados con genes presentes en *Salmonella* sp. y contiene un fragmento de menor tamaño que posee mayor similitud de secuencia con ciertas especies de *Pseudomonas* (ver Tabla Suplementaria S2). El gen de replicación del pMC6 es muy similar al descrito en el plásmido presente en *Salmonella enterica* cepa SL483 (identificado en GenBank como *unnamed plasmid*). Sin embargo, el contenido promedio de G + C (contenido de guanina y citosina) de este género es de aproximadamente 52%, que es considerablemente más alto que el contenido promedio de G + C de pMC6 (aprox. 45,7%), lo que sugiere que este plásmido se ha originado a partir de un género diferente. El plásmido pMC6 también contiene los genes para tres cisteinil-tRNA sintetisas. La presencia de estos genes podría ser beneficiosa para el crecimiento de la bacteria huésped bajo condiciones de limitación de cisteína.

Por último, es interesante remarcar que en ninguno de los 6 replicones se han identificado genes que puedan ser utilizados como marcadores (por ejemplo genes de resistencia a antibióticos o metales pesados) para seleccionar estos elementos en ensayos *in vitro*, limitando su análisis, en una primera instancia, al análisis *in silico* mediante el uso de herramientas bioinformáticas.

V.6. Discusión

En los últimos años se han difundido un gran número de artículos estudiando la microbiología de distintos ambientes agrícolas, como plantas de biogás (Heeg, *et al.*, 2014, Roske, *et al.*, 2014), suelos de granjas en donde residen animales (Snell-Castro, *et al.*, 2005, Kumari, *et al.*, 2015) e incluso biofiltros (Verhagen, *et al.*, 2015). Estos estudios se han abordado mayormente mediante análisis metagenómicos que generalmente se basan en dos tipos de enfoques: **i)** estudios de tipo ecológico y/o taxonómico, basados en la secuenciación de amplicones de los genes ARNr 16S y ARNr 18S, y **ii)** bioprospección de microorganismos o de enzimas con potencial interés biotecnológico a partir de secuencias obtenidas por secuenciamiento de ADN total o plasmídico de comunidades enteras. El primero de los enfoques, permite determinar la composición y abundancia relativa de distintos microorganismos en un ambiente, mientras que el segundo implica un estudio más complejo que permite recabar información acerca del contenido génico global de los microorganismos que componen una población. Ambas aproximaciones incluyen la extracción del material genético de comunidades microbianas y una etapa posterior de secuenciamiento.

Si bien el estudio del contenido génico de una comunidad permite identificar la presencia de plásmidos y otros MGEs, pocos son los estudios que se han centrado específicamente en el secuenciamiento de los elementos móviles y análisis de su contenido en comunidades microbianas de ambientes agrícolas. En el presente trabajo, hemos decidido focalizarnos en el *pool* de genes móviles que son compartidos entre bacterias presentes en una comunidad bacteriana de un biofiltro utilizado en la remoción de pesticidas. Dado que se trata de un ambiente enriquecido en compuestos

altamente tóxicos para los microorganismos, sería esperable que los MGEs jueguen un rol importante en la adaptación y supervivencia de los mismos bajo esta presión de selección. Por esta razón hemos elegido el material extracromosomal para ser secuenciado y analizado en detalle. Es importante resaltar que nuestra estrategia general de trabajo nos brindó la ventaja de obtener información de plásmidos de alto peso molecular (muy difícil de acceder con otras metodologías), teniendo la posibilidad de recuperar y almacenar físicamente dichos plásmidos en sus huéspedes originales.

Si bien nuestro estudio no incluye la totalidad de los elementos extracromosomales presentes en la comunidad entera del BPS, todas las técnicas conocidas y empleadas, incluso aquéllas independientes de cultivo, tienen sus limitaciones, impidiendo la obtención de la totalidad de los plásmidos de un ambiente.

Las metodologías de purificación y secuenciamiento utilizadas han demostrado ser eficientes en cuanto a la cantidad y calidad de secuencias recolectadas. La obtención de *contigs* de gran longitud (hasta 76 Kb aproximadamente) es producto de la alta calidad y cantidad de lecturas individuales que fueron generadas por la plataforma Illumina. El ensamblado de las lecturas permitió obtener la secuencia completa de 6 replicones, aunque sus tamaños resultaron relativamente pequeños (~4-40 Kb) si se los compara con los tamaños estimados por la técnica de lisis *in situ* (hasta 1.500 Kb aproximadamente, ver Capítulo III, sección III.2). Otro punto que merece ser remarcado es la identificación *in silico* de plásmidos que no habían sido visualizados previamente mediante la técnica de lisis *in situ*; tal es el caso de los plásmidos pMC3, pMC4 y pMC5, que tienen un tamaño < 10 Kb.

Como hemos mencionado anteriormente, la mayor parte de los ensamblados que obtuvimos no pudieron circularizarse en elementos cerrados. Esto es normal en secuenciamientos genómicos bacterianos. El hecho de no haber obtenido un número mayor de plásmidos cerrados puede deberse a múltiples factores; en primer lugar, algunas secuencias o replicones podrían estar menos representadas que otras en la muestra. Por otro lado, la presencia de secuencias repetitivas (secuencias de inserción, transposones, etc.) dificulta el ensamblado correcto de los *contigs*, e incluso puede

introducir errores derivando en la formación de quimeras artificiales. Los ensamblajes lineales, al igual que los plásmidos cerrados anteriormente presentados, mostraron diversos tamaños, coberturas de secuenciamiento y grados de similitud con otros replicones ya secuenciados. Contar con la información contenida en ellos, nos permitirá avanzar en un análisis exhaustivo de la información presente en el plasmidoma de la colección bacteriana de estudio. Tal como ocurrió para los elementos cerrados, en el capítulo siguiente abarcaremos y discutiremos el análisis por BLASTp realizado sobre los ORFs predichos en ensamblajes lineales.

Finalmente, no es sorprendente la presencia de una alta proporción (~36%) de ORFs identificados que codificarían para proteínas hipotéticas con homólogos en otros organismos y ORFans (genes predichos codificantes de proteínas que no tienen similitud estadísticamente significativa para ninguna otra secuencia de proteína, es decir, que no parecen ser conservadas o ampliamente distribuidos). Una observación similar fue reportada en estudios previos realizados a partir del secuenciamiento de plásmidos purificados de aislamientos obtenidos a partir de una muestra de lodo activado (Sentchilo *et al*, 2013), y lo mismo ocurre cuando se analiza la composición génica global de comunidades microbianas (Frias-Lopez, *et al.*, 2008). Esta observación nos permite concluir que los plásmidos no sólo proporcionan a sus huéspedes procariontes funciones accesorias conocidas de valor ecológico o adaptativas, sino que también podrían ser considerados como una fuente importante de información genética hasta ahora desconocida.

CAPÍTULO VI

**ANÁLISIS BIOINFORMÁTICO DEL SET DE DATOS
OBTENIDO A PARTIR DEL SECUENCIAMIENTO
MASIVO DEL *POOL* DE PLÁSMIDOS**

VI.1. Introducción

Los plásmidos pueden ser muy diversos en cuanto a su tamaño y contenido. Hasta el día de hoy, se han secuenciado y caracterizado miles de estos replicones a partir de aislamientos o comunidades de distintos ambientes. Muchos de los plásmidos que se encuentran depositados en las bases de datos han sido aislados, purificados y secuenciados por métodos tradicionales, principalmente a partir de técnicas de aislamiento endógeno y exógeno. La reciente implementación de tecnologías de secuenciamiento de alto rendimiento (de gran eficiencia y bajos costos) ha permitido el desarrollo de numerosas estrategias basadas en la metagenómica e implementarlas en conjunto con otras técnicas, como por ejemplo las técnicas de TRACA y MDA (mencionadas en la introducción de este trabajo), que permiten la captura, purificación y secuenciación de plásmidos.

El secuenciamiento de los plásmidos puede proveer información muy valiosa acerca de su composición y biología. Tal información puede utilizarse para, por ejemplo, determinar qué rango de huéspedes pueden habitar (mediante el análisis de los módulos de replicación), su estabilidad, movilidad, si portan genes que confieren resistencia a antimicrobianos u otras moléculas o compuestos, si poseen genes de virulencia, si permiten al huésped catabolizar algún compuesto, establecer algún tipo de interacción simbiótica o proveer de alguna función adaptativa a los microorganismos que los contienen. Si bien los fenotipos deben ser corroborados experimentalmente, y muchas veces son dependientes del *background* genético de la célula hospedadora, la presencia de genes particulares puede dar un indicio de las características y del rol biológico que confieren los plásmidos a las bacterias que los contienen. Sin embargo, muchos de los genes identificados pueden no tener una función asignada, lo cual dificulta el análisis de estos replicones. Se ha reportado que los plásmidos contienen la mayor fracción de ADN no conocido, siendo en general clasificables menos de un 50% de las secuencias, tanto a nivel nucleotídico como a nivel aminoacídico (Sentchilo *et al.*, 2013). En consonancia con los párrafos precedentes, si bien el secuenciamiento es una herramienta poderosa para el

conocimiento de las secuencias que los plásmidos portan, experimentos adicionales deben ser realizados para avanzar en el estudio del rol biológico y ecológico de los mismos.

En este capítulo describiremos los resultados del análisis de las secuencias obtenidas a partir del ADN de los plásmidos purificados de los aislamientos de la colección. El análisis lo realizamos en base a la comparación de los productos de traducción de todos los ORFs predichos (blastx, <http://www.ncbi.nlm.nih.gov/BLAST/>) contra la base de datos de proteínas. Nos centraremos en establecer qué tipo de replicones portan los aislamientos y sus posibles capacidades de transferencia y conjugativas; analizaremos si poseen genes de resistencia a antibióticos y metales, si contienen otros MGEs como integrones y secuencias de inserción y, finalmente, considerando el ambiente del cual provienen los aislamientos, indagaremos si poseen genes involucrados en la degradación de compuestos xenobióticos e hidrocarburos aromáticos.

VI.2. Identificación de genes asociados a funciones plasmídicas

La primera parte del análisis del set de datos obtenido del secuenciamiento de los plásmidos de la colección está enfocada en el estudio de genes estrictamente asociados a plásmidos, esto es, aquéllos involucrados en la replicación (*rep*), partición-estabilización (*par-sta*) y movilidad (*mob-tra*) plasmídicas. En una primera instancia se realizó un análisis general a partir de las lecturas crudas del secuenciamiento. Si bien estas lecturas contienen, como se explicó en el capítulo anterior, una contaminación considerable de ADN cromosomal, el análisis de funciones específicas de plásmidos puede realizarse sobre el total de las lecturas. Estas funciones fueron analizadas a partir de la comparación de las secuencias aminoacídicas inferidas de las lecturas individuales con los dominios y familias de proteínas depositados en la base de datos de Pfam. El análisis fue llevado a cabo con la plataforma MGX (Bielefeld, Alemania) la cual permite el procesamiento de millones de lecturas a la vez. De las 2.065.817 lecturas generadas en el secuenciamiento, 2.055.622 generaron *matches* contra la

base de datos de Pfam, lo cual implica que un 99,5% de las lecturas se corresponderían, en principio, con algún dominio o familia de proteínas. De estas 2.055.622 lecturas, 368.055 resultaron relacionadas a alguna de las funciones plasmídicas mencionadas previamente, indicando que al menos un 18% de las secuencias crudas representan genes estrictamente plasmídicos. A partir de la identificación de estas lecturas, junto con la anotación automática generada por la plataforma GenDB, se realizó una búsqueda y clasificación de todos los genes *rep*, *par-sta* y *mob-tra*, los cuales se listan en las Tablas Suplementarias S3, S4 y S5, respectivamente, y se describen en las siguientes secciones.

VI.2.1. Búsqueda y clasificación de los genes que codifican para proteínas involucradas en la replicación

En la base de datos de Pfam existen 14 familias de proteínas distintas asociadas a la replicación plasmídica. De estas 14 familias, 10 fueron identificadas en las lecturas crudas, dando una idea de la diversidad de plásmidos presentes en los aislamientos representativos seleccionados. Estos Pfams fueron asignados a 31 genes *rep* distintos (ver Tabla VI.1 y Tabla Suplementaria S3, para más información). Además, el ensamblado y anotación automáticos de la plataforma GenDB permitió identificar 6 genes *rep* extra que no contienen ningún Pfam estrictamente asociado a replicación. De estos 6 genes, 4 contienen un Pfam perteneciente a la familia de dominios del tipo hélice-giro-hélice (HTH_36) y las dos restantes, si bien no contienen ningún dominio Pfam de replicación asociado, poseen más de 87% de identidad a nivel aminoácido con alguna proteína de replicación depositada en la base de datos de GenBank (ver Tabla VI.1 y Tabla Suplementaria S3).

La mayoría de los genes *rep* detectados pudieron ser identificados en plásmidos cuyos huéspedes han sido descritos y comprenden distintas especies de alfa, beta y gammaproteobacterias, firmicutes y actinobacterias (Tabla VI.1), lo cual se condice con la taxonomía de los aislamientos de la colección (ver Capítulo III; Tabla III.1). Algunos de los plásmidos asociados a las proteínas de replicación identificadas contienen genes catabólicos, como el pCAR1 (Maeda, *et al.*, 2003), pND6-1 (Li, *et al.*, 2004), pSTY

(Kohler, *et al.*, 2013), pDK1 (Kunz & Chapman, 1981), y genes de resistencia a antibióticos, como el pMET-1 (Soler Bistue, *et al.*, 2008), RP4 (Pansegrau, *et al.*, 1994), pAHH01 (Han, *et al.*, 2012), pLE321 (Götz, *et al.*, 1996) y pTB11 (Tennstedt, *et al.*, 2005), lo cual sugiere que los plásmidos de la colección podrían contener algunas

Tabla VI.1. ORFs que contienen un Pfam asociado a replicación plasmídica

Número de Pfam	ID Pfam	Descripción	Proteínas encontradas	Organismo	% de identidad con la proteína más relacionada
pfam01051	Rep_3	Proteína iniciadora de la replicación	6	<i>Providencia alcalifaciens</i>	64%
				<i>Salmonella enterica</i>	99%
				<i>Pseudomonas putida</i>	100%
				<i>Pseudomonas taiwanensis</i>	71%
				<i>Salmonella enterica</i>	99%
				<i>Pseudomonas taiwanensis</i>	99%
pfam02486	Rep_trans	Factor de iniciación de la replicación	5	<i>Acidovorax citrulli</i>	56%
				<i>Streptococcus parasanguinis</i>	97%
				<i>Acidovorax delafieldii</i>	48%
				<i>Streptococcus parasanguinis</i>	97%
				<i>Acidovorax citrulli</i>	56%
pfam11800	RP-C_C	Región C-terminal de la proteína de replicación C	5	<i>Ochrobactrum anthropi</i>	83%
				<i>Ochrobactrum anthropi</i>	89%
				<i>Ochrobactrum rhizosphaerae</i>	99%
				<i>Sinorhizobium meliloti</i>	94%
				<i>Ochrobactrum anthropi</i>	88%
pfam03428	RP-C	Dominio N-terminal de la proteína de replicación C	4	<i>Ochrobactrum rhizosphaerae</i>	88%
				<i>Sinorhizobium meliloti</i>	94%
				<i>Ochrobactrum rhizosphaerae</i>	95%
				<i>Ochrobactrum rhizosphaerae</i>	94%
pfam04796	RepA_C	Proteína RepA plasmídica	3	Bacteria no cultivable	62%
				Bacteria no cultivable	53%
				<i>Enterobacteriaceae</i>	100%
pfam07042	trfA	Proteína TrfA	2	<i>Thauera terpenica</i>	68%
				<i>Pseudomonadales</i>	100%
pfam01446	Rep_1	Proteína de replicación	2	<i>Bacillus pumilus</i>	94%
				<i>Bacillus pumilus</i>	98%
pfam03090	Replicase	Familia replicasa	2	<i>Pseudomonas sp.</i>	90%
				<i>Aeromonas hydrophila</i>	99%
pfam06504	RepC	Proteína de replicación C (RepC)	1	<i>Pseudomonas aeruginosa</i>	96%
pfam02387	IncFII_repA	Familia de proteínas RepA IncFII	1	<i>Klebsiella pneumoniae</i>	96%
pfam13730	HTH_36	Proteína de replicación putativa RepA	4	<i>Agrobacterium tumefaciens</i>	100%
				<i>Ochrobactrum intermedium</i>	100%
				<i>Ochrobactrum rhizosphaerae</i>	99%
				Plásmido pM3	99%
Sin Pfam asociado a replicación			2	<i>Aeromonas hydrophila</i>	71%
				<i>Bacillus cereus</i>	100%

ID: código de identificación en la base de datos de Pfam

funciones de este tipo.

El parecido entre las proteínas de replicación identificadas y el mejor *hit* depositado en GenBank fue variable, comprendiendo valores de entre 46% y 100% de identidad (Tabla VI.1). De los 37 genes que codifican para proteínas Rep, 28 presentaron entre 80-100% de identidad con el *hit* más cercano, mientras que las 9 restantes variaron entre 46-79% de identidad. Esto muestra la diversidad de genes *rep* encontrados, poniendo en evidencia la presencia de distintos tipos de replicones plasmídicos presentes en los aislamientos de la colección del biofiltro. Asimismo, la identificación de genes que codifican para proteínas de replicación que no están muy estrechamente relacionadas con aquellas previamente descritas, nos permite ampliar nuestra comprensión de la diversidad de módulos de replicación plasmídica.

VI.2.2. Búsqueda y clasificación de los genes que codifican para proteínas involucradas en la partición y estabilidad plasmídica

Las funciones de estabilidad y partición incluyen aquellos genes involucrados en el mantenimiento y herencia de los plásmidos durante el proceso de la división celular. Los dos mecanismos más importantes responsables de la partición y estabilización plasmídica son la muerte postsegregacional (o *segregational killing*) y los sistemas de partición activa (*active partitioning systems*), los cuales estuvieron ampliamente representados en el set de datos de los plásmidos del biofiltro. Se identificaron 128 genes del *tipo par-sta* que podrían participar de estos mecanismos. En total, 21 de las 23 familias asociadas a este tipo de funciones depositadas en la base de datos de Pfam fueron identificadas, incluyendo las familias ParBc (PF02195) y CbiA (PF01656) que fueron además las más abundantes (ver Tabla suplementaria S4).

VI.2.3. Búsqueda y clasificación de genes que codifican para proteínas involucradas en la movilidad plasmídica

Las funciones de movilización (*tra*, referidas a aquellas presentes en plásmidos movilizables, es decir, la relaxasa y otras proteínas accesorias) y conjugación (*mob*, referidas a los genes involucrados en la formación del aparato conjugativo: proteína

acopladora T4CP y sistema de formación del par de conjugación MPF) fueron las más abundantes en el set de datos general, comprendiendo el 12% de los *hits* a nivel de lecturas individuales. Estos *hits* corresponden a secuencias plasmídicas cuya distribución taxonómica representa principalmente los filos Proteobacteria (de las clases alfa, beta, gamma), Firmicutes y Actinobacteria. Cuando se analizó la presencia de estos genes en los datos ensamblados, se encontró un total de 254 genes pertenecientes a estas dos categorías, los cuales se distribuyeron en 33 familias de Pfam distintas (ver tabla suplementaria S5).

Respecto a los replicones cerrados, sólo dos de ellos poseen el gen de la relaxasa, mientras que el plásmido pMC3 codifica sólo para la relaxasa, lo cual indica que el mismo sería movilizable, el plásmido pMC6 contiene los componentes del relaxosoma, T4CP y MPF, indicando que el mismo sería conjugativo (ver Capítulo V).

Los genes asociados a movilidad son un punto interesante a ser analizado, dado que la presencia de éstos puede ser estimativa de la capacidad de movilizarse y propagarse que tienen los plásmidos. A continuación, con motivo de estudiar más en profundidad las funciones asociadas a movilidad en los plásmidos recuperados de la colección de aislamientos del BPS, realizamos un análisis a partir del estudio de la diversidad y abundancia de los genes que codifican para la enzima relaxasa y otras del MPF, como se ha establecido previamente por Garcillán-Barcia y colaboradores (Garcillán-Barcia, *et al.*, 2009).

VI.2.3.1. Clasificación de las secuencias presentes en los plásmidos de la colección de acuerdo a la relaxasa y al MPF

Las relaciones filogenéticas entre las relaxasas disponibles en las bases de datos han permitido realizar una clasificación de los plásmidos transmisibles en 8 familias MOB (Francia, *et al.*, 2004, Garcillan-Barcia, *et al.*, 2009). De la misma manera, se ha podido clasificar a los plásmidos autotransmisibles (conjugativos) en 8 familias, de acuerdo al MPF (Smillie, *et al.*, 2010, Guglielmini, *et al.*, 2014).

Basados en estas clasificaciones, con motivo de realizar un análisis profundo de los plásmidos de la colección del BPS, analizamos las relaciones filogenéticas de las proteínas del tipo VirB4 (que componen el MPF), para las cuales fueron identificados 13 genes (ver Tabla Suplementaria S5; identificadas con el Pfam PF03135 ID: CagE_TrbE_VirB) y las relaxasas, para las cuales se utilizaron los 19 genes identificados en el set de datos plasmídico. Los genes que codifican para las proteínas del tipo VirB4 y relaxasa incluidos en este análisis se listan en la sección II.3.8.2 de Materiales y Métodos).

Como puede observarse en la Figura VI.1, todas las proteínas VirB4 identificadas

pertenecen al tipo MPF_T. Dicho grupo se caracteriza por incluir hasta once proteínas (VirB1-VirB11) que participan en el ensamblado del canal secreción y en la biogénesis del *pilus* conjugativo (Trokter et al., 2014, Cabezon et al., 2015). Los plásmidos que contienen sistemas MPF_T se encuentran ampliamente distribuidos en todos los filos de proteobacterias y se caracterizan por conjugar eficientemente en superficies sólidas pero no en líquidas (Smillie, et al., 2010). El análisis de las 19 relaxasas identificadas nos permitió agruparlas en 4 grupos MOB distintos: 2 de ellas pertenecen a la familia MOB_F, 4 a la familia MOB_Q, una a la MOB_V y 12 a la familia MOB_P (Figuras VI.2 y VI.3).

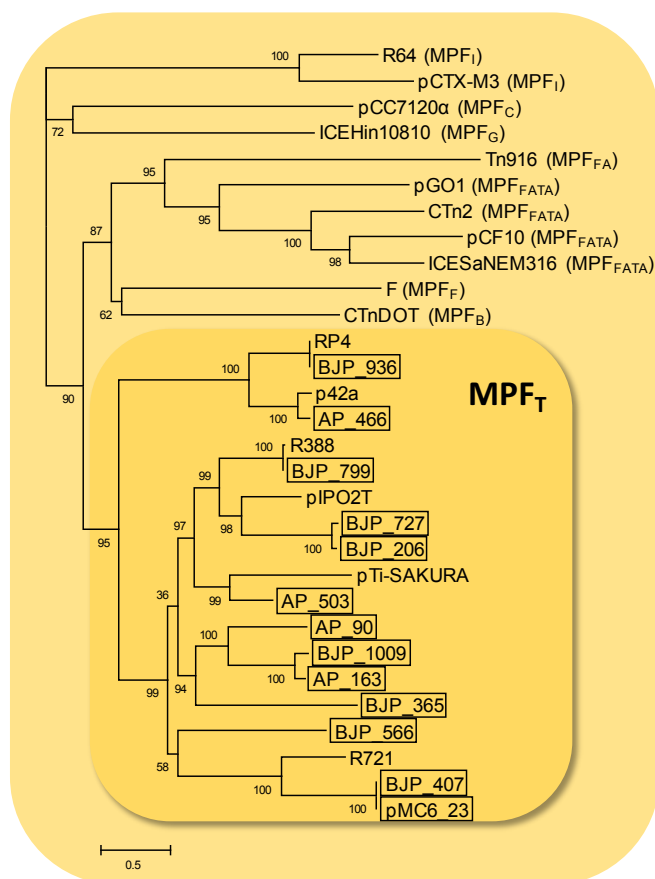


Figura VI.1. Análisis filogenético de las proteínas VirB4 encontradas en los plásmidos de la colección. El análisis incluye 13 proteínas pertenecientes al grupo CagE-TrbE-VirB (ver tabla suplementaria S5) y los representativos de los 8 sistemas MPF definidos en Guglielmini y colaboradores (2014). El *cluster* MPF_T está resaltado en naranja oscuro. Las proteínas del tipo VirB4 del set de datos plasmídico están recuadradas en negro.

Las relaxasas MOB_Q fueron agrupadas en dos grupos monofiléticos (o *clusters*) (Figura VI.2.A). La proteína AP_265 fue agrupada con las relaxasas conjugativas de plásmidos de *Rhizobium/Agrobacterium* en la subfamilia MOB_{Q2}. Las relaxasas MOB_{Q2} contienen un dominio helicasa C-terminal homólogo al de las relaxasas MOB_{F1} (Perez-Mendoza, *et al.*, 2005). Estas relaxasas están involucradas en el proceso de conjugación plasmídica y son diferentes de las proteínas del tipo VirD2 de MOB_{P2} que realizan la transferencia del T-ADN de bacterias hacia células de la planta (Garcillán-Barcia, *et al.*, 2009).

Las proteínas BJP_6246, BJP_4415 y BJP_1723 fueron agrupadas en un clado que no ha sido clasificado por Garcillán-Barcia (Garcillán-Barcia, *et al.*, 2009). Este clado incluye relaxasas de diversos huéspedes de alfa y gamma Proteobacterias, ya sea bacterias rizosféricas (como el plásmido pRL11 de *Rhizobium leguminosarum*), fitopatógenos (como el pXcaw19 de *Xanthomonas citri*), o bacterias con capacidades degradativas versátiles (como el plásmido pAMI2 de *Paracoccus aminophilus*).

La familia de relaxasas más abundante en la naturaleza es la de MOB_P (Garcillán-Barcia, *et al.*, 2009, Smilie, *et al.*, 2010). La mayoría de las relaxasas encontradas en el set de datos plasmídico pertenecen a esta familia (proteínas BJP_1498, AP_669, BJP_441, BJP_847, BJP_634, BJP_2740, AP_394, AP_113, BJP_237, BJP_387, BJP_198 y pMC6_1). En la Figura VI.3.A se observa la localización de estas proteínas en los distintos subgrupos de MOB_P. Una simplificación de cada uno de los agrupamientos (*clusters*) en donde se encuentran dichas proteínas se muestra en la Figura VI.3.B. Como puede verse en esta figura, las relaxasas BJP_1498 y AP_669 son prácticamente idénticas a la relaxasa MOB_{P11} del plásmido RP4, del prototipo IncP-1, y la relaxasa BJP_441 pertenece a otra subfamilia MOB_{P1}, el agrupamiento MOB_{P13}. Otras relaxasas MOB_P se clasificaron dentro de otras subfamilias caracterizadas: MOB_{P6} incluye a las proteínas BJP_387, BJP_198 y pMC6_1 (esta última corresponde a la relaxasa del plásmido completamente secuenciado pMC6) mientras que la subfamilia MOB_{P7} contiene a la proteína BJP_237. El resto de las relaxasas MOB_P (BJP_847, AP_113, BJP_2740 y AP_394) se localizaron en diferentes clados no caracterizados.

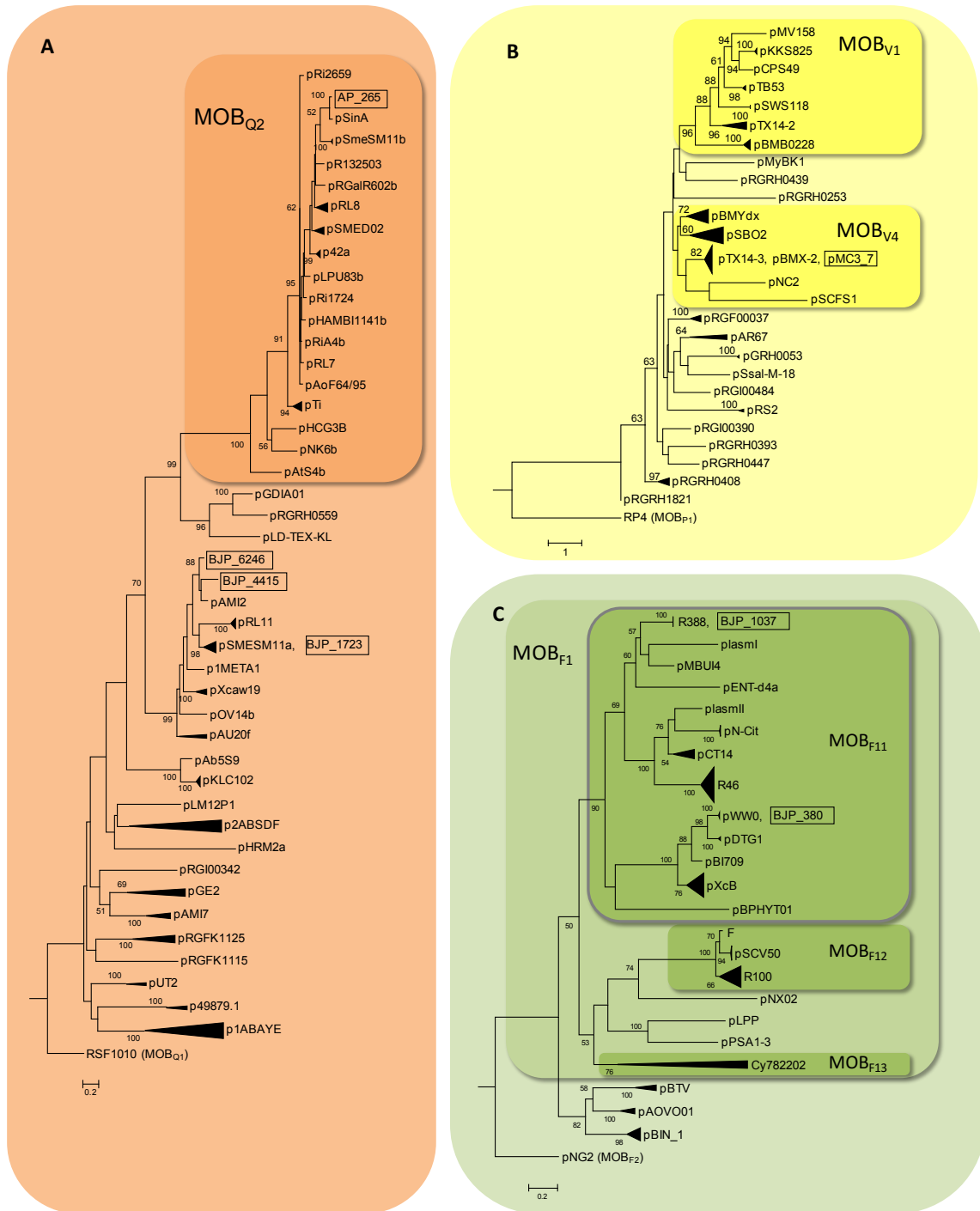
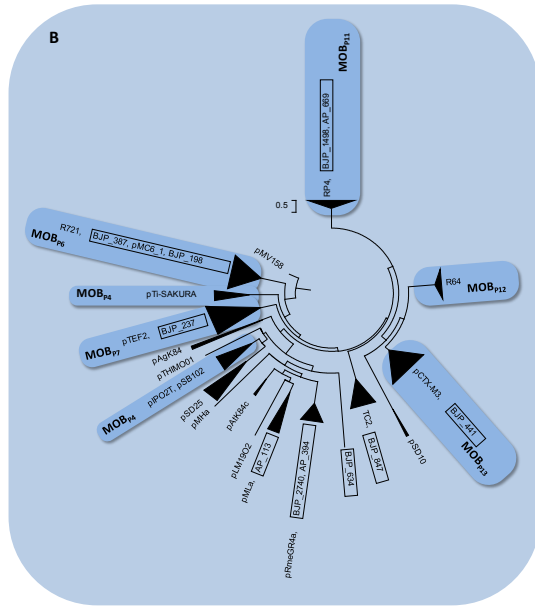
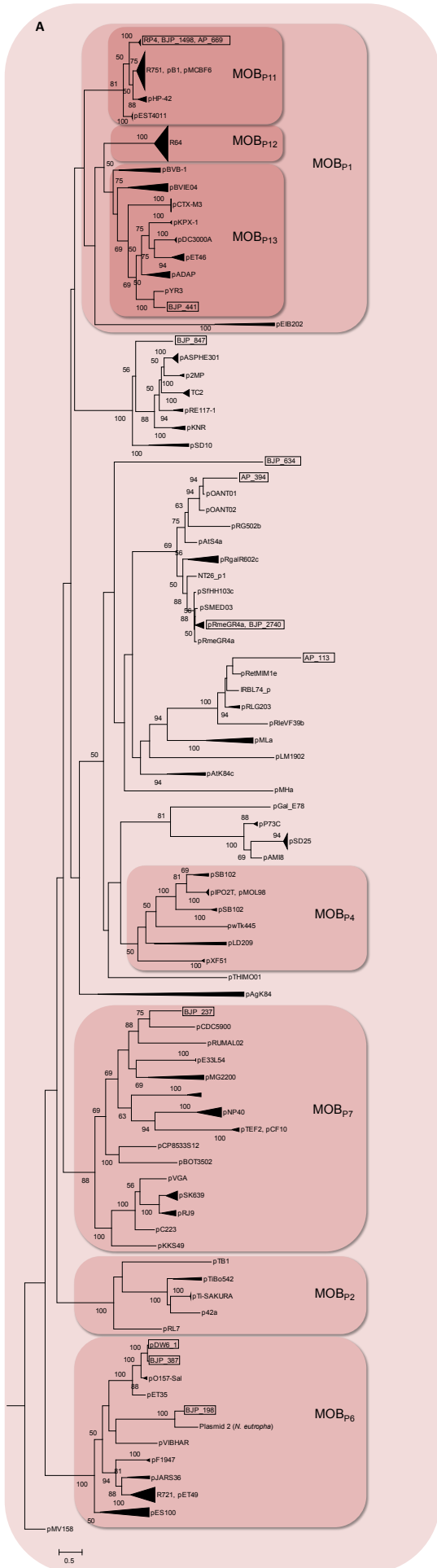


Figura VI.2. Filogenia de las relaxasas identificadas en el set de datos plasmídico. Se muestran las filogenias MOB_Q (A), MOB_V (B) y MOB_F (C), para las cuales se usaron las relaxasas de los plásmidos RSF1010 (MOB_{Q1}), RP4 (MOB_{P1}) y pNG2 (MOB_{F2}), respectivamente, como *outgroup* para construir el árbol. Sólo los valores de *bootstrap* >50% se indican en los nodos de los árboles. Las proteínas del set de datos plasmídico están recuadradas en negro.



VI.3. Filogenia de las relaxasas MOB_p identificadas en el set de datos plasmídico. Sólo los valores de *bootstrap* >50% se indican en los nodos de los árboles. Las proteínas del set de datos plasmídico están recuadradas en negro. Se usó la relaxasa del plásmido pMV158 (MOB_V) como *outgroup*. En la Figura B, se muestra una representación simplificada del árbol completo, que se muestra en la Figura A.

Como puede observarse en la Figura VI.3.A, la proteína BJP_634 es una relaxasa singular que no posee homólogos cercanos. Un punto importante que merece ser mencionado es que todas las relaxasas MOB_p encontradas en el set de datos están incluidas en subfamilias pobladas por plásmidos autotransmisibles en lugar de movilizables. La relaxasa pMC3_7 (presente en el plásmido cerrado pMC3) fue la única agrupada en la familia MOB_v, más precisamente en el subgrupo MOB_{v4} (Figura VI.2.A). Este tipo de relaxasas se encuentran generalmente en plásmidos movilizables pequeños hospedados en bacterias pertenecientes a los géneros *Bacillus* y *Streptococcus*. Una característica relevante de los plásmidos movilizables MOB_v es que tienen la capacidad de ser transferidos por una amplia variedad de plásmidos *helper* (Lorenzo-Díaz, *et al.*, 2014).

Por último, las dos relaxasas MOB_f encontradas en el set de datos plasmídico se localizaron en la subfamilia MOB_{f11} (Figura VI.2.C). La relaxasa BJP_1037 es idéntica a las de los plásmidos IncW (como el R388 y el R7K), mientras que la BJP_380 es altamente similar a las relaxasas de los plásmidos IncP-9. Los plásmidos IncP-9 abundan en ambientes contaminados (Krasowiak *et al.*, 2002, Dealtry *et al.*, 2014) y generalmente contienen genes implicados en la degradación de xenobióticos (como los plásmidos pWW0 and NAH7 de *Pseudomonas*).

VI.3. Análisis del contenido plasmídico: genes accesorios comúnmente presentes en plásmidos

Además del estudio de las funciones estrictamente relacionadas a plásmidos, resulta importante conocer qué genes se encuentran formando parte de los esqueletos plasmídicos. Como se mencionó previamente en esta Tesis, es frecuente encontrar genes de resistencia a antibióticos y metales pesados, así como aquéllos involucrados en el catabolismo de ciertos compuestos orgánicos en plásmidos y otros MGEs. Por este motivo, y teniendo en cuenta que previamente hemos detectado en esta colección plásmidos capaces de conferir resistencias a antibióticos y metales pesados, implementamos la búsqueda y análisis de estos genes. Estos análisis fueron

realizados *in silico*, utilizando las secuencias ensambladas cuya naturaleza se determinó de origen plasmídico luego del filtrado de las secuencias cromosomales.

VI.3.1. Presencia de genes asociados con resistencia a antibióticos

La anotación automática en GenDB reveló la presencia de 112 posibles genes de resistencia a antibióticos (ARGs: *Antibiotic Resistance Genes*) y otros compuestos antimicrobianos, confiriendo resistencia a tetraciclinas, macrólidos, β -lactamas, aminoglicósidos, bleomicina, fosmidomicina, bacitracina, fenicoles y acriflavina (Figura VI.4 y Tabla Suplementaria S6). Estos resultados indican la presencia de una gran cantidad y variedad de ARGs en los plásmidos. Con el fin de identificar y clasificar estos ARGs, realizamos un análisis comparando estos genes con todos aquellos depositados en la base de datos ARDB (*Antibiotic Resistance Data Base*). Esta base de datos centraliza la información y contiene un compendio de 23.137 genes de resistencia pertenecientes a 380 tipos que confieren resistencia a 249 antibióticos distintos (<http://arbd.cbc.umd.edu/>) (Liu & Pop, 2009). De los 112 posibles ARGs identificados, 72 pudieron ser asignados a algún gen específico (Figura VI.5).

Los ARGs más abundantes en nuestro set de datos fueron los sistemas de resistencia/eflujo de **multidrogas** (41% del total). Estos sistemas funcionan como transportadores de compuestos tóxicos y son capaces de conferir resistencia a una gran variedad de antibióticos y otros como acriflavina, bromuro de etidio y otros compuestos cuaternarios de amonio, vía expulsión de la droga hacia el exterior celular. Estos genes se encuentran muchas veces en regulones u operones, por lo cual para conferir cierta resistencia suelen requerir de la presencia y expresión de otros genes. En las secuencias ensambladas, se pudieron identificar 3 operones de este tipo: *marABCR*, *mdtIJ* y *mdtON*, mientras que el resto de los genes se encontraron individualmente en *contigs*. Esto indicaría que no están formando parte de operones, o bien puede ser consecuencia de la imposibilidad de detectar los genes vecinos a causa de la longitud insuficiente de algunos *contigs* en los que se encuentran.

Respecto a las **β -lactamasas**, se identificaron 16 genes que pudieron clasificarse en 8 variedades distintas de acuerdo a la base de datos ARDB (ver Figura VI.5 y Tabla

Suplementaria S6). Estos genes mostraron una identidad a nivel proteína con el mejor *hit* que fue variable (entre 40% y 100%), sugiriendo la presencia de posibles nuevas variantes de genes que codifican para lactamasas. Las lactamasas resultan de gran interés desde un punto de vista clínico, ya que los β -lactámicos son la clase de antibióticos más utilizada (Brolund & Sandegren, 2015), y se considera a los plásmidos los principales vehículos para la diseminación de estas resistencias.

Los **aminoglicósidos** están dentro de los antibióticos de amplio espectro más utilizados y su resistencia es conferida principalmente por enzimas que modifican la droga vía fosforilación, adenilación o acetilación (Mingeot-Leclercq, *et al.*, 1999, Kotra, *et al.*, 2000). Se encontraron dos posibles genes que codifican para fosfotransferasas, de los cuales uno corresponde al gen *aphA*, presente en el plásmido pRK2. Además, encontramos que los ORFs adyacentes a este gen presentan homología con el módulo *traABC* del mismo plásmido.

Respecto a los **fenicoles**, se encontraron 3 genes de resistencia a estos antimicrobianos, de los cuales dos de ellos (*cmIA9* y *cmIA6*) son similares a otros previamente reportados en la región variable de algunos integrones de clase 1 encontrados en distintos géneros bacterianos (Levings, *et al.*, 2008, Perez, *et al.*, 2014).

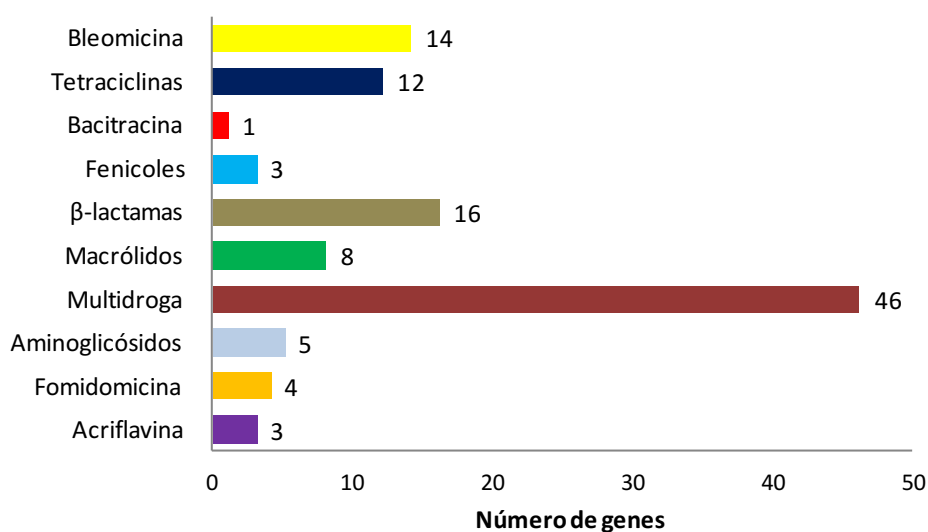


Figura VI.4. Cantidad y variedad de genes de resistencia a antibióticos y acriflavina identificados en el análisis *in silico* realizado con el conjunto de datos de secuencia.

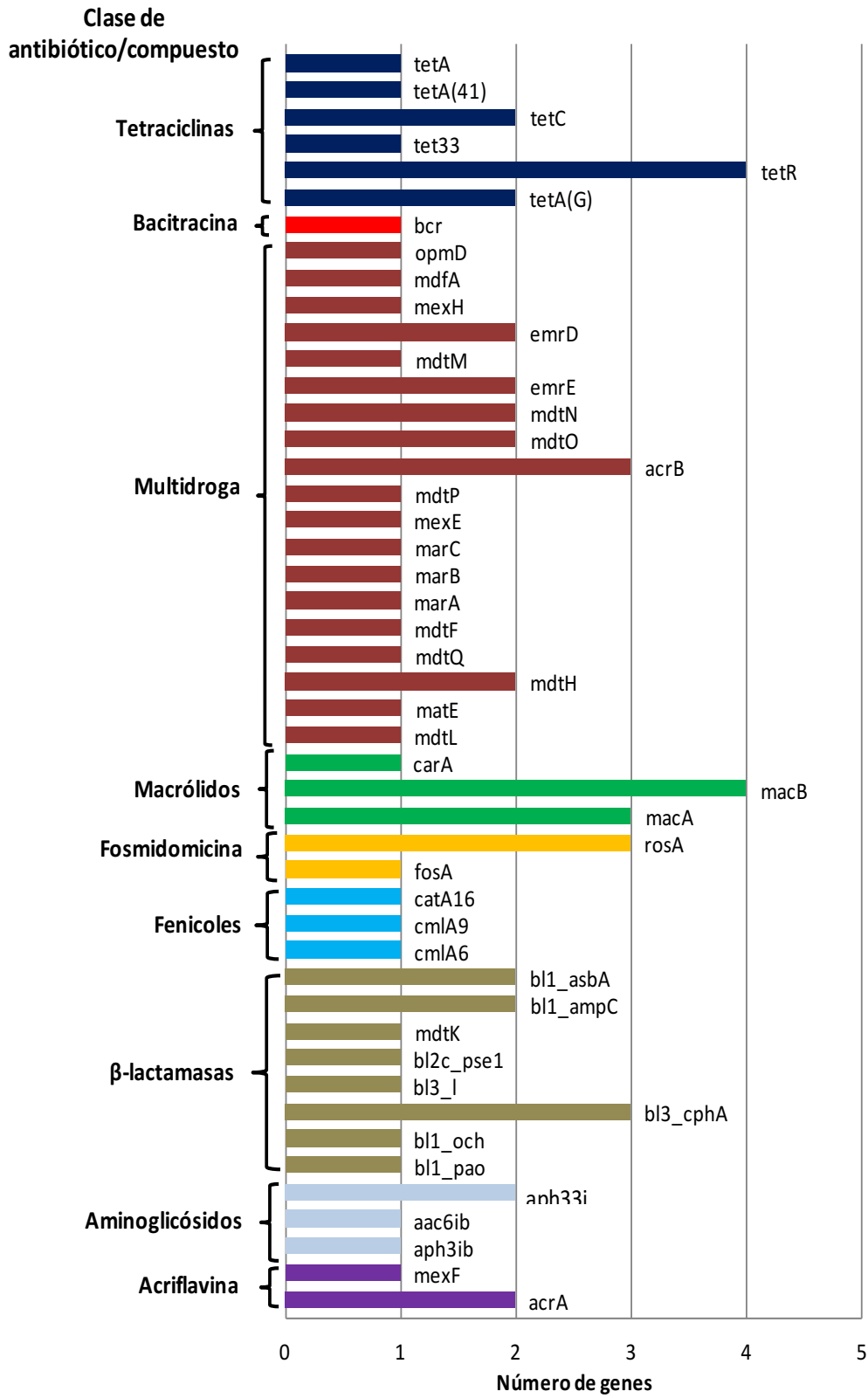


Figura VI.5. Clasificación de los genes de resistencia a antibióticos y acriflavina identificados en el análisis *in silico* realizado con el conjunto de datos de secuencia

embargo, el análisis realizado sobre los *contigs* que los contienen no detectó la presencia del gen *intl1* acompañando a estos genes.

Los genes de resistencia a **macrólidos** y **glicopéptidos** mostraron homología con otros previamente identificados en muestras ambientales. El gen *macB*, que codifica para un sistema de eflujo específico de macrólidos, fue detectado en el set de datos plasmídico. Los genes *mac* pueden ser transferidos fácilmente de un huésped a otro (Roberts, 2008), ya que son usualmente adquiridos por plásmidos (Liu, *et al.*, 2007) y otros MGEs.

Finalmente, encontramos distintos genes de resistencia a **tetraciclinas**, presentando todos ellos un alto porcentaje de identidad con otros previamente descritos (*tetA*, *tetC*, *tet33*, y *tetR*) en distintas bacterias.

El análisis *in silico* aquí realizado indica que las bacterias aisladas del BPS constituyen un importante reservorio de genes de resistencia a antibióticos, evidenciando la presencia de parte de un arsenal adaptativo de degradación de compuestos con actividad antimicrobiana que forman parte de la química de la competencia y adaptación a distintos nichos, y que son comúnmente encontrados en bacterias de suelo.

VI.3.2. Presencia de genes de resistencia a metales

La presencia de genes que codifican para proteínas que confieren resistencia a metales fue evidenciada por la anotación automática generada en la plataforma de GenDB. En total, pudimos identificar 78 genes asociados a la resistencia de distintos metales; todos ellos fueron manualmente corroborados con la base de datos de GenBank (ver Tabla Suplementaria S7). El análisis de los *contigs* permitió la detección de un *cluster* de genes *mer* insertados en un transposón truncado. Este *cluster* contiene el gen que codifica para una proteína reguladora (MerR), una proteína que transporta el ion Hg^{+2} (MerT), una proteína periplasmática de unión a Hg^{+2} (MerP), una reductasa (MerA), una proteína reguladora secundaria (MerD) y proteínas de función desconocida (MerE y ORF2). Se han descrito una gran cantidad de transposones que

contienen determinantes de resistencia a mercurio (*mer*) distribuidos tanto en aislamientos clínicos como en bacterias ambientales (Misra, 1992, Mindlin, *et al.*, 2001). Este *cluster* identificado en el set de datos plasmídico es similar a los encontrados previamente en el transposón *Tn4378* del megaplásmido pMOL28 de *Cupriavidus metallidurans* CH34 (Taghavi, *et al.*, 1997), en el *Tn501* del plásmido pVS1 de *Pseudomonas aeruginosa* (Brown *et al.*, 1986) en el plásmido de resistencia a mercurio pPB de *Pseudomonas stutzeri* (Reniero, *et al.*, 1998), y en el *Tn501* del plásmido pJP4 (Trefault, *et al.*, 2004). Además de este *cluster*, se identificaron otros genes de resistencia a mercurio. También encontramos genes asociados al transporte/resistencia de arsénico, cobalto, zinc, cobre, cadmio, hierro, níquel, telurito, selenio y manganeso (ver Tabla Suplementaria S7), indicando una gran cantidad y variedad de genes involucrados en el flujo de cationes metálicos presentes en los plásmidos.

VI.3.3. Identificación de genes asociados a la degradación de compuestos xenobióticos e hidrocarburos aromáticos

Teniendo en cuenta que la colección de aislamientos bacterianos fue aislada a partir de un biofiltro en el que son depositados diversos pesticidas, sería esperable que los microorganismos que habitan este ambiente, al estar bajo una presión de selección constante, desarrollen la capacidad de sobrevivir en tales condiciones. Como los plásmidos juegan un rol esencial en la adaptación bacteriana a distintos cambios ambientales, podría esperarse que los genes asociados a la tolerancia/degradación de xenobióticos estén localizados en estos MGEs. En general, es difícil de atribuir a genes particulares las funciones específicas involucradas en la degradación de pesticidas y otros xenobióticos, ya que la degradación de estos compuestos suele ser compleja y llevarse a cabo mediante una serie de reacciones químicas que implican diversas enzimas, cuyos genes que las codifican, en muchos casos no han sido descritos. Sin embargo, los genes involucrados en algunas vías de degradación han sido bien caracterizados. En este marco, hemos podido llevar a cabo la búsqueda *in silico* de ciertos genes que han sido vinculados a la degradación de determinados compuestos xenobióticos. Para ello, se utilizó la anotación automática generada por la plataforma

GenDB (*contigs* ensamblados de origen plasmídico) y también se utilizaron genes depositados en la base de datos de GenBank, con los cuales se realizó una búsqueda de similitud de secuencia con la herramienta BLASTp en la plataforma de GenDB. En la Tabla Suplementaria S8 se muestra un listado con los genes de referencia (Fang et al 2014) utilizados para la búsqueda por BLASTp.

Este análisis nos permitió identificar dos posibles genes (*tfdA* y *tfdB*) que codifican para las enzimas principales requeridas en la degradación de ácido 2,4-diclorofenoxiacético (2,4-D), el gen que codifica para la enzima responsable del primer paso en la degradación de atrazina (*atzA*), otros dos genes que codifican para las subunidades α y β de la enzima benzoato 1,2-dioxigenasa requerida para la degradación de 3-fluorobenzoato, y dos genes cuyos productos enzimáticos participan en la degradación de dibenzofurán (ver Tabla VI.2). También hemos detectado los genes únicos requeridos para generar resistencia a metilviológenos (transportadores del tipo MFS: *Major Facilitator System*) y para la degradación de paratión.

Además de estos genes, investigamos la presencia de aquéllos involucrados en la degradación de hidrocarburos aromáticos. Esta búsqueda permitió la identificación de algunos genes que permiten la utilización de xileno/tolueno/benceno (Tabla VI.2). Dos *contigs* mostraron alta similitud con el plásmido pWW0 de *Pseudomonas putida*. Este plásmido codifica para las rutas metabólicas que degradan tolueno, xileno y sus alcoholes y carboxilatos derivados (Assinder & Williams, 1990, Harayama & Reik, 1990). Uno de los *contigs* identificados tiene una longitud de 9 Kb y contiene 4 genes involucrados en la rutas superiores de degradación (*xyIC-xyIM-xyIA-xilB*); otro *contig* de 2Kb contiene 3 genes pertenecientes a la ruta inferior de degradación (*xyII-xyIH-xyIS*) de estos compuestos aromáticos.

VI.3.4. Presencia de otros MGEs en los plásmidos

Como hemos descrito en la introducción de este trabajo, los plásmidos suelen estar asociados a otros MGEs como integrones, secuencias de inserción y transposones. Como consecuencia de los resultados obtenidos en el *screening* por PCR realizado para integrones de clase 1, en el cual 6 de los 35 aislamientos de la colección

Tabla VI.2. Lista de posibles genes involucrados en la degradación de xenobióticos e hidrocarburos aromáticos

Sustrato	ID en GenDB	Primer similitud en el análisis BLASTp	Organismo	Cobertura	E-value	Máxima identidad	Número de acceso GenBank	Observaciones (Segunda similitud/Predicción)
ácido 2,4-diclorofenoxiacético (2,4-D)	BJP_2678	2,4-diclorofenoxiacetato dioxigenasa	<i>Burkholderia dolosa</i>	99%	1,00E-71	44%	WP_006764062.1	gen <i>tfdA</i>
	BJP_155	dioxigenasa del catabolismo de taurina	<i>Acidovorax</i> sp.	97%	2,00E-130	62%	GAD20643.1	gen <i>tfdA</i>
	BJP_2488	fenol 2-monooxigenasa	<i>Acinetobacter</i> sp.	97%	0.0	61%	KEC84934.1	gen <i>tfdB</i>
	BJP_10655	taurina dioxigenasa dependiente de alfa-cetoglutarato	<i>Achromobacter xylooxidans</i>	88%	2,00E-108	68%	YP_008028310.1	gen <i>tfdA</i>
Atrazina	BJP_9495	clorohidrolasa	<i>Paenibacillus</i> sp.	98%	2,00E-41	96%	ETT37468.1	gen <i>atzA</i>
	BJP_2294	oxidoreductasa	<i>Variovorax</i> sp.	98%	0.0	85%	WP_028248695.1	benzoato 1,2-dioxigenasa, subunidad alfa
3-Fluorobenzoato	BJP_2293	subunidad beta de la 1,2-dioxigenasa aromática	<i>Variovorax paradoxus</i>	99%	9,00E-87	83%	WP_019654664.1	ND
	BJP_10023	6-clorohidroquinol-1,2-dioxygenasa	<i>Acidovorax</i> sp.	98%	4,00E-170	83%	WP_007855039.1	ND
Dibenzofurano	BJP_4909	2,3-dihidroxi-2,3-dihidrofnylpropionato deshidrogenasa	<i>Escherichia coli</i>	100%	0.0	99%	POCI32.1	2,3-dihidroxi-2,3-dihidrofnylpropionato deshidrogenasa
	BJP_6277	glioxalasa	<i>Methylocella silvestris</i>	98%	4,00E-137	64%	WP_012590495.1	2,3-dihidroxi-bifeny 1,2-dioxigenasa
	AP_378	2-hidroxi-6-oxo-6-fenilhexa-2,4-dienoato hidrolasa	<i>Pseudomonas</i> sp.	100%	3,00E-179	86%	KFJ90076.1	2-hidroxi-6-oxo-6-fenilhexa-2,4-dienoato hidrolasa
Paratión	BJP_3668	fosfotriesterasa	<i>Escherichia coli</i>	100%	0.0	99%	WP_000007005.1	ND
metil viológeno	BJP_8262	proteína hipotética	alpha proteobacterium	96%	4,00E-167	99%	WP_017502939.1	proteína SmvA, resistencia a metil viológenos
	BJP_3584	proteína de membrana	<i>Enterobacter</i> sp.	99%	0.0	100%	YP_008098281.1	proteína SmvA, resistencia a metil viológenos
	BJP_4968	superfamilia de proteínas mayor facilitador	<i>Comamonas testosteroni</i>	95%	3,00E-59	65%	YP_003277211.1	Proteína de resistencia a metil viológenos

Tabla VI.2 (continuación)

Sustrato	ID en GenDB	Primer similitud en el análisis BLASTp	Organismo	Covertura	E-value	Máxima identidad	Número de acceso GenBank	Observaciones (Segunda similitud/Predicción)
	BJP_5604	4-oxalocrotonato decarboxilasa	<i>Pseudomonas putida</i>	99%	0.0	99%	NP_542860.1	ND
	BJP_5605	proteína permeasa transportador ABC	<i>Pseudomonas putida</i>	98%	8,00E-36	100%	NP_542859.1	4-Oxalocrotonato tautomerasa
	BJP_5606	proteína regulatoria	<i>Pseudomonas putida</i>	99%	2,00E-97	100%	NP_542858.1	ND
	BJP_8726	4-oxalocrotonato decarboxilasa	<i>Burkholderia sp.</i>	99%	2,00E-121	79%	ACO92645.1	ND
	BJP_5748	transportador <i>mayor</i> <i>facilitador</i>	<i>Burkholderia terrae</i>	87%	2,00E-119	76%	WP_007588851.1	proteína transportadora de benzoato BenK
	BJP_5073	región Rieske (2Fe-2S)	<i>Dechloromonas aromatica</i>	88%	1,00E-26	54%	YP_287016.1	tolueno-4-monooxigenasa (complejo de la proteína efectora)
Xileno/ tolueno/ benceno	BJP_9384	ferredoxina-NAD(+) reductasa FAD/NAD(P)- proteína de unión	<i>Cupriavidus metallidurans</i>	95%	2,00E-66	41%	YP_583463.1	tolueno-4-monooxigenasa (complejo de la proteína efectora)
	BJP_10390	proteína hipotética	<i>Ramlibacter tataouinensis</i>	82%	3,00E-92	72%	YP_004617916.1	familia de proteína de tolerancia a tolueno
	BJP_1643	bencil alcohol deshidrogenasa	<i>Pseudomonas sp.</i>	99%	7,00E-108	97%	WP_028624487.1	gen <i>tdh</i>
	BJP_1644	bencil alcohol deshidrogenasa	<i>Pseudomonas putida</i>	99%	7,00E-94	98%	WP_004575908.1	gen <i>tdh</i>
	BJP_1645	bencil alcohol deshidrogenasa	<i>Pseudomonas putida</i>	99%	0.0	98%	NP_542888.1	gen <i>xy/C</i>
	BJP_1646	xileno monooxigenasa	<i>Pseudomonas sp.</i>	99%	0.0	94%	WP_028624485.1	gen <i>xy/IM</i>
	BJP_1647	componente de transferencia de electrones xileno monooxigenasa	<i>Pseudomonas putida</i>	99%	0.0	99%	YP_003617172.1	gen <i>xi/A</i>
	BJP_1648	alcohol deshidrogenasa	<i>Pseudomonas sp.</i>	99%	0.0	99%	WP_028624483.1	ND

ID: código de identificación en la plataforma GenDB.

ND: no disponibles.

resultaron positivos para *intl1* (ver Capítulo IV, sección IV.2.1), realizamos la búsqueda *in silico* del gen que codifica para la integrasa de los integrones clase 1 en los *contigs* de origen plasmídico. Sin embargo no fueron detectadas secuencias homólogas a este gen, lo cual resulta llamativo si tenemos en cuenta que en al menos 4 de los 6 aislamientos *intl1* positivos se determinó la localización plasmídica de este gen (Capítulo IV, sección IV.2.7). Estos resultados ponen en evidencia que: **i)** no todos los plásmidos presentes pudieron ser purificados, o **ii)** que no todas las secuencias plasmídicas están representadas en los datos obtenidos del secuenciamiento y/o, **iii)** que ciertas secuencias de origen plasmídico pudieron haber sido filtradas (con los parámetros utilizados) al tratar de eliminar la contaminación de ADN cromosomal. En este sentido, decidimos buscar el gen *intl1* en las secuencias clasificadas como “de origen cromosomal” que descartamos. Así, pudimos identificar un *contig* que contiene el gen *intl1* truncado junto a un gen *aadA* (que confiere resistencia a estreptomicina/espectinomicina), también truncado. Este pequeño *contig* (de ~700 pb de longitud), fue filtrado con los parámetros utilizados y eliminados del conjunto de datos plasmídicos. Dado que en 3 de los aislamientos positivos para *intl1* se demostró la presencia del gen *aadA* en la región adyacente a dicho gen (ver Capítulo IV, sección IV.2.5), se refuerza la hipótesis de que este *contig* pueda haber sido considerado de

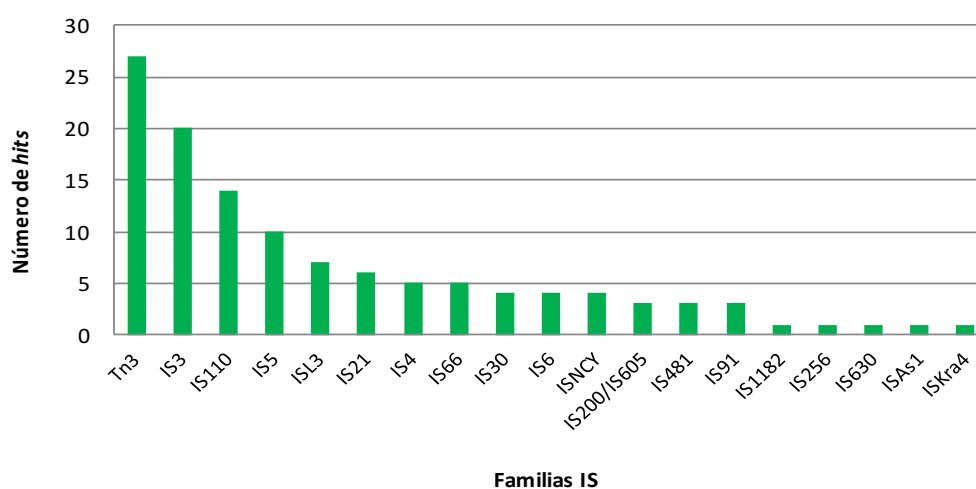


Figura VI.6. Familias de secuencias de inserción encontradas en el conjunto de secuencias plasmídicas. Se muestra el número de las ISs detectadas utilizando la base de datos ISfinder.

procedencia cromosomal y removido en el procedimiento de filtrado bioinformático del ADN. Independientemente de cuál sea la causa de la ausencia del gen *intl1* en las secuencias plasmídicas, estos resultados ponen en evidencia las limitaciones que los métodos de clasificación *in silico* de secuencias presentan, y nos permite suponer que otras secuencias de origen plasmídico también pueden haber sido removidas.

Las secuencias de inserción son MGEs simples que consisten en el gen que codifica para una transposasa (*tnpA*) requerida para su movilidad y repeticiones invertidas en los extremos. Además, estos elementos pueden transportar otros genes accesorios originando transposones. La comparación de los *contigs* plasmídicos contra la base de datos de transposones y secuencias de inserción ISfinder (Siguier, *et al.*, 2006), reveló un total de 120 coincidencias (*matches*), siendo la familia *Tn3* la más representada (27 *hits*). La familia *Tn3* es una de las más abundantes en la naturaleza y se encuentra ampliamente diseminada entre distintos grupos bacterianos, estando ampliamente ligada a la diseminación de genes de resistencia a antibióticos y genes catabólicos (Nicolas, *et al.*, 2015). Otras familias encontradas en nuestra base de datos fueron la *IS3* (20 *hits*) e *IS110* (14 *hits*), como puede verse en la Figura VI.6 y en la Tabla Suplementaria S9.

VI.4. Discusión

El conocimiento de la información que transportan los plásmidos resulta importante para determinar qué funciones o ventajas pueden proporcionar éstos a las bacterias que los hospedan, así como a la comunidad o consorcio bacteriano del que pueden formar parte. En este capítulo nos hemos centrado en el análisis bioinformático de los tipos de replicones y sus capacidades de diseminación (genes de movilidad), así como en la detección y clasificación de genes de resistencia a antibióticos y metales, y genes involucrados en la degradación de compuestos xenobióticos e hidrocarburos aromáticos. La diseminación de genes de resistencia a antimicrobianos, virulencia y biodegradación son ejemplos del impacto de la HGT en la

adaptación de las comunidades bacterianas bajo estrés, con consecuencias relevantes para la sociedad humana (de la Cruz y Davies, 2000).

Los plásmidos se han postulado como los principales vehículos de ADN que transfieren genes entre los cromosomas bacterianos, con la conjugación como el mecanismo de transferencia preponderante (Halary *et al.*, 2010, Kloesges *et al.*, 2011). Es importante remarcar que, si bien los plásmidos aquí caracterizados provienen de 35 aislamientos seleccionados, éstos conforman una parte del *pool* de elementos móviles de la comunidad de la que fueron obtenidos, por lo que estos replicones podrían hospedarse en otros géneros o especies bacterianas, incluso en aquellas no cultivables.

En este sentido, a partir del análisis de las proteínas de los módulos de replicación, se corroboró la presencia de plásmidos capaces de habitar una amplia diversidad de huéspedes. Además, la identificación de proteínas significativamente diferentes a las previamente identificadas y depositadas en las bases de datos (es decir, aquéllas que presentaron bajos porcentajes de identidad con sus homólogos más cercanos), sugiere la presencia de proteínas de replicación plasmídica aún inexploradas, extendiendo la diversidad de los módulos de replicación plasmídica. Por otro lado, la gran cantidad de genes de movilización encontrados sugiere la presencia de plásmidos con movilidad. Más aún, el análisis bioinformático nos permitió identificar 19 genes que codifican para relaxasas, las cuales pudieron ser asignadas a 4 familias MOB distintas, de acuerdo con la clasificación establecida por Garcillán-Barcia y colaboradores (2009). Además, se encontraron 13 proteínas del tipo VirB4, implicadas en la formación del MPF. La presencia de ambos tipos de proteínas (relaxasa y VirB4), junto con la gran cantidad de genes *mob-tra* identificados, refuerzan la idea de que los plásmidos obtenidos de la colección de aislamientos seleccionados del biofiltro presentan movilidad, pudiendo ser los mismos movilizables o autotransmisibles. En este sentido, estos datos se condicen con los resultados obtenidos previamente en los ensayos de conjugación bacteriana (Capítulo III, sección III.6), en los que se evidenció movilidad en al menos 7 replicones. Es importante mencionar también la presencia de un gran número de secuencias de inserción. Estos

elementos contribuyen ampliamente a la transferencia horizontal de material genético entre bacterias, y tienen una gran implicancia en la diseminación de genes de resistencia a antibióticos en el ambiente (Sheldon, 2005, Gootz, 2010). Todos estos datos, en conjunto, permiten inferir un flujo dinámico de información genética en la comunidad del biofiltro.

Por otro lado, el análisis *in silico* ha permitido la identificación de más de 100 genes de resistencia a antimicrobianos, confiriendo resistencia a 9 clases de antibióticos distintos. Estos resultados refuerzan los obtenidos en el capítulo III, en donde se determinó la tolerancia de los aislamientos a distintos antibióticos, y se corroboró la ubicación plasmídica para algunos de estos determinantes.

Por último, se pudo determinar bioinformáticamente la presencia de algunos genes cuyos productos proteicos están involucrados en la degradación de compuestos xenobióticos, como ácido 2,4-diclorofenoxiacético, atrazina, 3-fluorobenzoato, dibenzofurán, paratión y de resistencia a metil viológenos, así como en la degradación de hidrocarburos aromáticos como xileno, tolueno y benceno. Estas funciones resultan de particular interés, dado que están estrechamente relacionadas con el ambiente del cual los plásmidos fueron aislados, sugiriendo que los aislamientos en los que se encuentran podrían tener una participación activa en la degradación de estos compuestos. Un análisis similar al realizado en este trabajo fue realizado recientemente (Fang, *et al.*, 2014), en el cual se utilizaron herramientas bioinformáticas para detectar, a partir de datos metagenómicos, genes asociados a la degradación de este tipo de compuestos, permitiendo hacer inferencias sobre potenciales rutas degradativas en comunidades microbianas. Sin embargo, este tipo de estudios requiere ser complementado con ensayos de degradación *in vitro* para confirmar la capacidad de las bacterias que contienen estos genes de participar en la degradación de estos compuestos.

En resumen, en este capítulo hemos utilizado abordajes bioinformáticos para detectar y clasificar la información obtenida luego del secuenciamiento masivo del ADN plasmídico. Asociar las secuencias de ADN con posibles funciones biológicas

puede dar un indicio del rol que cumplen los plásmidos presentes en este nicho enriquecido en pesticidas.

CAPÍTULO VII

**BIOPROSPECCIÓN: BÚSQUEDA, CLONADO,
EXPRESIÓN Y CARACTERIZACIÓN DE ENZIMAS CON
POTENCIAL INTERÉS INDUSTRIAL**

VII.1. Introducción

La bioprospección puede ser definida como la búsqueda sistemática de componentes naturales y organismos completos de la biodiversidad con el fin de otorgarles un valor comercial para el desarrollo de productos (Feinsilver, 1996; Laird *et al.*, 2002; Castree, 2003). La bioprospección genera productos relacionados con industrias como la farmacéutica, biotecnológica, alimenticia, de agroinsumos y medicina, entre otras.

En este marco, las enzimas microbianas resultan de gran importancia en el desarrollo de bioprocesos industriales. Las mismas se emplean en una gran variedad de aplicaciones en diversas áreas, como en la obtención de pulpa y papel, cuero, textiles, detergentes, productos farmacéuticos, productos químicos, alimentos, bebidas, biocombustibles, alimentación y cuidado personal, entre otros (Kirchner & Tauch, 2003). Muchos procesos industriales presentan varias desventajas (desde un punto de vista práctico y económico), como por ejemplo la baja eficiencia catalítica, la falta de especificidad enantiomérica en la síntesis quiral, el uso de solventes que generan desechos orgánicos y contaminantes, la necesidad de altas temperaturas, bajos pHs y altas presiones, entre otras. Por su parte, las enzimas resultan más útiles para estas aplicaciones, ya que pueden funcionar en condiciones moderadas (como por ejemplo, presión atmosférica, temperaturas moderadas y pH cercano al fisiológico), no necesitan protección de grupos funcionales de sustratos, tienen generalmente una larga vida media, poseen una alta estereoselectividad, pueden acelerar la velocidad de las reacciones entre 5 y 17 órdenes de magnitud y pueden trabajar con sustratos artificiales (Marrs *et al.*, 1999; Johnson, 2013). Todas estas ventajas han llevado a reemplazar eficientemente ciertos pasos químicos y físicos en los procesos industriales por reacciones enzimáticas.

La diversidad microbiana junto con el avance de las técnicas moleculares modernas, como la genómica y la metagenómica, el desarrollo de sistemas de expresión y la posibilidad de sintetizar moléculas de ADN recombinante *in vitro*, han derivado en el descubrimiento de nuevas enzimas cuyas propiedades catalíticas

pueden ser incluso modificadas y mejoradas por estrategias basadas en evolución dirigida racional, semirracional o azarosa. En la actualidad, más de 500 productos son elaborados a partir del uso de enzimas (Kumar & Singh, 2013), y alrededor del 90% de ellas han sido modificadas por técnicas de ADN recombinante (Adrio, *et al.*, 2014).

Entre las herramientas más utilizadas para el descubrimiento de nuevas enzimas, se pueden mencionar las genómicas (exploración de los más de 3.500 genomas bacterianos secuenciados), los *screenings* metagenómicos y la búsqueda de actividades en microorganismos extremófilos. En particular, los *screenings* genómicos y metagenómicos, que son los que empleamos en este trabajo, se basan principalmente en aproximaciones funcionales o de similitud de secuencia. En esta última aproximación, se utilizan distintas herramientas bioinformáticas que permiten identificar regiones conservadas, dominios funcionales o estructurales o proteínas completas, las cuales son candidatas para realizar una posterior caracterización. Estas enzimas hipotéticas están sujetas a subsiguientes etapas de clonado, sobreexpresión, *screening* de sustratos y medida de actividad.

Entre las enzimas más utilizadas en procesos industriales, se pueden mencionar las proteasas, lipasas, amilasas, dehalogenasas, celulasas, quitinasas y lacasas. En este capítulo del trabajo, nos enfocaremos en la búsqueda *in silico* de algunas enzimas de interés industrial, realizando una selección de genes candidatos para su posterior clonado, expresión, purificación y estudio de sus parámetros cinéticos.

VII.2. Búsqueda bioinformática de enzimas de potencial interés industrial

Existe una gran variedad de enzimas que presentan distintas actividades que son utilizadas industrialmente. En este trabajo, nos enfocamos en la búsqueda *in silico* de quitinasas y lacasas, las cuales son ampliamente usadas en bioprocesos. A partir del set de datos de secuencias plasmídicas ensambladas se realizó la búsqueda de estas enzimas utilizando la plataforma GenDB, que permite buscar a partir de la anotación

automática generada por el programa. A continuación se describe la búsqueda de estas dos clases de enzimas.

VII.2.1. Búsqueda de genes que codifican para quitinasas

Las quitinasas (E.C 3.2.2.14) son enzimas capaces de hidrolizar quitina en sus componentes oligo y monoméricos. Los quitooligosacáridos, el dímero acetilquitobiosa y los monómeros de N-acetilglucosamina son de gran interés para la industria debido a su amplio rango de aplicaciones médicas, agrícolas e industriales (Bhattacharya *et al.*, 2007). Estas enzimas se han encontrado en una gran variedad de organismos como bacterias, hongos, plantas, artrópodos y humanos, y participan en numerosos procesos fisiológicos como nutrición, parasitismo, morfogénesis e inmunidad (Adrangi & Faramarzi, 2013). En el caso de las bacterias productoras de quitina, éstas se han encontrado en diversos ambientes como suelos, desechos, compost y aguas termales (Debaditya Bhattacharya, *et al.*, 2007). El interés comercial en la utilización de derivados de quitina ha llevado a la necesidad de buscar fuentes estables, seguras y baratas de enzimas quitinolíticas aisladas de organismos, como alternativa a los métodos químicos.

Para emprender la búsqueda de quitinasas en los plásmidos secuenciados, se utilizó la plataforma GenDB. El análisis realizado (ver materiales y Métodos, sección II.3.7) permitió identificar 3 ORFs con localización plasmídica que podrían codificar para quitinasas, los cuales presentaron un alto nivel de identidad (98-100%) con sus homólogos más cercanos cuando se corroboró su identidad en GenBank. En la tabla VII.1 se muestran los resultados de la búsqueda de identidad en GenBank a nivel aminoacídico.

En virtud a los resultados precedentes y dado que los tres ORFs encontrados que codifican para quitinasas presentaron un alto nivel de identidad (entre 98 y 100%) con otras quitinasas previamente descritas y caracterizadas, se decidió no avanzar en las etapas posteriores de clonado y medida de actividad de estas enzimas.

VII.2.2. Búsqueda de genes que codifican para lacasas

Las lacasas (EC 1.10.3.2) son polifenol oxidasas que catalizan la oxidación de una gran cantidad de sustratos orgánicos e inorgánicos, incluyendo mono, di y polifenoles, aminofenoles, metoxifenoles, aminos aromáticas y ascorbato (Madhavi & Lele, 2009). Estas enzimas constituyen el miembro más numeroso de la familia de las proteínas multicobre, y se encuentran ampliamente distribuidas en plantas, insectos, hongos y bacterias. La capacidad de las lacasas de actuar sobre una gran variedad de sustratos las convierte en biocatalizadores ampliamente utilizados en varios procesos biotecnológicos. Estas enzimas también tienen un potencial uso en biorremediación ambiental, ya que oxidan ciertos compuestos contaminantes como herbicidas y pesticidas, y pueden actuar como sistemas de purificación de aguas y suelos (Adrio, *et al.*, 2014).

La identificación *in silico* de ORFs que codifican para lacasas se realizó utilizando las mismas herramientas que se utilizaron para la identificación de quitinasas, junto con una búsqueda adicional utilizando un perfil del tipo HMM (*Hidden Markov Models*) (Ausec *et al.*, 2011), como se describe en la sección II.3.7 de Materiales y Métodos. La implementación de estas herramientas nos permitió encontrar dos ORFs

Tabla VII.1. Lacasas y quitinasas detectadas en el conjunto de secuencias de origen plasmídico

ID en GenDB	Descripción	Organismo	Cobertura*	E value*	Identidad*	N° de acceso*
Quitinasas						
BJP_1998	bifuncional quitinasa/lysozima	Proteobacteria	99%	0.0	100%	WP_000773156.1
BJP_5503	quitinasa	<i>Serratia marcescens</i>	99%	0.0	99%	WP_033639033.1
BJP_8683	quitinasa	<i>Serratia</i>	97%	0.0	98%	WP_012145450.1
Lacasas						
BJP_7118	Multicobre oxidasa tipo 2	<i>Delftia acidovorans</i>	99%	0.0	57%	YP_001564396.1
AP_188	Multicobre oxidasa tipo 2	<i>Ochrobactrum rhizosphaerae</i>	99%	0.0	100%	WP_024899901.1

*Estas comparaciones fueron realizadas con los productos de traducción de los respectivos ORFs.

ID: código de identificación en la plataforma GenDB.

que codifican para estas enzimas en nuestro set de secuencias (Tabla VII.1). El producto de traducción de uno de ellos (BJP_7118) presenta una baja similitud con su proteína homóloga más cercana (57% de identidad a nivel aminoacídico), una multicobre oxidasa de *Delftia acidovorans*. Esta oxidasa pertenece al grupo de las lacasas de tres dominios cupredoxina. Si bien en un principio su caracterización podría resultar interesante por su bajo nivel de similitud con la proteína más cercana depositada en la base de datos de GenBank, la comparación *in silico* de ambas proteínas puso en evidencia que esta proteína está truncada (carece de aproximadamente 100 aminoácidos en el extremo N-terminal), por lo cual probablemente su producto de expresión no sea funcional. Esta lacasa hipotética se encuentra ubicada en un *contig* pequeño que sólo contiene parcialmente a este ORF, lo cual podría indicar que este gen no se encuentra truncado naturalmente, sino que a causa del tamaño del *contig*, sólo hemos podido acceder de forma incompleta a la secuencia del gen.

En el momento en el que comenzamos con la búsqueda y análisis de genes que codifican para lacasas, el segundo ORF identificado (AP_188) mostró un 90% de identidad con la proteína homóloga más cercana depositada en GenBank, que está anotada como una multicobre oxidasa de 3 dominios cupredoxina presente en *Ochrobactrum anthropi*. Sin embargo, una búsqueda reciente en esta misma base de datos reveló una identidad del ORF AP_188 del 100% con la misma enzima, de *Ochrobactrum rhizosphaerae* (ver Tabla VII.1). Sin embargo, ninguna de estas proteínas homólogas ha sido caracterizada. La proteína AP_188, con potencial actividad lacasa, nos resultó de particular interés ya que se han reportado previamente lacasas pertenecientes al género *Ochrobactrum* con buena actividad sobre distintos sustratos y capaces de mantener su actividad enzimática en un rango amplio de pH, donde otras lacasas clásicas pierden su capacidad biocatalizadora (Yang, *et al.*, 2012; Rahman, *et al.*, 2013). Por las razones expuestas, y contando en nuestro banco de datos con su secuencia completa, decidimos concentrarnos en su caracterización molecular, clonado y expresión para avanzar luego en el estudio de sus propiedades cinéticas. De

aquí en adelante, nos referiremos al ORF AP_188 como *lac188* y a su producto de traducción como Lac188.

VII.2.2.1. Estructura y secuencia de la lacasa Lac188

El ORF *lac188* posee 1.509 nucleótidos que codifican para una lacasa hipotética de 502 aminoácidos, cuyo peso molecular estimado es de 53,68 kDa. El análisis *in silico* reveló que la misma contiene tres dominios conservados cobre-oxidasa (cupredoxina) pertenecientes a los Pfams PF07732, PF00394 y PF07731 (Figura VII.1). También se evaluó la presencia de un posible péptido señal utilizando el servidor SignalP4.1 (<http://www.cbs.dtu.dk/services/SignalP/>), el cual predijo un corte entre el aminoácido número 20 y 21 de la proteína. Este dato es importante a la hora de realizar el clonado del gen: en consecuencia decidimos que era conveniente su remoción. Esta decisión se basó en que generalmente la remoción exacta del péptido señal es esencial para el plegado correcto de los dominios Cu-oxidasa y nos planteamos recuperar la proteína del citoplasma bacteriano.

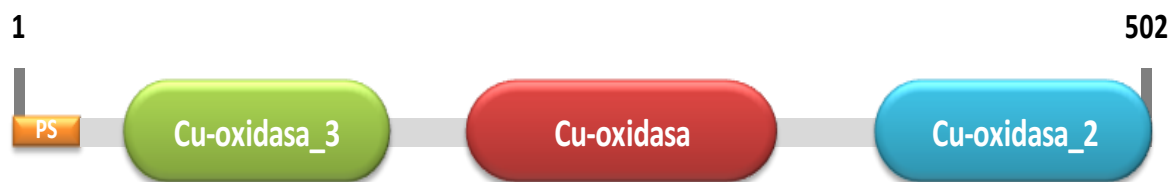


Figura VII.1. Esquema de la estructura de la lacasa Lac188. Se observan los tres dominios cupredoxina: Cu-oxidasa_3 (PF07732), Cu-oxidasa (PF00394) y Cu-oxidasa_2 (PF07731). Se indica el péptido señal (PS) en el extremo N-terminal de la proteína.

VII.2.2.2. Elección del sistema de expresión

Si bien lacasas fúngicas se han utilizado más ampliamente en procesos industriales, siendo sobreexpresadas heterológamente en *Saccharomyces cerevisiae*, *Trichoderma reesei*, *Pichia pastoris*, *Aspergillus nidulans*, etc., varias lacasas bacterianas se han sobreexpresado en *Escherichia coli* (Kunamneni *et al.*, 2008). Sin embargo, los niveles de enzima obtenidos están muchas veces lejos de los rendimientos exigidos en las aplicaciones industriales.

Para lograr la sobreexpresión del gen y obtener elevadas cantidades de proteína, se decidió utilizar un sistema de expresión heterólogo empleando *E. coli* como cepa de expresión y un plásmido de tipo pET. Este tipo de plásmidos incluye un promotor transcripcional del bacteriófago T7.

Por otro lado, las cepas de *E. coli* empleadas contienen el gen de la ARN polimerasa del fago T7, la cual es producida mediante la inducción con IPTG a través del promotor *lacUV5*. Además, algunos de estos vectores de tipo pET tienen una gran cantidad de sitios de restricción que facilitan el clonado del gen y permiten la adición de una cola de histidinas a la proteína, facilitando su posterior purificación por cromatografía de afinidad. En consecuencia, para la expresión de la lacasa *lac188* se decidió utilizar el vector pET22b(+) de Novagen (Figura Suplementaria S11), el cual posee un promotor T7, y permite la adición de una cola de histidinas al extremo C-terminal de la proteína. En la figura VII.2 se muestra el esquema final de la construcción pET22b(+)-*lac188*. En cuanto a la cepa utilizada para la expresión, se eligió la *E. coli* BL21(DE3)Star, que permite la obtención de grandes cantidades de proteína exógena (porque tiene mutado el gen *rne31* y esto genera un aumento en la estabilidad del ARNm) y es de crecimiento rápido (Materiales y Métodos, Tabla II.1).

VII.2.2.3. Obtención de una construcción portadora del gen de la lacasa *lac188*

Para el diseño de los *primers* se tuvo en cuenta la exclusión de los primeros 60 nucleótidos del gen (que codifican para el péptido señal predicho), así como del codón de *stop*, para realizar una fusión traduccional que permita la incorporación de 6 histidinas en el extremo C-terminal de la proteína.

Dado que los datos de secuencias plasmídicas provienen del *pool* de plásmidos de la colección, no se conoce en principio qué aislamiento es el portador del gen de la lacasa *lac188*. Para identificarlo, amplificarlo completamente y clonarlo, se realizó el diseño de los *primers* *lac80Ndel/lac80Ncol* (ver Materiales y Métodos, Tabla II.4) basado en la secuencia identificada *in silico*. Con este juego de *primers* se realizó un *screening* por PCR utilizando como ADN molde extracciones

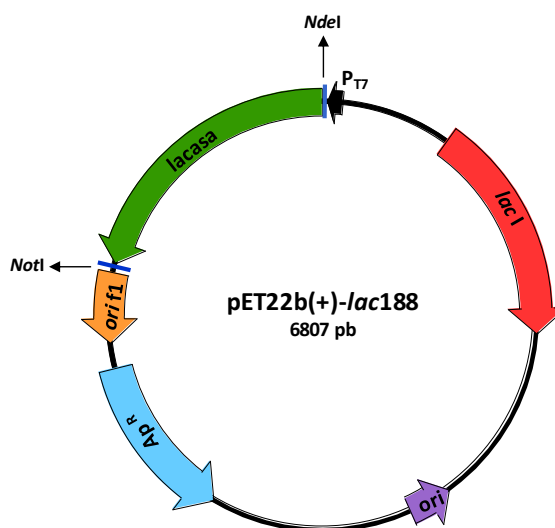


Figura VII.2. Esquema de la construcción pET22b(+)-*lac188*. Se muestran los elementos principales del vector pET22b(+): *ori f1* (origen del fago f1), *ori* (origen de replicación plasmídica), *lacI* (gen que codifica para el represor LacI), Ap^R (gen de resistencia a ampicilina), P_{T7} (promotor T7) y el gen clonado *lac188*. Se muestran los sitios *NdeI* y *NotI*, utilizados en el clonado del gen.

de ADN total de los aislamientos. En todas las condiciones de amplificación ensayadas, la corrida electroforética en gel de agarosa mostró una gran cantidad de bandas inespecíficas en todos los aislamientos, imposibilitando la identificación del aislamiento que porta el fragmento de interés. Esto puede deberse a que la lacasa en cuestión pertenece al tipo de las oxidasas, las cuales son ubicuas y abundantes en los genomas bacterianos, dificultando la detección del microorganismo en cuestión. En consecuencia de los resultados obtenidos, y aprovechando la posibilidad de realizar la optimización de la secuencia en relación a la frecuencia relativa del uso de codones para expresar la enzima en *E. coli* y así mejorar el rendimiento de la producción de la proteína, se procedió a sintetizar químicamente la secuencia optimizada de *lac188* (GenScript), adicionándole secuencias para enzimas de restricción *NdeI* y *NotI* que facilitarán en una etapa posterior el subclonado del mismo. A partir del gen sintetizado, el cual fue en una primera etapa clonado en el vector pUC57 (ver Figura Suplementaria S12), luego se realizó el subclonado del gen en el vector pET22b(+) y se

transformó la construcción resultante (denominada pET22b(+)-*lac188*) en la cepa de *E. coli* BL21 (DE3)Star.

VII.2.2.4. Expresión heteróloga de la enzima Lac188

A fin de confirmar que el gen *lac188* clonado en el plásmido pET22b(+) codifica para una enzima con actividad multicobre oxidasa, se realizaron ensayos para poner a punto las condiciones de expresión en la cepa de *E. coli* BL21(DE3)Star, como se indica en Materiales y Métodos, sección II.2.16.1. En un primer paso, se comparó la expresión de la proteína de interés a dos temperaturas (25 °C y 37 °C) y distintos tiempos de inducción, con el agregado de IPTG como inductor (Figura VII.3). A partir de este ensayo, se determinó que a 37°C, se produjo mayor cantidad de la proteína recombinante que a 25°C.

Se ha reportado que la adición de iones de cobre facilita el plegamiento adecuado de la mayoría de las lacasas estudiadas (Koschorreck *et al.*, 2008; Gunne *et*

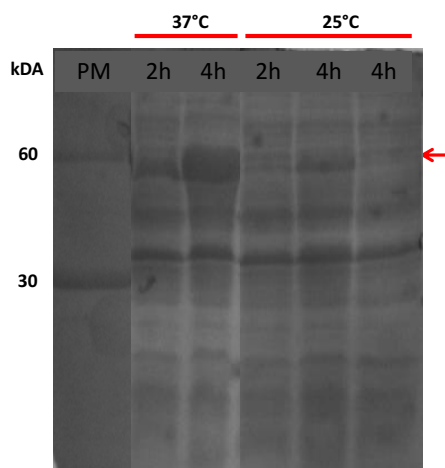


Figura VII.3. Análisis mediante SDS-PAGE de la expresión de Lac188 a distintas temperaturas y tiempos de inducción. En la figura se muestran las proteínas totales en un gel de 10% SDS-PAGE, tras inducir con IPTG 0,4 mM. La cepa con la construcción recombinante se creció en medio LB líquido a 25 °C y 37 °C y 200 rpm. Cuando se alcanzó una DO_{600} de 0,6 se realizó la inducción. Las células se recolectaron 1 y 5 horas post-inducción con 0,4 mM IPTG.

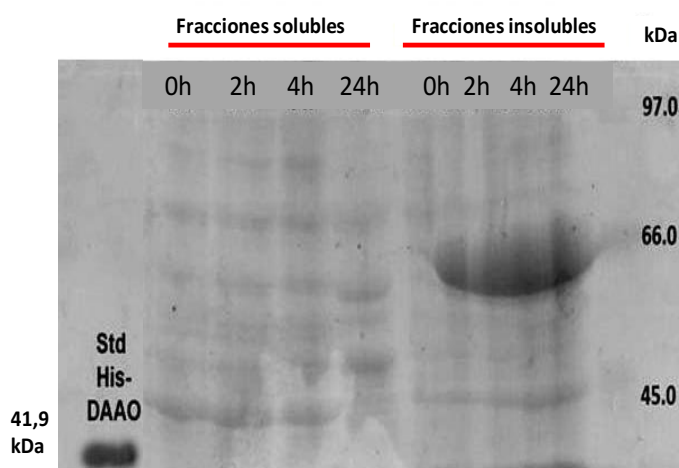


Figura VII.4. Análisis mediante SDS-PAGE de la expresión de Lac188. En la figura se muestra un gel de 10% SDS-PAGE, en el que se indican las fracciones soluble e insoluble. La cepa con la construcción recombinante se creció en medio LB líquido a 37 °C y 200 rpm. Cuando se alcanzó una $DO_{600} = 0,6$ se realizó la inducción por agregado de 0,4 mM IPTG y 0,25 mM $CuSO_4$. Las células se recolectaron luego de 0, 2, 4 y 24 horas, luego de incubar (a 37°C) sin agitación. Std: proteína estándar de referencia (41,9 kDa).

al., 2013) y la incorporación de cobre en el sitio activo. Por otro lado, ha sido descrito que la interrupción de la agitación del cultivo luego de la adición de IPTG es crucial, ya que genera condiciones de baja concentración de oxígeno en el medio, lo cual facilita la incorporación de cobre (Duraó *et al.*, 2008) En consecuencia, con el objetivo de optimizar la expresión de esta lacasa, se estudiaron distintas condiciones. Se analizó la producción de la proteína con y sin la adición de CuSO_4 0,25 mM al medio y se compararon diversos regímenes de agitación luego de la adición de IPTG y distintos tiempos de incubación. Para la mayoría de las condiciones se observó una banda correspondiente al peso molecular de la lacasa sintética fusionada al *tag* de histidinas (52 kDa), aunque sólo pudo visualizarse en las fracciones insolubles, mientras que en la fracción soluble no se observó la banda del tamaño correspondiente. Esto indicaría que la misma podría localizarse en cuerpos de inclusión.

En la Figura VII.4 se muestra un gel desnaturalizante SDS-PAGE en donde pueden verse las proteínas totales en las distintas fracciones, cuando las células fueron crecidas en medio LB a 37°C y a diferentes tiempos de inducción, y luego de la adición de IPTG 0,4 mM y CuSO_4 0,25 mM se cesó la agitación. Los resultados hasta aquí obtenidos, indican que la proteína Lac188 es mayoritariamente acumulada en la fracción insoluble como proteína inactiva. Sin embargo, el análisis por

Western blot permitió detectar la presencia de la Lac188 en la fracción soluble cuando la inducción se realizó con IPTG 0,4 mM y CuSO_4 0,25 mM por 24 horas y a 37°C (Figura VII.5). En estas condiciones, y en la fracción soluble que corresponde a 24 horas de inducción, pudo detectarse actividad lacasa cuando se utilizó ABTS (ácido 2,2'-azino

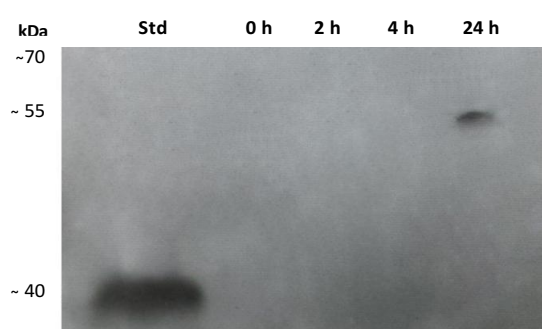


Figura VII.5. Análisis por *Western blot* de las fracciones solubles de *E. coli* BL21(DE3)Star portando la construcción pET22b(+)-*lac188* utilizando un anticuerpo conjugado anti- His-Tag y empleando revelado por quimioluminiscencia. Se indican los distintos tiempos de incubación a los cuales las células fueron colectadas. Std: proteína estándar de referencia (9,2 µg; 41,9 kDa).

bis(3-etilbenzotiazolina-6-sulfónico)) como sustrato. La actividad enzimática calculada correspondió a 3.5 UI/l de cultivo y 0.0161 UI/g de biomasa.

VII.2.2.5. Purificación preliminar de la Lac188

Para realizar la purificación de la Lac188, se partió de 1,5 l de cultivo bacteriano de la cepa de *E. coli* BL21(DE3)Star portadora de la construcción pET22b(+)-lac188. A partir de este volumen se obtuvieron 6 g de *pellet* celular y, luego de la rotura celular, la fracción soluble se pasó por una columna de cromatografía de afinidad (Materiales y Métodos, sección II.2.16.3). Todos los *buffers* fueron preparados con Na₂SO₄ en lugar de NaCl, ya que se ha demostrado que los iones cloruro inhiben la actividad lacasa (Fang *et al.*, 2012, Ausec *et al.*, 2015) y la elusión de la proteína fue llevada a cabo empleando concentraciones crecientes de imidazol, hasta 250 mM. Como era de esperarse, dada la presencia mayoritaria de la proteína en la fracción insoluble, el rendimiento fue bajo (estimado en promedio 7.1 µg/g de células, correspondiente a 30-100 µg/l de cultivo). A pesar de ello, la proteína de tamaño esperado (52 KDa) se observó en geles SDS-PAGE teñidos con azul de Coomassie, y además se detectó en ensayos de *Western blot* empleando anticuerpo monoclonal anti-His.

VII.2.2.6. Evaluación preliminar de la actividad de Lac188

Para confirmar la funcionalidad de la enzima purificada, se analizó la actividad de Lac188 utilizando algunos sustratos típicos de lacasas: ABTS, DMP (2,6-dimetoxifenol), tirosina, pirogalol, pirocatecol, ácido vanílico, siringaldazina, guayacol, L-DOPA (L-3,4-dihidroxifenilalanina), ácido ferúlico, siringaldehído y K₄Fe(CN)₆ en las concentraciones que se indican en la sección II.2.16.5 de Materiales y Métodos, en *buffer* sodio acetato pH 5,0.

Tabla VII.2. Actividad de Lac188 frente a distintos sustratos.

Sustrato	UI mg ⁻¹
DMP	10,4
ABTS	1.25
Pirogalol	3.2
Pirocatecol	0.85
Ácido vanílico	0.87

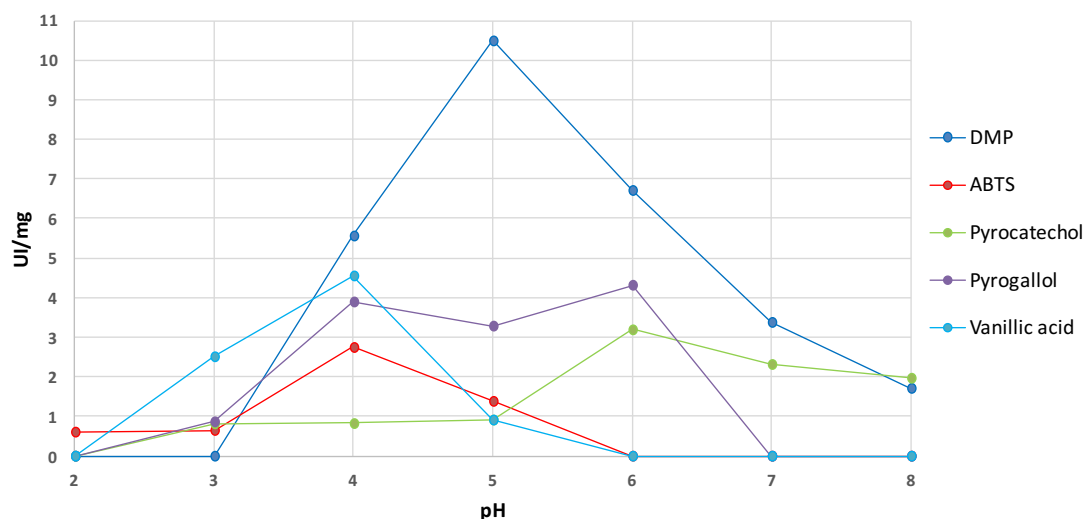


Figura VII.6. Actividad enzimática de la Lac188 a diversos pHs empleando distintos sustratos. La actividad de Lac188 fue investigada a 25 °C para los 5 sustratos distintos especificados.

De este *screening* preliminar, se pudo determinar que la enzima fue activa sobre 5 de los 12 sustratos ensayados, que son los que se muestran en la tabla VII.2. La unidad enzimática (UI) fue definida como la cantidad de Lac188 requerida para oxidar 1 μmol de sustrato por minuto en condiciones saturantes. Como puede verse, el DMP es el sustrato sobre el cual la lacasa Lac188 presentó una mayor actividad enzimática.

Dado que el pH óptimo de las lacasas depende de los sustratos utilizados, se procedió a evaluar la actividad de la Lac188 a distintos valores de pH empleando los 5 sustratos en los que previamente la enzima presentó actividad catalítica (Figura VII.6). Como puede verse, la actividad específica máxima se observó para el sustrato DMP a pH 5,0, siendo ésta de 10.4 U/mg. Para los sustratos restantes, los pHs de mayor actividad fueron de 4 para el ácido vanílico y ABTS, 6 para el pirocatecol, y un rango de 4-6 para el pirogallol. Al igual que sucede con la mayoría de las lacasas bacterianas estudiadas, el pH óptimo de actividad es ácido.

VII.3. Discusión

En este capítulo hemos presentado la búsqueda *in silico* de genes con potencial interés biotecnológico, detallando la selección, clonado, puesta a punto de las

condiciones óptimas para su expresión, purificación y estudio de los parámetros cinéticos de una enzima con potencial interés industrial o ambiental. Comenzamos con un *screening* bioinformático de las secuencias obtenidas a partir del ADN de los plásmidos presentes en los aislamientos de la colección bacteriana en estudio, el cual nos permitió identificar una posible lacasa (*lac188*) cuya secuencia resultó completa.

Las lacasas más utilizadas industrialmente son de origen fúngico; sin embargo, esto no se debe a la ineficiencia de las lacasas bacterianas, sino a su más reciente descubrimiento y exploración. Asimismo, a pesar de los cientos de genes de lacasas que podemos encontrar al realizar una búsqueda en bancos de datos, muchos de ellos se corresponden con secuencias parciales o genes putativos de lacasas, por lo que el número desciende considerablemente cuando se trata de lacasas cuyos productos proteicos han sido caracterizados de manera experimental. Un punto importante para remarcar aquí es la localización plasmídica del gen que codifica para la lacasa Lac188. Si bien en un trabajo reciente se ha realizado un análisis minucioso de lacasas bacterianas, demostrándose una amplia distribución de estas enzimas en distintos genomas y su gran abundancia en replicones plasmídicos (Ausec, *et al.*, 2011), pocas lacasas de origen plasmídico han sido caracterizadas en cuanto a su actividad frente a distintos sustratos (Ihssen *et al.*, 2015). En los últimos años algunas lacasas bacterianas han sido estudiadas en cuanto a su actividad, pero no se ha identificado el o los genes que las codifican (Phugare *et al.*, 2011) o bien provienen de *screenings* de actividad realizados sobre bibliotecas metagenómicas (Ye *et al.*, 2010) o *screenings* de secuencias a partir de bibliotecas metagenómicas (Fang, *et al.*, 2012) o búsquedas *in silico* seguidas de un posterior clonado y medida de actividad (Ausec *et al.*, 2015; Ihssen *et al.*, 2015). Por lo tanto su localización permanece desconocida, por lo que no podemos descartar que algunos de estos genes tengan una localización plasmídica.

La síntesis y el clonado de la lacasa *lac188* en el vector de expresión pET22b(+) nos permitió avanzar en el estudio de las condiciones necesarias para la expresión de este gen. La expresión de la proteína fue detectada en ensayos donde la inducción se realizó en presencia de Cu^{2+} . Este requisito es similar al previamente descrito para la

expresión de los genes de lacasa de basidiomicetos que es regulada por la presencia de cobre (Cu^{2+}) en el medio de cultivo, así como en el caso de *T. versicolor* (Collins & Dobson, 1997), *T. pubescens* (Galhaup & Haltrich, 2001), diferentes lacasas de *P. sajorajyu* (Soden & Dobson, 2001) o lacasas producidas por *P. ostreatus* (Faraco *et al*, 2003).

Es frecuente que la sobreexpresión de proteínas heterólogas en *E. coli* dé como resultado la formación de cuerpos de inclusión, es decir, agregados citoplasmáticos insolubles de dicha proteína (Gillings, 2014, Gillings, *et al.*, 2015). Esto suele ocurrir cuando las bacterias son forzadas a producir una proteína en concentraciones elevadas, y se traduce en una disminución o ausencia de la proteína en cuestión en la fracción soluble. Frente a esta situación, pueden modificarse las condiciones de inducción y/o purificación para obtener un mayor rendimiento, aunque esto no dé siempre buenos resultados. Si bien la proporción de proteína Lac188 en la fracción soluble fue pequeña, fue suficiente para avanzar en la purificación y determinar su actividad frente a 5 sustratos distintos de los 12 ensayados y analizar la actividad enzimática a diversos valores de pH. Es importante mencionar que la proteína no fue purificada hasta homogeneidad. En consecuencia, serán necesarios más ensayos y/u otras técnicas adicionales para lograr un grado de purificación adecuado de la enzima funcionalmente activa para avanzar en posteriores análisis de caracterización *in vitro*.

Finalmente, es interesante recalcar aquí lo poderoso que pueden resultar los análisis *in silico* realizados a partir de datos de secuencias genómicas de distinto origen, ya que permiten detectar la presencia de genes cuyos productos enzimáticos pueden presentar una actividad de interés. Si bien existen y son aún utilizadas una gran cantidad de técnicas basadas en *screening* de actividades *in vitro*, el secuenciamiento masivo de genomas u otras moléculas como plásmidos y virus, así como la existencia y disponibilidad de herramientas computacionales, han llevado al descubrimiento y posterior caracterización de una gran cantidad de enzimas que hoy son utilizadas eficientemente en distintos procesos industriales. Otros estudios similares al realizado aquí han abordado búsquedas *in silico* que les han permitido identificar enzimas del tipo lacasa en distintos genomas bacterianos y clonarlas en

vectores de expresión para medir actividad sobre distintos sustratos, encontrando resultados positivos (Ihssen, *et al.*, 2015, Ausec, *et al.*, 2015). Estudios *in silico* más sofisticados han sido realizados con el objetivo de identificar, clasificar e incluso predecir posibles sustratos de interés tanto industrial como ambiental (Suresh, *et al.*, 2008). Todos estos estudios, junto a los resultados obtenidos en este trabajo, ponen en evidencia la eficiencia y el poder que presentan los métodos basados en *screenings* metagenómicos y el uso de herramientas bioinformáticas en el descubrimiento de nuevas enzimas de posible interés comercial o ambiental para el desarrollo de productos.

CAPÍTULO VIII

**DISEÑO DE UNA ESTRATEGIA Y CONSTRUCCIÓN DE
LAS HERRAMIENTAS MOLECULARES NECESARIAS
PARA LA CAPTURA DE PLÁSMIDOS**

VIII.1. Introducción

La obtención y caracterización de plásmidos de comunidades microbianas así como de aislamientos bacterianos específicos, puede resultar muchas veces problemática, tediosa o imposible. Si bien existen una gran cantidad de métodos que han sido utilizados eficientemente en la obtención y caracterización de plásmidos (Capítulo I, sección I.2.2.1.3), todos presentan ciertas desventajas o complicaciones técnicas, derivando en limitaciones propias de cada método. Una de estas limitantes, y quizá la más relevante al estudiar plásmidos ambientales, reside en el tamaño de los replicones que pueden ser obtenidos. Como ya hemos mencionado, los plásmidos pueden tener desde unas pocas hasta cientos de miles de pares de bases. Sin embargo, son pocas las metodologías que permiten la captura de plásmidos de alto peso molecular; por ello a la hora de estudiar el plasmidoma de un determinado ambiente existen marcadas limitaciones. Por ejemplo, las técnicas de TRACA y aislamiento endógeno están acotadas a replicones pequeños, ya que estas metodologías se ven limitadas por el paso de transformación bacteriana del ADN. La misma limitación de tamaño existe para la técnica de amplificación de plásmidos por MDA, en la cual los plásmidos de gran tamaño suelen romperse durante el paso de extracción de ADN plasmídico y, en consecuencia, son removidos por exonucleasas empleadas en etapas posteriores. La técnica de aislamiento exógeno permite la captura de plásmidos de alto peso molecular, pero sólo si éstos son movilizables o conjugativos, poseen un marcador de selección particular y un origen de replicación funcional en la cepa receptora, lo cual restringe drásticamente el rango de plásmidos que pueden ser estudiados mediante estos métodos.

Además del tamaño plasmídico existen otros factores que determinan el éxito de las metodologías empleadas. Por ejemplo, todos los métodos que requieren de una etapa de ruptura celular para la extracción del material genético extracromosomal (esto es, todos los conocidos excepto la técnica de aislamiento exógeno), pueden resultar deficientes en ciertos tipos de bacterias que son más difíciles de lisar o cuyas preparaciones plasmídicas son difíciles o imposibles de obtener. Cuando se trata de

una extracción plasmídica masiva a partir de una comunidad microbiana, este problema se ve reflejado en una escasa o nula recuperación de algunos replicones. En este sentido, aquellos elementos que se encuentran en bacterias difíciles de lisar, o en microorganismos que están en baja proporción en la muestra, o que se encuentran en bajo número de copias por célula, generalmente no estarán presentes en el conjunto de plásmidos aislados.

Si bien las metodologías existentes junto a los avances en el secuenciamiento de genomas bacterianos han permitido caracterizar una gran cantidad de plásmidos, las limitaciones antes mencionadas en muchas ocasiones dificultan, e incluso imposibilitan, la obtención de material plasmídico. En vista de estas limitaciones, y con el objetivo de aumentar el número y variedad de replicones que puedan ser estudiados, en este capítulo de la Tesis nos enfocamos en el diseño y desarrollo de una estrategia que pueda ser utilizada para obtener plásmidos diversos, independiente de la movilidad de los mismos, del rango de huésped, del tamaño y de que transporten marcadores de selección específicos.

VIII.2. Diseño de una estrategia para la captura de plásmidos: etapas y fundamentos

La estrategia que generamos posee dos etapas consecutivas. En una primera etapa se realiza la marcación de el/los plásmidos que se pretenden recuperar (independiente del cultivo de la cepa portadora) a los que llamamos plásmidos *target*. Con este fin utilizamos un transposón modificado que permitirá en una segunda etapa su movilización hacia y selección en una cepa de *E. coli* (hospedadora) u otra cepa determinada, en la cual el plásmido pueda ser fácilmente purificado para luego caracterizarlo y/o secuenciarlo.

Para avanzar en la estrategia delineada necesitamos construir un transposón que contenga: **i)** un origen de transferencia plasmídica (*oriT*), que permita su movilidad hacia la cepa receptora marcada; **ii)** un origen de replicación plasmídica (*oriV*), que permita su replicación en la cepa receptora; y **iii)** un marcador de selección, que

permita identificar o seleccionar la cepa receptora. La inserción de este transposón en cualquier posición del plásmido *target* le permitirá al plásmido modificado (independientemente de su composición génica) moverse desde la bacteria hospedadora natural hacia otra bacteria hospedadora conocida y marcada en la que podrá replicar. La cepa receptora del plásmido modificado (*target*+transposón) podrá ser identificada gracias a los marcadores aportados por el transposón.

VIII.2.1. Diseño y construcción del vector p-RESC

El vector p-RESC (plásmido de rescate, Figura VIII.1) fue generado a partir del esqueleto del vector pBAM1 (Martínez-García, *et al.*, 2011) sobre el cual se realizaron sucesivas modificaciones, como se muestra en la Figura VIII.2. El pBAM1 posee un minitransposón que contiene un gen de resistencia a kanamicina, flanqueado por dos secuencias de inserción. Por fuera del transposón se

ubica el gen que codifica para la transposasa (*tnpA*, rediseñada para dotarla de hiperactividad (Goryshin & Reznikoff, 1998)) responsable de la transposición del transposón, un gen de resistencia a ampicilina (Ap^R), un origen de transferencia plasmídica (*oriT*) y un origen de replicación (*oriV*-R6K), el cual permite la replicación plasmídica en un grupo acotado de bacterias. Para avanzar en la construcción del p-RESC, el *oriV*-R6K y *oriT* que se encuentran por fuera del transposón fueron removidos, y se adicionaron el *oriV* (del pSM10) y el *oriT* (del pUC19mob) dentro del transposón. Además, en el esqueleto del vector y por fuera del transposón se insertó un sitio de reconocimiento para la

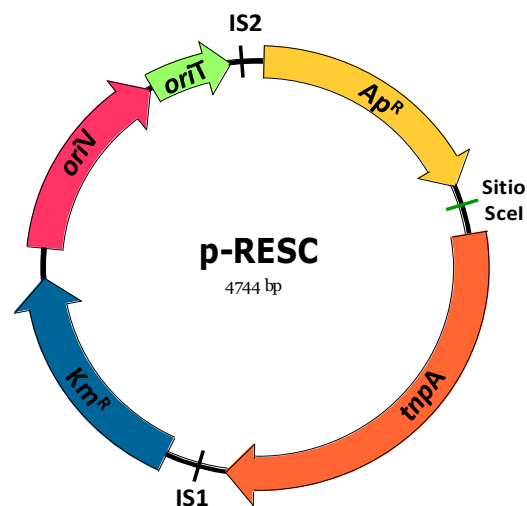


Figura VIII.1. Esquema del vector p-RESC. Km^R : gen de resistencia a kanamicina; Ap^R : gen de resistencia a ampicilina; *oriV*: origen de replicación; *oriT*: origen de transferencia; *tnpA*: gen que codifica para la transposasa; IS1 y IS2: secuencias de inserción.

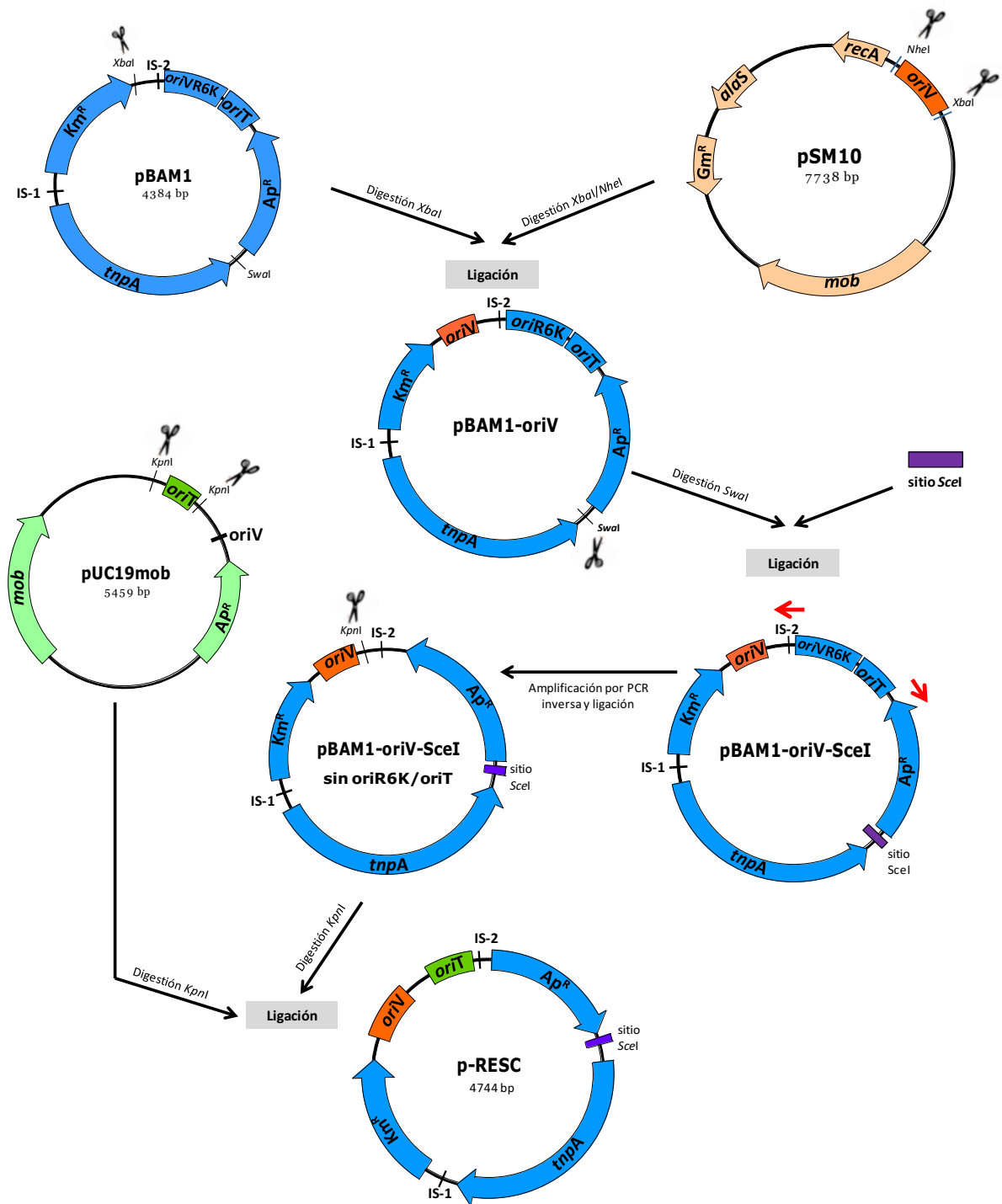


Figura VIII.2. Esquema general de los pasos realizados para la construcción del vector p-RESC. Se detallan los sitios de corte utilizados en las digestiones, la ubicación de *primers* de PCR (flechas rojas) y los elementos más relevantes de los vectores utilizados.

endonucleasa doméstica (*homing endonuclease*) I-SceI (que reconoce una secuencia de 18 pb) que servirá en etapas posteriores ya que la probabilidad de encontrar un sitio de corte I-SceI en un genoma es de, aproximadamente, una vez cada 20 veces el tamaño del genoma humano. En consecuencia sitios de reconocimiento para SceI usualmente no están presentes en genomas procariotas.

Los clonados sucesivos realizados para llegar a la obtención de la construcción p-RESC se detallan en Materiales y Métodos, sección II.2.14.

Finalizada la construcción del p-RESC, el mismo fue secuenciado (utilizando los *primers* que se muestran en la Tabla II.4 de Materiales y Métodos) para chequear que no se hayan generado errores en la secuencia de ADN tras los sucesivos pasos de clonado y amplificación por PCR.

VIII.2.2.Estrategia delineada para la captura de plásmidos

El procedimiento metodológico consiste en tres etapas consecutivas que comprenden dos eventos de conjugación correlativos que se detallan a continuación y se esquematizan en la Figura VIII.3.

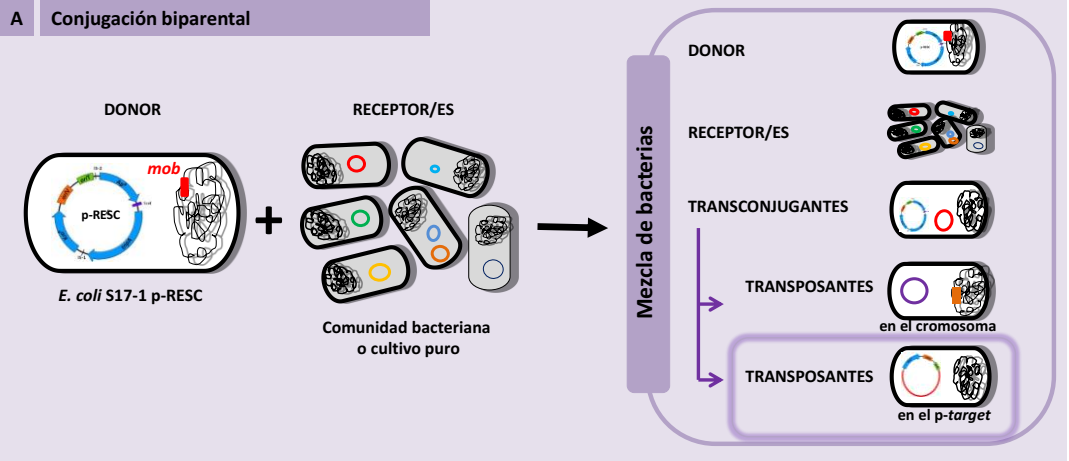
1. La primera etapa, corresponde a una conjugación biparental que tiene como objetivo introducir en la bacteria portadora del plásmido *target* el vector p-RESC que contiene el transposón con los elementos previamente mencionados. Para ello, se cuenta con la *E. coli* S17-1 p-RESC. Esta cepa contiene integrado en su cromosoma los genes necesarios para la formación del *pilus*, por lo que el p-RESC. Gracias al *oriT* que posee el p-RESC, puede moverse vía conjugación hacia la/las bacterias que contienen el/los plásmidos *target*. Esta conjugación puede ser realizada tanto entre *E. coli* S17-1 p-RESC y un cultivo puro de otro aislamiento bacteriano conteniendo uno o más plásmidos, o bien con las bacterias presentes en una comunidad bacteriana en conjunto (con o sin previo cultivo), de la cual podrían aislarse uno o más plásmidos. Una vez ocurrida la conjugación, el transposón del p-RESC (que contiene los elementos necesarios para la transposición) puede insertarse al azar en cualquier parte del

genoma de la bacteria huésped, incluso en plásmidos, confiriéndoles el *oriT*, el *oriV* y el gen de resistencia a kanamicina (Figura VIII.3.A).

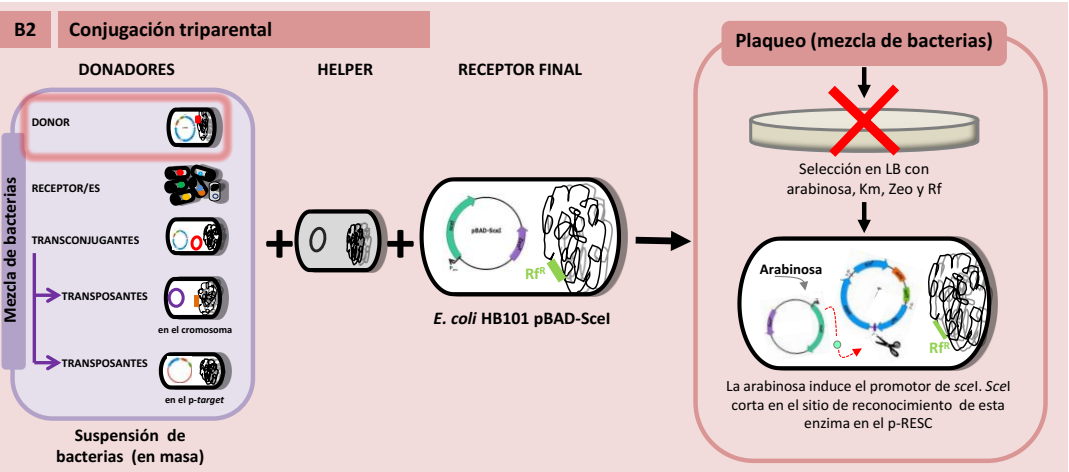
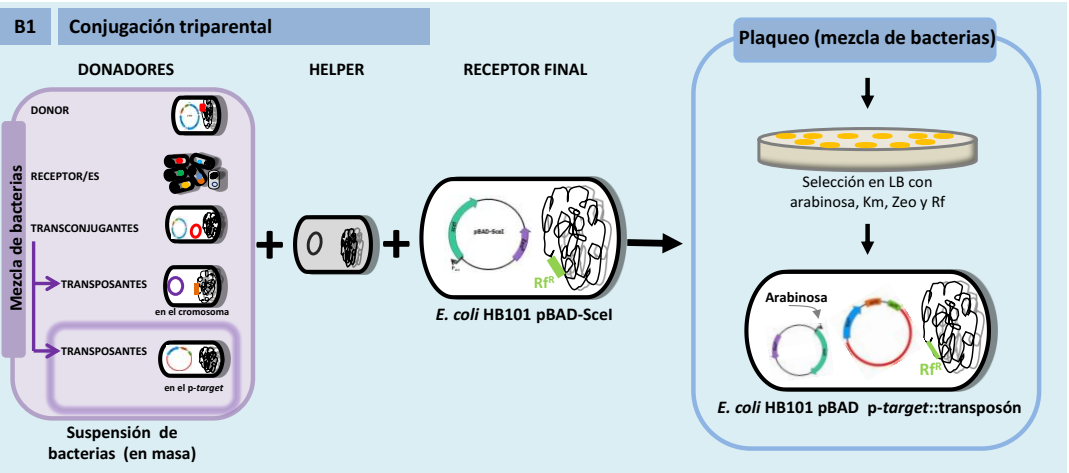
2. La segunda etapa, corresponde a una segunda conjugación en presencia de un plásmido *helper*, que contiene los genes necesarios para movilizar al plásmido *target* que adquirió el transposón hacia una bacteria receptora final (*E. coli* HB101 pBAD-Scel, resistente a rifampicina y zeocina), a la cual le hemos incorporado el vector pBAD-Scel, que contiene el gen que codifica para la megaendonucleasa *Scel* bajo un promotor finamente regulado e inducible por arabinosa (P_{ara}) y un gen de resistencia a zeocina. La selección de los transconjugantes obtenidos luego de esta segunda conjugación se lleva a cabo en medio LB conteniendo rifampicina, kanamicina y arabinosa (Figura VIII.3.B1). La rifampicina permite el crecimiento selectivo de la *E. coli* HB101; la kanamicina permite seleccionar bacterias portadoras de plásmidos que hayan adquirido el transposón, y la arabinosa activa la expresión del gen la endonucleasa *Scel*, enzima que cliva al plásmido p-RESC en caso de que éste sea recibido por la cepa receptora final *E. coli* HB101 pBAD-Scel (Figura VIII.3.B2). La adición de arabinosa en esta etapa es clave para descartar la presencia del p-RESC en la cepa receptora final y evitar la aparición de resultados falsos positivos (o fondo), ya que el vector p-RESC posee resistencia a kanamicina y es capaz de replicar en esta cepa.

Figura VIII.3. Esquema general del procedimiento del sistema utilizado para la captura de plásmidos. En una primera etapa (A) se transfiere el vector p-RESC (que contiene el transposón) a las bacterias que portan los plásmidos que quieren capturarse (*p-target*) por conjugación biparental. En una segunda etapa, mediante una conjugación triparental se busca movilizar los plásmidos (marcados) hacia una bacteria receptora final (*E.coli* que contiene la construcción pBAD-Scel) (B1). Los transconjugantes se seleccionan en medio conteniendo arabinosa a fin de evitar recuperar colonias (Km^R , Amp^R , Zeo^R , Rf^R) en las que el vector p-RESC pasó directamente a la cepa receptora final (B2). La tercera etapa (C) corresponde al *screening*. Una vez realizada la selección bacteriana en medio LB conteniendo kanamicina, rifampicina y arabinosa, las colonias obtenidas deberán ser chequeadas para corroborar que efectivamente posean la construcción buscada (plásmido *target* + transposón) en la cepa receptora final.

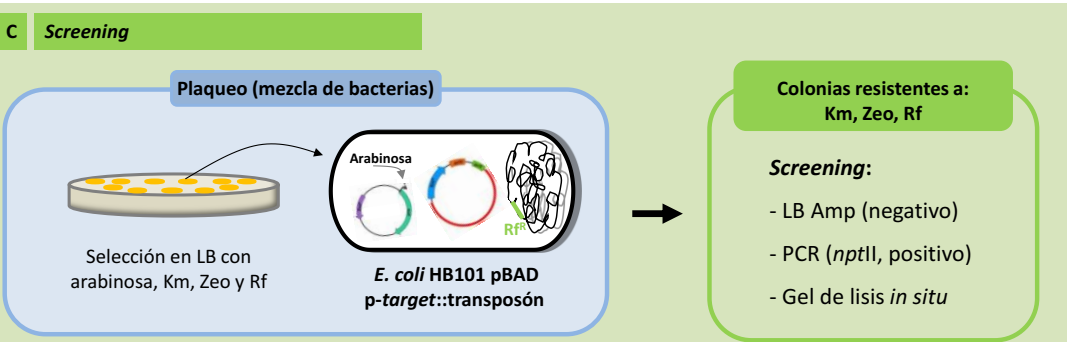
Primera etapa



Segunda etapa



Tercera etapa



De esta manera, la estrategia desarrollada nos permitirá tener en una cepa de clonado, de crecimiento rápido y fácil manipulación, uno o más plásmidos que luego podrán ser fácilmente purificados, secuenciados y caracterizados.

3- La tercera etapa corresponde al *screening*. Una vez realizada la selección bacteriana en medio LB conteniendo kanamicina, rifampicina y arabinosa, las colonias obtenidas deberán ser chequeadas para corroborar que efectivamente posean la construcción buscada (plásmido *target* + transposón) en la cepa receptora final (Figura VIII.3.C). Los siguientes controles ayudarán para avanzar en el análisis de las colonias obtenidas:

i) Analizar el crecimiento de las colonias obtenidas en medio conteniendo ampicilina; aquellos clones que crecieran en este medio serán portadores del vector p-RESC, o podrán contener un plásmido *target* capturado que posea un gen que confiere resistencia a este antibiótico o bien podrán ser resistentes espontáneas.

ii) Analizar el contenido plasmídico de las colonias obtenidas mediante la técnica de lisis *in situ* de Eckhardt (Materiales y Métodos, II.2.4.3). Esta técnica permitirá distinguir los clones que posean el plásmido p-RESC de aquéllos que contengan el plásmido *target* con el transposón insertado, e incluso hacer una estimación de su tamaño.

VIII.3. Evaluación de la funcionalidad de las herramientas que emplearemos en la estrategia delineada

Hasta aquí hemos descrito el diseño, construcción y mapa físico del plásmido que contiene al nuevo transposón. El uso de la nueva herramienta en estudios de captura de plásmidos requerirá entonces verificar que, luego de la introducción del *oriV* y el *oriT*, el nuevo minitransposón mantiene sus propiedades de transposición.

VIII.3.1. Evaluación de la frecuencia de transposición del transposón diseñado

Estudios previos determinaron la frecuencia de transposición del transposón

original, siendo ésta del orden de 10^{-3} (transposiciones / célula recipiente) cuando se empleó la cepa de *E. coli* HB101 como bacteria receptora (Lozano, *et al.*, 2013). Mediante ensayos de transposición realizados en el laboratorio con la versión modificada del transposón, se corroboró que la frecuencia de transposición del mismo resultó coincidente con la previamente descrita, por lo que las modificaciones realizadas sobre el transposón no han modificado su funcionalidad.

VIII.3.2. Evaluación de la actividad de *Scel*: control de la escisión del p-RESC como consecuencia de la expresión de *Scel* inducida por la presencia de arabinosa

En el vector pBAD-*Scel* el promotor P_{ara} se induce en presencia de L-arabinosa 0,5% p/v, lo que permite activar la expresión génica de la endonucleasa *Scel*. En la estrategia diseñada la expresión de la proteína *Scel* en la segunda etapa es necesaria para evitar la presencia de la construcción p-RESC en la cepa receptora de *E. coli* HB101 pBAD-*Scel*. Para chequear la funcionalidad del sistema de inducción por arabinosa se diseñó un ensayo que se explica a continuación y se esquematiza en la Figura VIII.4. Se realizó una conjugación biparental entre *E. coli* S17-1 conteniendo el p-RESC y *E. coli* HB101 conteniendo el pBAD-*Scel*, y la suspensión bacteriana fue colocada en una caja de Petri con medio LB sólido conteniendo o no arabinosa. Para evaluar si la preinducción del gen *scel* mejora la eficiencia de corte del p-RESC, en uno de los casos se preincubó HB101 pBAD-*Scel* en medio con arabinosa, para encender el sistema de inducción del promotor P_{ara} antes del ingreso del p-RESC a la célula receptora final. Luego, las células se resuspendieron en medio líquido para homogeneizar el cultivo conteniendo rifampicina y zeocina para favorecer el crecimiento de la cepa receptora, y conteniendo o no arabinosa (como se muestra en la Figura VII.4). Finalmente, se realizaron diluciones seriadas de estos cultivos y recuentos (Materiales y Métodos, sección II.1.12) en medio de similar composición a la del cultivo líquido anterior, conteniendo además ampicilina, para seleccionar aquellos transconjugantes que hayan adquirido el vector p-RESC. Los resultados obtenidos fueron los siguientes: en la placa control (Figura VIII.4.A) se observó crecimiento hasta la dilución 10^{-6} inclusive, mientras que cuando el sistema se indujo con arabinosa se

observó una reducción del crecimiento en tres órdenes de magnitud (10^{-3}) cuando no se preincubó la *E. coli* HB101 pBAD-Scel con arabinosa (Figura VIII.4.B) y de 4 órdenes de magnitud cuando sí hubo preincubación con arabinosa (Figura VIII.4.C), viéndose crecimiento en este caso hasta la dilución 10^{-2} . Estos resultados indican que existe un fuerte control en la regulación del promotor P_{ara} . Si bien es deseable que no haya crecimiento alguno en presencia de arabinosa, el resultado encontrado podría atribuirse a que el número de copias del vector p-RESC por célula es alto, y en consecuencia su eliminación por acción de la endonucleasa *Scel* no es completa. Hemos repetido el ensayo con diversas concentraciones de arabinosa, pero no se han obtenido mejores resultados independientemente de las condiciones de inducción diversas del promotor P_{ara} analizadas.

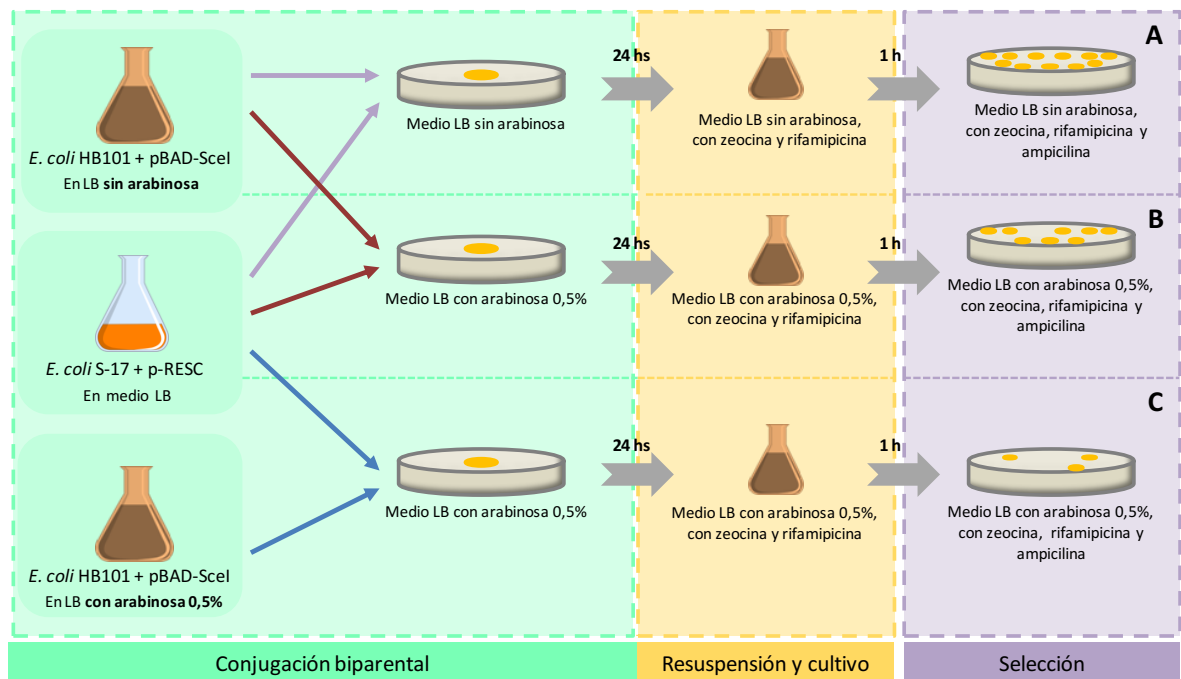


Figura VIII.4. Esquema de los ensayos realizados a fin de evaluar la inducción por arabinosa y la expresión de *Scel*. **A)** control: conjugación entre S17-1 p-RESC y HB101 pBAD-Scel sin inducción por arabinosa. En la placa se observa el crecimiento de todos los transcojugantes (*E. coli* HB101 pBAD-Scel que adquirieron el p-RESC). **B)** Inducción con arabinosa sin preinducción del cultivo de *E. coli* HB101 pBAD-Scel. **C)** Inducción con arabinosa con preinducción de la endonucleasa en el cultivo de *E. coli* HB101 pBAD-Scel. Los recuentos se realizaron por el método de la gota (el número de colonias en el esquema es sólo ilustrativo).

VIII.4. Validación del sistema de captura de plásmidos mediada por el vector p-RESC

En un último ensayo se procedió a desarrollar el protocolo completo planteado para la captura de plásmidos. Si bien el sistema fue diseñado con el objetivo de aislar plásmidos de comunidades bacterianas y de aislamientos bacterianos que porten plásmidos desconocidos, en una primera instancia se utilizaron plásmidos de tamaño y características conocidas (poseen un marcador de selección positivo), a fin de simplificar el *screening* y la validación del sistema de captura diseñado.

VIII.4.1. Elección de plásmidos *target* para la evaluación del sistema

A fin de evaluar el funcionamiento de la metodología diseñada, se escogieron dos plásmidos *target* de características diversas: **i)** pLT06, el cual se encuentra en una cepa de *E. coli* y tiene un tamaño de 8 Kb, disponemos de su secuencia, posee un gen de resistencia a cloranfenicol (Cm^{R}), un origen de replicación termosensible (que le permite replicar a 28 °C o temperaturas inferiores pero no puede hacerlo a 37 °C) y no contiene *oriT*; **ii)** pAci39, el cual se encuentra en un aislamiento clínico de *Enterobacter* sp. que ha sido caracterizado en nuestro laboratorio (Salto, Tesis en redacción). El plásmido pAci39 posee un tamaño estimado de 40 Kb y confiere resistencia a plata (Ag^{R}), aunque se desconoce su mapa físico y secuencia. La elección del plásmido pLT06 se debe, pese a su tamaño pequeño, a la imposibilidad de moverse por sí mismo y de replicar a una cierta temperatura, por lo cual sería esperable que, si luego de realizar el protocolo experimental diseñado se evidencia su presencia en la cepa huésped HB101 pBAD-Scel, ésta sería consecuencia de la incorporación del transposón. En cuanto al plásmido pAci39, su elección se debe principalmente a su mayor tamaño, a la existencia de un gen marcador (Ag^{R}) y a su imposibilidad de movilizarlo en ensayos previos realizados en el laboratorio, por lo cual su movilidad se atribuiría a la presencia del transposón aportado por el p-RESC.

VIII.4.2. Prueba de la estrategia diseñada

Para realizar la prueba del sistema de captura plasmídica diseñado se procedió

como se indica a continuación, empleando como plásmidos *target* pLT06 (contenido en *E. coli* Top10) o el pAci39 (contenido en *Enterobacter* sp.).

1. En una primera etapa se realizaron dos conjugaciones biparentales independientes: 1) entre *E. coli* S17-1 p-RESC y *Enterobacter* sp. pAci39, 2) entre *E. coli* S17-1 p-RESC y *E. coli* Top10 pLT06. Las suspensiones bacterianas de cada conjugación (y los controles de ambos cultivos) se incubaron en medio LB a 28 °C durante 24 horas.

2. Luego se realizó la segunda conjugación (triparental) entre la suspensión bacteriana resultante de cada primer conjugación con la cepa *E. coli* HB101 pBAD-Scel y el plásmido pRK2013 como *helper*. Las conjugaciones (y los controles correspondientes a cada cultivo), se incubaron en medio LB con arabinosa a 28 °C durante 24 horas.

3. Finalmente, la mezcla de bacterias se suspendieron en medio LB suplementado con arabinosa, rifampicina y kanamicina, en donde se incubaron durante 6 horas. De estos cultivos se realizaron diluciones seriadas y recuentos en placas de LB suplementado con rifampicina, kanamicina, arabinosa, zeocina, más cloranfenicol o AgNO₃ (dependiendo de si el *target* utilizado fue el plásmido pLT06 o el pAci39, respectivamente) y conteniendo (o no) ampicilina.

VIII.4.3. Resultados obtenidos

De acuerdo con la metodología utilizada, debería esperarse que aquellas colonias que crecen en medio conteniendo kanamicina, rifampicina y cloranfenicol o AgNO₃ (dependiendo del plásmido empleado) contengan el plásmido *target* con el transposón insertado en alguna posición de su esqueleto. Por el contrario, en aquellas placas que fueron suplementadas con ampicilina, no debería observarse crecimiento, ya que el gen que confiere esta resistencia está presente en el p-RESC. Sin embargo, al realizar el *screening* mencionado (para los dos plásmidos ensayados) observamos en las placas con y sin ampicilina, crecimiento en el plaqueo directo y en la dilución 10⁻¹. Los resultados de este procedimiento fueron similares en 3 ensayos independientes realizados. Estos resultados indican que todas las colonias obtenidas contendrían aún

el vector p-RESC. Esta hipótesis se corroboró realizando diversas amplificaciones por PCR en las que se utilizaron *primers* que amplifican tanto fragmentos internos del transposón como regiones del esqueleto del plásmido p-RESC. Los resultados obtenidos evidencian que alguna de las etapas de la estrategia desarrollada no ha funcionado correctamente.

Dado que previamente se chequeó la inducción de la expresión de la endonucleasa *SceI* y la frecuencia de transposición del transposón construido resultó del mismo orden en *E. coli* a la descrita por Lozano y colaboradores (2013) podemos suponer que la etapa crítica de esta estrategia corresponde al paso de transposición. Posiblemente, la limitación no esté dada por el evento de transposición en sí, sino por la probabilidad (baja) de que el transposón se inserte en el plásmido *target*, en vez de hacerlo en el cromosoma bacteriano. Si se considera que el cromosoma de *E. coli* posee $\sim 4 \times 10^6$ pb de longitud y los plásmidos *target* ensayados tienen ~ 40.000 pb (pAci39) y ~ 8.000 pb (pLT06), la probabilidad que al azar el transposón se inserte en el plásmido en lugar de hacerlo en el cromosoma es de $1/100$ y $1/500$, respectivamente. Por lo tanto al considerar en forma conjunta la frecuencia de cada conjugación, la frecuencia del evento de transposición y la probabilidad de que el transposón se incorpore al azar en el plásmido, podemos entender nuestra imposibilidad de obtener colonias conteniendo los plásmidos *target* con el transposón insertado bajo las condiciones experimentales ensayadas.

VIII.5. Discusión

En este capítulo hemos presentado el diseño de una nueva herramienta que tiene como finalidad capturar plásmidos de distintos ambientes o de aislamientos bacterianos y tenerlos en una cepa huésped cultivable de la cual además sea fácil obtener ADN plasmídico en buena cantidad y calidad para abordar su secuenciamiento y/o su manipulación con herramientas moleculares. Si bien esta herramienta fue diseñada con el objetivo de capturar plásmidos de comunidades microbianas, en una

primera etapa hemos intentado validar la estrategia diseñada utilizando cultivos bacterianos puros a fin de facilitar la puesta a punto del experimento.

Para discutir los resultados obtenidos con esta estrategia, es importante analizar las limitaciones que la misma presenta. Los resultados hasta aquí presentados indican algunas falencias en el sistema, centradas principalmente en la etapa de transposición, y generadas principalmente por la diferencia de tamaños entre los plásmidos y el cromosoma. Sin embargo, como se mencionó anteriormente, esta estrategia fue diseñada con el objetivo de capturar principalmente plásmidos de alto peso molecular, que resultan difíciles de obtener con otros métodos. Si se tiene en cuenta que los plásmidos ambientales pueden tener hasta cientos de miles de pares de bases, la probabilidad de que el transposón se inserte en estos replicones en lugar de hacerlo en el cromosoma es mucho mayor que en los casos ensayados. Además, a medida que aumente el número de copias del plásmido *target* por célula, esta probabilidad se incrementaría. Si bien en este trabajo hemos presentado resultados obtenidos con plásmidos *target* relativamente pequeños, a futuro planeamos utilizar plásmidos de mayor tamaño para evaluar la herramienta diseñada.

Asimismo, si consideramos emplear esta herramienta en comunidades microbianas enteras, debemos tener en cuenta otras consideraciones referidas a la presencia de distintos tipos de bacterias. Los transposones poseen distintas eficiencias de transposición de acuerdo a la bacteria en la que se encuentran. De hecho, para el transposón de partida se ha observado que la frecuencia de transposición en otras especies bacterianas como *Sinorhizobium* y *Burkholderia* es mucho menor que en *E. coli* (Lozano *et al.*, 2013). Sumado a la frecuencia de la transposición, debe considerarse que la expresión del gen de la transposasa requiere del reconocimiento del promotor de dicho gen, por lo cual depende de la maquinaria de la bacteria hospedadora. También debe tenerse en cuenta que la conjugación bacteriana, si bien es muy promiscua, puede no ocurrir entre ciertos grupos de bacterias. Esto implica que ciertos plásmidos presentes en algunas especies bacterianas no podrían ser capturados.

Sin embargo, y a pesar de las limitaciones mencionadas que implican una reducción del número de plásmidos que potencialmente pueden ser capturados mediante el uso de esta estrategia, debe tenerse en cuenta que nuestro objetivo no es obtener un mayor número de plásmidos en relación al que puede obtenerse con otras herramientas, sino mejorar la posibilidad de capturar plásmidos de alto peso molecular que no poseen marcadores de selección conocidos y que con otras metodologías, clásicas o modernas, no han podido ser capturados. Por último, es importante remarcar que contamos en nuestro laboratorio con el vector p-RESC, cuya estructura ha sido verificada por análisis de secuencia, el cual será sujeto a futuras modificaciones dirigidas a mejorar la eficiencia del sistema.

CAPÍTULO IX

DISCUSIÓN GENERAL Y CONCLUSIONES

IX.1 discusión general

Para avanzar en la comprensión de la ecología evolutiva general de las bacterias, es importante estudiar la movilidad de los MGEs y genes plasmídicos (De la Cruz & Davies, 2000). Una de las estrategias de adaptación y evolución más aceptadas para el mundo procarionta es la transferencia horizontal de material genético. Por ejemplo, la evolución de la resistencia a antibióticos ha demostrado que, bajo una fuerte presión selectiva, ésta puede observarse en tiempos que son relativamente cortos (Mazel & Davies, 1999). La secuenciación completa de genomas bacterianos ha demostrado claramente que existe una gran proporción de genes bacterianos adquiridos por HGT (More, *et al.*, 1996). En muchas especies se ha encontrado que existe una gran proporción de genes transferidos horizontalmente cuyas secuencias a menudo pueden atribuirse a restos de MGEs como plásmidos, fagos o transposones (Ochman, *et al.*, 2000). Por ejemplo en *H. pylori* se ha demostrado que existen cambios considerables en la arquitectura del genoma que han ocurrido de forma natural con la participación de elementos repetidos (Romero, *et al.*, 1999).

Es escasa la información disponible respecto de la incidencia de movilización de plásmidos dentro de una población/especie de bacterias, así como la información respecto de la diversidad de los sistemas conjugativos. La frecuencia con la que existen los sistemas conjugativos en una determinada especie de bacterias, sus características funcionales (frecuencias, compatibilidades para la movilización cruzada) y su caracterización molecular, representan instancias obligadas para avanzar con sustento en el conocimiento del modo en que la transferencia horizontal conjugativa participa en: **i)** los fenómenos de adaptación de las bacterias a sus diferentes nichos, **ii)** la evolución genómica a mediano y largo plazo, y **iii)** el eventual impacto de inocular cepas con fines agrícolas o biorremediadores que puedan contener plásmidos.

En nuestro laboratorio estamos abocados a la búsqueda sistemática y caracterización molecular de la información contenida en plásmidos, particularmente en plásmidos transmisibles involucrados en la HGT por conjugación. En este sentido, y en este trabajo de Tesis Doctoral, hemos presentado la obtención y caracterización

fenotípica y genotípica de una amplia colección de bacterias portadoras de plásmidos de alto peso molecular aisladas a partir de un BPS utilizado para la remoción de pesticidas, centrándonos en el estudio de los plásmidos que éstos transportan. La elección del BPS (ubicado en Kortrijk, Bélgica) como modelo de estudio, se sustentó en análisis moleculares previos (en colaboración con el grupo de la Dra. Kornellia Smalla, Alemania) que pusieron en evidencia la diversidad y abundancia tanto a nivel de la microbiota presente como (y principalmente) de los MGEs detectados en este biofiltro (Dunon, *et al.*, 2013; Dealtry, *et al.*, 2014a y Dealtry, *et al.*, 2014b). Como ya hemos mencionado, los BPSs permiten la remoción de compuestos tóxicos por su degradación, la cual es llevada a cabo por la microbiota presente. Pese a su importancia agrícola y ambiental, poco se sabe de la dinámica de estos sistemas, aunque algunos estudios recientes han comenzado a investigarlo (Tortella, *et al.*, 2013, Verhagen, *et al.*, 2015, Pinto, *et al.*, 2016) incluyendo el presente trabajo.

Bajo la hipótesis de que los plásmidos juegan un rol esencial en la evolución y adaptación microbiana a distintos nichos biológicos acarreado frecuentemente genes que les otorgan a los microorganismos la capacidad de metabolizar distintos compuestos, nos hemos enfocado en el estudio de bacterias cultivables que contienen elementos extracromosomales circulares y cerrados de alto peso molecular (Capítulo III). En este sentido, el análisis del contenido plasmídico de los aislamientos (mediante la técnica de lisis *in situ*) fue una herramienta poderosa que nos permitió, en primer lugar, obtener 75 aislamientos portadores de 1 o más plásmidos y, en segundo lugar, seleccionar 35 de ellos como aislamientos representativos, los cuales corresponden a las clases Alfa, Beta y Gammaproteobacteria, Actinobacteria, Bacilli y Sphingobacteria, y están representados por 13 géneros bacterianos distintos. Si bien es cierto que los microorganismos cultivables representan tan sólo una pequeña proporción de los microorganismos totales de cualquier nicho biológico (Amann *et al.*, 1995), el estudio de comunidades microbianas completas sólo puede ser abordado con herramientas exclusivamente ómicas, esto es, secuenciamiento masivo del ADN total, del ADN extracromosomal o bien del ARNm y/o análisis de las proteínas expresadas en los microorganismos de la comunidad. Sin embargo, estas técnicas metagenómicas no

permiten realizar caracterizaciones fenotípicas de aislamientos individuales, y poseen otras limitaciones. En cuanto al estudio y caracterización del plasmidoma, el abordaje metagenómico sólo permite realizar análisis *in silico* del contenido genético global de los plásmidos recuperados de la comunidad, pero no permite comprobar su rol biológico ni determinar el fenotipo que estos plásmidos les confieren a sus huéspedes, y requieren un gran esfuerzo de secuenciamiento para llegar a detectar plásmidos que están presentes en muy bajo número de copias o en bacterias poco representadas en una muestra. En este sentido, la identificación *in silico* de genes de resistencia a antibióticos y metales y de genes involucrados en la degradación de compuestos xenobióticos o hidrocarburos aromáticos, sólo nos permite inferir su presencia. El análisis de la abundancia de estos genes, como dependen en gran medida de la abundancia de los aislamientos portadores, del número de copias de cada plásmido por célula, de los métodos de extracción de ADN plasmídico y del proceso de secuenciación en sí mismo, puede resultar poco preciso y poco confiable.

El diseño experimental elegido, si bien presenta ciertas limitaciones, ha resultado conveniente para avanzar en los objetivos establecidos. En este sentido, contar con una colección de bacterias portadoras de plásmidos nos ha permitido complementar los estudios bioinformáticos con caracterizaciones bioquímicas y funcionales del tipo de actividades asociadas a diferentes genes del genoma plasmídico móvil.

Una vez establecida la colección bacteriana, en una primera instancia, a través de experimentos clásicos, realizamos una caracterización fenotípica de los aislamientos, con la cual pudimos determinar las MICs que los mismos presentan frente a distintos antibióticos y metales pesados (Capítulo III). Si bien este análisis no permite determinar si los genes que confieren resistencia se encuentran ubicados en el cromosoma bacteriano o en los plásmidos que portan, la tolerancia diferencial que presentaron algunos aislamientos de un mismo género y especie frente a un antimicrobiano o metal indicaría que dichos genes podrían estar ubicados en plásmidos u otros MGEs. Adicionalmente, los ensayos de conjugación en los que se

emplearon como dadores los aislamientos de la colección, nos permitieron atribuirle localización plasmídica a ciertos genes de resistencia poniendo además en evidencia la movilidad de algunos plásmidos de la colección.

En una etapa posterior de este trabajo nos hemos enfocado en la purificación y secuenciamiento del material plasmídico de la colección bacteriana de estudio (en cooperación con el CeBiTec, Alemania) (Capítulo V), a fin de recopilar información sobre el contenido de los elementos extracromosomales circulares recuperados. La estrategia empleada en este trabajo de tesis ha permitido obtener y abordar el análisis genético molecular del contenido plasmídico. La gran cantidad de lecturas crudas obtenidas por el secuenciamiento del *pool* de plásmidos a partir de la plataforma Illumina MiSeq y el posterior ensamblado de las mismas (más de 9 Mpb no redundantes consideradas de origen plasmídico), nos ha permitido realizar un análisis exhaustivo del contenido plasmídico y recabar una gran cantidad de información genética con la cual pudimos acceder a: **i)** la obtención de la secuencia completa y el mapa físico de 6 elementos circulares (Capítulo V), **ii)** analizar qué tipo de replicones están presentes en los aislamientos de la colección bacteriana (Capítulo VI), **iii)** identificar y clasificar genes de resistencia a antibióticos y metales pesados (Capítulo VI); **iv)** analizar si estos MGEs transportan genes involucrados en la degradación de compuestos xenobióticos e hidrocarburos aromáticos; (Capítulo VI), y **v)** identificar enzimas con potencial interés industrial y ambiental (Capítulo VII). Además, y como producto del trabajo presentado en esta Tesis, contamos con las secuencias plasmídicas anotadas en una plataforma informática de fácil manejo y con la colección completa de plásmidos intactos preservados en sus hospedadores originales y en stocks glicerol a -80 °C. Este material, además de representar un depósito genético, es un recurso valioso que contiene un conjunto de plásmidos ambientales que conservan su estructura y por lo tanto, con el tiempo se podrán utilizar en sus hospedadores naturales (supuestos) o en otros, para realizar *screenings* funcionales. Los genes de interés, identificados a partir de la secuenciación y el análisis bioinformático, por lo tanto, estarán físicamente disponibles en los *stocks* conservados.

Como ya hemos mencionado, a partir del secuenciamiento de los plásmidos se pudieron identificar genes de resistencia a antibióticos, lo cual corrobora la hipótesis de una localización plasmídica de muchos de ellos. En este sentido, trabajos previos han reportado la presencia de genes de resistencia a antibióticos y metales pesados en plásmidos provenientes de ambientes contaminados con pesticidas y otros compuestos tóxicos (Anjum *et al.*, 2012; Sharma y Thakur, 2009). Hoy en día es bien sabido que los genes de resistencia a antibióticos se encuentran presentes en todo tipo de ambientes, incluso aquéllos en los que no se ha registrado contaminación antropogénica (Allen, *et al.*, 2009, Berglund, 2015). Desde un punto de vista clínico, el estudio de la diseminación de genes de resistencia a antibióticos en el ambiente resulta de gran relevancia ya que existe fuerte evidencia de que muchos genes de resistencia que se encuentran en aislamientos clínicos se han originado en el ambiente (Perry & Wright, 2013).

La cantidad y diversidad de proteínas involucradas en la replicación plasmídica identificadas (Capítulo VI) nos da una idea de la gran variedad de los plásmidos presentes en la colección del BPS analizado. Algunos estudios similares realizados sobre replicones aislados de distintos ambientes se han focalizado en el análisis de proteínas de replicación plasmídica, ya que éstas permiten hacer inferencias acerca de los huéspedes bacterianos que pueden habitar estas moléculas (Schlüter *et al.*, 2008, Jorjensen *et al.*, 2014). Es importante remarcar que hemos podido identificar proteínas pertenecientes a plásmidos cuyos huéspedes bacterianos pertenecen a distintos filos y clases, que abarcan tanto bacterias Gram-positivas como Gram-negativas, lo cual pone en evidencia la diversidad de los replicones caracterizados. Atendiendo a que el moviloma transita diversas bacterias (tanto cultivables como no cultivables) el tráfico de estos elementos y su capacidad de replicar en una célula hospedadora aparece como una estrategia potencialmente importante para la recolección y ensayo de funciones presentes en otras bacterias sometidas a condiciones similares y alcanzadas por el mismo moviloma (valioso por ser portador de información adaptativa al mismo nicho y susceptible de ser compartida).

Otro aspecto importante que fue analizado detalladamente es la presencia de determinantes de movilidad (como el gen de la relaxasa y otras proteínas accesorias involucradas en la formación del aparato conjugativo). El análisis *in silico* nos permitió identificar 19 ORFs que codifican para relaxasas en el set de datos plasmídico, las cuales fueron asignadas a 4 familias MOB (MOB_F, MOB_O, MOB_V y MOB_P) de acuerdo con la clasificación previamente establecida por el grupo del Dr. Fernando De La Cruz ((Francia, *et al.*, 2004, Garcillan-Barcia, *et al.*, 2009, Smillie, *et al.*, 2010, Guglielmini, *et al.*, 2011). También se identificaron 13 proteínas del tipo VirB4 (formadoras del MPF), todas ellas del grupo MPF_T. Estos resultados sugieren que posiblemente habría más plásmidos movilizables que conjugativos en los aislamientos analizados, debido a la relación entre el número de secuencias codificantes para relaxasas y proteínas VirB4 encontradas, resultados que se condicen con los obtenidos en los ensayos de conjugación bi y triparentales detallados en el Capítulo III. Los plásmidos que poseen movilidad resultan muy interesantes desde un punto de vista evolutivo; al poder dispersarse fácilmente entre distintas bacterias, éstos generan un intercambio dinámico de material genético, resultando en la adquisición, o a veces pérdida, de información. Desde un punto de vista ambiental, la movilidad de los plásmidos en comunidades microbianas puede proporcionar a distintas bacterias una gran cantidad de ventajas adaptativas, como la capacidad de degradar una fuente de carbono, la capacidad de colonizar un determinado huésped o la tolerancia a un compuesto presente en dicho ambiente. Por otro lado, desde un punto de vista clínico, la movilidad puede representar un riesgo para la salud humana, ya que los plásmidos juegan un rol importante en la diseminación de resistencia a agentes microbianos (Bennett, 2008). Por estas razones, el estudio de las capacidades de transferencia que presentan los plásmidos es un punto en el que la comunidad científica ha puesto particular atención.

El análisis *in silico* de la información residente en los plásmidos reveló además la presencia de un gran número de genes relacionados con la resistencia a metales, así como de genes involucrados en la degradación de hidrocarburos aromáticos y compuestos xenobióticos. En el proceso evolutivo, algunas bacterias han desarrollado

o adquirido mecanismos para sortear los efectos nocivos generados por los compuestos tóxicos. Esta tolerancia puede darse de manera indirecta a través de la extrusión del contaminante al medio extracelular mediante bombas de eflujo o por degradación del compuesto tóxico a través de rutas metabólicas (Segura *et al.*, 2014). El análisis de genes de resistencia a compuestos xenobióticos recibió una atención especial, ya que los plásmidos fueron obtenidos a partir de aislamientos que provienen de un ambiente enriquecido en pesticidas de estructura química diversa. Para el estudio de rutas metabólicas específicas, es necesario contar con mecanismos para organizar las secuencias en base a su anotación. La mayor limitación que encontramos son las bases de datos disponibles para este fin (SEED, COG, KEGG, etc.) que están incompletas y mal organizadas. Generalmente, estas bases de datos se encuentren sesgadas hacia las categorías metabólicas mejor caracterizadas y representadas en todos los genomas. Asimismo, existen numerosos errores en la forma en la que están construidas las categorías, por ejemplo porque no incluyen ciertas enzimas clave, ya que faltan rutas metabólicas o porque incluyen enzimas que no son específicas del metabolismo correspondiente a una categoría. A pesar de ello, la búsqueda bioinformática que hemos realizado nos ha permitido encontrar algunos genes cuyos productos de expresión están implicados en la degradación de ciertos pesticidas como 2,4-D, atrazina, paratión y fluorobenzoato, además de bombas de eflujo implicadas en la tolerancia a metil viológenos. Varios de estos genes han sido previamente identificados en plásmidos y otros MGEs (Sakai *et al.* 2014, Changey *et al.* 2011, Liu *et al.* 2005, Sebastianelli *et al.* 2007), por lo cual su localización plasmídica no es inesperada. Más aún, los plásmidos que transportan genes implicados en la tolerancia a estos compuestos juegan un rol importante en la diseminación de estas funciones en el ambiente. La presencia de estos genes en MGEs puede verse como una ventaja desde un punto de vista adaptativo para las bacterias que los adquieren, y su diseminación en el ambiente podría resultar favorable desde un punto de vista ecológico, considerando que la degradación de estos compuestos deriva en una disminución de la contaminación ambiental. El hecho de haber identificado plásmidos portadores de genes posiblemente implicados en la degradación de compuestos xenobióticos, podría

proveer de una herramienta biotecnológica importante para la remediación de suelos agrícolas o aguas contaminadas. A partir del estudio de plásmidos de estas características se han planteado estrategias de “bioaumento génico”, el cual consiste en introducir en un ambiente bacterias conteniendo plásmidos conjugativos catabólicos a fin de estimular la HGT de los mismos hacia microorganismos endógenos, proporcionándoles a éstos una mayor capacidad para adaptarse y sobrevivir en un nicho contaminado (Top, *et al.*, 2002). Sin embargo, y a pesar de que la bioremediación es una alternativa de bajo costo, suele no ser exitosa a consecuencia de varios factores, entre ellos la complejidad de las interacciones microbianas, la necesidad de diseñar consorcios microbianos selectos capaces de un cometabolismo específico, la presencia de depredadores naturales de microbios, la textura y el tipo de suelo, la alta concentración y amplia variedad de pesticidas (Maloney, 2001). En este contexto, los datos obtenidos en nuestro trabajo deberán ser complementados con ensayos de tolerancia/degradación *in vitro* e *in vivo* para determinar si los genes identificados realmente proporcionan el fenotipo esperado a los aislamientos que los contienen y para avanzar en la identificación de los factores ambientales necesarios para una degradación óptima.

El análisis de las secuencias plasmídicas nos permitió identificar varias CDSs que codifican para polipéptidos de tipo lacasas (oxidasas multicobre) de potencial aplicación industrial (Capítulo VII). Las lacasas son ampliamente utilizadas en procesos biotecnológicos por su capacidad de actuar sobre una gran cantidad de sustratos (Giardina, *et al.*, 2010, Pezzella, *et al.*, 2015), y también se ha planteado su uso en estrategias de descontaminación de ambientes debido a su capacidad de oxidar ciertos compuestos como herbicidas y pesticidas (Adrio, *et al.*, 2014). Con el propósito de analizar su potencial aplicación, se seleccionó la secuencia del gen *lac188* (que codifica para una posible lacasa de 3 dominios). En una primera etapa recurrimos a herramientas de biología sintética para obtener la secuencia nucleotídica de *lac188* optimizada y luego ésta se clonó en un vector de expresión para avanzar en la purificación y caracterización del producto funcional de *lac188* en *E. coli*. Ensayos realizados en colaboración con el grupo de la Dra. Flavia Marinelli (Italia), permitieron

demostrar su actividad sobre 5 sustratos distintos: ABTS, pirogalol, pirocatecol, ácido vanílico y DMP, actuando preferencialmente sobre este último. Si bien se ha reportado previamente que este tipo de enzimas posee actividad sobre distintos sustratos (Adrio, *et al.*, 2014), no hay informes de lacasas que presenten una actividad específica alta que provengan de plásmidos. Sin embargo, se sabe que estas proteínas, si bien se encuentran mayoritariamente en el cromosoma bacteriano, pueden estar presentes en replicones plasmídicos (Ausec, *et al.*, 2011). La evidencia recolectada indica que se requerirán ensayos extras para mejorar el rendimiento de proteína soluble y el grado de purificación de la misma que permitan avanzar en la caracterización de los parámetros cinéticos y rangos de estabilidad de esta enzima.

Si bien la metodología presentada en este trabajo de Tesis para abordar el secuenciamiento y estudio de los plásmidos nos permitió obtener megabases de información, realizar un análisis profundo a partir de las secuencias y poder recuperar plásmidos intactos, posee sus limitaciones. Como ya hemos mencionado, otros métodos disponibles basados en metagenómica tienen la gran ventaja de que permiten la obtención de un gran número de plásmidos a partir de una comunidad bacteriana, incluso a partir de aquellas bacterias que no pueden cultivarse en el laboratorio; pero su principal desventaja reside en la imposibilidad de recuperar físicamente dichos plásmidos en una célula, impidiendo tanto su caracterización fenotípica como el conocimiento de la bacteria que lo hospedaba. En vista de las limitaciones que presenta el abordaje experimental del estudio de los plásmidos presentes en bacterias ambientales, hemos diseñado un vector (p-RESC) como parte de una nueva herramienta molecular, a fin de ampliar el espectro de plásmidos que pueden ser capturados de un ambiente. Nuestra estrategia, propone movilizar los plásmidos (presentes en los cultivos bacterianos puros o comunidades bacterianas) hacia una cepa hospedadora de laboratorio facilitando su posterior purificación y, en consecuencia, las siguientes etapas de estudio. La estrategia diseñada se basa en la incorporación de un transposón al plásmido *target*, que aporta la capacidad de moverse a una cepa huésped de laboratorio, replicar en ella y un gen de resistencia que permite seleccionarla. Atendiendo a los resultados obtenidos (Capítulo VIII) se

necesitará concentrar esfuerzos para avanzar hacia futuras mejoras, ya sea en el p-RESC que hemos construido o bien en las sucesivas etapas del protocolo experimental diagramado. En este marco, la evidencia colectada servirá como punto de partida para abordar futuros ensayos a fin de validar el sistema de captura plasmídica diseñado.

Los resultados que hemos presentado en este trabajo de Tesis han abordado aspectos bioquímicos, genómicos, evolutivos y genéticos a fin de caracterizar una diversa colección de plásmidos, así como a los aislamientos que los contienen.

IX.2. Perspectivas

Los resultados presentados han aportado nueva evidencia de las características moleculares y funcionales de los plásmidos presentes en las cepas seleccionadas de la colección. La información recogida podrá ser utilizada sobre una base racional para la construcción de nuevos vectores, nuevas herramientas o bien para la búsqueda de otros genes u operones que resulten de interés.

Nuestra base de datos se podrá integrar o combinar con bases de datos de transposones y secuencias de otros plasmidomas. Esto generaría un recurso valioso que potencialmente se podría utilizar para responder múltiples preguntas diferentes en proyectos donde se requerirá una marcada interacción en las áreas de biología molecular de MGEs, ecología microbiana y bioinformática para obtener el máximo provecho del conjunto de datos generados.

Finalmente, está claro que nuestra comprensión básica de los factores que influyen en la difusión y el destino de los genes transportados en plásmidos (y otros MGEs) necesita ser mejorada. Esto incluye avanzar en la comprensión de las condiciones que estimulan los mecanismos de conjugación y transferencia de genes en diversos ecosistemas, el rango de huéspedes de los MGEs y el rango de expresión de sus genes, la aptitud competitiva de los transconjugantes, y el mantenimiento de los MGEs tanto en presencia como en ausencia de presión de selección o de un

determinado sustrato. A futuro, el éxito de estrategias que impliquen la liberación de bacterias o la bioaumentación mediada por plásmidos, así como la utilización de bioproductos derivados de los mismos, dependerá en gran medida del aumento de los conocimientos básicos en estas áreas.

MATERIAL SUPLEMENTARIO

S1. Alineamiento de las 6 integrasas secuenciadas a partir de los aislamientos del biofiltro

```

BF_19 -----
BF_42 ---CCCAAAAAACAGTCATAACAAGCCATGAAAACCGCCAATTGCGCCGTTACCACCGCTG
BF_44 GCTCCCAAAAAACAGTCATAACAAGCCATGAAAACCGCCAATTGCGCCGTTACCACCGCTG
BF_27 -----CAAAAAACAGTCATAACAAGCCATGAAAACCGCCAATTGCGCCGTTACCACCGCTG
BF_02 -----AAAAACAGTCATAACAAGCCATGAAAACCGCCAATTGCGCCGTTACCACCGCTG
BF_25 -----CAAAAAACAGTCATAACAAGCCATGAAAACCGCCAATTGCGCCGTTACCACCGCTG

```

```

BF_19 -----
BF_42 CGTTCGGTCAAGGTTCTGGACCAGTTGCGTGAGCGCATACGCTACTTGCATTACAGTTTA
BF_44 CGTTCGGTCAAGGTTCTGGACCAGTTGCGTGAGCGCATACGCTACTTGCATTACAGTTTA
BF_27 CGTTCGGTCAAGGTTCTGGACCAGTTGCGTGAGCGCATACGCTACTTGCATTACAGTTTA
BF_02 CGTTCGGTCAAGGTTCTGGACCAGTTGCGTGAGCGCATACGCTACTTGCATTACAGTTTA
BF_25 CGTTCGGTCAAGGTTCTGGACCAGTTGCGTGAGCGCATACGCTACTTGCATTACAGTTTA

```

```

BF_19 -----
BF_42 CGAACCGAACAGGCTTATGTCCACTGGGTTTCGTGCCTTCATCCGTTTCCACGGTGTGCGT
BF_44 CGAACCGAACAGGCTTATGTCCACTGGGTTTCGTGCCTTCATCCGTTTCCACGGTGTGCGT
BF_27 CGAACCGAACAGGCTTATGTCCACTGGGTTTCGTGCCTTCATCCGTTTCCACGGTGTGCGT
BF_02 CGAACCGAACAGGCTTATGTCCACTGGGTTTCGTGCCTTCATCCGTTTCCACGGTGTGCGT
BF_25 CGAACCGAACAGGCTTATGTCCACTGGGTTTCGTGCCTTCATCCGTTTCCACGGTGTGCGT

```

```

BF_19 -----
BF_42 CACCCGGCAACCTTGGGCAGCAGCGAAGTCGAGGCATTTCTGTCTGGCTGGCGAACGAG
BF_44 CACCCGGCAACCTTGGGCAGCAGCGAAGTCGAGGCATTTCTGTCTGGCTGGCGAACGAG
BF_27 CACCCGGCAACCTTGGGCAGCAGCGAAGTCGAGGCATTTCTGTCTGGCTGGCGAACGAG
BF_02 CACCCGGCAACCTTGGGCAGCAGCGAAGTCGAGGCATTTCTGTCTGGCTGGCGAACGAG
BF_25 CACCCGGCAACCTTGGGCAGCAGCGAAGTCGAGGCATTTCTGTCTGGCTGGCGAACGAG

```

```

BF_19 -----
BF_42 CGCAAGGTTTTCGGTCTCCACGCATCGTCAGGCATTGGCGGCCTTGCTGTTCTTCTACGGC
BF_44 CGCAAGGTTTTCGGTCTCCACGCATCGTCAGGCATTGGCGGCCTTGCTGTTCTTCTACGGC
BF_27 CGCAAGGTTTTCGGTCTCCACGCATCGTCAGGCATTGGCGGCCTTGCTGTTCTTCTACGGC
BF_02 CGCAAGGTTTTCGGTCTCCACGCATCGTCAGGCATTGGCGGCCTTGCTGTTCTTCTACGGC
BF_25 CGCAAGGTTTTCGGTCTCCACGCATCGTCAGGCATTGGCGGCCTTGCTGTTCTTCTACGGC

```

```

BF_19 -----
BF_42 AAGGTGCTGTGCACGGATCTGCCCTGGCTTCAGGAGATCGGAAGACCTCGGCCGTCGCGG
BF_44 AAGGTGCTGTGCACGGATCTGCCCTGGCTTCAGGAGATCGGAAGACCTCGGCCGTCGCGG
BF_27 AAGGTGCTGTGCACGGATCTGCCCTGGCTTCAGGAGATCGGAAGACCTCGGCCGTCGCGG
BF_02 AAGGTGCTGTGCACGGATCTGCCCTGGCTTCAGGAGATCGGAAGACCTCGGCCGTCGCGG
BF_25 AAGGTGCTGTGCACGGATCTGCCCTGGCTTCAGGAGATCGGAAGACCTCGGCCGTCGCGG

```

```

BF_19 -----
BF_42 CGCTTGCCGGTGGTGCTGACCCCGGATGAAGTGGTTCGCATCCTCGGTTTTCTGGAAGGC
BF_44 CGCTTGCCGGTGGTGCTGACCCCGGATGAAGTGGTTCGCATCCTCGGTTTTCTGGAAGGC
BF_27 CGCTTGCCGGTGGTGCTGACCCCGGATGAAGTGGTTCGCATCCTCGGTTTTCTGGAAGGC
BF_02 CGCTTGCCGGTGGTGCTGACCCCGGATGAAGTGGTTCGCATCCTCGGTTTTCTGGAAGGC
BF_25 CGCTTGCCGGTGGTGCTGACCCCGGATGAAGTGGTTCGCATCCTCGGTTTTCTGGAAGGC

```

```

BF_19 -----
BF_42 GAGCATCGTTTTGTTTCGCCAGCTTCTGTATGGAACGGGCATGCGGATCAGTGAGGGTTTG
BF_44 GAGCATCGTTTTGTTTCGCCAGCTTCTGTATGGAACGGGCATGCGGATCAGTGAGGGTTTG
BF_27 GAGCATCGTTTTGTTTCGCCAGCTTCTGTATGGAACGGGCATGCGGATCAGTGAGGGTTTG
BF_02 GAGCATCGTTTTGTTTCGCCAGCTTCTGTATGGAACGGGCATGCGGATCAGTGAGGGTTTG
BF_25 GAGCATCGTTTTGTTTCGCCAGCTTCTGTATGGAACGGGCATGCGGATCAGTGAGGGTTTG

```

BF_25 GAGCATCGTTTTGTTGCGCCAGCTTCTGTATGGAACGGGCATGCGGATCAGTGAGGGTTTG

BF_19 -----CTGGATTTTCGATCACGGCAGCATCATCGTGCGGGAGGGCAAG
BF_42 CAACTGCGGGTCAAGGATCTGGATTTTCGATCACGGCAGCATCATCGTGCGGGAGGGCAAG
BF_44 CAACTGCGGGTCAAGGATCTGGATTTTCGATCACGGCAGCATCATCGTGCGGGAGGGCAAG
BF_27 CAACTGCGGGTCAAGGATCTGGATTTTCGATCACGGCAGCATCATCGTGCGGGAGGGCAAG
BF_02 CAACTGCGGGTCAAGGATCTGGATTTTCGATCACGGCAGCATCATCGTGCGGGAGGGCAAG
BF_25 CAACTGCGGGTCAAGGATCTGGATTTTCGATCACGGCAGCATCATCGTGCGGGAGGGCAAG

BF_19 GGCTCCAAGGATCGGGCCTTGATGTTACCCGAGAGCTTGGCACCCAGCCTGCGCGAGCAG
BF_42 GGCTCCAAGGATCGGGCCTTGATGTTACCCGAGAGCTTGGCACCCAGCCTGCGCGAGCAG
BF_44 GGCTCCAAGGATCGGGCCTTGATGTTACCCGAGAGCTTGGCACCCAGCCTGCGCGAGCAG
BF_27 GGCTCCAAGGATCGGGCCTTGATGTTACCCGAGAGCTTGGCACCCAGCCTGCGCGAGCAG
BF_02 GGCTCCAAGGATCGGGCCTTGATGTTACCCGAGAGCTTGGCACCCAGCCTGCGCGAGCAG
BF_25 GGCTCCAAGGATCGGGCCTTGATGTTACCCGAGAGCTTGGCACCCAGCCTGCGCGAGCAG

BF_19 CTGTGCGGTGCACGGGCATGGTGGCTGAAGGACCAGGCCGAGGGCCGACGCGGCGTTGCG
BF_42 CTGTGCGGTGCACGGGCATGGTGGCTGAAGGACCAGGCCGAGGGCCGACGCGGCGTTGCG
BF_44 CTGTGCGGTGCACGGGCATGGTGGCTGAAGGACCAGGCCGAGGGCCGACGCGGCGTTGCG
BF_27 CTGTGCGGTGCACGGGCATGGTGGCTGAAGGACCAGGCCGAGGGCCGACGCGGCGTTGCG
BF_02 CTGTGCGGTGCACGGGCATGGTGGCTGAAGGACCAGGCCGAGGGCCGACGCGGCGTTGCG
BF_25 CTGTGCGGTGCACGGGCATGGTGGCTGAAGGACCAGGCCGAGGGCCGACGCGGCGTTGCG

BF_19 CTTCCCGACGCCCTTGAGCGGAAGTATCCGCGCGCCGGGCATTCTGGCCGTGGTTCTGG
BF_42 CTTCCCGACGCCCTTGAGCGGAAGTATCCGCGCGCCGGGCATTCTGGCCGTGGTTCTGG
BF_44 CTTCCCGACGCCCTTGAGCGGAAGTATCCGCGCGCCGGGCATTCTGGCCGTGGTTCTGG
BF_27 CTTCCCGACGCCCTTGAGCGGAAGTATCCGCGCGCCGGGCATTCTGGCCGTGGTTCTGG
BF_02 CTTCCCGACGCCCTTGAGCGGAAGTATCCGCGCGCCGGGCATTCTGGCCGTGGTTCTGG
BF_25 CTTCCCGACGCCCTTGAGCGGAAGTATCCGCGCGCCGGGCATTCTGGCCGTGGTTCTGG

BF_19 GTTTTTGCGCAGCACACGCATTCGACCGATCCACGGAGCGGTGTCGTGCGTCGCCATCAC
BF_42 GTTTTTGCGCAGCACACGCATTCGACCGATCCACGGAGCGGTGTCGTGCGTCGCCATCAC
BF_44 GTTTTTGCGCAGCACACGCATTCGACCGATCCACGGAGCGGTGTCGTGCGTCGCCATCAC
BF_27 GTTTTTGCGCAGCACACGCATTCGACCGATCCACGGAGCGGTGTCGTGCGTCGCCATCAC
BF_02 GTTTTTGCGCAGCACACGCATTCGACCGATCCACGGAGCGGTGTCGTGCGTCGCCATCAC
BF_25 GTTTTTGCGCAGCACACGCATTCGACCGATCCACGGAGCGGTGTCGTGCGTCGCCATCAC

BF_19 ATGTATGACCAGACCTTTTCAGCGCGCCTTCAAACGTGCCGTAGAACAAGCAGGCATCACG
BF_42 ATGTATGACCAGACCTTTTCAGCGCGCCTTCAAACGTGCCGTAGAACAAGCAGGCATCACG
BF_44 ATGTATGACCAGACCTTTTCAGCGCGCCTTCAAACGTGCCGTAGAACAAGCAGGCATCACG
BF_27 ATGTATGACCAGACCTTTTCAGCGCGCCTTCAAACGTGCCGTAGAACAAGCAGGCATCACG
BF_02 ATGTATGACCAGACCTTTTCAGCGCGCCTTCAAACGTGCCGTAGAACAAGCAGGCATCACG
BF_25 ATGTATGACCAGACCTTTTCAGCGCGCCTTCAAACGTGCCGTAGAACAAGCAGGCATCACG

BF_19 AAGCCCGCCACACCGCAACCCCTCCGCCACTCGTTTCGCGACGGCT
BF_42 AAGCCCGCCACACCGCAACCCCTCCGCCACTCGTTTCGCGACGGCT
BF_44 AAGCCCGCCACACCGCAACCCCTCCGCCACTCGTTTCGCGACGGCT
BF_27 AAGCCCGCCACACCGCAACCCCTCCGCCACTCGTTTCGCGACGGCT
BF_02 AAGCCCGCCACACCGCAACCCCTCCGCCACTCGTTTCGCGACGGCT
BF_25 AAGCCCGCCACACCGCAACCCCTCCGCCACTCGTTTCGCGACGGCT

S2. Genes identificados en los seis replicones completos

ORF	Descripción (major hit BLASTp en NCBI)	Organismo	Covertura	e-value	Máx ident	Número de acceso
pMC1						
1	No significant similarity found					
2	hypothetical protein	<i>Ochrobactrum anthropi</i>	98%	5,00E-36	79%	WP_036587968.1
3	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	98%	2,00E-32	80%	WP_036568669.1
4	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	96%	6,00E-31	67%	WP_024900284.1
5	C-5 cytosine-specific DNA methylase	<i>Ochrobactrum rhizosphaerae</i>	99%	0.0	81%	WP_024900288.1
6	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	42%	2,00E-13	58%	WP_024899497.1
7	repressor protein C	<i>Ochrobactrum intermedium</i>	97%	2,00E-98	60%	WP_036565897.1
8	hypothetical protein	<i>Mesorhizobium</i> sp.	48%	2,00E-15	59%	ESZ60509.1
9	hypothetical protein	<i>Hoeflea phototrophica</i>	96%	3,00E-58	60%	WP_007196931.1
10	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	77%	4,00E-22	77%	WP_024900293.1
11	hypothetical protein	<i>chrobactrum rhizosphaerae</i>	98%	5,00E-49	80%	WP_024900294.1
12	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	78%	7,00E-139	78%	WP_036568721.1
13	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	99%	7,00E-32	41%	WP_024900296.1
14	No significant similarity found					
15	DNA methylase	<i>Ochrobactrum rhizosphaerae</i>	77%	6,00E-136	84%	WP_036568723.1
16	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	97%	1,00E-42	78%	WP_024900298.1
17	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	99%	1,00E-82	75%	WP_024900299.1
18	hypothetical protein	<i>Ochrobactrum anthropi</i>	99%	1,00E-147	58%	WP_012092184.1
19	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	98%	8,00E-77	62%	WP_024899506.1
20	No significant similarity found					
21	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	77%	2,00E-57	89%	WP_036568726.1
22	hypothetical protein	<i>Sinorhizobium</i> sp.	85%	3,00E-18	72%	WP_037421955.1
23	hypothetical protein	<i>Starkeya novella</i>	64%	5,00E-25	51%	WP_013164853.1
24	terminase	<i>Sphingobium</i> sp.	94%	0.0	54%	WP_037530719.1
25	portal protein	<i>Ochrobactrum rhizosphaerae</i>	95%	0.0	97%	WP_024900305.1
26	peptidase	<i>Ochrobactrum rhizosphaerae</i>	99%	3,00E-177	95%	WP_024900306.1

Tabla S2 (continuación)

ORF	Descripción (major hit BLASTp en NCBI)	Organismo	Coertura	e-value	Máx ident	Número de acceso
27	nucleoid-structuring protein H-NS	<i>Ochrobactrum rhizosphaerae</i>	99%	0.0	90%	WP_024900307.1
28	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	99%	1,00E-75	72%	WP_024900308.1
29	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	48%	1,00E-61	97%	WP_036568728.1
30	phage head-tail adapter protein	<i>Ochrobactrum rhizosphaerae</i>	73%	8,00E-68	91%	WP_024900310.1
31	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	83%	3,00E-86	90%	WP_024900312.1
32	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	98%	3,00E-98	95%	WP_024900314.1
33	hypothetical protein	<i>Ochrobactrum</i> sp.	99%	1,00E-77	85%	WP_036596028.1
34	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	17%	2,00E-76	96%	WP_036568734.1
35	hypothetical protein	<i>Ochrobactrum anthropi</i>	61%	3,00E-10	76%	WP_041544905.1
36	hypothetical protein	<i>Ochrobactrum anthropi</i>	95%	1,00E-17	84%	WP_041544905.1
37	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	99%	3,00E-136	92%	WP_024900317.1
38	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	99%	2,00E-131	93%	WP_024900318.1
39	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	99%	5,00E-72	80%	WP_024900319.1
40	fibronectin	<i>Ochrobactrum rhizosphaerae</i>	91%	0.0	86%	WP_024900320.1
41	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	32%	3,00E-79	65%	WP_036568737.1
42	hypothetical protein	<i>Ochrobactrum intermedium</i>	42%	1,00E-68	43%	WP_031353039.1
43	hypothetical protein	<i>Mesorhizobium loti</i>	98%	1,00E-89	54%	WP_010915306.1
44	No significant similarity found					
45	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	93%	7,00E-152	85%	WP_036568757.1
46	hypothetical protein	<i>Ochrobactrum intermedium</i>	96%	7,00E-27	58%	WP_025089589.1
47	hypothetical protein	<i>Ochrobactrum anthropi</i>	91%	1,00E-39	73%	WP_029376117.1
48	hypothetical protein	<i>Agrobacterium vitis S4</i>	98%	3,00E-33	40%	ACM36644.1
49	hypothetical protein	<i>Devosia</i> sp.	97%	4,00E-21	51%	WP_035102542.1
50	hypothetical protein	<i>Sinorhizobium meliloti</i>	81%	8,00E-52	38%	WP_017274933.1
51	integrase	<i>Proteobacteria</i>	99%	7,00E-167	60%	WP_035026333.1
pMC2						
1	hypothetical protein	<i>Delftia</i> sp.	93%	7,00E-18	57%	WP_013803144.1
2	hypothetical protein	<i>Acidovorax radidis</i>	93%	4,00E-06	46%	WP_010465941.1
3	hypothetical protein	<i>Burkholderia xenovorans</i>	94%	3,00E-59	62%	WP_011487586.1
4	hypothetical protein	<i>Alicyclophilus denitrificans</i>	56%	3,00E-04	45%	WP_013516292.1

Tabla S2 (continuación)

ORF	Descripción (major hit BLASTp en NCBI)	Organismo	Covertura	e-value	Máx ident	Número de acceso
5	hypothetical protein	<i>Polaromonas</i> sp.	85%	7,00E-10	40%	WP_007874681.1
6	parB-like partition protein	<i>Acidovorax citrulli</i>	76%	1,00E-133	46%	WP_011794736.1
7	No significant similarity found					
8	No significant similarity found					
9	No significant similarity found					
10	No significant similarity found					
11	hypothetical protein	<i>Delftia acidovorans</i>	52%	3,00E-04	38%	WP_016447887.1
12	hypothetical protein	<i>Mesorhizobium</i> sp.	84%	1,00E-23	36%	WP_027142925.1
13	hypothetical protein	<i>Parasutterella excrementihominis</i>	50%	9,00E-13	45%	WP_040603707.1
14	hypothetical protein	<i>Alcaligenes faecalis</i>	93%	1,00E-14	47%	WP_035267837.1
15	No significant similarity found					
16	hypothetical protein	<i>Delftia acidovorans</i>	89%	3,00E-51	65%	WP_016454567.1
17	hypothetical protein	<i>Acidovorax citrulli</i>	75%	1,00E-20	62%	WP_041827888.1
18	conjugal transfer protein TraC	<i>Pseudomonas mendocina</i>	88%	4,00E-88	31%	WP_041773113.1
19	hypothetical protein	<i>Hylemonella gracilis</i>	75%	2,00E-21	57%	WP_035608219.1
20	No significant similarity found					
21	hypothetical protein	<i>Delftia acidovorans</i>	98%	3,00E-108	85%	EPD44780.1
22	hypothetical protein	<i>elftia acidovorans</i>	98%	0.0	85%	WP_016445316.1
23	HK97 family phage portal protein	<i>Delftia</i> sp.	98%	0.0	65%	EZP51836.1
24	peptidase	<i>Delftia acidovorans</i>	99%	6,00E-124	80%	WP_034399341.1
25	HK97 family phage major capsid protein	<i>Comamonas testosteroni</i>	99%	0.0	85%	WP_034349600.1
26	hypothetical protein CTATCC11996_18657 Comamonas testosteroni	<i>Delftia</i> sp.	26%	9,00E-06	58%	EZP51833.1
27	No significant similarity found					
28	phage head-tail adaptor	<i>Marinomonas phage</i>	92%	8,00E-04	38%	YP_006560248.1
29	bacteriophage TP901-1 ORF40-like protein	<i>Burkholderiales bacterium</i>	96%	5,00E-49	55%	BAP88849.1
30	No significant similarity found					
31	hypothetical protein	<i>Ruegeria</i> sp.	81%	7,00E-23	45%	YP_613609.1
32	hypothetical protein	<i>Novosphingobium</i> sp.	86%	0.004	28%	WP_007678920.1

Tabla S2 (continuación)

ORF	Descripción (major hit BLASTp en NCBI)	Organismo	Covertura	e-value	Máx ident	Número de acceso
33	hypothetical protein	<i>Massilia</i> sp.	64%	9,00E-08	41%	WP_036176014.1
34	tail protein	<i>seudomonas aeruginosa</i>	17%	9,00E-33	40%	WP_034067959.1
35	carbohydrate-binding protein	<i>Acidovorax citrulli</i>	98%	3,00E-65	39%	WP_011794764.1
36	hypothetical protein	<i>Sulfurospirillum cavolei</i>	34%	6,00E-08	33%	WP_041963463.1
37	hypothetical protein	<i>elftia tsuruhatensis</i>	40%	2,00E-05	41%	KEH10790.1
38	hypothetical protein	<i>Alicyclophilus denitrificans</i>	69%	4,00E-07	43%	WP_013518562.1
39	hypothetical protein	<i>Acidovorax citrulli</i>	26%	1,00E-04	36%	ABM32217.1
40	No significant similarity found					
41	hypothetical protein	<i>Massilia alkalitolerans</i>	45%	2,00E-12	58%	WP_027864255.1
42	hypothetical protein	<i>Hermiinimonas</i> sp.	55%	1,00E-07	52%	WP_025915811.1
43	glycoside hydrolase	<i>Comamonas testosteroni</i>	99%	2,00E-64	69%	WP_043005020.1
44	hypothetical protein	<i>Alicyclophilus denitrificans</i>	96%	3,00E-20	50%	WP_013518568.1
45	hypothetical protein	<i>Delftia acidovorans</i>	98%	2,00E-41	63%	WP_034395290.1
46	hypothetical protein	<i>Nevskia ramosa</i>	52%	3,00E-39	53%	WP_033418158.1
47	hypothetical protein	<i>Alicyclophilus denitrificans</i>	85%	3,00E-31	77%	WP_013518572.1
48	No significant similarity found					
49	No significant similarity found					
50	integrase	<i>Comamonas testosteroni</i>	98%	0.0	69%	WP_034371762.1
pMC3						
1	No significant similarity found					
2	hypothetical protein	<i>Bacillus cereus</i> group	98%	4,00E-13	100%	WP_000182548.1
3	No significant similarity found					
4	No significant similarity found					
5	hypothetical protein	<i>Bacillus cereus</i>	98%	1,00E-36	100%	WP_001038660.1
6	hypothetical protein	<i>Bacillus cereus</i>	57%	0.004	83%	EPC06203.1
7	mobilization protein	<i>Bacillus cereus</i>	99%	0.0	98%	WP_000901534.1
8	replication protein	<i>Bacillus cereus</i>	99%	5,00E-145	99%	WP_000170172.1
9	hypothetical protein	<i>Bacillus cereus</i>	69%	3,00E-09	62%	EDZ48787.1
10	hypothetical protein	<i>Bacillus cereus</i>	97%	4,00E-16	87%	WP_033719991.1
11	hypothetical protein	<i>Bacillus cereus</i>	98%	2,00E-29	100%	EEL58551.1
12	Cold shock protein	<i>Bacillus cereus</i>	98%	9,00E-39	100%	WP_000434813.1
13	hypothetical protein	<i>Bacillus cereus</i>	75%	4,00E-11	76%	WP_000903438.1
14	Barstar (barnase inhibitor) Paenibacillus curdlanolyticus YK9	<i>Paenibacillus curdlanolyticus</i>	98%	4,00E-57	53%	EFM10710.1

Tabla S2 (continuación)

ORF	Descripción (major hit BLASTp en NCBI)	Organismo	Coertura	e-value	Máx ident	Número de acceso
15	hypothetical protein	<i>Bacillus cereus</i>	98%	1,00E-58	97%	WP_002166780.1
16	hypothetical protein	<i>Bacillus cereus</i>	57%	1.4	61%	WP_000601186.1
pMC4						
1	hypothetical protein	<i>Acidovorax</i> sp.	81%	4,00E-09	71%	ABM40964.1
2	hypothetical protein	<i>Acidovorax</i> sp.	86%	2,00E-135	61%	ABM40963.1
3	No significant similarity found					
4	Resolvase	<i>Comamonas testosteroni</i>	98%	2,00E-67	58%	WP_034378873.1
5	hypothetical protein	<i>Acidovorax ebreus</i>	54%	4,00E-07	49%	ACM33035.1
6	No significant similarity found					
7	hypothetical protein	<i>Comamonas testosteroni</i>	49%	0.002	68%	WP_034357633.1
8	replication initiation factor	<i>Acidovorax citrulli</i>	86%	4,00E-104	56%	ABM32965.1
9	hypothetical protein	<i>Ralstonia pickettii</i>	89%	1,00E-16	46%	WP_012435524.1
10	No significant similarity found					
11	hypothetical protein	<i>Acidovorax</i> sp.	99%	2,00E-05	42%	WP_041835882.1
12	No significant similarity found					
pMC5						
1	hypothetical protein	<i>Streptococcus parasanguinis</i>	99%	0.0	99%	ETD10351.1
2	hypothetical protein	<i>Streptococcus parasanguinis</i>	99%	9,00E-75	99%	WP_023920331.1
3	Putative ribonucleotide transport ATP-binding protein mkl	<i>Streptococcus parasanguinis</i>	98%	4,00E-43	92%	WP_023920329.1
4	hypothetical protein	<i>Streptococcus agalactiae</i>	88%	6,00E-50	75%	WP_000711035.1
5	replication initiation factor	<i>Streptococcus parasanguinis</i>	99%	0.0	97%	WP_023920333.1
6	TetR family No significant similarity found					
pMC6						
1	Relaxase	<i>Escherichia coli</i>	99%	0.0	99%	AEB22177.1
2	hypothetical protein	<i>Salmonella enterica</i>	99%	1,00E-85	100%	WP_001157192.1
3	hypothetical protein	<i>Salmonella enterica</i>	99%	1,00E-85	100%	WP_001157192.1
4	conjugal transfer protein TraL	<i>Salmonella enterica</i>	99%	0.0	100%	WP_001051252.1
5	hypothetical protein	<i>Enterobacteriaceae</i>	84%	4 E-108	100%	WP_006882065.1
6	hypothetical protein	<i>Enterobacteriaceae</i>	99%	2,00E-78	100%	WP_006882064.1
7	hypothetical protein	<i>Escherichia coli</i>	98%	4,00E-45	97%	WP_021545900.1
8	hypothetical protein	<i>Enterobacteriaceae</i>	99%	3,00E-165	99%	WP_006882056.1

Tabla S2 (continuación)

ORF	Descripción (major <i>hit</i> BLASTp en NCBI)	Organismo	Coertura	e-value	Máx ident	Número de acceso
9	hypothetical protein	<i>Enterobacteriaceae</i>	99%	0.0	99%	WP_000246370.1
10	hypothetical protein	<i>Enterobacteriaceae</i>	99%	6,00E-74	100%	WP_006882053.1
11	hypothetical protein	<i>Escherichia coli</i>	98%	3,00E-28	100%	AEB22188.1
12	hypothetical protein	<i>Enterobacteriaceae</i>	98%	2,00E-38	100%	WP_006882049.1
13	hypothetical protein	<i>Enterobacteriaceae</i>	99%	7,00E-140	100%	WP_006882046.1
14	putative conjugal transfer protein traA	<i>Salmonella enterica</i>	99%	0.0	100%	WP_000421118.1
15	hypothetical protein	<i>Enterobacteriaceae</i>	98%	1,00E-51	100%	WP_000570802.1
16	hypothetical protein	<i>Salmonella enterica</i>	71%	1,00E-11	63%	WP_024139611.1
17	plasmid replication protein	<i>Enterobacteriaceae</i>	90%	0.0	99%	WP_006882108.1
18	transcriptional regulator	<i>Enterobacteriaceae</i>	99%	2,00E-116	99%	WP_000288043.1
19	hypothetical protein	<i>Enterobacteriaceae</i>	89%	3,00E-82	100%	WP_000902259.1
20	conjugal transfer protein TraB	<i>Enterobacteriaceae</i>	99%	2,00E-164	100%	WP_000241899.1
21	conjugal transfer protein TraC	<i>Enterobacteriaceae</i>	99%	8,00E-51	100%	WP_006882103.1
22	conjugal transfer protein TraD	<i>Enterobacteriaceae</i>	99%	3,00E-78	100%	WP_000796685.1
23	conjugal transfer protein TraE	<i>Salmonella enterica</i>	97%	0.0	99%	WP_000106190.1
24	Putative conjugal transfer protein TraF	<i>Escherichia coli</i>	99%	8,00E-148	100%	AEB22148.1
25	Putative conjugal transfer protein TraF	<i>Enterobacteriaceae</i>	98%	4,00E-32	100%	WP_001128573.1
26	conjugal transfer protein traG	<i>Enterobacteriaceae</i>	99%	1,00E-171	100%	WP_000476773.1
27	conjugal transfer protein traH	<i>Enterobacteriaceae</i>	99%	0.0	100%	WP_000722128.1
28	conjugal transfer protein traI	<i>Salmonella enterica</i>	99%	0.0	99%	WP_000999413.1
29	secretion system protein E	<i>Enterobacteriaceae</i>	99%	0.0	100%	WP_001005926.1
30	conjugal transfer protein traK	<i>Salmonella enterica</i>	99%	0.0	100%	WP_001177100.1
31	arginine transportr	<i>Enterobacteriaceae</i>	99%	1,00E-64	100%	WP_000848516.1
32	hypothetical protein	<i>Enterobacteriaceae</i>	99%	8,00E-92	100%	WP_000368397.1
33	DNA topoisomerase III	<i>Salmonella enterica</i>	99%	0.0	99%	WP_000604074.1
34	DNA-binding protein	<i>Enterobacteriaceae</i>	99%	3,00E-69	100%	WP_000004210.1
35	hypothetical protein	<i>Salmonella enterica</i>	99%	8,00E-74	99%	WP_000083087.1
36	cysteinyl-tRNA synthetase	<i>Pseudomonas syringae</i>	85%	2,00E-44	95%	WP_032698189.1
37	cysteinyl-tRNA synthetase	<i>Pseudomonas resinovorans</i>	99%	8,00E-91	97%	WP_041770341.1
38	hypothetical protein	<i>Pseudomonas aeruginosa</i>	97%	6,00E-38	89%	WP_033957408.1
39	hypothetical protein	<i>Pseudomonas alcaligenes</i>	42%	2,00E-45	67%	WP_021220100.1

Tabla S2 (continuación)

ORF	Descripción (major <i>hit</i> BLASTp en NCBI)	Organismo	Coertura	e-value	Máx ident	Número de acceso
40	thiol:disulfide interchange protein	<i>Pseudomonas oleovorans</i>	94%	2,00E-52	80%	WP_037051403.1
41	thiol:disulfide interchange protein	<i>Pseudomonas oleovorans</i>	89%	2,00E-174	78%	WP_037051403.1
42	histidinol phosphatase	<i>Pseudomonas oleovorans</i>	94%	6,00E-33	84%	WP_037051219.1
43	GntR family transcriptional regulator	<i>Pseudomonas oleovorans</i>	66%	0.0	80%	WP_037051217.1
44	histidinol phosphatase	<i>Pseudomonas oleovorans</i>	98%	6,00E-135	77%	WP_037051219.1

S3. Genes de replicación plasmídica identificados *in silico*

ID en GenDB	Descripción (major hit BLASTp en NCBI)	Organismo	Cover-tura	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BF_Rep1	single-stranded DNA-binding protein	<i>Legionella pneumophila</i>	91%	5,00E-36	50%	WP_02722343 6.1	SSB	00436
BF_Rep2	single-stranded DNA-binding protein	<i>Legionella shakespearei</i>	97%	2,00E-35	49%	WP_01857713 5.1	SSB	00436
BF_Rep3	single-stranded DNA-binding protein	<i>Ottowia thiooxydans</i>	83%	9,00E-74	96%	WP_02860433 6.1	SSB	00436
BF_Rep4	single-strand DNA-binding protein	<i>Pseudomonas putida</i>	99%	1E-96	99%	WP_01310089 0.1	SSB	00436
BF_Rep5	single-stranded DNA-binding protein	<i>Ochrobactrum intermedium</i>	99%	6E-108	93%	WP_02508973 7.1	SSB	00436
BF_Rep6	single-stranded DNA-binding protein	<i>Microbacterium yannicii</i>	90%	5E-38	52%	WP_01918113 5.1	SSB	00436
BF_Rep7	single-stranded DNA-binding protein, partial	<i>Pseudomonas syringae</i>	93%	2E-20	86%	EFW77643.1	SSB	00436
BF_Rep8	single-strand DNA-binding protein	<i>Acinetobacter baumannii</i>	84%	1E-65	99%	WP_02964243 5.1	SSB	00436
BF_Rep9	single-strand DNA-binding protein	<i>Pseudomonas putida</i>	98%	7E-29	59%	WP_01310089 0.1	SSB	00436
BF_Rep10	single-stranded DNA-binding protein	<i>Salmonella enterica</i>	99%	5E-78	100%	WP_01219644 3.1	SSB	00436
BF_Rep11	initiator RepB protein	<i>Providencia alcalifaciens</i>	88%	1,00E-99	64%	EUD00416.1	Rep_3	01051
BF_Rep12	initiator RepB protein	<i>Salmonella enterica</i>	95%	0.0	99%	WP_00081086 6.1	Rep_3	01051
BF_Rep13	hypothetical protein	<i>Pseudomonas putida</i>	99%	8,00E-180	100%	WP_01115437 2.1	Rep_3	01051
BF_Rep14	hypothetical protein	<i>Pseudomonas taiwanensis</i>	99%	0.0	99%	ESW38296.1	Rep_3	01051
BF_Rep15	initiator RepB protein	<i>Salmonella enterica</i>	90%	0.0	99%	WP_00081086 6.1	Rep_3	01051
BF_Rep16	hypothetical protein	<i>Pseudomonas taiwanensis</i>	94%	0.0	71%	ESW38296.1	Rep_3	01051

Tabla S3 (continuación)

ID en GenDB	Descripción (major hit BLASTp en NCBI)	Organismo	Cover-tura	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BF_Rep17	replication initiation factor	<i>Acidovorax citrulli</i>	86%	4,00E-104	56%	ABM32965.1	Rep_trans	02486
BF_Rep18	hypothetical protein	<i>Streptococcus parasanguinis</i>	99%	0.0	97%	WP_023920333.1	Rep_trans	02486
BF_Rep19	replication initiation factor	<i>Acidovorax delafieldii</i>	97%	2E-65	48%	WP_005794747.1	Rep_trans	02486
BF_Rep20	hypothetical protein	<i>Streptococcus parasanguinis</i>	99%	0.0	97%	WP_023920333.1	Rep_trans	02486
BF_Rep21	replication initiation factor	<i>Acidovorax citrulli</i>	86%	4E-104	56%	WP_011795501.1	Rep_trans	02486
BF_Rep22	RepA: plasmid replication protein	uncultured bacterium	91%	9,00E-152	62%	WP_031943359.1	RepA_C	04796
BF_Rep23	RepA: plasmid replication protein	uncultured bacterium	90%	1,00E-115	53%	WP_031943359.1	RepA_C	04796
BF_Rep24	replication initiation protein	<i>Enterobacteriaceae</i>	99%	0.0	100%	WP_012196453.1	RepA_C	04796
BF_Rep25	transcriptional regulator	<i>Pseudomonas aeruginosa</i>	99%	1E-103	100%	WP_011205825.1	KfrA_N	11740
BF_Rep26	hypothetical protein	<i>Pseudomonas putida</i>	90%	4E-58	46%	WP_004574783.1	KfrA_N	11740
BF_Rep27	hypothetical protein	<i>Serratia marcescens</i>	99%	9E-79	89%	KFF80982.1	KfrA_N	11740
BF_Rep28	KfrA	<i>Salmonella enterica</i>	99%	1E-132	100%	WP_012443523.1	KfrA_N	11740
BF_Rep29	hypothetical protein	<i>Pseudomonas</i> sp.	99%	3E-171	100%	WP_004577307.1	KfrA_N	11740
BF_Rep30	Rep	<i>Bacillus pumilus</i>	99%	7,00E-162	94%	AAB71488.1	Rep_1	01446
BF_Rep31	rep	<i>Bacillus pumilus</i>	98%	5,00E-180	98%	WP_011116799.1	Rep_1	01446
BF_Rep32	replication protein A	<i>Pseudomonas</i> sp.	96%	8,00E-49	90%	WP_032881486.1	Replicase	03090
BF_Rep33	RepA	<i>Aeromonas hydrophila</i>	99%	2,00E-86	99%	WP_014343694.1	Replicase	03090
BF_Rep34	replication protein C	<i>Ochrobactrum rhizosphaerae</i>	99%	0.0	88%	WP_024899937.1	RP-C	03428
BF_Rep35	Replication protein C	<i>Sinorhizobium meliloti</i>	99%	0.0	94%	WP_015241539.1	RP-C	03428

Tabla S3 (continuación)

ID en GenDB	Descripción (major hit BLASTp en NCBI)	Organismo	Cover-tura	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BF_Rep36	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	98%	9,00E-93	94%	WP_02489992 0.1	RP-C	03428]
BF_Rep37	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	87%	2E-80	95%	WP_03656839 1.1	RP-C	03428
BF_Rep38	replication initiation protein	<i>Ochrobactrum anthropi</i>	99%	0.0	83%	WP_01198293 4.1	RP-C_C	11800
BF_Rep39	replication initiation protein	<i>Ochrobactrum anthropi</i>	99%	0.0	89%	WP_01198287 8.1	RP-C_C	11800
BF_Rep40	plasmid replication protein RepC	<i>Ochrobactrum rhizosphaerae</i>	99%	0.0	99%	WP_02489795 4.1	RP-C_C	11800
BF_Rep41	replication initiation protein	<i>Sinorhizobium meliloti</i>	99%	0.0	94%	WP_01727086 8.1	RP-C_C	11800
BF_Rep42	replication initiation protein	<i>Ochrobactrum anthropi</i>	99%	2,00E-139	88%	WP_01198287 8.1	RP-C_C	11800
BF_Rep43	replication protein C	<i>Pseudomonas aeruginosa</i>	96%	0.0	96%	ETD44099.1	RepC	06504
BF_Rep44	RepA	<i>Klebsiella pneumoniae</i>	96%	2,00E-113	79%	WP_01245710 7.1	IncFII_rep A	02387
BF_Rep45	DNA replication terminus site-binding protein	<i>Escherichia coli</i>	97%	2E-87	99%	WP_00144458 7.1	Ter	05472
BF_Rep46	replication terminus site binding protein	<i>Pseudomonas sp.</i>	99%	0.0	100%	WP_01107791 3.1	Ter	05472
BF_Rep47	TrfA family protein	<i>Thauera terpenica</i>	91%	6,00E-143	68%	WP_02125058 8.1	trfA	07042
BF_Rep48	hypothetical protein	<i>Pseudomonadales</i>	99%	0.0	100%	WP_00108227 9.1	trfA	07042
BF_Rep49	putative replicating protein RepA	<i>Agrobacterium tumefaciens</i>	100%	2,00E-173	100%	CDN96155.1	HTH_36	13730
BF_Rep50	Rep	<i>Ochrobactrum intermedium</i>	100%	2,00E-137	99%	WP_00647347 6.1	HTH_36	13730
BF_Rep51	replication protein	<i>Ochrobactrum rhizosphaerae</i>	99%	0.0	80%	WP_02489992 0.1	HTH_36	13730

Tabla S3 (continuación)

ID en GenDB	Descripción (major hit BLASTp en NCBI)	Organismo	Cover-tura	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BF_Rep52	putative Rep	Plasmid pM3	99%	1,00E-119	100%	AAD46123.1	HTH_36	13730
BF_Rep53	plasmid replication protein RepB	<i>Aeromonas hydrophila</i>	71%	6,00E-24	87%	WP_02494412 4.1		Sin dominios putativos conservados
BF_Rep54	hypothetical protein	<i>Bacillus cereus</i>	100%	3,00E-155	99%	WP_00017017 2.1		Sin dominios putativos conservados

S4. Genes de partición y estabilización plasmídica identificados *in silico*

ID en GenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_459	cobyrinic acid a,c-diamide synthase	<i>Nitrosomonas eutropha</i>	1E-104	65%	WP_011630666.1		
BJP_247	hypothetical protein	<i>Paenibacillus</i> sp.	9,00E-176	94%	GAK43422.1		
BJP_9586	partitioning protein	<i>Salmonella enterica</i>	7,00E-131	91%	YP_002332254.1		
BJP_65	partitioning protein	<i>Pseudomonas</i>	0.0	83%	WP_004574780.1		
BJP_5270	hypothetical protein	<i>Sinorhizobium meliloti</i>	5,00E-143	95%	WP_032492664.1		
AP_397	chromosome partitioning protein	<i>Ochrobactrum rhizosphaerae</i>	2,00E-163	97%	WP_024899911.1		
BJP_928	partitioning protein	<i>Pseudomonas</i>	0.0	100%	WP_004574780.1		
BJP_194	cobyrinic acid a,c-diamide synthase	<i>Nitrosomonas eutropha</i>	8,00E-120	72%	WP_011630666.1		
AP_16	chromosome partitioning protein ParA	<i>Ochrobactrum anthropi</i>	0.0	95%	EXL01616.1		
BJP_882	ParAF protein	<i>Frankia</i> sp.	4,00E-47	47%	WP_032490082.1		
BJP_6594	chromosome partitioning protein	<i>Cellulosimicrobium cellulans</i>	0.0	90%	WP_024841573.1	CbiA	01656
BJP_1333	chromosome partitioning protein ParA	<i>Ochrobactrum intermedium</i>	0.0	96%	WP_025091213.1		
BJP_784	plasmid replication protein	<i>Thauera terpenica</i>	2,00E-95	60%	WP_021250593.1		
AP_413	chromosome partitioning protein ParA	<i>Ochrobactrum rhizosphaerae</i>	0.0	99%	WP_024899918.1		
AP_462	RepA	<i>Sinorhizobium</i> sp.	0.0	99%	WP_015647669.1		
BJP_1253	hypothetical protein	<i>Klebsiella oxytoca</i>	4,00E-151	97%	WP_023321975.1		
AP_658	IncC partitioning protein, short form	<i>Pseudomonas aeruginosa</i>	0.0	100%	WP_011645034.1		
BJP_3797	hypothetical protein	<i>Pseudomonas veronii</i>	2,00E-63	88%	WP_032804263.1		
BJP_361	putative ParA	Plasmid pM3	6,00E-161	99%	AAD46128.1		
BJP_1033	StbB	<i>Enterobacteriaceae</i>	8,00E-167	100%	WP_012443534.1		
BJP_384	plasmid stabilization protein	<i>Pseudomonas putida</i>	2,00E-171	99%	WP_028691950.1		

Tabla S.4 (continuación)

ID en GenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_2596	chromosome partitioning protein ParA	<i>Mesorhizobium alhagi</i>	0.0	76%	WP_008840498.1		
BJP_624	hypothetical protein	<i>Pseudomonas putida</i>	0.0	100%	WP_011154370.1		
BJP_201	hypothetical protein	<i>Nitrosomonas eutropha</i>	3,00E-146	85%	WP_011630708.1		
BJP_5126	plasmid stability protein StbB	<i>Pseudomonas</i> sp.	2,00E-120	70%	WP_032881477.1		
BJP_3572	putative plasmid stability/partitioning protein	<i>Providencia stuartii</i>	4,00E-98	71%	WP_036944392.1		
BJP_732	hypothetical protein	<i>Nitrosomonas eutropha</i>	6,00E-125	79%	WP_011630708.1		
BJP_554	hypothetical protein	<i>Burkholderia phytofirmans</i>	2,00E-73	52%	WP_012430992.1		
AP_110	chromosome partitioning protein	<i>Ochrobactrum anthropi</i>	4,00E-177	97%	WP_036588675.1		
BJP_6544	hypothetical protein	<i>Cellulosimicrobium cellulans</i>	9,00E-105	97%	WP_034605270.1		
BJP_1829	plasmid stabilization protein	<i>Photorhabdus luminescens</i>	6,00E-44	81%	WP_036780593.1		
BJP_3240	hypothetical protein	<i>Ochrobactrum</i> sp.	3,00E-55	87%	WP_021585839.1		
	plasmid stabilization protein ParE	<i>Sinorhizobium meliloti</i>	3,00E-31	55%	WP_017270034.1		
AP_402	RelE toxin	<i>Ochrobactrum</i>	3,00E-62	100%	WP_011982866.1		
BJP_1830	hypothetical protein	<i>Photorhabdus luminescens</i>	2,00E-77	82%	WP_036780591.1	ParE_toxin	05016
BJP_8969	hypothetical protein	<i>Xanthomonas campestris</i>	3,00E-60	99%	WP_010378961.1		
BJP_8847	RelE toxin	<i>Aeromonas hydrophila</i>	4,00E-55	97%	WP_017411297.1		
BJP_2512	hypothetical protein	<i>Pseudomonas aeruginosa</i>	2,00E-31	96%	WP_023102300.1		
BJP_5251	hypothetical protein	<i>Serratia marcescens</i>	4,00E-21	88%	WP_033645615.1		
BJP_8968	hypothetical protein	<i>Proteobacteria</i>	7,00E-24	100%	WP_010378963.1		
BJP_8969	hypothetical protein	<i>Xanthomonas campestris</i>	3,00E-60	99%	WP_010378961.1	PtaRNA1	12703
BJP_360	putative parB	Plasmid pM3	0.0	98%	AAD46127.1		
pMC6_9	hypothetical protein	<i>Escherichia coli</i>	0.0	99%	AEB22186.1		
BJP_395	hypothetical protein	<i>Escherichia coli</i>	0.0	99%	AEB22186.1	ParBc	02195
BJP_64	parB-like partition protein	<i>Pseudomonas</i>	8,00E-165	69%	WP_023662218.1		

Tabla S.4 (continuación)

ID en GenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
AP_783	chromosome partitioning protein ParB	<i>Sinorhizobium fredii</i>	0.0	95%	WP_014330980.1		
BJP_257	parB-like partition protein	<i>Acidovorax citrulli</i>	1,00E-133	46%	WP_011794736.1		
96	parB-like partition protein	<i>Acidovorax citrulli</i>	1,00E-133	46%	WP_011794736.1		
BJP_96	hypothetical protein	<i>Pseudomonas</i>	4,00E-112	56%	WP_023662083.1		
BJP_1833	hypothetical protein	<i>Xenophilus azovorans</i>	0.0	56%	WP_038215288.1		
AP_461	chromosome partitioning protein ParB	<i>Ochrobactrum rhizosphaerae</i>	0.0	99%	WP_024897953.1		
BJP_1332	chromosome partitioning protein ParB	<i>Ochrobactrum rhizosphaerae</i>	0.0	99%	WP_024897953.1		
AP_17	replication protein B	<i>Ochrobactrum anthropi</i>	0.0	87%	WP_011983024.1		
BJP_3886	plasmid stabilization protein	<i>Sinorhizobium meliloti</i>	1,00E-60	80%	CDH80775.1		
BJP_2597	chromosome partitioning protein ParB	<i>Mesorhizobium alhagi</i>	2,00E-106	54%	WP_040589407.1		
BJP_192	parB-like partition proteins	<i>Nitrosomonas eutropha</i>	9,00E-161	71%	WP_011630668.1		
BJP_461	parB-like partition proteins	<i>Nitrosomonas eutropha</i>	1,00E-175	70%	WP_011630668.1		
AP_412	chromosome partitioning protein ParB	<i>Ochrobactrum rhizosphaerae</i>	0.0	98%	WP_024899919.1		
BJP_783	plasmid-partitioning protein ParB	<i>Xylella fastidiosa</i>	3,00E-67	50%	WP_010895251.1		
BJP_929	chromosome partitioning protein ParB	<i>Pseudomonas</i>	0.0	100%	WP_023662218.1		
BJP_619	putative replication/partitioning-related protein	<i>Pseudomonas putida</i>	0.0	100%	WP_011600728.1		
BJP_1227	putative chromosome partitioning protein	<i>Pseudomonas putida</i>	0.0	99%	WP_011600744.1		
AP_48	peptide transporter	<i>Rhizobiales</i>	0.0	93%	WP_011982954.1		
AP_249	ParBC	<i>Sinorhizobium</i> sp.	0.0	94%	WP_015647731.1		
BJP_2957	Rac prophage; conserved protein	<i>Escherichia coli</i>	1,00E-56	100%	WP_001350510.1		
BJP_5052	hypothetical protein	<i>Cellulosimicrobium cellulans</i>	0.0	87%	WP_034605271.1		

Tabla S.4 (continuación)

ID en GenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_360	putative ParB	Plasmid pM3	0.0	98%	AAD46127.1		
BJP_96	hypothetical protein	<i>Pseudomonas</i>	4,00E-112	56%	WP_023662083.1		
BJP_783	plasmid partitioning protein parB	<i>Xylella fastidiosa</i>	3,00E-67	50%	WP_010895251.1		
BJP_192	parB-like partition proteins chromosome	<i>Nitrosomonas eutropha</i>	9,00E-161	71%	WP_011630668.1	KorB	08535
BJP_929	partitioning protein ParB	<i>Pseudomonas</i>	0.0	100%	WP_023662218.1		
BJP_623	putative partitioning protein	<i>Pseudomonas putida</i>	0.0	100%	WP_011600726.1		
BJP_1252	hypothetical protein	<i>Hafnia paralvei</i>	0.0	83%	WP_039187478.1		
BJP_39	hypothetical protein	<i>Pseudomonas</i> sp.	0.0	74%	WP_039614031.1		
BJP_2857	hypothetical protein	<i>Pseudomonas</i> sp.	0.0	100%	WP_039614031.1		
BJP_10603	recombinase	<i>Yokenella regensburgei</i>	2,00E-98	95%	WP_038255478.1	StbA	06406
BJP_5808	recombinase	<i>Kosakonia radicincitans</i>	1,00E-84	93%	WP_007374358.1		
BJP_5697	recombinase	<i>Enterobacter cloacae</i>	1,00E-45	94%	AHE72662.1		
BJP_805	DNA-binding protein, hns family	<i>Enterobacteriaceae</i>	2,00E-59	99%	WP_012414174.1	Histone_HNS	00816
BJP_617	putative transcriptional regulator protein	<i>Pseudomonas putida</i>	1,00E-80	100%	WP_011600730.1		

S5. Genes de movilidad plasmídica identificados *in silico*

ID en DenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_441	hypothetical protein	<i>Yersinia ruckeri</i>	0.0	73%	WP_004720 847.1		
BJP_1498	hypothetical protein	<i>Sphingobium</i> sp.	0.0	100%	KFL47726.1		
pMC6_1	relaxase	<i>Escherichia coli</i>	0.0	99%	AEB22177.1		
BJP_387	relaxase	<i>Escherichia coli</i>	0.0	99%	AEB22177.1		
AP_669	Tral	Cloning vector pDMK3	3,00E- 97	99%	ACO50702.1		
BJP_2740	conjugal transfer protein TraA	<i>Sinorhizobium</i> <i>fredii</i>	0.0	91%	WP_014330 857.1		
BJP_237	hypothetical protein	<i>Paenibacillus</i> sp.	0.0	85%	GAK43394.1		
BJP_198	hypothetical protein	<i>Nitrosomonas</i> <i>eutropha</i>	0.0	79%	WP_011630 664.1	Relaxase	03432
BJP_634	endonuclease	<i>Rhodobacterac</i> <i>eae bacterium</i>	3,00E- 34	33%	WP_023849 101.1		
AP_394	relaxase	<i>Ochrobactrum</i> <i>anthropi</i>	0.0	94%	WP_011982 859.1		
BJP_847	hypothetical protein (relaxase)	<i>Rhodococcus</i> <i>fascians</i>	2,00E- 70	45%	WP_032393 683.1		
AP_668	plasmid conjugative transfer protein	Shuttle vector pMTL82151	7,00E- 59	100%	ACR43888.1		
AP_113	hypothetical protein	<i>Agrobacterium</i> <i>tumefaciens</i>	0.0	95%	WP_035226 598.1		
BJP_735	hypothetical protein	<i>Nitrosomonas</i> <i>eutropha</i>	3,00E- 07	86%	WP_011630 664.1		
BJP_796	Trwl	<i>Salmonella</i> <i>enterica</i>	1,00E- 135	99%	WP_012443 542.1		
BJP_369	type VI secretion protein	<i>Pseudomonas</i> <i>mosselii</i>	1,00E- 174	99%	WP_028691 936.1		
BJP_4586	conjugal transfer protein TrbL	<i>Sinorhizobium</i> <i>medicae</i>	0.0	99%	WP_018210 781.1		
pMC6_14	conjugal transfer protein traA	<i>Salmonella</i> <i>enterica</i>	0.0	100%	WP_000421 118.1	TrbL	04610
BJP_400	conjugal transfer protein traA	<i>Salmonella</i> <i>enterica</i>	0.0	100%	WP_000421 118.1		
AP_166	conjugal transfer protein TrbL	<i>Ochrobactrum</i> <i>anthropi</i>	0.0	99%	WP_011982 853.1		
AP_469	TrbL	<i>Sinorhizobium</i> sp.	1,00E- 169	98%	WP_015647 759.1		
AP_94	TrbL/VirB6 plasmid conjugal transfer protein	<i>Rhizobium</i> sp.	0.0	91%	WP_022557 206.1		

Tabla S.5 (continuación)

ID en DenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_724	TrbL/VirB6 plasmid conjugal transfer protein	<i>Nitrosomonas eutropha</i>	0.0	78%	WP_011630700.1		
BJP_209	TrbL/VirB6 plasmid conjugal transfer protein	<i>Nitrosomonas eutropha</i>	3,00E-158	63%	WP_011630700.1		
BJP_942	TrbL protein	<i>Proteobacteria</i>	3,00E-132	100%	WP_011205794.1		
AP_672	conjugal transfer protein TrbL	<i>Ochrobactrum rhizosphaerae</i>	3,00E-175	90%	WP_024896844.1		
BJP_569	hypothetical protein	<i>Xenophilus azovorans</i>	4,00E-39	42%	WP_038214259.1		
BJP_4240	No significant similarity found.						
BJP_798	TrwJ	<i>Salmonella enterica</i>	3,00E-160	99%	WP_012443544.1		
BJP_366	hypothetical protein	<i>Pseudomonas mosselii</i>	5,00E-117	98%	WP_036986392.1		
AP_165	conjugal transfer protein	<i>Ochrobactrum anthropi</i>	4,00E-147	99%	WP_011982852.1		
BJP_4588	conjugal transfer protein	<i>Sinorhizobium medicae</i>	2,00E-121	98%	WP_018210782.1		
BJP_941	conjugal transfer protein TrbJ	<i>Proteobacteria</i>	2,00E-105	100%	WP_011205792.1		
BJP_567	hypothetical protein	<i>Xenophilus azovorans</i>	4,00E-67	53%	WP_038214266.1		
pMC6_24	Putative conjugal transfer protein TraF	<i>Escherichia coli</i>	7,00E-148	100%	AEB22148.1		
BJP_408	Putative conjugal transfer protein TraF	<i>Escherichia coli</i>	3,00E-156	99%	AEB22148.1		
BJP_726	type IV secretion system family protein	<i>Nitrosomonas eutropha</i>	6,00E-129	75%	WP_011630702.1	T4SS	07996
AP_92	conjugal transfer protein	<i>Ochrobactrum anthropi</i>	9,00E-166	94%	WP_036588726.1		
BJP_207	type IV secretion system family protein	<i>Nitrosomonas eutropha</i>	1,00E-110	75%	WP_011630702.1		
BJP_4587	hypothetical protein	<i>Sinorhizobium medicae</i>	2,00E-35	98%	WP_011971091.1		
BJP_1012	conjugal transfer protein	<i>Sinorhizobium meliloti</i>	4,00E-09	95%	WP_028004229.1		
BJP_797	Eex (entry exclusion)	<i>Salmonella enterica</i>	3,00E-48	100%	WP_012443543.1		
BJP_1010	lytic transglycosylase	<i>Sinorhizobium medicae</i>	0.0	99%	WP_018210783.1		
AP_467	TrbJ	<i>Sinorhizobium</i> sp.	5,00E-163	96%	WP_015647760.1		
AP_502	P-type DNA transfer protein VirB5	<i>Ochrobactrum</i> sp.	5,00E-109	86%	WP_029928324.1		

Tabla S.5 (continuación)

ID en DenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_1038	conjugal transfer protein	<i>Providencia rettgeri</i>	0.0	100%	WP_012414 161.1		
BJP_91	general secretion pathway protein E	<i>Pseudomonas putida</i>	0.0	64%	WP_020190 101.1		
BJP_2881	P-type DNA transfer ATPase VirB11	<i>Sinorhizobium meliloti</i>	0.0	97%	WP_015241 453.1		
BJP_373	type VI secretion protein	<i>Pseudomonas mosselii</i>	0.0	97%	WP_028691 940.1		
BJP_574	hypothetical protein	<i>Xenophilus azovorans</i>	9,00E-138	59%	WP_038200 604.1		
pMC6_29	secretion system protein E	<i>Enterobacteriac eae</i>	0.0	100%	WP_001005 926.1		
BJP_412	secretion system protein E	<i>Enterobacteriac eae</i>	0.0	100%	WP_001005 926.1		
BJP_713	type II/IV secretion system family protein	<i>Yersinia ruckeri</i>	0.0	61%	KGA44905.1		
BJP_719	type II secretion system protein E	<i>Nitrosomonas eutropha</i>	0.0	84%	WP_011630 696.1		
BJP_705	conjugal transfer protein	<i>Yersinia ruckeri</i>	0.0	85%	WP_004720 814.1		
AP_170	type VI secretion protein	<i>Ochrobactrum anthropi</i>	0.0	98%	WP_011982 857.1		
AP_677	type IV secretion protein VirB11	<i>Ochrobactrum intermedium</i>	0.0	86%	WP_025091 591.1	T2SSE	00437
AP_463	conjugal transfer protein TrbB	<i>Rhizobium giardinii</i>	0.0	99%	WP_018326 380.1		
BJP_214	type II secretion system protein E	<i>Nitrosomonas eutropha</i>	0.0	90%	WP_011630 696.1		
BJP_4489	general secretory pathway component, cryptic	<i>Enterobacteriac eae</i>	0.0	100%	WP_001219 894.1		
BJP_597	phytochrome sensor protein	<i>Pseudomonas</i>	0.0	99%	WP_023662 235.1		
AP_6	Phosphonate ABC transporter ATP-binding protein	<i>Rhizobiales</i>	0.0	100%	WP_006473 175.1		
AP_98	type VI secretion protein	<i>Ochrobactrum anthropi</i>	0.0	98%	WP_036588 717.1		
BJP_4418	type II secretion system protein E	<i>Ottowia thiooxydans</i>	0.0	90%	WP_028605 692.1		
BJP_8409	conjugal transfer ATPase TrbB	<i>Proteobacteria</i>	0.0	100%	WP_011205 784.1		
BJP_4419	protein transporter hofB	<i>Escherichia coli</i>	0.0	100%	WP_016241 104.1		
BJP_374	putative mating pair formation protein	<i>Pseudomonas mosselii</i>	0.0	91%	WP_028691 941.1		

Tabla S.5 (continuación)

ID en DenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_2966	phosphate ABC transporter ATP-binding protein	<i>Pseudomonas</i>	0.0	99%	WP_003118 432.1		
BJP_2882	conjugal transfer protein TraG	<i>Sinorhizobium meliloti</i>	3,00E-79	97%	WP_017265 125.1		
AP_223	PhnC	<i>Sinorhizobium</i> sp.	0.0	100%	WP_015647 706.1		
BJP_9333	twitching motility protein PilT	<i>Ottowia thiooxydans</i>	1,00E-69	96%	WP_028601 942.1		
BJP_9603	hypothetical protein	<i>Serratia marcescens</i>	2,00E-110	99%	WP_016927 991.1		
AP_742	ABC transporter related protein	<i>Acidovorax</i> sp.	1,00E-156	85%	EHL22328.1		
BJP_2707	ABC transporter	<i>Comamonas</i> sp.	0.0	81%	WP_027011 233.1		
BJP_10165	cell division protein FtsE	<i>Paenibacillus</i>	2,00E-164	99%	WP_024631 580.1		
BJP_4134	sugar ABC transporter ATP-binding protein	<i>Delftia</i> sp.	4,00E-106	65%	WP_034364 857.1		
BJP_10723	branched-chain amino acid ABC transporter ATP-binding protein	<i>Acidovorax</i> sp.	9,00E-68	67%	WP_008905 448.1		
BJP_5636	nickel ABC transporter ATPase	<i>Enterobacteriac eae</i>	0.0	100%	WP_000173 631.1		
BJP_5231	conjugal transfer protein TraG	<i>Sinorhizobium</i> sp.	0.0	96%	WP_037425 374.1		
AP_171	conjugal transfer protein TraG	<i>Ochrobactrum rhizosphaerae</i>	0.0	97%	WP_024899 892.1		
BJP_9254	conjugal transfer protein TraG	<i>Paracoccus</i> sp.	1,00E-171	73%	WP_010399 298.1		
pMC6_30	conjugal transfer protein TraK	<i>Salmonella enterica</i>	0.0	100%	WP_001177 100.1		
BJP_413	conjugal transfer protein TraK	<i>Salmonella enterica</i>	0.0	100%	WP_001177 100.1		
BJP_718	TRAG family protein	<i>Nitrosomonas eutropha</i>	0.0	90%	WP_011630 695.1	T4SS- DNA_transf	02534
BJP_215	TRAG family protein	<i>Nitrosomonas eutropha</i>	0.0	91%	WP_011630 695.1		
BJP_231	hypothetical protein	<i>Paenibacillus</i> sp.	0.0	92%	GAK43401.1		
BJP_7004	conjugal transfer protein TraG	<i>Agrobacterium tumefaciens</i>	0.0	94%	WP_011199 111.1		
BJP_744	hypothetical protein	<i>Leifsonia aquatica</i>	0.0	62%	WP_039921 200.1		
AP_262	conjugal transfer protein TraG	<i>Ochrobactrum rhizosphaerae</i>	0.0	97%	WP_024897 928.1		

Tabla S.5 (continuación)

ID en DenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_2882	conjugal transfer protein TraG	<i>Sinorhizobium meliloti</i>	3,00E-79	97%	WP_017265 125.1		
BJP_2000	hypothetical protein	<i>Sphingobium</i>	1,00E-76	85%	WP_020817 825.1		
AP_99	conjugal transfer protein TraG	<i>Ochrobactrum anthropi</i>	0.0	89%	WP_036588 715.1		
BJP_630	conjugal transfer protein TraG	<i>Ochrobactrum rhizosphaerae</i>	1,00E-28	75%	WP_024899 892.1		
BJP_7005	hypothetical protein	<i>Rhizobium</i> sp.	3,00E-16	65%	WP_018900 447.1		
BJP_1499	conjugal transfer coupling protein TraG	<i>Proteobacteria</i>	0.0	100%	WP_011205 817.1		
BJP_7753	hypothetical protein	<i>Bacillus</i> sp.	0.0	92%	WP_021483 182.1		
BJP_381	trwB protein	<i>Pseudomonas putida</i> group	0.0	99%	WP_028691 947.1		
BJP_1036	TrwB	<i>Salmonella enterica</i>	0.0	100%	WP_012443 537.1		
BJP_137	conjugal transfer protein TraG	<i>Pseudomonas</i> sp.	0.0	88%	WP_039614 010.1		
BJP_1262	conjugal transfer protein TraG	<i>Pseudomonas</i>	0.0	100%	WP_028699 870.1	TrwB_AAD_bind	10412
BJP_558	hypothetical protein	<i>Xenophilus azovorans</i>	0.0	64%	WP_038213 399.1		
BJP_136	hypothetical protein	<i>Pseudomonas</i>	4,00E-144	53%	WP_023662 248.1		
BJP_2265	hypothetical protein	<i>Pseudomonas</i>	0.0	100%	WP_023662 248.1		
BJP_794	TrwG	<i>Salmonella enterica</i>	5,00E-170	100%	WP_012443 541.1		
BJP_370	mating pair formation protein	<i>Pseudomonas mosselii</i>	5,00E-158	96%	WP_028691 937.1		
BJP_722	VirB8 family protein	<i>Nitrosomonas eutropha</i>	3,00E-103	66%	WP_011630 699.1		
pMC6_26	conjugal transfer protein	<i>Enterobacteriac eae</i>	1,00E-171	100%	WP_000476 773.1		
BJP_409	conjugal transfer protein	<i>Enterobacteriac eae</i>	1,00E-171	100%	WP_000476 773.1	VirB8	04335
AP_95	type VI secretion protein	<i>Agrobacterium tumefaciens</i>	4,00E-157	96%	WP_035226 642.1		
BJP_211	VirB8 family protein	<i>Nitrosomonas eutropha</i>	8,00E-128	81%	WP_011630 699.1		
BJP_571	hypothetical protein	<i>Xenophilus azovorans</i>	2,00E-80	51%	WP_038214 256.1		
AP_470	TrbF	<i>Sinorhizobium</i> sp.	5,00E-156	99%	WP_015647 758.1		

Tabla S.5 (continuación)

ID en DenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
AP_674	conjugal transfer protein TraJ	<i>Ochrobactrum anthropi</i>	2,00E-126	92%	WP_010660 269.1		
BJP_792	conjugal transfer protein	<i>Enterobacteriac eae</i>	5,00E-21	100%	WP_012414 163.1		
BJP_795	trwH	<i>Enterobacteriac eae</i>	2,00E-24	98%	WP_012196 427.1		
BJP_937	conjugal transfer protein TrbF	<i>Proteobacteria</i>	2,00E-167	100%	WP_011205 788.1		
BJP_2879	conjugal transfer protein TrbG	<i>Sinorhizobium</i> sp.	8,00E-170	97%	WP_037425 383.1		
BJP_1040	conjugal transfer protein	<i>Enterobacteriac eae</i>	6,00E-138	100%	WP_012414 163.1		
BJP_371	hypothetical protein	<i>Pseudomonas mosselii</i>	0.0	99%	WP_028691 938.1		
AP_168	conjugal transfer protein TrbG	<i>Ochrobactrum rhizosphaerae</i>	8,00E-156	97%	WP_024899 889.1		
pMC6_27	conjugal transfer protein TraH	<i>Enterobacteriac eae</i>	0.0	100%	WP_000722 128.1		
BJP_410	conjugal transfer protein TraH	<i>Enterobacteriac eae</i>	0.0	100%	WP_000722 128.1		
BJP_721	conjugal transfer protein TrbG/VirB9/CagX	<i>Nitrosomonas eutropha</i>	7,00E-118	72%	WP_011630 698.1	CagX	03524
BJP_212	conjugal transfer protein TrbG/VirB9/CagX	<i>Nitrosomonas eutropha</i>	9,00E-136	82%	WP_011630 698.1		
AP_471	Conjugal transfer protein TrbG	<i>Rhizobium giardinii</i>	4,00E-177	99%	WP_026203 418.1		
BJP_572	hypothetical protein	<i>Xenophilus azovorans</i>	5,00E-97	58%	WP_038214 254.1		
AP_96	Conjugal transfer protein TrbG	<i>Ochrobactrum anthropi</i>	0.0	97%	EXL02641.1		
AP_675	type IV secretion protein VirB9	<i>Ochrobactrum</i>	2,00E-159	88%	WP_006468 255.1		
BJP_938	conjugal transfer protein TrbG	<i>Proteobacteria</i>	0.0	100%	WP_011645 028.1		
BJP_792	conjugal transfer protein	<i>Enterobacteriac eae</i>	5,00E-21	100%	WP_012414 163.1		
BJP_799	TrwK	<i>Salmonella enterica</i>	0.0	100%	WP_012443 545.1		
BJP_365	putative mating pair formation protein	<i>Pseudomonas putida</i>	0.0	97%	WP_011005 966.1		
BJP_1009	type VI secretion protein	<i>Sinorhizobium medicae</i>	0.0	99%	WP_018210 784.1	CagE_TrbE_VirB	03135
AP_163	type VI secretion protein	<i>Ochrobactrum anthropi</i>	0.0	98%	WP_011982 850.1		
BJP_727	type IV secretion/conjugal transfer ATPase	<i>Nitrosomonas eutropha</i>	0.0	89%	WP_011630 703.1		

Tabla S.5 (continuación)

ID en DenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_407	hypothetical protein	<i>Escherichia coli</i>	0.0	99%	EQV00987.1		
pMC6_23	conjugal transfer protein traE type IV	<i>Salmonella enterica</i>	0.0	99%	WP_000106190.1		
BJP_206	secretion/conjugal transfer ATPase	<i>Nitrosomonas eutropha</i>	0.0	93%	WP_011630703.1		
AP_466	conjugal transfer protein TrbE	<i>Agrobacterium tumefaciens</i>	0.0	98%	WP_035227071.1		
BJP_566	hypothetical protein	<i>Xenophilus azovorans</i>	0.0	47%	WP_038214269.1		
AP_90	ATPase	<i>Ochrobactrum anthropi</i>	0.0	98%	WP_036588729.1		
AP_503	transporter	<i>Ochrobactrum anthropi</i>	0.0	88%	WP_010660274.1		
BJP_936	TrbE protein	<i>Ralstonia pickettii</i>	0.0	89%	YP_006963073.1		
BJP_1037	TrwC	<i>Salmonella enterica</i>	0.0	99%	WP_012443538.1	TrwC	08751
BJP_380	ATPase AAA	<i>Pseudomonas mosselii</i>	0.0	99%	WP_028691946.1		
BJP_1039	conjugal transfer protein	<i>Providencia rettgeri</i>	0.0	99%	WP_012414162.1		
BJP_2880	conjugal transfer protein Trbl	<i>Sinorhizobium medicae</i>	0.0	99%	WP_011971095.1		
AP_169	conjugal transfer protein Trbl	<i>Ochrobactrum anthropi</i>	0.0	95%	WP_011982856.1		
AP_97	conjugal transfer protein Trbl	<i>Ochrobactrum anthropi</i>	0.0	92%	WP_036588719.1		
pMC6_28	conjugal transfer protein tral	<i>Salmonella enterica</i>	0.0	99%	WP_000999413.1		
BJP_411	conjugal transfer protein tral	<i>Salmonella enterica</i>	0.0	99%	WP_000999413.1		
BJP_372	putative mating pair formation protein	<i>Pseudomonas mosselii</i>	0.0	97%	WP_028691939.1	Trbl	03743
BJP_573	hypothetical protein	<i>Xenophilus azovorans</i>	1,00E-76	59%	WP_038200862.1		
BJP_143	hypothetical protein	<i>Pseudomonas putida</i>	2,00E-165	65%	WP_016486961.1		
BJP_720	conjugation Trbl family protein	<i>Nitrosomonas eutropha</i>	7,00E-171	72%	WP_011630697.1		
AP_676	type IV secretion protein VirB10	<i>Shinella zoogloeoides</i>	0.0	78%	WP_023517144.1		
BJP_213	conjugation Trbl family protein	<i>Nitrosomonas eutropha</i>	1,00E-174	82%	WP_011630697.1		
BJP_1268	conjugal transfer protein TraB	<i>Pseudomonas taiwanensis</i>	0.0	100%	WP_023662243.1		

Tabla S.5 (continuación)

ID en DenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_940	conjugal transfer protein TrbI	<i>Proteobacteria</i>	0.0	100%	WP_011205791.1		
BJP_647	proQ/FINO family protein	<i>Ochrobactrum anthropi</i>	2,00E-59	62%	AIK40840.1	ProQ	04352
BJP_5540	FinO	<i>Pectobacterium</i> sp.	4,00E-23	54%	WP_014701991.1		
BJP_363	putative TrbC	Plasmid pM3	2,00E-103	98%	AAD46129.1		
BJP_1007	VIRB2 type IV secretion	<i>Sinorhizobium meliloti</i>	2,00E-67	100%	WP_017265136.1		
BJP_801	TrwL	<i>Salmonella enterica</i>	5,00E-52	100%	WP_012443547.1		
BJP_802	KorA	<i>Salmonella enterica</i>	1,00E-56	100%	WP_012443548.1		
AP_161	VIRB2 type IV secretion family protein	<i>Ochrobactrum anthropi</i>	2,00E-56	97%	WP_011982848.1		
BJP_204	hypothetical protein	<i>Nitrosomonas eutropha</i>	3,00E-42	89%	WP_011630705.1	TrbC	04956
BJP_1006	lytic transglycosylase	<i>Sinorhizobium meliloti</i>	0.0	92%	WP_013845317.1		
BJP_8408	conjugal transfer protein TrbC	<i>Proteobacteria</i>	1,00E-52	100%	WP_011205785.1		
BJP_730	lytic transglycosylase, catalytic	<i>Nitrosomonas eutropha</i>	1,00E-109	75%	WP_011630706.1		
BJP_203	lytic transglycosylase, catalytic	<i>Nitrosomonas eutropha</i>	2,00E-110	79%	WP_011630706.1		
BJP_728	type IV secretory pathway, VirB3 family protein	<i>Nitrosomonas eutropha</i>	2,00E-33	91%	WP_011630704.1		
BJP_800	trwM protein	<i>Salmonella enterica</i>	1,00E-65	97%	WP_012196422.1		
BJP_364	putative mating pair formation protein	<i>Pseudomonas putida</i>	8,00E-61	100%	WP_011005967.1		
AP_162	type IV secretory pathway VirB3 family protein	<i>chrobactrum anthropi</i>	6,00E-55	99%	WP_011982849.1		
BJP_205	type IV secretory pathway, VirB3 family protein	<i>Nitrosomonas eutropha</i>	2,00E-53	93%	WP_011630704.1	VirB3	05101
BJP_728	type IV secretory pathway, VirB3 family protein	<i>Nitrosomonas eutropha</i>	2,00E-33	91%	WP_011630704.1		
BJP_1008	type IV secretory pathway, VirB3	<i>Sinorhizobium meliloti</i>	4,00E-64	100%	WP_020479309.1		
AP_465	conjugal transfer protein TrbD	<i>Ochrobactrum rhizosphaerae</i>	8,00E-49	98%	WP_024897950.1		

Tabla S.5 (continuación)

ID en DenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
AP_88	conjugal transfer protein TrbC	<i>Ochrobactrum anthropi</i>	1,00E-74	100%	EXL02649.1		
AP_504	type IV secretion protein VirB3	<i>Ochrobactrum</i>	1,00E-67	89%	WP_006468248.1		
BJP_1723	Ti-type conjugative transfer relaxase TraA	<i>Rhizobium leguminosarum</i>	0.0	81%	EJC77937.1		
BJP_6246	conjugal transfer protein TraA	<i>Ochrobactrum intermedium</i>	0.0	99%	WP_036566255.1		
AP_265	Dtr system oriT relaxase	<i>Agrobacterium tumefaciens</i>	0.0	95%	WP_035227113.1	MobA_MobL	03389
BJP_4415	conjugal transfer protein TraA	<i>Agrobacterium tumefaciens</i>	0.0	78%	WP_026109554.1		
BJP_1361	Rac prophage; predicted tail fiber protein	<i>Escherichia coli</i>	0.0	100%	AIZ27845.1		
BJP_2253	TraG family bacterial conjugation protein	<i>Arthrobacter arilaitensis</i>	0.0	64%	WP_013350309.1	TraG-D_C	12696
BJP_444	hypothetical protein	<i>Yersinia ruckeri</i>	0.0	81%	EEP98795.1		
BJP_744	hypothetical protein	<i>Leifsonia aquatica</i>	0.0	62%	WP_039921200.1		
BJP_3931	ATPase involved in chromosome partitioning-like protein	<i>Rhizobium</i> sp.	1,00E-129	84%	CCF22397.1	PRKCSH	07915
BJP_152	hypothetical protein	<i>Pseudomonas</i>	0.0	67%	WP_023662237.1	TraG_N	07916
BJP_4168	hypothetical protein	<i>Pseudomonas</i>	0.0	100%	WP_023662237.1		
BJP_2547	membrane protein	<i>Parvibaculum lavamentivorans</i>	0.0	95%	WP_012112346.1		
BJP_6103	2,3-dehydroadipyl-CoA hydratase	<i>Serratia marcescens</i>	5,00E-129	96%	WP_033639967.1		
BJP_588	hypothetical protein	<i>Pseudomonas</i> sp.	0.0	90%	WP_039614051.1		
BJP_84	conjugal transfer protein traN	<i>Pseudomonas syringae</i>	0.0	73%	WP_032708792.1	TraN	06986
BJP_1942	mating pair stabilization protein	<i>Pseudomonas putida</i>	0.0	100%	WP_033040477.1		
BJP_150	conjugal transfer protein	<i>Pseudomonas monteilii</i>	0.0	78%	WP_028699874.1		
BJP_7057	conjugal transfer protein traH	<i>Pseudomonas</i>	5,00E-68	99%	WP_023662238.1	TraH	06122
BJP_4167	conjugal transfer protein traH	<i>Pseudomonas</i>	2,00E-168	100%	WP_023662238.1		
BJP_7055	conjugal transfer protein traH	<i>Pseudomonas</i>	9,00E-36	100%	WP_023662238.1		
BJP_790	hypothetical protein	<i>Thauera terpenica</i>	1,00E-45	54%	WP_021250285.1	TrbM	07424

Tabla S.5 (continuación)

ID en DenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_458	TrbM family protein	<i>Nitrosomonas eutropha</i>	5,00E-80	70%	WP_011630 665.1		
BJP_195	TrbM family protein	<i>Nitrosomonas eutropha</i>	4,00E-93	74%	WP_011630 665.1		
BJP_375	hypothetical protein	<i>Pseudomonas mosselii</i>	2,00E-62	96%	WP_028691 942.1		
BJP_804	KikA	<i>Salmonella enterica</i>	8,00E-145	99%	WP_012443 550.1		
pMC6_31	arginine transporter	<i>Enterobacteriac eae</i>	1,00E-64	100%	WP_000848 516.1		
BJP_414	arginine transporter	<i>Enterobacteriac eae</i>	1,00E-64	100%	WP_000848 516.1		
BJP_376	endonuclease	<i>Pseudomonas mosselii</i>	2,00E-120	94%	WP_028691 943.1		
BJP_943	conjugal transfer protein TrbM	<i>Pseudomonas aeruginosa</i>	7,00E-129	100%	WP_011205 795.1		
AP_507	hypothetical protein	<i>Ochrobactrum rhizosphaerae</i>	3,00E-91	72%	WP_024899 142.1		
BJP_791	hypothetical protein	<i>Alicyclophilus sp. CRZ1</i>	5,00E-80	81%	WP_019372 870.1		
pMC6_32	hypothetical protein	<i>Salmonella enterica</i>	8,00E-92	100%	WP_000368 397.1		
BJP_415	hypothetical protein	<i>Salmonella enterica</i>	4,00E-92	100%	YP_0021449 47.1		
AP_155	hypothetical protein	<i>Ochrobactrum anthropi</i>	1,00E-132	97%	YP_0013730 86.1		
AP_103	hypothetical protein	<i>Ochrobactrum anthropi</i>	7,00E-114	92%	WP_036588 691.1		
BJP_146	F-pilin subunit assembly into extended F pili	<i>Pseudomonas monteilii</i>	0.0	85%	WP_028699 872.1		
BJP_1271	sex pilus assembly protein	<i>Pseudomonas</i>	0.0	99%	WP_023662 242.1	TraC_F_IV	11130
BJP_227	TrsE protein	<i>Paenibacillus peoriae</i>	0.0	96%	WP_010349 924.1		
BJP_734	mobilisation protein	<i>Nitrosomonas eutropha</i>	4,00E-69	85%	WP_011630 663.1		
pMC6_2	hypothetical protein	<i>Salmonella enterica</i>	9,00E-86	100%	WP_001157 192.1		
BJP_388	hypothetical protein	<i>Salmonella enterica</i>	9,00E-86	100%	WP_001157 192.1		
BJP_199	hypothetical protein	<i>Nitrosomonas eutropha</i>	8,00E-69	83%	WP_011630 663.1	MobC	05713
AP_395	mobilisation protein	<i>Ochrobactrum anthropi</i>	4,00E-84	95%	WP_011982 860.1		
BJP_2739	mobilization protein	<i>Sinorhizobium/ Ensifer group</i>	1,00E-43	97%	WP_034859 608.1		

Tabla S.5 (continuación)

ID en DenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_848	putative mobilisation protein	<i>Propionibacterium freudenreichii</i>	3,00E-32	59%	CEH05296.1		
BJP_198	hypothetical protein	<i>Nitrosomonas eutropha</i>	0.0	79%	WP_011630664.1		
BJP_2740	conjugal transfer protein TraA	<i>Sinorhizobium fredii</i>	0.0	91%	WP_014330857.1		
AP_394	relaxase	<i>Ochrobactrum anthropi</i>	0.0	94%	WP_011982859.1		
BJP_454	hypothetical protein	<i>Nitrosomonas eutropha</i>	0.0	77%	WP_011630664.1		
BJP_5273	mobilisation protein	<i>Sinorhizobium fredii</i>	3,00E-07	82%	WP_037461848.1		
BJP_735	hypothetical protein	<i>Nitrosomonas eutropha</i>	3,00E-07	86%	WP_011630664.1		
BJP_140	hypothetical protein	<i>Pseudomonas</i>	2,00E-39	90%	WP_016486958.1		
BJP_1265	hypothetical protein	<i>Pseudomonas</i>	6,00E-44	99%	WP_016486958.1	TraL	07178
BJP_139	hypothetical protein	<i>Pseudomonas putida</i>	2,00E-66	66%	WP_020190088.1		
BJP_1264	flagellar motor protein MotB	<i>Pseudomonas</i>	2,00E-119	100%	WP_023662245.1		
BJP_141	hypothetical protein	<i>Pseudomonas</i>	3,00E-105	69%	WP_016486959.1	TraE	05309
BJP_1266	hypothetical protein	<i>Pseudomonas</i>	4,00E-153	100%	WP_016486959.1		
BJP_894	mobilization protein	<i>Bacillus cereus</i>	0.0	98%	WP_000901534.1	Mob_Pre	01076
pMC3_7	mobilization protein	<i>Bacillus cereus</i>	0.0	98%	WP_000901534.1		
AP_477	hypothetical protein	<i>Rhizobium giardinii</i>	1,00E-59	95%	WP_018326248.1	Prok-TraM	09228
AP_478	TraR	<i>Rhizobiaceae</i>	4,00E-170	100%	WP_015647751.1		
AP_263	conjugal transfer protein TraD	<i>Agrobacterium tumefaciens</i>	7,00E-37	94%	WP_035227115.1	TraD	06412
BJP_1361	Rac prophage; predicted tail fiber protein	<i>Escherichia coli</i>	0.0	100%	AIZ27845.1		
AP_264	conjugal transfer protein TraC	<i>Agrobacterium tumefaciens</i>	1,00E-38	93%	WP_035227185.1	TraC	07820
BJP_7783	MbeD/MobD	<i>Serratia</i>	6,00E-39	100%	WP_004928976.1		
BJP_6195	mobilisation protein	<i>Proteobacteria</i>	2,00E-44	99%	WP_000137445.1	MbeD_MobD	04899
BJP_1361	Rac prophage; predicted tail fiber protein	<i>Escherichia coli</i>	0.0	100%	AIZ27845.1		

Tabla S.5 (continuación)

ID en DenDB	Descripción (mejor hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	ID de Pfam	Pfam
BJP_7784	membrane protein	<i>Serratia marcescens</i>	1,00E-92	100%	WP_033639 556.1		
BJP_4419	protein transporter hofB	<i>Escherichia coli</i>	0.0	100%	WP_016241 104.1	T2SSE_N	05157
BJP_144	membrane lipoprotein lipid attachment site	<i>Pseudomonas</i> sp.	8,00E-94	82%	WP_023383 789.1		
BJP_1269	membrane lipoprotein lipid attachment site	<i>Pseudomonas</i> sp.	4,00E-114	96%	WP_023383 789.1	TraV	09676
BJP_149	conjugal transfer protein	<i>Pseudomonas putida</i>	6,00E-170	84%	WP_029615 335.1		
BJP_7056	conjugal transfer protein	<i>Pseudomonas</i>	2,00E-174	100%	WP_031325 012.1	TraF	13728

S6. Lista de los genes de resistencia a antibióticos identificados en el set de datos plasmídico

ID en GenDB	Descripción (Mejor hit BLASTp en NCBI)	Clase de resistencia a antibiótico	Nombre del gen	Organismo	E-value	Máx ident	Número de acceso
AP_369	family drug resistance transporter	Multidrogas	<i>emrD</i>	<i>Pseudomonas fluorescens</i>	0.0	83%	YP_00520886 8.1
BJP_10004	Multidrug resistance efflux pump	Multidrogas	<i>mexH</i>	<i>Stenotrophomonas maltophilia</i>	2,00E-70	74%	WP_0054158 00.1
BJP_10347	antibiotic ABC transporter permease	Multidrogas		<i>Serratia marcescens</i>	4,00E-25	100%	WP_0169293 17.1
BJP_10349	antibiotic ABC transporter permease	Multidrogas		<i>Serratia marcescens</i>	7,00E-138	100%	WP_0169293 17.1
BJP_10488	multidrug transporter	Multidrogas	<i>opmD</i>	<i>Variovorax paradoxus</i>	2,00E-40	44%	WP_0207233 53.1
BJP_10538	beta-lactamase/D-alanine carboxypeptidase	β-lactamas	<i>bl1_ampC</i>	<i>Escherichia coli</i>	0.0	100%	NP_418574.1
BJP_10562	Multidrug translocase MdfA, partial	Multidrogas	<i>mdfA</i>	<i>Escherichia coli</i>	7,00E-109	100%	WP_0013197 91.1
BJP_10749	streptomycin 3''-kinase	Aminoglicósidos	<i>aph33ib</i>	<i>Xanthomonas campestris</i>	0.0	100%	YP_364056.1
BJP_10828	multidrug resistance protein MdtH	Multidrogas	<i>mdtH</i>	<i>Serratia marcescens</i>	2,00E-24	100%	WP_0049296 33.1
BJP_1085	bleomycin resistance protein	Bleomicina		<i>Enterobacter asburiae</i>	2E-48	65%	YP_00482703 5.1
BJP_10930	multidrug transporter	Multidrogas		<i>Escherichia coli</i>	2,00E-81	100%	WP_0014428 52.1
BJP_10953	streptomycin 3''-kinase	Aminoglicósidos	<i>aph33ib</i>	<i>Xanthomonas campestris</i>	0.0	100%	YP_364056.1
BJP_10956	tetracycline resistance protein TetA	Tetraciclinas	<i>tetA</i>	<i>Paenibacillus elgii</i>	4,00E-61	47%	WP_0104926 84.1
BJP_10957	glyoxalase/bleomycin resistance protein/dioxygenase	Bleomicina		<i>Serratia marcescens</i>	1,00E-85	100%	YP_00740693 0.1
BJP_10964	macrolide export ATP-binding/permease MacB	Macrólidos	<i>macB</i>	<i>Delftia acidovorans</i>	1,00E-178	84%	WP_0164529 30.1
BJP_1449	glyoxalase/bleomycin resistance protein/dioxygenase family protein	Bleomicina		<i>Pseudomonas fluorescens</i>	0.0	97%	YP_00632673 4.1

Tabla S6 (continuación)

ID en GenDB	Descripción (Mejor hit BLASTp en NCBI)	Clase de resistencia a antibiótico	Nombre del gen	Organismo	E-value	Máx ident	Número de acceso
BJP_1470	tetracycline resistance protein, class G (TETA(G))	Tetraciclinas	<i>tetA(G)</i>	<i>Acinetobacter baumannii</i>	0.0	99%	YP_00171536 9.1
BJP_1472	tetracycline repressor protein class G	Tetraciclinas	<i>tetA(G)</i>	<i>Acinetobacter baumannii</i>	9,00E-140	100%	YP_00171537 0.1
BJP_1473	chloramphenicol and florfenicol resistance protein (CmlA)	Fenicoles	<i>cmlA9</i>	<i>Acinetobacter baumannii</i>	0.0	100%	YP_00171537 1.1
BJP_1477	TetR family transcriptional regulator	Tetraciclinas	<i>tetR</i>	<i>Leucobacter chromiiresistens</i>	1,00E-133	100%	WP_0101547 54.1
BJP_1479	tetracycline resistance protein	Tetraciclinas	<i>tet33</i>	<i>Leucobacter chromiiresistens</i>	0.0	99%	WP_0101547 55.1
BJP_1564	aminoglycoside phosphotransferase	Aminoglicósidos		<i>Kribbella flavida</i>	4,00E-51	51%	YP_00337903 9.1
BJP_1577	multidrug transporter	Multidrogas		<i>Ochrobactrum anthropi</i>	2,00E-178	98%	WP_0106603 79.1
BJP_1578	MFS transporter	Multidrogas		<i>Ochrobactrum anthropi</i>	0.0	94%	WP_0106603 78.1
BJP_1773	beta-lactamase	β -lactamas	<i>bl3_cphA</i>	<i>Serratia fonticol</i>	3,00E-131	77%	AAF09244.1
BJP_1786	class B beta-lactamase	β -lactamas	<i>bl3_I</i>	<i>uncultured bacterium</i>	6,00E-48	62%	AGD93225.1
BJP_1946	multidrug transporter	Multidrogas		<i>Hafnia alvei</i>	2,00E-92	52%	WP_0040925 79.1
BJP_1947	Putative Na ⁺ -driven multidrug efflux pump, partial	Multidrogas		<i>Moritella sp.</i>	3,00E-16	57%	WP_0060313 86.1
BJP_2046	multidrug efflux system protein	Multidrogas	<i>mdtL</i>	<i>Escherichia coli</i>	0.0	100%	NP_418166.1
BJP_2274	aminoglycoside 3'-phosphotransferase	Aminoglicósidos	<i>aph3ib</i>	<i>uncultured bacterium</i>	5,00E-99	100%	YP_112402.1
BJP_2281	multidrug transporter MatE	Multidrogas	<i>matE</i>	<i>Variovorax paradoxus</i>	2,00E-152	67%	WP_0189061 31.1
BJP_2342	efflux protein	Multidrogas		<i>Burkholderia gladioli</i>	5,00E-72	54%	YP_00436254 8.1
BJP_2845	multidrug resistance efflux transporter conferring overexpression resistance to norfloxacin and enoxacin	Multidrogas	<i>mdtH</i>	<i>Escherichia coli</i>	0.0	100%	NP_415583.4
BJP_2917	multidrug resistance outer membrane protein MdtQ	Multidrogas	<i>mdtQ</i>	<i>Escherichia coli</i>	0.0	100%	YP_00173108 1.1

Tabla S6 (continuación)

ID en GenDB	Descripción (Mejor hit BLASTp en NCBI)	Clase de resistencia a antibiótico	Nombre del gen	Organismo	E-value	Máx ident	Número de acceso
BJP_3020	multidrug transporter	Multidrogas	<i>mdtF</i>	<i>Escherichia coli</i>	6,00E-142	99%	WP_0000248 91.1
BJP_3282	multiple antibiotic resistance protein MarC	Multidrogas	<i>marC</i>	<i>Escherichia coli</i>	4,00E-154	99%	WP_0015270 63.1
BJP_3283	DNA-binding transcriptional repressor of multiple antibiotic resistance	Multidrogas		<i>Escherichia coli</i>	7,00E-99	99%	NP_416047.4
BJP_3284	Multiple antibiotic resistance protein marA	Multidrogas	<i>marA</i>	<i>Shigella flexneri</i>	1,00E-89	100%	YP_00572725 2.1
BJP_3285	multiple antibiotic resistance protein MarB	Multidrogas	<i>marB</i>	<i>Escherichia coli</i>	1,00E-32	96%	YP_00613371 1.1
BJP_3342	multidrug RND transporter, membrane fusion protein MexP	Multidrogas	<i>mexE</i>	<i>Pseudomonas fluorescens</i>	3,00E-171	83%	YP_00632448 4.1
BJP_3394	antibiotic transport system permease	Multidrogas		<i>Delftia acidovorans</i>	5,00E-33	70%	WP_0164456 23.1
BJP_3395	antibiotic transport system permease	Multidrogas		<i>Delftia acidovorans</i>	3,00E-173	82%	WP_0164532 67.1
BJP_3397	beta-lactamase	β -lactamas	<i>bl2c_pse1</i>	<i>Acidovorax</i> sp.	1,00E-23	48%	WP_0078551 23.1
BJP_3434	antibiotic transport system permease	Multidrogas		<i>Delftia acidovorans</i>	3,00E-69	82%	WP_0164532 68.1
BJP_3453	metallo-beta-lactamase	β -lactamas	<i>bl3_cphA</i>	<i>Acidovorax</i> sp.	6,00E-175	75%	GAD21457.1
BJP_3557	Multidrug resistance protein, partial [Escherichia coli	Multidrogas		<i>Escherichia coli</i>	3,00E-105	99%	YP_00613955 8.1
BJP_3558	multidrug resistance protein Y [Escherichia coli	Multidrogas		<i>Escherichia coli</i>	0.0	99%	WP_0016135 51.1
BJP_3829	beta-lactamase-like	β -lactamas	<i>bl3_l</i>	<i>Polaromonas</i> sp.	2,00E-137	71%	YP_552021.1
BJP_4008	beta-lactamase	β -lactamas	<i>mdtK</i>	<i>Burkholderia</i> sp.	0.0	58%	WP_0083459 09.1
BJP_4048	hydrophobe/amphiphile efflux-1 (HAE1) family transporter	Multidrogas		<i>Delftia</i> sp.	0.0	84%	YP_00448874 8.1
BJP_4120	tetracycline resistance regulatory protein TetR	Tetraciclinas	<i>tetR</i>	<i>Aeromonas salmonicida</i>	2,00E-67	100%	NP_387462.1
BJP_4122	tetracycline resistance protein, partial	Tetraciclinas	<i>tetC</i>	<i>Pasteurella dagmatis</i>	5,00E-81	100%	WP_0057656 49.1

Tabla S6 (continuación)

ID en GenDB	Descripción (Mejor hit BLASTp en NCBI)	Clase de resistencia a antibiótico	Nombre del gen	Organismo	E-value	Máx ident	Número de acceso
BJP_4123	tetracycline resistance protein, class C, partial	Tetraciclinas	<i>tetC</i>	<i>Escherichia</i> sp.	2,00E-150	100%	WP_0024310 94.1
BJP_4183	multidrug transporter	Multidrogas		<i>Escherichia coli</i>	4,00E-48	99%	WP_0014103 77.1
BJP_4744	glyoxalase/bleomycin resistance protein/dioxygenase	Bleomicina		<i>Serratia proteamaculans</i>	2,00E-61	56%	YP_00147853 3.1
BJP_4808	beta-lactamase	β -lactamas	<i>bl1_pao</i>	<i>Pandoraea</i> sp.	8,00E-136	61%	WP_0108038 89.1
BJP_4809	glutathione transferase	Fosmidomina	<i>fosA</i>	<i>Comamonas testosteroni</i>	5,00E-59	67%	WP_0030682 11.1
BJP_4998	beta-lactamase	β -lactamas	<i>bl1_ampC</i>	<i>Streptomyces</i> sp.	4,00E-60	41%	WP_0087430 65.1
BJP_5116	outer membrane factor of efflux pump	Multidrogas	<i>mdtP</i>	<i>Serratia marcescens</i>	0.0	99%	YP_00740767 0.1
BJP_5279	acriflavine resistance protein E	Acriflavina	<i>acrA</i>	<i>Escherichia coli</i>	0.0	99%	WP_0016260 61.1
BJP_5281	multidrug efflux system protein, partial	Multidrogas	<i>acrB</i>	<i>Escherichia coli</i>	8,00E-22	100%	WP_0013208 83.1
BJP_5319	bleomycin resistance protein	Bleomicina		<i>Pusillimonas noertemannii</i>	8,00E-135	67%	WP_0175226 78.1
BJP_5515	macrolide transporter subunit MacA, partial	Macrólidos	<i>macA</i>	<i>Escherichia coli</i>	5,00E-58	99%	WP_0007464 44.1
BJP_5614	Bcr/CflA subfamily drug resistance transporter	Multidrogas		<i>Cupriavidus necator</i>	3,00E-37	74%	WP_0065769 66.1
BJP_5700	multidrug transporter	Multidrogas		<i>Planctomyces maris</i>	9,00E-62	54%	WP_0026441 59.1
BJP_6026	glyoxalase/bleomycin resistance protein/dioxygenase superfamily protein	Bleomicina		<i>Burkholderia cenocepacia</i>	8,00E-107	81%	YP_00223029 9.1
BJP_6093	fosmidomycin resistance protein	Fosmidomina	<i>rosA</i>	<i>Enterobacter aerogenes</i>	1,00E-146	73%	YP_00459279 9.1
BJP_6101	glyoxalase/bleomycin resistance protein/dioxygenase	Bleomicina		<i>Pseudomonas</i> sp.	2,00E-53	56%	WP_0096199 19.1
BJP_6173	beta-lactamase	β -lactamas	<i>bl1_asbA</i>	<i>Serratia marcescens</i>	2,00E-122	99%	WP_0169277 73.1
BJP_6253	beta-lactamase domain-containing protein	β -lactamas	<i>bl3_cphA</i>	<i>Polaromonas naphthalenivorans</i>	2,00E-141	70%	YP_980965.1

Tabla S6 (continuación)

ID en GenDB	Descripción (Mejor hit BLASTp en NCBI)	Clase de resistencia a antibiótico	Nombre del gen	Organismo	E-value	Máx ident	Número de acceso
BJP_6622	multidrug resistance protein MdtO	Multidrogas	<i>mdtO</i>	<i>Escherichia coli</i>	1,00E-85	100%	WP_0015442 72.1
BJP_6623	efflux transporter, RND family, MFP subunit	Multidrogas	<i>mdtN</i>	<i>Escherichia coli</i>	0.0	99%	WP_0016386 38.1
BJP_6895	acriflavin resistance protein	Acriflavina	<i>mexF</i>	<i>Variovorax paradoxus</i>	6,00E-165	64%	YP_00294548 9.1
BJP_7098	chloramphenicol resistance protein	Fenicoles	<i>cmlA6</i>	<i>Klebsiella pneumoniae</i>	0.0	99%	YP_00133881 2.1
BJP_7385	TetR family transcriptional regulator	Tetraciclinas	<i>tetR</i>	<i>Paenibacillus polymyxa</i>	8,00E-87	64%	YP_00394816 2.1
BJP_7397	beta-lactamase:copper amine oxidase-like protein	β -lactamas	<i>bl1_asbA</i>	<i>Acetivibrio cellulolyticus</i>	3,00E-65	32%	WP_0102488 13.1
BJP_749	penicillin-binding protein	β -lactamas	<i>bl1_och</i>	<i>Ochrobactrum intermedium</i>	0.0	92%	WP_0064705 59.1
BJP_7501	beta-lactamase	β -lactamas		<i>Cupriavidus taiwanensis</i>	2,00E-67	71%	WP_0180041 45.1
BJP_7527	chloramphenicol acetyltransferase	Fenicoles	<i>catA16</i>	<i>Streptococcus pneumoniae</i>	2,00E-55	61%	CAO82974.1
BJP_7974	multidrug transporter	Multidrogas		<i>Escherichia coli</i>	4,00E-48	99%	WP_0014103 77.1
BJP_7982	multidrug ABC transporter ATP-binding protein, partial	Multidrogas		<i>Escherichia coli</i>	6,00E-69	98%	WP_0202408 56.1
BJP_8071	glyoxalase/bleomycin resistance protein/dioxygenase	Bleomicina		<i>Delftia acidovorans</i>	2,00E-102	76%	YP_00156107 2.1
BJP_8089	multidrug transporter	Multidrogas	<i>emre</i>	<i>Serratia marcescens</i>	4,00E-40	100%	WP_0169274 46.1
BJP_8090	small multidrug resistance protein	Multidrogas	<i>emrE</i>	<i>Serratia proteamaculans</i>	3,00E-72	91%	YP_00147899 5.1
BJP_8121	Macrolide export ATP-binding/permease protein MacB, partial	Macrólidos	<i>macB</i>	<i>Escherichia coli</i>	0.0	100%	WP_0003165 35.1
BJP_8267	glyoxalase	Bleomicina		<i>Serratia marcescens</i>	8,00E-107	100%	WP_0169287 90.1
BJP_8309	multidrug transporter	Multidrogas	<i>mdtM</i>	<i>Escherichia coli</i>	0.0	99%	WP_0202364 18.1
BJP_8388	beta-lactamase	β -lactamas		<i>Spirosoma spitsbergense</i>	5,00E-65	40%	WP_0206077 45.1

Tabla S6 (continuación)

ID en GenDB	Descripción (Mejor hit BLASTp en NCBI)	Clase de resistencia a antibiótico	Nombre del gen	Organismo	E-value	Máx ident	Número de acceso
BJP_8406	membrane protein	Macrólidos	<i>macA</i>	<i>Escherichia coli</i>	4,00E-128	99%	WP_001384391.1
BJP_8407	ABC transporter ATP-binding protein	Macrólidos	<i>macB</i>	<i>Escherichia coli</i>	1,00E-146	99%	WP_000188121.1
BJP_8592	acriflavin resistance protein A, partial	Acriflavine	<i>acrA</i>	<i>Escherichia coli</i>	6,00E-170	100%	YP_668452.1
BJP_8593	multidrug efflux system protein AcrB	Multidrogas	<i>acrB</i>	<i>Salmonella enterica</i>	3,00E-19	95%	WP_001717341.1
BJP_8654	fosmidomycin resistance protein	Fosmidomicina	<i>rosA</i>	<i>Shigella sonnei</i>	8,00E-143	100%	YP_309479.1
BJP_8685	multidrug resistance protein MdtN	Multidrogas	<i>mdtN</i>	<i>Serratia marcescens</i>	3,00E-40	98%	WP_004941492.1
BJP_8992	TetA(41)	Tetraciclinas	<i>tetA(41)</i>	<i>Serratia marcescens</i>	8,00E-134	97%	AAP93922.1
BJP_8993	TetR family transcriptional regulator	Tetraciclinas	<i>tetR</i>	<i>Serratia marcescens</i>	8,00E-58	96%	YP_007406306.1
BJP_9086	bicyclomycin/multidrug efflux system macrolide export	Bacitracin	<i>bcr</i>	<i>Serratia marcescens</i>	0.0	100%	YP_007407161.1
BJP_9089	ATP-binding/permease MacB	Macrólidos	<i>macB</i>	<i>Delftia acidovorans</i>	3,00E-51	82%	WP_016452930.1
BJP_9091	macrolide-specific efflux protein MacA	Macrólidos	<i>macA</i>	<i>Delftia acidovorans</i>	8,00E-104	72%	WP_016450742.1
BJP_9128	glyoxalase	Bleomicina		<i>Serratia marcescens</i>	0.0	99%	WP_016930047.1
BJP_9174	aminoglycoside resistance protein	Aminoglicósidos	<i>aac6ib</i>	<i>Alcanivorax pacificus</i>	4,00E-90	56%	WP_008736168.1
BJP_9307	ABC transporter ATP-binding protein	Macrólidos	<i>carA</i>	<i>Serratia marcescens</i>	1E-166	99%	WP_016928844.1
BJP_9313	TetR family transcriptional regulator	Tetraciclinas		<i>Serratia proteamaculans</i>	9,00E-112	84%	YP_001479369.1
BJP_9390	Fosmidomycin resistance protein	Fosmidomicina	<i>rosA</i>	<i>Serratia marcescens</i>	8,00E-31	99%	WP_016928828.1
BJP_9514	multidrug efflux system protein MdtO	Multidrogas	<i>mdtO</i>	<i>Serratia marcescens</i>	0.0	99%	WP_004941490.1
BJP_9677	metal-dependent hydrolases of the beta-lactamase superfamily I	β -lactamas		<i>uncultured bacterium</i>	8,00E-102	62%	AGC71534.1
BJP_9755	glyoxalase/bleomycin resistance protein/dioxygenase	Bleomicina		<i>Cupriavidus metallidurans</i>	1,00E-14	48%	YP_585917.1

Tabla S6 (continuación)

ID en GenDB	Descripción (Mejor hit BLASTp en NCBI)	Clase de resistencia a antibiótico	Nombre del gen	Organismo	E-value	Máx ident	Número de acceso
BJP_9756	glyoxalase/bleomycin resistance protein/dioxygenase Bcr/CflA subfamily	Bleomicina		<i>Cupriavidus</i> sp.	4,00E-80	70%	WP_008651425.1
BJP_9759	drug resistance transporter	Multidrogas	<i>emrD</i>	<i>Cupriavidus necator</i>	8,00E-61	59%	WP_006576966.1
BJP_9772	hydrophobe/amphiphile efflux-1 (HAE1) family transporter	Multidrogas	<i>acrB</i>	<i>Delftia</i> sp.	5,00E-46	94%	YP_004488748.1
BJP_9824	bleomycin resistance protein	Bleomicina		<i>Serratia plymuthica</i>	2,00E-44	91%	WP_006319297.1
BJP_9888	bleomycin resistance protein	Bleomicina		<i>Serratia plymuthica</i>	2,00E-44	91%	WP_006319297.1

S7. Lista de los genes de resistencia metales encontrados en el set de datos plasmídico

ID en GenDB	Description (major hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	Metal
BJP_756	zinc efflux system	<i>Mesorhizobium</i> sp.	1,00E-163	82%	WP_006329619.1	zinc
BJP_774	RND family efflux transporter MFP subunit	<i>Acidovorax</i> sp.	2,00E-71	77%	YP_986110.1	cobalto, zinc
BJP_775	outer membrane efflux protein	<i>Acidovorax</i> sp.	3,00E-168	73%	YP_986109.1	cobalto, zinc
BJP_776	cobalto-zinc-cadmium resistance protein	<i>Cupriavidus metallidurans</i>	1,00E-14	42%	YP_145592.1	cobalto, zinc, cadmio
BJP_791	Cd(II)/Pb(II)-responsive transcriptional regulator	<i>Burkholderiales bacterium</i>	2,00E-61	64%	WP_009552846.1	cadmio, plomo
BJP_1604	tellurite, selenium methyltransferase, SAM-dependent; tellurite, selenium resistance protein	<i>Escherichia coli</i>	1,00E-143	100%	NP_415947.1	telurito, selenio
BJP_1605	tellurite resistance protein TehA	<i>Escherichia coli</i>	0.0	100%	WP_000222935.1	telurito
BJP_1810	MerR protein	<i>Pseudomonas</i> sp.	2,00E-100	100%	NP_862494.1	mercurio
BJP_1811	MerF protein	<i>Pseudomonas</i> sp.	2,00E-66	99%	NP_862493.1	mercurio
BJP_1812	MerP protein	<i>Pseudomonas</i> sp.	3,00E-38	100%	NP_862492.1	mercurio
BJP_1813	mercuric reductase	<i>Plasmid pMCBF1</i>	3,00E-51	100%	AAAY97939.1	mercurio
BJP_1814	mercuric reductase	<i>Klebsiella pneumoniae</i>	0.0	99%	EQA20972.1	mercurio
BJP_1815	MerD protein	<i>Pseudomonas</i> sp.	2,00E-56	100%	NP_862489.1	mercurio
BJP_1816	MerE protein	<i>Pseudomonas</i> sp.	3,00E-48	100%	NP_862488.1	mercurio
BJP_2103	MerR	<i>Pseudomonas alcaligenes</i>	2,00E-83	100%	YP_025338.1	mercurio
BJP_2104	mercuric transport protein	<i>Pseudomonas stutzeri</i>	2,00E-21	100%	YP_001173902.1	mercurio
BJP_2105	Mercuric transport protein periplasmic component precursor	<i>Thiomonas</i> sp.	5,00E-35	85%	YP_003622579.1	mercurio
BJP_2192	MerR family transcriptional regulator	<i>Idiomarina loihiensis</i>	3,00E-93	99%	YP_155021.1	mercurio
BJP_2193	CDF family heavy metal/H(+) antiporter	<i>Burkholderia xenovorans</i>	0.0	99%	YP_552434.1	metales pesados
BJP_3085	arsenic transporter	<i>Ochrobactrum anthropi</i>	3,00E-27	98%	WP_010660310.1	arsénico
BJP_3083	ArsR family transcriptional regulator	<i>Ochrobactrum anthropi</i>	4,00E-42	94%	WP_010660309.1	arsénico
BJP_3082	ArsR family transcriptional regulator	<i>Ochrobactrum anthropi</i>	6,00E-61	97%	WP_010660309.1	arsénico
BJP_3081	ArsR family transcriptional regulator	<i>Ochrobactrum</i> sp.	2,00E-73	92%	WP_007881800.1	arsénico

Tabla S7 (continuación)

ID en GenDB	Description (major hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	Metal
BJP_3505	cation diffusion facilitator family transporter	<i>Delftia acidovorans</i>	7,00E-152	75%	YP_001561495.1	cobalto, zinc, cadmio
BJP_4019	C4-dicarboxylate transporter/malic acid transport protein	<i>Leptothrix cholodnii</i>	5,00E-67	52%	YP_001792807.1	telurito
BJP_4103	RND transporter	<i>Acidovorax</i> sp.	2,00E-103	60%	WP_008903848.1	
BJP_4404	arsenical resistance protein ArsH	<i>Afipia clevelandensis</i>	4,00E-122	75%	WP_002713267.1	arsénico
BJP_4405	ArsC2	<i>Ochrobactrum tritici</i>	2,00E-87	94%	ABF48395.1	arsénico
BJP_4406	arsenic transporter	<i>Ochrobactrum anthropi</i>	5,00E-50	99%	WP_010660310.1	arsénico
BJP_4486	MerR family transcriptional regulator	<i>Pseudomonas mandelii</i>	7,00E-70	91%	WP_019582817.1	mercurio
BJP_5212	zinc/iron permease	<i>Polaromonas naphthalenivorans</i>	1,00E-63	69%	YP_981719.1	zinc, hierro
BJP_5275	cobalto-zinc-cadmium resistance protein CzcC	<i>Delftia acidovorans</i>	2,00E-29	54%	WP_016448635.1	cobalto, zinc, cadmio
BJP_5274	two component heavy metal response transcriptional regulator	<i>Delftia acidovorans</i>	2,00E-122	77%	YP_001562121.1	metales pesados
BJP_5551	arsenite transporter	<i>Escherichia coli</i>	3,00E-119	99%	YP_006129959.1	arsénico
BJP_5552	Putative arsenate reductase-like protein	<i>Escherichia coli</i>	1,00E-27	100%	POCF86.1	arsénico
BJP_5636	nickel transporter subunit	<i>Escherichia coli</i>	0.0	100%	NP_417937.1	nickel
BJP_5637	nickel ABC transporter ATP-binding protein	<i>Escherichia coli</i>	2,00E-107	100%	WP_001187280.1	nickel
BJP_5726	nickel/cobalto efflux system rcnA	<i>Escherichia coli</i>	9,00E-169	99%	WP_001685159.1	nickel, cobalto
BJP_5737	heavy metal translocating P-type ATPase	<i>Polaromonas naphthalenivorans</i>	5,00E-46	65%	YP_984001.1	metales pesados
BJP_5741	MerR family transcriptional regulator	<i>Ramlibacter tataouinensis</i>	4,00E-63	69%	YP_004618893.1	mercurio
BJP_5772	C4-dicarboxylate ABC transporter	<i>Pseudomonas</i> sp.	2,00E-154	89%	WP_008059900.1	telurito
BJP_5874	heavy metal efflux pump, CzcA family protein	<i>Escherichia coli</i>	0.0	99%	WP_001740261.1	metales pesados
BJP_7145	ArsR family transcriptional regulator	<i>Ochrobactrum</i> sp.	1,00E-68	88%	WP_007881800.1	arsénico
BJP_7147	arsenic transporter	<i>Ochrobactrum anthropi</i>	4,00E-27	96%	WP_010660310.1	arsénico
BJP_7474	heavy metal efflux protein	<i>Acidovorax</i> sp.	2,00E-65	81%	ABE73718.1	metales pesados
BJP_7559	ArsR family transcriptional regulator	<i>Limnohabitans</i> sp.	5,00E-43	71%	WP_019429746.1	arsénico

Tabla S7 (continuación)

ID en GenDB	Description (major hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	Metal
BJP_7932	nickel transporter ATP-binding protein NikD, partial	<i>Escherichia coli</i>	6,00E-49	100%	WP_001319468.1	nickel
BJP_7931	nickel ABC transporter permease	<i>Escherichia coli</i>	7,00E-172	99%	WP_001008967.1	nickel
BJP_7930	nickel transporter permease NikB	<i>Shigella flexneri</i>	2,00E-49	100%	YP_690824.1	nickel
BJP_8188	mercuric transport periplasmic protein	<i>Acidobacterium capsulatum</i>	1,00E-23	55%	YP_002754327.1	mercurio
BJP_8187	mercury transporter MerT	<i>Oceanibaculum indicum</i>	5,00E-08	34%	WP_008946214.1	mercurio
BJP_8384	putative transcriptional regulator MerR	<i>uncultured bacterium</i>	6,00E-57	100%	NP_858038.1	mercurio
BJP_8597	outer membrane efflux protein	<i>Alicyclophilus denitrificans</i>	0.0	82%	YP_004388576.1	cobre
BJP_8917	zinc import ATP-binding protein ZnuC	<i>Escherichia coli</i>	3,00E-103	100%	WP_001571550.1	zinc
BJP_8954	merR family transcriptional regulator	<i>Pseudomonas sp.</i>	8,00E-30	89%	YP_007030358.1	mercurio
BJP_8992	merR family transcriptional regulator	<i>Pseudomonas sp.</i>	8,00E-30	89%	YP_007030358.1	mercurio
BJP_10150	iron transporter FeoB	<i>Acidovorax sp.</i>	3,00E-72	76%	WP_020227043.1	hierro
AP_187	zinc ABC transporter	<i>Paracoccus sp.</i>	0.0	94%	WP_010398614.1	zinc
AP_227	ArsR	<i>Sinorhizobium sp.</i>	7,00E-68	99%	YP_007974287.1	arsénico
AP_229	ArsC	<i>Sinorhizobium sp.</i>	4,00E-126	100%	YP_007974288.1	arsénico
AP_230	ArsC	<i>Sinorhizobium sp.</i>	3,00E-99	100%	YP_007974289.1	arsénico
AP_231	ArsB	<i>Sinorhizobium sp.</i>	0.0	100%	YP_007974290.1	arsénico
AP_234	ArsH	<i>Sinorhizobium sp.</i>	2,00E-161	100%	YP_007974293.1	arsénico
AP_235	ArsR	<i>Sinorhizobium sp.</i>	6,00E-60	100%	YP_007974294.1	arsénico
AP_601	MerR transcriptional regular	<i>uncultured bacterium</i>	6,00E-94	92%	ADD63292.1	mercurio
AP_602	mercuric transport protein	<i>Shigella flexneri</i>	4,00E-54	94%	NP_052882.1	mercurio
AP_603	mercury uptake protein	<i>Delftia acidovorans</i>	3,00E-36	98%	AAP88280.1	mercurio
AP_604	mercury transporter MerC	<i>Ralstonia sp.</i>	1,00E-73	96%	WP_004636756.1	mercurio
AP_605	mercuric reductase	<i>uncultured bacterium</i>	0.0	97%	AFN66362.1	mercurio
BJP_676	ArsR family transcriptional regulator	<i>Mesorhizobium alhagi</i>	1,00E-51	73%	WP_008834809.1	arsénico
BJP_4060	ferric uptake regulator family protein	<i>Escherichia coli</i>	2,00E-101	99%	WP_001708323.1	hierro
BJP_8916	high-affinity zinc uptake system membrane protein znuB	<i>Escherichia coli</i>	3,00E-117	99%	WP_001413405.1	zinc
BJP_4781	heavy metal sensor histidine kinase	<i>Pseudomonas sp.</i>	6,00E-166	67%	WP_020797496.1	metales pesados

Tabla S7 (continuación)

ID en GenDB	Description (major hit BLASTp en NCBI)	Organismo	E value	Máx ident	Número de acceso	Metal
BJP_7694	heavy metal sensor signal transduction histidine kinase	<i>Comamonas testosteroni</i>	2,00E-60	59%	YP_003278516.1	metales pesados
BJP_8034	ferric uptake regulator family protein	<i>Verminephrobacter eiseniae</i>	3,00E-64	78%	YP_998128.1	hierro
AP_31	metal transport system ATP-binding protein	<i>Sinorhizobium fredii</i>	2,00E-52	71%	YP_006397726.1	zinc, manganeso
BJP_7226	ArsC	<i>Sinorhizobium sp.</i>	3,00E-99	100%	YP_007974289.1	arsénico
AP_188	copper oxidase	<i>Ochrobactrum rhizosphaerae</i>	0.0	100%	WP_024899901.1	cobre

S8. Genes utilizados para realizar la búsqueda de funciones de degradación de xenobióticos

Sustrato	Gen(es)	Enzima(s)	Número de acceso/ referencia
Metil viológenos	<i>smvA</i>	Major facilitator superfamily transporter	AAL20492/ WP_003234942
Linurón	<i>hilA</i>	hydrolase	AGF25452
	<i>libA</i>	hydrolase	AEO20132
	<i>puhA</i>	hydrolase	EU851877
	<i>puhB</i>	hydrolase	EU851876
Ácido 2,4.diclorofenoxiacético (2,4-D)	<i>tfdA</i>	dioxygenase, alpha-ketoglutarate-dependent	YP_006963028
	<i>tfdB</i>	2,4-dichlorophenol hydroxylase transcriptional regulator, 2,4-D oxygenase large subunit of oxygenase, 2,4-D oxygenase small subunit of oxygenase, 2,4-D transporter, 2,4-D oxygenase ferredoxin	YP_006963026
	<i>cadRABKC</i>		AB062679
1,4-Diclorobenceno	<i>tcbA</i>	chlorobenzene dioxygenase	WP_032488666
	<i>tcbB</i>	cis-chlorobenzene dihydrodiol dehydrogenase	ABR26229
	<i>tcbC</i>	catechol 1,2-dioxygenase	AAD13625
Dibenzofurán	<i>dfdA</i> cluster	dibenzofuran 4,4a-dioxygenase alpha and beta subunits and ferredoxin	AB373750
	<i>dfdBC</i> cluster	two hypothetical proteins, tetR-type regulator, 2,2',3-trihydroxybiphenyl 1,2-dioxygenase, 2-hydroxy-6-(2-hydroxyphenyl)-6-oxo-2,4-hexadienoic acid hydrolase	AB373749
	<i>bphBCD</i> cluster	dihydrodiol dehydrogenase, 2,3-dihydroxybiphenyl dioxygenase, 2-hydroxy-6-phenylhexa-2,4-dienoic acid hydrolase	AJ539227
Endosulfán	<i>esd</i>	FMNH ₂ -dependent mono-oxygenase	AF537302
	<i>dszA</i>	monooxygenase	AAU14820
	<i>dszB</i>	Substrate binding domain of 2'-hydroxybiphenyl-2-sulfinate desulfinate	AAU14821
	<i>dszC</i>	sulfur acquisition oxidoreductase	AAU14822
3-Fluorobenzoato	<i>benA</i>	benzene 1,2-dioxygenase	WP_043251427
	<i>xylL</i>	1,6-dihydroxycyclohexa-2,4-diene-1-carboxylatedehydrogenase	CDF83450
4-Fluorobenzoato	<i>benA</i>	benzene 1,2-dioxygenase	WP_043251427
	<i>xylZ</i>	toluate 1,2-dioxygenase	CDF83449

Tabla S8 (continuación)

Sustrato	Gen(es)	Enzima(s)	Número de acceso/ referencia
Paratión Atrazina	<i>catA3</i>	catechol 1,2-dioxygenase	CDF83454
	<i>catB</i>	muconate cycloisomerase 1	CDF83452
	<i>clcD</i>	dienelactone hydrolase	CDF83124
	<i>opd</i>	organophosphate-degrading protein	YP_009089179
Hexaclorociclohexano (HCH)	<i>triA, atzA, atzB, atzC, atzD, trzB, thc, apobec</i>	complete pathway	Database constructed by Fang. <i>et al.</i> (Fang, <i>et al.</i> , 2014)
	<i>dha, cbd, rdt, mog, rdg, dhc, dcn, dog, ccd, dhg</i>	complete pathway	Database constructed by Fang, <i>et al.</i> (Fang, <i>et al.</i> , 2014)
Diclorodifeniltricloroetano (DDT)	<i>dhc, ods, rrat, cpo, rdh, doa, sds, ort, dhg, hdl, hdg</i>	complete pathway	Database constructed by Fang, <i>et al.</i> (Fang, <i>et al.</i> , 2014)

S9. Mejores *hits* generados a partir de la búsqueda BLASTp en ISfinder

Nombre de la secuencia de inserción	Familia	Origen
ISUnCu4	IS30	Plasmid QKH54
TnAs2	Tn3	<i>Aeromonas salmonicida</i>
ISStma17	IS3	<i>Stenotrophomonas maltophilia</i>
ISSso4	IS21	<i>Shigella sonnei</i>
ISSso4	IS21	<i>Shigella sonnei</i>
ISRssp4	IS30	<i>Roseovarius sp.</i>
IS2020	IS6	<i>Brucella melitensis</i>
ISEc1	IS3	<i>Escherichia fergusonii</i>
ISSfl3	IS66	<i>Shigella flexneri</i>
TnShfr1	Tn3	<i>Shewanella frigidimarina</i>
IS869	IS5	<i>Agrobacterium tumefaciens</i>
IS869	IS5	<i>Agrobacterium tumefaciens</i>
ISPa16	IS4	<i>Pseudomonas aeruginosa</i>
ISCte2	IS3	<i>Comamonas testosteroni</i>
ISKpn25	ISL3	<i>Klebsiella pneumoniae</i>
ISMex9	IS110	<i>Methylobacterium extorquens</i>
ISSba14	Tn3	<i>Shewanella baltica</i>
ISKpn26	IS5	<i>Klebsiella pneumoniae</i>
ISHne5	IS110	<i>Hyphomonas neptunium</i>
ISRtr2	IS3	<i>Corynebacterium jeikeium</i>
ISRtr2	IS3	<i>Corynebacterium jeikeium</i>
ISRssp4	IS30	<i>Roseovarius sp.</i>
ISVsa3	IS91	<i>Vibrio salmonicida</i>
ISYps3	Tn3	<i>Yersinia pseudotuberculosis</i>
ISYps3	Tn3	<i>Yersinia pseudotuberculosis</i>
IS609	IS200/IS605	<i>Escherichia coli</i>
IS609	IS200/IS605	<i>Escherichia coli</i>
ISPa38	Tn3	<i>Pseudomonas aeruginosa</i>
ISMex40	IS5	<i>Methylobacterium extorquens</i>
ISRsp9	IS6	<i>Rhizobium sp.</i>
ISMex22	Tn3	<i>Methylobacterium extorquens</i>
ISXc4	Tn3	<i>Xanthomonas campestris</i>
ISAb30	IS200/IS605	<i>Acinetobacter baumannii</i>
ISPre1	IS5	<i>Pseudomonas resinovorans</i>
ISXca1	IS3	<i>Xanthomonas campestris</i>
ISPa42	Tn3	<i>Pseudomonas aeruginosa</i>
ISPa42	Tn3	<i>Pseudomonas aeruginosa</i>
ISXc4	Tn3	<i>Xanthomonas campestris</i>
ISPPu12	ISL3	<i>Pseudomonas putida</i>
ISPPu12	ISL3	<i>Pseudomonas putida</i>

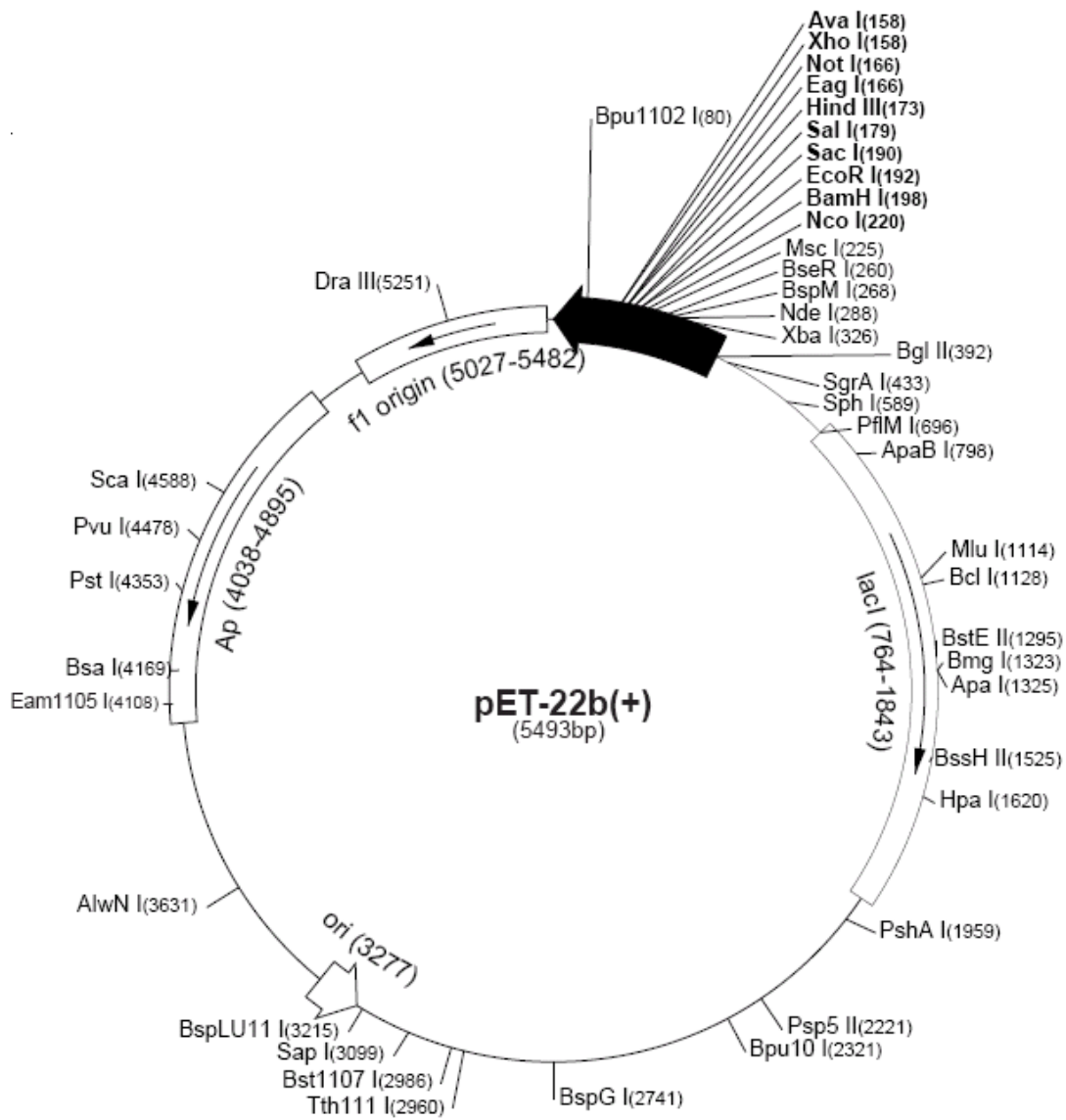
Tabla S9 (continuación)

Nombre de la secuencia de inserción	Familia	Origen
ISPpu12	ISL3	<i>Pseudomonas putida</i>
ISPst9	ISL3	<i>Pseudomonas stutzeri</i>
ISRel8	IS66	<i>Rhizobium etli</i>
ISRel8	IS66	<i>Rhizobium etli</i>
ISPsy24	IS3	<i>Pseudomonas syringae</i>
ISXax1	IS256	<i>Xanthomonas axonopodis</i>
TnAs1	Tn3	<i>Aeromonas salmonicida</i>
ISSwi1	Tn3	<i>Salmonella wien</i>
TnAs2	Tn3	<i>Aeromonas salmonicida</i>
ISThsp9	Tn3	<i>Salmonella wien</i>
ISPsy30	Tn3	<i>Pseudomonas syringae</i>
ISXc4	Tn3	<i>Xanthomonas campestris</i>
ISAzo2	IS91	<i>Weeksella zoohelcum</i>
ISMex22	Tn3	<i>Methylobacterium extorquens</i>
ISMtsp16	IS481	<i>Methylobacterium sp.</i>
IS15DIV	IS6	<i>Salmonella typhimurium</i>
ISAav1	IS21	<i>Acidovorax avenae</i>
ISAav1	IS21	<i>Acidovorax avenae</i>
ISRso21	ISL3	<i>Ralstonia solanacearum</i>
ISMmg1	IS110	<i>Magnetospirillum magnetotacticum</i>
ISAtu3	IS3	<i>Agrobacterium tumefaciens</i>
ISSpo6	IS630	<i>Silicibacter pomeroyi</i>
ISMmg1	IS110	<i>Magnetospirillum magnetotacticum</i>
ISVsa17	ISNCY	<i>Aliivibrio salmonicida</i>
IS1394	IS30	<i>Pseudomonas alcaligenes</i>
ISRel10	ISNCY	<i>Rhizobium etli</i>
ISVsa17	ISNCY	<i>Aliivibrio salmonicida</i>
ISAzs29	Tn3	<i>Azospirillum sp.</i>
ISGme8	IS110	<i>Geobacter metallireducens</i>
ISPps1	IS91	<i>Pseudomonas huttiensis</i>
IS693	IS66	<i>Sinorhizobium medicae</i>
ISAzs17	Tn3	<i>Azospirillum sp.</i>
IS1071	Tn3	<i>Alcaligenes sp</i>
ISRsp12	Tn3	<i>Rhizobium sp.</i>
ISCARN95	IS21	Metagenomic data
ISCARN24	IS21	Metagenomic data
ISPlu15	ISNCY	<i>Photorhabdus luminescens</i>
ISShes11	Tn3	<i>Shewanella sp.</i>
ISRsp18	IS5	<i>Rhizobium sp.</i>
ISEc11	IS110	<i>Escherichia coli</i>
ISWpi13	IS110	<i>Wolbachia pipientis</i>
ISAzo28	IS110	<i>Azoarcus sp.</i>
IS1492	IS110	<i>Pseudomonas putida</i>

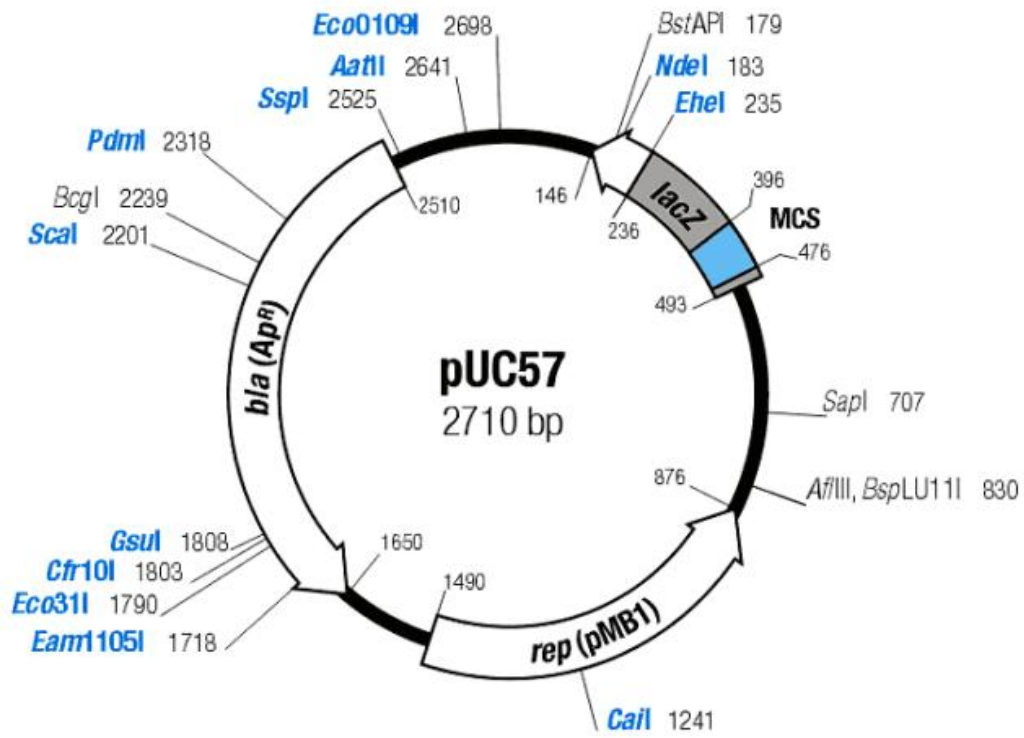
Tabla S9 (continuación)

Nombre de la secuencia de inserción	Familia	Origen
ISKpn1	IS3	<i>Klebsiella pneumoniae</i>
ISKpn1	IS3	<i>Klebsiella pneumoniae</i>
ISRle11	IS1182	<i>Rhizobium leguminosarum</i>
ISRme12	IS3	<i>Ralstonia metallidurans</i>
IS150	IS3	<i>Escherichia coli</i>
ISAzvi6	IS4	<i>Azotobacter vinelandii</i>
ISEc26	ISAs1	<i>Escherichia coli</i>
ISMdi9	IS110	<i>Methylobacterium dichloromethanicum</i>
ISSpr1	IS3	<i>Serratia proteamaculans</i>
ISPsy9	IS3	<i>Pseudomonas syringae</i>
ISJs1	IS481	<i>Janibacter sp.</i>
IS1271	IS4	<i>Enterobacter agglomerans</i>
ISPfl1	IS110	<i>Pseudomonas fluorescens</i>
IS421	IS4	<i>Escherichia coli</i>
ISSba14	Tn3	<i>Shewanella baltica</i>
ISPPu15	IS66	<i>Pseudomonas putid</i>
ISCTe5	IS5	<i>Comamonas testosteroni</i>
ISStma6	IS110	<i>Stenotrophomonas maltophilia</i>
IS401	IS3	<i>Burkholderia cepacia</i>
ISAAv4	IS3	<i>Acidovorax avenae</i>
ISPsy11	IS3	<i>Pseudomonas syringae</i>
ISPsy11	IS3	<i>Pseudomonas syringae</i>
ISHne2	IS481	<i>Hyphomonas neptunium</i>
ISPsy28	IS3	<i>Pseudomonas syringae</i>
ISThsp9	Tn3	<i>Salmonella wien</i>
ISPPu10	IS110	<i>Pseudomonas putida</i>
IS5D	IS5	<i>Escherichia coli</i>
ISAch1	IS5	<i>Achromobacter sp.</i>
ISPPu10	IS110	<i>Pseudomonas putida</i>
ISMet1	ISKra4	<i>Methylocystis sp.</i>
TnAs1	Tn3	<i>Aeromonas salmonicida</i>
ISPPa9	IS6	<i>Paracoccus pantotrophus</i>
ISMex24	ISL3	<i>Methylobacterium extorquens</i>
ISSod9	Tn3	<i>Shewanella oneidensis</i>
IS903	IS5	<i>Escherichia coli</i>
ISPPu4	IS3	<i>Pseudomonas putida</i>

S10. Mapa del vector pET22b(+)



S11. Mapa del vector pUC57



Bibliografía

Adrangi S & Faramarzi MA (2013) From bacteria to human: a journey into the world of chitinases. *Biotechnology advances* **31**: 1786-1795.

Adrio JL & Demain AL (2014). Microbial enzymes: tools for biotechnological processes. *Biomolecules*, *4*(1), 117-139.

Ahangarkani F, Rajabnia R, Shahandashti EF, Bagheri M & Ramez M (2015) Frequency of Class 1 Integron in Escherichia Coli Strains Isolated from Patients with Urinary Tract Infections in North of Iran. *Materia Socio-Medica* **27**: 10-12.

Alvarez-Martinez CE & Christie PJ (2009) Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* **73**: 775-808.

Allen HK, Moe LA, Rodbumrer J, Gaarder A & Handelsman J (2009) Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J* **3**: 243-251.

Amann RI, Ludwig W & Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143-169.

Anjum R, Grohmann E & Malik A (2012) Exogenous isolation of conjugative plasmids from pesticide contaminated soil. *World J Microbiol Biotechnol* **28**: 567-574.

Aravind L & Koonin EV (2001) A natural classification of ribonucleases. *Methods Enzymol* **341**: 3-28.

Arber W (2000) Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol Rev* **24**: 1-7.

Arduino SM, Catalano M, Orman BE, Roy PH & Centron D (2003) Molecular epidemiology of orf513-bearing class 1 integrons in multiresistant clinical isolates from Argentinean hospitals. *Antimicrob Agents Chemother* **47**: 3945-3949.

Arduino SM, Quiroga MP, Ramirez MS, *et al.* (2012) Transposons and integrons in colistin-resistant clones of *Klebsiella pneumoniae* and *Acinetobacter baumannii* with epidemic or sporadic behaviour. *J Med Microbiol* **61**: 1417-1420.

Assinder SJ & Williams PA (1990) The TOL plasmids: determinants of the catabolism of toluene and the xylenes. *Adv Microb Physiol* **31**: 1-69.

Ausec L, Zakrzewski M, Goesmann A, Schlüter A & Mandic-Mulec I (2011) Bioinformatic analysis reveals high diversity of bacterial genes for laccase-like enzymes. *PLoS One* **6**: e25724.

Ausec L, Črnigoj M, Šnajder M, Ulrich NP & Mandic-Mulec I (2015) Characterization of a novel high-pH-tolerant laccase-like multicopper oxidase and its sequence diversity in *Thioalkalivibrio* sp. *Applied microbiology and biotechnology* **99**: 9987-9999.

Bahl MI, Hansen LH, Licht TR & Sorensen SJ (2007) Conjugative transfer facilitates stable maintenance of IncP-1 plasmid pKJK5 in *Escherichia coli* cells colonizing the gastrointestinal tract of the germfree rat. *Appl Environ Microbiol* **73**: 341-343.

Bahl MI, Burmolle M, Meisner A, Hansen LH & Sorensen SJ (2009) All IncP-1 plasmid subgroups, including the novel epsilon subgroup, are prevalent in the influent of a Danish wastewater treatment plant. *Plasmid* **62**: 134-139.

Bhattacharya D, Nagpure A & Gupta RK (2007) Bacterial chitinases: properties and potential. *Critical reviews in biotechnology* **27**: 21-28.

Banerjee J, Mishra N & Dhas Y (2015) Metagenomics: A new horizon in cancer research. *Meta Gene* **5**: 84-89.

Barbolla R, Catalano M, Orman BE, *et al.* (2004) Class 1 integrons increase trimethoprim-sulfamethoxazole MICs against epidemiologically unrelated *Stenotrophomonas maltophilia* isolates. *Antimicrob Agents Chemother* **48**: 666-669.

Barbolla R, Catalano M, Orman BE, *et al.* (2004) Class 1 integrons increase trimethoprim-sulfamethoxazole MICs against epidemiologically unrelated *Stenotrophomonas maltophilia* isolates. *Antimicrobial agents and chemotherapy* **48**: 666-669.

Bengtsson-Palme J, Boulund F, Fick J, Kristiansson E & Larsson DG (2014) Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India. *Front Microbiol* **5**: 648.

Berdy J (2012) Thoughts and facts about antibiotics: where we are now and where we are heading. *J Antibiot (Tokyo)* **65**: 385-395.

Berglund B (2015) Environmental dissemination of antibiotic resistance genes and correlation to anthropogenic contamination with antibiotics. *Infect Ecol Epidemiol* **5**: 28564.

Beringer JE (1974) R factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol* **84**: 188-198.

Binh CT, Heuer H, Kaupenjohann M & Smalla K (2008) Piggery manure used for soil fertilization is a reservoir for transferable antibiotic resistance plasmids. *FEMS Microbiol Ecol* **66**: 25-37.

Binh CT, Heuer H, Kaupenjohann M & Smalla K (2009) Diverse *aadA* gene cassettes on class 1 integrons introduced into soil via spread manure. *Res Microbiol* **160**: 427-433.

Binh CTT, Heuer H, Kaupenjohann M & Smalla K (2009) Diverse *aadA* gene cassettes on class 1 integrons introduced into soil via spread manure. *Research in Microbiology* **160**: 427-433.

Bleicher A, Schofl G, Rodicio Mdel R & Saluz HP (2013) The plasmidome of a *Salmonella enterica* serovar Derby isolated from pork meat. *Plasmid* **69**: 202-210.

Boto L (2009) Horizontal gene transfer in evolution: facts and challenges. *Proc Biol Sci* **277**: 819-827.

Boyer HW & Roulland-Dussoix D (1969) A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J Mol Biol* **41**: 459-472.

- Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* **72**: 248-254.
- Brolund A & Sandegren L (2015) Characterization of ESBL disseminating plasmids. *Infect Dis (Lond)* 1-8.
- Brolund A, Franzen O, Melefors O, Tegmark-Wisell K & Sandegren L (2013) Plasmidome-analysis of ESBL-producing escherichia coli using conventional typing and high-throughput sequencing. *PLoS ONE* **8**: e65793.
- Brom S, Girard L, Garcia-de los Santos A, Sanjuan-Pinilla JM, Olivares J & Sanjuan J (2002) Conservation of plasmid-encoded traits among bean-nodulating Rhizobium species. *Appl Environ Microbiol* **68**: 2555-2561.
- Brown Kav A, Benhar I & Mizrahi I (2013) A method for purifying high quality and high yield plasmid DNA for metagenomic and deep sequencing approaches. *J Microbiol Methods* **95**: 272-279.
- Brown NL, Misra TK, Winnie JN, Schmidt A, Seiff M & Silver S (1986) The nucleotide sequence of the mercuric resistance operons of plasmid R100 and transposon Tn501: further evidence for mer genes which enhance the activity of the mercuric ion detoxification system. *Mol Gen Genet* **202**: 143-151.
- Cabezón E, Ripoll-Rozada J, Peña A, de la Cruz F & Arechaga I (2015) Towards an integrated model of bacterial conjugation. *FEMS Microbiol Rev* **39**: 81-95.
- Castree N (2003) Bioprospecting: from theory to practice (and back again). *Transactions of the Institute of British Geographers* **28**: 35-55.
- Changey F, Devers-Lamrani M, Rouard N & Martin-Laurent F (2011) In vitro evolution of an atrazine-degrading population under cyanuric acid selection pressure: Evidence for the selective loss of a 47kb region on the plasmid ADP1 containing the atzA, B and C genes. *Gene*, 490(1), 18-25.
- Carter A (2000) How pesticides get into water - and proposed reduction measures. *Pesticide Outlook* **11**: 149-156.
- Cascales E & Christie PJ (2003) The versatile bacterial type IV secretion systems. *Nat Rev Microbiol* **1**: 137-149.
- Castillo Mdel P, Torstensson L & Stenstrom J (2008) Biobeds for environmental protection from pesticide use--a review. *J Agric Food Chem* **56**: 6206-6219.
- Cowan DA, Ramond JB, Makhalanyane TP & De Maayer P (2015) Metagenomics of extreme environments. *Curr Opin Microbiol* **25**: 97-102.
- Curtis TP, Sloan WT & Scannell JW (2002) Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci U S A* **99**: 10494-10499.

Cherry JR & Fidantsef AL (2003) Directed evolution of industrial enzymes: an update. *Curr Opin Biotechnol* **14**: 438-443.

Chistoserdova L (2014) Is metagenomics resolving identification of functions in microbial communities? *Microb Biotechnol* **7**: 1-4.

Choi MK, Kim KD, Ahn KM, Shin DH, Hwang JH, Seong CN & Ka JO (2009) Genetic and phenotypic diversity of parathion-degrading bacteria isolated from rice paddy soils. *J Microbiol Biotechnol* **19**: 1679-1687.

Collins PJ & Dobson A (1997) Regulation of laccase gene transcription in *Trametes versicolor*. *Applied and Environmental Microbiology* **63**: 3444-3450.

D'Costa VM, Griffiths E & Wright GD (2007) Expanding the soil antibiotic resistome: exploring environmental diversity. *Curr Opin Microbiol* **10**: 481-489.

Dagan T, Artzy-Randrup Y & Martin W (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci U S A* **105**: 10039-10044.

Dahlberg C, Linberg C, Torsvik VL & Hermansson M (1997) Conjugative plasmids isolated from bacteria in marine environments show various degrees of homology to each other and are not closely related to well-characterized plasmids. *Appl Environ Microbiol* **63**: 4692-4697.

Davison J (1999) Genetic exchange between bacteria in the environment. *Plasmid* **42**: 73-91.

de la Cruz F & Davies J (2000) Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol* **8**: 128-133.

Dealtry S, Holmsgaard PN, Dunon V, *et al.* (2014) Shifts in abundance and diversity of mobile genetic elements after the introduction of diverse pesticides into an on-farm biopurification system over the course of a year. *Appl Environ Microbiol* **80**: 4012-4020.

Dealtry S, Ding GC, Weichelt V, *et al.* (2014) Cultivation-independent screening revealed hot spots of IncP-1, IncP-7 and IncP-9 plasmid occurrence in different environmental habitats. *PLoS ONE* **9**: e89922.

del Papa MF, Balague LJ, Sowinski SC, *et al.* (1999) Isolation and characterization of alfalfa-nodulating rhizobia present in acidic soils of central argentina and uruguay. *Appl Environ Microbiol* **65**: 1420-1427.

Delcher AL, Harmon D, Kasif S, White O & Salzberg SL (1999) Improved microbial gene identification with GLIMMER. *Nucleic acids research* **27**: 4636-4641.

Delmont TO, Prestat E, Keegan KP, *et al.* (2012) Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J* **6**: 1677-1687.

Deng Y, Bao X, Ji L, *et al.* (2015) Resistance integrons: class 1, 2 and 3 integrons. *Ann Clin Microbiol Antimicrob* **14**: 45.

Dennis JJ (2005) The evolution of IncP catabolic plasmids. *Curr Opin Biotechnol* **16**: 291-298.

- Di Conza JA & Gutkind GO (2010) [Integrans: gene collectors]. *Rev Argent Microbiol* **42**: 63-78.
- Dib JR, Wagenknecht M, Farias ME & Meinhardt F (2015) Strategies and approaches in plasmidome studies-uncovering plasmid diversity disregarding of linear elements? *Front Microbiol* **6**: 463.
- Domingues S, da Silva GJ & Nielsen KM (2012) Integrans: Vehicles and pathways for horizontal dissemination in bacteria. *Mob Genet Elements* **2**: 211-223.
- Drell T, Lutsar I, Stsepetova J, *et al.* (2014) The development of gut microbiota in critically ill extremely low birth weight infants assessed with 16S rRNA gene based sequencing. *Gut Microbes* **5**: 304-312.
- Dunon V, Sniegowski K, Bers K, Lavigne R, Smalla K & Springael D (2013) High prevalence of IncP-1 plasmids and IS1071 insertion sequences in on-farm biopurification systems and other pesticide-polluted environments. *FEMS Microbiol Ecol* **86**: 415-431.
- Durao P, Chen Z, Fernandes AT, Hildebrandt P, Murgida DH, Todorovic S, Pereira MM, Melo EP & Martins LO (2008) Copper incorporation into recombinant CotA laccase from *Bacillus subtilis*: characterization of fully copper loaded enzymes. *JBIC Journal of Biological Inorganic Chemistry* **13**: 183-193.
- Eckhardt T (1978) A rapid method for the identification of plasmid desoxyribonucleic acid in bacteria. *Plasmid* **1**: 584-588.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.
- Ericsson AC & Franklin CL (2015) Manipulating the Gut Microbiota: Methods and Challenges. *ILAR J* **56**: 205-217.
- Escobar JS, Klotz B, Valdes BE & Agudelo GM (2014) The gut microbiota of Colombians differs from that of Americans, Europeans and Asians. *BMC Microbiol* **14**: 311.
- Fang H, Cai L, Yang Y, Ju F, Li X, Yu Y & Zhang T (2014) Metagenomic analysis reveals potential biodegradation pathways of persistent pesticides in freshwater and marine sediments. *Sci Total Environ* **470-471**: 983-992.
- Fang ZM, Li TL, Chang F, *et al.* (2012) A new marine bacterial laccase with chloride-enhancing, alkaline-dependent activity and dye decolorization ability. *Bioresour Technol* **111**: 36-41.
- Faraco V, Giardina P & Sannia G (2003) Metal-responsive elements in *Pleurotus ostreatus* laccase gene promoters. *Microbiology* **149**: 2155-2162.
- Feinsilver JM (1996) Prospección de la biodiversidad: potencialidades para los países en desarrollo. *Revista de la CEPAL*.
- Fernandez-Lopez R, Garcillan-Barcia MP, Revilla C, Lazaro M, Vielva L & de la Cruz F (2006) Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS Microbiol Rev* **30**: 942-966.

- Figurski DH & Helinski DR (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.
- Finan TM, Kunkel B, De Vos GF & Signer ER (1986) Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. *Journal of Bacteriology* **167**: 66-72.
- Francia MV, Varsaki A, Garcillan-Barcia MP, Latorre A, Drainas C & de la Cruz F (2004) A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiol Rev* **28**: 79-100.
- Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW & Delong EF (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* **105**: 3805-3810.
- Frisli T, Haverkamp TH, Jakobsen KS, Stenseth NC & Rudi K (2013) Estimation of metagenome size and structure in an experimental soil microbiota from low coverage next-generation sequence data. *J Appl Microbiol* **114**: 141-151.
- Frost LS & Koraimann G (2010) Regulation of bacterial conjugation: balancing opportunity with adversity. *Future Microbiol* **5**: 1057-1071.
- Frost LS, Leplae R, Summers AO & Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3**: 722-732.
- Galhaup C & Haltrich D (2001) Enhanced formation of laccase activity by the white-rot fungus *Trametes pubescens* in the presence of copper. *Applied Microbiology and Biotechnology* **56**: 225-232.
- Garcillan-Barcia MP, Francia MV & de la Cruz F (2009) The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol Rev* **33**: 657-687.
- Gaze WH, Abdoulsam N, Hawkey PM & Wellington EMH (2005) Incidence of Class 1 Integrons in a Quaternary Ammonium Compound-Polluted Environment. *Antimicrobial Agents and Chemotherapy* **49**: 1802-1807.
- Giardina P, Faraco V, Pezzella C, Piscitelli A, Vanhulle S & Sannia G (2010) Laccases: a never-ending story. *Cellular and Molecular Life Sciences* **67**: 369-385.
- Gillings MR (2014) Integrons: past, present, and future. *Microbiol Mol Biol Rev* **78**: 257-277.
- Gillings MR, Holley MP, Stokes H & Holmes AJ (2005) Integrons in *Xanthomonas*: a source of species genome diversity. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 4419-4424.
- Gillings MR, Xuejun D, Hardwick SA, Holley MP & Stokes HW (2009) Gene cassettes encoding resistance to quaternary ammonium compounds: a role in the origin of clinical class 1 integrons? *ISME J* **3**: 209-215.
- Gillings MR, Gaze WH, Pruden A, Smalla K, Tiedje JM & Zhu Y-G (2015) Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J* **9**: 1269-1279.

- Gillings MR, Gaze WH, Pruden A, Smalla K, Tiedje JM & Zhu YG (2015) Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J* **9**: 1269-1279.
- Giovannoni SJ, DeLong EF, Olsen GJ & Pace NR (1988) Phylogenetic group-specific oligodeoxynucleotide probes for identification of single microbial cells. *J Bacteriol* **170**: 720-726.
- Gogarten JP, Doolittle WF & Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* **19**: 2226-2238.
- Goldenfeld N & Woese C (2007) Biology's next revolution. *Nature* **445**: 369.
- Gootz TD (2010) The global problem of antibiotic resistance. *Crit Rev Immunol* **30**: 79-93.
- Goryshin IY & Reznikoff WS (1998) Tn5 in vitro transposition. *Journal of Biological Chemistry* **273**: 7367-7374.
- Gotz A, Pukall R, Smit E, *et al.* (1996) Detection and characterization of broad-host-range plasmids in environmental bacteria by PCR. *Appl Environ Microbiol* **62**: 2621-2628.
- Greated A, Lambertsen L, Williams PA & Thomas CM (2002) Complete sequence of the IncP-9 TOL plasmid pWW0 from *Pseudomonas putida*. *Environ Microbiol* **4**: 856-871.
- Green ED, Watson JD & Collins FS (2015) Human Genome Project: Twenty-five years of big biology. *Nature* **526**: 29-31.
- Guerra B, Soto S, Helmuth R & Mendoza MC (2002) Characterization of a self-transferable plasmid from *Salmonella enterica* serotype typhimurium clinical isolates carrying two integron-borne gene cassettes together with virulence and drug resistance genes. *Antimicrob Agents Chemother* **46**: 2977-2981.
- Guglielmini J, Quintais L, Garcillan-Barcia MP, de la Cruz F & Rocha EP (2011) The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet* **7**: e1002222.
- Guglielmini J, Neron B, Abby SS, Garcillan-Barcia MP, de la Cruz F & Rocha EP (2014) Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res* **42**: 5715-5727.
- Gunne M, Al-Sultani D & Urlacher VB (2013) Enhancement of copper content and specific activity of CotA laccase from *Bacillus licheniformis* by coexpression with CopZ copper chaperone in *E. coli*. *Journal of biotechnology* **168**: 252-255.
- Halary S, Leigh JW, Cheaib B, Lopez P & Baptiste E (2010) Network analyses structure genetic diversity in independent genetic worlds. *Proc Natl Acad Sci U S A* **107**: 127-132.
- Han JE, Kim JH, Choresca CH, Jr., Shin SP, Jun JW, Chai JY & Park SC (2012) Prevalence of tet gene and complete genome sequencing of tet gene-encoded plasmid (pAHH01) isolated from *Aeromonas* species in South Korea. *J Appl Microbiol* **112**: 631-638.

Handelsman J, Rondon MR, Brady SF, Clardy J & Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**: R245-249.

Harayama S & Rekik M (1990) The meta cleavage operon of TOL degradative plasmid pWWO comprises 13 genes. *Mol Gen Genet* **221**: 113-120.

Harmer CJ & Hall RM (2015) The A to Z of A/C plasmids. *Plasmid* **80**: 63-82.

Harrison E & Brockhurst MA (2012) Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends Microbiol* **20**: 262-267.

Hartley RW (1989) Barnase and barstar: two small proteins to fold and fit together. *Trends Biochem Sci* **14**: 450-454.

Heeg K, Pohl M, Sontag M, Mumme J, Klocke M & Nettmann E (2014) Microbial communities involved in biogas production from wheat straw as the sole substrate within a two-phase solid-state anaerobic digestion. *Syst Appl Microbiol* **37**: 590-600.

Hefford MA, Kobayashi Y, Allard SE, Forster RJ & Teather RM (1997) Sequence analysis and characterization of pOM1, a small cryptic plasmid from *Butyrivibrio fibrisolvens*, and its use in construction of a new family of cloning vectors for *Butyrivibrios*. *Appl Environ Microbiol* **63**: 1701-1711.

Herigstad B, Hamilton M, Heersink J (2001) How to optimize the drop plate method for enumerating bacteria. *Journal of Microbiological Methods*, **44**: 121-129.

Heuer H & Smalla K (2012) Plasmids foster diversification and adaptation of bacterial populations in soil. *FEMS Microbiol Rev* **36**: 1083-1104.

Howe AC, Jansson JK, Malfatti SA, Tringe SG, Tiedje JM & Brown CT (2014) Tackling soil diversity with the assembly of large, complex metagenomes. *Proc Natl Acad Sci U S A* **111**: 4904-4909.

Hoyle DV, Davison HC, Knight HI, *et al.* (2006) Molecular characterisation of bovine faecal *Escherichia coli* shows persistence of defined ampicillin resistant strains and the presence of class 1 integrons on an organic beef farm. *Vet Microbiol* **115**: 250-257.

Hynes MF, Simon R & Puhler A (1985) The development of plasmid-free strains of *Agrobacterium tumefaciens* by using incompatibility with a *Rhizobium meliloti* plasmid to eliminate pAtC58. *Plasmid* **13**: 99-105.

Hynes MF, Simon R, Müller P, Niehaus K, Labes M & Pühler A The two megaplasmids of *Rhizobium meliloti* are involved in the effective nodulation of alfalfa. *Molecular and General Genetics MGG* **202**: 356-362.

Ihsen J, Reiss R, Luchsinger R, Thöny-Meyer L & Richter M (2015) Biochemical properties and yields of diverse bacterial laccase-like multicopper oxidases expressed in *Escherichia coli*. *Scientific reports* **5**.

Ilangovan A, Connery S & Waksman G (2015) Structural biology of the Gram-negative bacterial conjugation systems. *Trends Microbiol* **23**: 301-310.

Izmalkova T, Sazonova OI, Sokolov SL, Kosheleva IA & Boronin AM (2005) [Incompatibility group P-7 plasmids responsible for biodegradation of naphthalene and salicylate in fluorescent pseudomonads]. *Mikrobiologiya* **74**: 342-348.

Izmalkova TY, Sazonova O, Sokolov S, Kosheleva I & Boronin A (2005) The P-7 incompatibility group plasmids responsible for biodegradation of naphthalene and salicylate in fluorescent pseudomonads. *Microbiology* **74**: 290-295.

Jechalke S, Schreiter S, Wolters B, Dealtry S, Heuer H & Smalla K (2013) Widespread dissemination of class 1 integron components in soils and related ecosystems as revealed by cultivation-independent analysis. *Front Microbiol* **4**: 420.

Johnson KA (2013) A century of enzyme kinetic analysis, 1913 to 2013. *FEBS Lett* **587**: 2753-2766.

Johnson CM & Grossman AD (2015) Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annu Rev Genet*.

Jonas DA, Elmadfa I, Engel KH, *et al.* (2001) Safety considerations of DNA in food. *Ann Nutr Metab* **45**: 235-254.

Jones BV & Marchesi JR (2007) Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Nat Methods* **4**: 55-61.

Jones BV, Sun F & Marchesi JR (2010) Comparative metagenomic analysis of plasmid encoded functions in the human gut microbiome. *BMC Genomics* **11**: 46.

Jorgensen TS, Kiil AS, Hansen MA, Sorensen SJ & Hansen LH (2014) Current strategies for mobilome research. *Front Microbiol* **5**: 750.

Jouanin L, De Lajudie P, Bazetoux S & Huguet T (1981) DNA sequence homology in *Rhizobium meliloti* plasmids. *Mol Gen Genet* **182**: 189-195.

Ju F, Guo F, Ye L, Xia Y & Zhang T (2014) Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. *Environ Microbiol Rep* **6**: 80-89.

Karanasios E, Tsiropoulos NG & Karpouzias DG (2012) On-farm biopurification systems for the depuration of pesticide wastewaters: recent biotechnological advances and future perspectives. *Biodegradation* **23**: 787-802.

Kazama H, Hamashima H, Sasatsu M & Arai T (1998) Distribution of the antiseptic-resistance gene *qacE delta 1* in gram-positive bacteria. *FEMS Microbiol Lett* **165**: 295-299.

Kergourlay G, Taminiau B, Daube G & Champomier Verges MC (2015) Metagenomic insights into the dynamics of microbial communities in food. *Int J Food Microbiol* **213**: 31-39.

- Kirchner O & Tauch A (2003) Tools for genetic engineering in the amino acid-producing bacterium *Corynebacterium glutamicum*. *J Biotechnol* **104**: 287-299.
- Kloesges T, Popa O, Martin W & Dagan T (2011) Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. *Mol Biol Evol* **28**: 1057-1074.
- Krasowiak R, Smalla K, Sokolov S, Kosheleva I, Sevastyanovich Y, Titok M & Thomas CM (2002) PCR primers for detection and characterisation of IncP-9 plasmids. *FEMS Microbiol Ecol* **42**: 217-225.
- Kodzius R & Gojobori T (2015) Marine metagenomics as a source for bioprospecting. *Mar Genomics*.
- Kohler KA, Ruckert C, Schatschneider S, *et al.* (2013) Complete genome sequence of *Pseudomonas* sp. strain VLB120 a solvent tolerant, styrene degrading bacterium, isolated from forest soil. *J Biotechnol* **168**: 729-730.
- Kopf A & Bica M & Kottmann R, *et al.* (2015) The ocean sampling day consortium. *Gigascience* **4**: 27.
- Koschorreck K, Richter SM, Ene AB, Roduner E, Schmid RD & Urlacher VB (2008) Cloning and characterization of a new laccase from *Bacillus licheniformis* catalyzing dimerization of phenolic acids. *Appl Microbiol Biotechnol* **79**: 217-224.
- Kotra LP, Haddad J & Mobashery S (2000) Aminoglycosides: perspectives on mechanisms of action and resistance and strategies to counter resistance. *Antimicrob Agents Chemother* **44**: 3249-3256.
- Kunamneni A, Plou FJ, Ballesteros A & Alcalde M (2008) Laccases and their applications: a patent review. *Recent patents on biotechnology* **2**: 10-24.
- Kristiansson E, Fick J, Janzon A, *et al.* (2011) Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements. *PLoS ONE* **6**: e17038.
- Kumar A & Singh S (2013) Directed evolution: tailoring biocatalysts for industrial applications. *Crit Rev Biotechnol* **33**: 365-378.
- Kumari P, Choi HL & Sudiarto SI (2015) Assessment of Bacterial Community Assembly Patterns and Processes in Pig Manure Slurry. *PLoS ONE* **10**: e0139437.
- Kunz DA & Chapman PJ (1981) Isolation and characterization of spontaneously occurring TOL plasmid mutants of *Pseudomonas putida* HS1. *J Bacteriol* **146**: 952-964.
- Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.
- Laird SA, Wynberg R & Laird SA (2002) Institutional policies for biodiversity research. *Biodiversity and Traditional Knowledge, Equitable Partnerships in Practice*, Londres, Earthscan Publications Ltd 39-76.

- Lawley TD, Klimke WA, Gubbins MJ & Frost LS (2003) F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett* **224**: 1-15.
- Lawrence JG (2005) Horizontal and vertical gene transfer: the life history of pathogens. *Contrib Microbiol* **12**: 255-271.
- Layton AC, Sanseverino J, Wallace W, Corcoran C & Sayler GS (1992) Evidence for 4-chlorobenzoic acid dehalogenation mediated by plasmids related to pSS50. *Appl Environ Microbiol* **58**: 399-402.
- Lederberg J (1952) Cell genetics and hereditary symbiosis. *Physiol Rev* **32**: 403-430.
- Leverstein-Van Hall MA, Paauw A, Box AT, Blok HE, Verhoef J & Fluit AC (2002) Presence of integron-associated resistance in the community is widespread and contributes to multidrug resistance in the hospital. *J Clin Microbiol* **40**: 3038-3040.
- Levesque C, Piche L, Larose C & Roy PH (1995) PCR mapping of integrons reveals several novel combinations of resistance genes. *Antimicrobial agents and chemotherapy* **39**: 185-191.
- Levine M (1918) Differentiation of B. Coli and B. Aerogenes on a Simplified Eosin-Methylene Blue Agar. *The Journal of Infectious Diseases* **23**: 43-47.
- Levings RS, Djordjevic SP & Hall RM (2008) SGI2, a relative of Salmonella genomic island SGI1 with an independent origin. *Antimicrob Agents Chemother* **52**: 2529-2537.
- Li LL, Norman A, Hansen LH & Sorensen SJ (2012) Metamobilomics--expanding our knowledge on the pool of plasmid encoded traits in natural environments using high-throughput sequencing. *Clin Microbiol Infect* **18 Suppl 4**: 5-7.
- Li PL, Everhart DM & Farrand SK (1998) Genetic and sequence analysis of the pTIC58 trb locus, encoding a mating-pair formation system related to members of the type IV secretion family. *J Bacteriol* **180**: 6164-6172.
- Li W, Shi J, Wang X, *et al.* (2004) Complete nucleotide sequence and organization of the naphthalene catabolic plasmid pND6-1 from Pseudomonas sp. strain ND6. *Gene* **336**: 231-240.
- Liu B & Pop M (2009) ARDB--Antibiotic Resistance Genes Database. *Nucleic Acids Res* **37**: D443-447.
- Liu YF, Wang CH, Janapatla RP, Fu HM, Wu HM & Wu JJ (2007) Presence of plasmid pA15 correlates with prevalence of constitutive MLS(B) resistance in group A streptococcal isolates at a university hospital in southern Taiwan. *J Antimicrob Chemother* **59**: 1167-1170.
- Loftie-Eaton W & Rawlings DE (2012) Diversity, biology and evolution of IncQ-family plasmids. *Plasmid* **67**: 15-34.
- Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J & Pallen MJ (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* **30**: 434-439.

Lopes GV, Michael GB, Cardoso M & Schwarz S (2014) Identification and characterization of *Salmonella enterica* subsp. *enterica* serovar Derby isolates carrying a new aadA26 gene cassette in a class 1 integron obtained at pig slaughterhouses. *FEMS Microbiol Lett* **356**: 71-78.

Lorenzo-Díaz F, Fernández-López C, Garcillán-Barcia MP & Espinosa M (2014) Bringing them together: Plasmid pMV158 rolling circle replication and conjugation under an evolutionary perspective. *Plasmid* **74**: 15-31.

Lozano MJ, Salas ME, Giusti Mde L, Martini MC, Lopez JL, Salto I, Del Papa MF, Pistorio M & Lagares A (2013) Novel tnpR-based transposable promoter traps suitable for RIVET studies in different gram-negative bacteria. *J Microbiol Methods* **93**: 9-11.

Llosa M, Bolland S & de la Cruz F (1991) Structural and functional analysis of the origin of conjugal transfer of the broad-host-range IncW plasmid R388 and comparison with the related IncN plasmid R46. *Mol Gen Genet* **226**: 473-483.

Ma Y, Paulsen IT & Palenik B (2012) Analysis of two marine metagenomes reveals the diversity of plasmids in oceanic environments. *Environ Microbiol* **14**: 453-466.

Ma YF, Wu JF, Wang SY, *et al.* (2007) Nucleotide sequence of plasmid pCNB1 from *Comamonas* strain CNB-1 reveals novel genetic organization and evolution for 4-chloronitrobenzene degradation. *Appl Environ Microbiol* **73**: 4477-4483.

Macconkey A (1905) Lactose-Fermenting Bacteria in Faeces. *J Hyg (Lond)* **5**: 333-379.

Madhavi V & Lele S (2009) Laccase: properties and applications. *BioResources* **4**: 1694-1717.

Maeda K, Nojiri H, Shintani M, Yoshida T, Habe H & Omori T (2003) Complete nucleotide sequence of carbazole/dioxin-degrading plasmid pCAR1 in *Pseudomonas resinovorans* strain CA10 indicates its mosaicity and the presence of large catabolic transposon Tn4676. *J Mol Biol* **326**: 21-33.

Malik A, Celik EK, Bohn C, Bockelmann U, Knobel K & Grohmann E (2008) Detection of conjugative plasmids and antibiotic resistance genes in anthropogenic soils from Germany and India. *FEMS Microbiol Lett* **279**: 207-216.

Marrs B, Delagrave S & Murphy D (1999) Novel approaches for discovering industrial enzymes. *Curr Opin Microbiol* **2**: 241-245.

Martinez-Garcia E, Calles B, Arevalo-Rodriguez M & de Lorenzo V (2011) pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiol* **11**: 38.

Martinez B, Tomkins J, Wackett LP, Wing R & Sadowsky MJ (2001) Complete nucleotide sequence and organization of the atrazine catabolic plasmid pADP-1 from *Pseudomonas* sp. strain ADP. *J Bacteriol* **183**: 5684-5697.

Martínez B, Tomkins J, Wackett LP, Wing R & Sadowsky MJ (2001) Complete nucleotide sequence and organization of the atrazine catabolic plasmid pADP-1 from *Pseudomonas* sp. strain ADP. *J Bacteriol* **183**: 5684-5697.

Martini MC, Albicoro FJ, Nour E, *et al.* (2015) Characterization of a collection of plasmid-containing bacteria isolated from an on-farm biopurification system used for pesticide removal. *Plasmid* **80**: 16-23.

Maxam AM & Gilbert W (1977) A new method for sequencing DNA. *Proc Natl Acad Sci U S A* **74**: 560-564.

Mazel D (2006) Integrons: agents of bacterial evolution. *Nat Rev Microbiol* **4**: 608-620.

Mazel D, Dychinco B, Webb VA & Davies J (2000) Antibiotic resistance in the ECOR collection: integrons and identification of a novel aad gene. *Antimicrob Agents Chemother* **44**: 1568-1574.

Mazel D & Davies J (1999) Antibiotic resistance in microbes. *Cellular and Molecular Life Sciences CMLS* **56**: 742-754.

Mell JC & Redfield RJ (2014) Natural competence and the evolution of DNA uptake specificity. *J Bacteriol* **196**: 1471-1483.

Metzker ML (2005) Emerging technologies in DNA sequencing. *Genome Res* **15**: 1767-1776.

Meyer F, Goesmann A, McHardy AC, *et al.* (2003) GenDB--an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31**: 2187-2195.

Michael GB, Cardoso M & Schwarz S (2005) Class 1 integron-associated gene cassettes in *Salmonella enterica* subsp. *enterica* serovar Agona isolated from pig carcasses in Brazil. *J Antimicrob Chemother* **55**: 776-779.

Miller J (1972) Experiments in Molecular Genetics. *Cold Spring Harbor, NY*.

Miller J (1972) Experiments in Molecular Genetics. *Cold Spring Harbor Laboratory*.

Mindlin S, Kholodii G, Gorlenko Z, *et al.* (2001) Mercury resistance transposons of gram-negative environmental bacteria and their classification. *Res Microbiol* **152**: 811-822.

Mingeot-Leclercq MP, Glupczynski Y & Tulkens PM (1999) Aminoglycosides: activity and resistance. *Antimicrob Agents Chemother* **43**: 727-737.

Misra TK (1992) Bacterial resistances to inorganic mercury salts and organomercurials. *Plasmid* **27**: 4-16.

More MI, Finger LD, Stryker JL, Fuqua C, Eberhard A & Winans SC (1996) Enzymatic synthesis of a quorum-sensing autoinducer through use of defined substrates. *Science* **272**: 1655-1658.

Morey M, Fernandez-Marmiesse A, Castineiras D, Fraga JM, Couce ML & Cocho JA (2013) A glimpse into past, present, and future DNA sequencing. *Mol Genet Metab* **110**: 3-24.

Moyer CL, Dobbs FC & Karl DM (1994) Estimation of diversity and community structure through restriction fragment length polymorphism distribution analysis of bacterial 16S rRNA genes from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. *Appl Environ Microbiol* **60**: 871-879.

- Mulder N & Apweiler R (2007) InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol Biol* **396**: 59-70.
- Muyzer G, de Waal EC & Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**: 695-700.
- Nagahari K & Sakaguchi K (1978) RSF1010 plasmid as a potentially useful vector in *Pseudomonas* species. *J Bacteriol* **133**: 1527-1529.
- Nakamura Y, Itoh T, Matsuda H & Gojobori T (2004) Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet* **36**: 760-766.
- Nakatsu C, Ng J, Singh R, Straus N & Wyndham C (1991) Chlorobenzoate catabolic transposon Tn5271 is a composite class I element with flanking class II insertion sequences. *Proc Natl Acad Sci U S A* **88**: 8312-8316.
- Nardelli M, Scalzo PM, Ramírez MS, Quiroga MP, Cassini MH & Centrón D (2012) Class 1 Integrons in Environments with Different Degrees of Urbanization. *PLoS One* **7**: e39223.
- Nicolas E, Lambin M, Dandoy D, Galloy C, Nguyen N, Oger CA & Hallet B (2015) The Tn3-family of Replicative Transposons. *Microbiology spectrum*, **3** (4).
- Nikokar I, Tishayar A, Flakiyan Z, *et al.* (2013) Antibiotic resistance and frequency of class 1 integrons among *Pseudomonas aeruginosa*, isolated from burn patients in Guilan, Iran. *Iranian Journal of Microbiology* **5**: 36-41.
- Nojiri H, Shintani M & Omori T (2004) Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. *Appl Microbiol Biotechnol* **64**: 154-174.
- Norman A, Riber L, Luo W, Li LL, Hansen LH & Sorensen SJ (2014) An improved method for including upper size range plasmids in metamobilomes. *PLoS ONE* **9**: e104405.
- Novick RP (1987) Plasmid incompatibility. *Microbiol Rev* **51**: 381-395.
- Ochman H, Lawrence JG & Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299-304.
- Ono A, Miyazaki R, Sota M, Ohtsubo Y, Nagata 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.
- Orman BE, Pineiro SA, Arduino S, *et al.* (2002) Evolution of multiresistance in nontyphoid salmonella serovars from 1984 to 1998 in Argentina. *Antimicrob Agents Chemother* **46**: 3963-3970.
- Palenik B, Ren Q, Tai V & Paulsen IT (2009) Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol* **11**: 349-359.

- Pansegrau W, Lanka E, Barth PT, *et al.* (1994) Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. *J Mol Biol* **239**: 623-663.
- Paulsen IT, Littlejohn TG, Radstrom P, Sundstrom L, Skold O, Swedberg G & Skurray RA (1993) The 3' conserved segment of integrons contains a gene associated with multidrug resistance to antiseptics and disinfectants. *Antimicrob Agents Chemother* **37**: 761-768.
- Penades JR, Chen J, Quiles-Puchalt N, Carpena N & Novick RP (2015) Bacteriophage-mediated spread of bacterial virulence genes. *Curr Opin Microbiol* **23**: 171-178.
- Perez-Mendoza D, Sepulveda E, Pando V, *et al.* (2005) Identification of the *rctA* gene, which is required for repression of conjugative transfer of rhizobial symbiotic megaplasmids. *J Bacteriol* **187**: 7341-7350.
- Perez-Valera E, Goberna M & Verdu M (2015) Phylogenetic structure of soil bacterial communities predicts ecosystem functioning. *FEMS Microbiol Ecol* **91**.
- Perez F, Hujer AM, Marshall SH, *et al.* (2014) Extensively drug-resistant *Pseudomonas aeruginosa* isolates containing blaVIM-2 and elements of Salmonella genomic island 2: a new genetic resistance determinant in Northeast Ohio. *Antimicrob Agents Chemother* **58**: 5929-5935.
- Perry JA & Wright GD (2013) The antibiotic resistance "mobilome": searching for the link between environment and clinic. *Front Microbiol* **4**: 138.
- Pezzella C, Guarino L & Piscitelli A (2015) How to enjoy laccases. *Cellular and Molecular Life Sciences* **72**: 923-940.
- Phan MD & Wain J (2008) IncHI plasmids, a dynamic link between resistance and pathogenicity. *J Infect Dev Ctries* **2**: 272-278.
- Phugare SS, Kalyani DC, Patil AV & Jadhav JP (2011) Textile dye degradation by bacterial consortium and subsequent toxicological analysis of dye and dye metabolites using cytotoxicity, genotoxicity and oxidative stress studies. *Journal of Hazardous Materials* **186**: 713-723.
- Phuong Hoa PT, Nonaka L, Hung Viet P & Suzuki S (2008) Detection of the *sul1*, *sul2*, and *sul3* genes in sulfonamide-resistant bacteria from wastewater and shrimp ponds of north Vietnam. *Sci Total Environ* **405**: 377-384.
- Pinto AP, Rodrigues SC, Caldeira AT & Teixeira DM (2016) Exploring the potential of novel biomixtures and *Lentinula edodes* fungus for the degradation of selected pesticides. Evaluation for use in biobed systems. *Sci Total Environ* **541**: 1372-1381.
- Post V, Recchia GD & Hall RM (2007) Detection of gene cassettes in Tn402-like class 1 integrons. *Antimicrobial agents and chemotherapy* **51**: 3467-3468.
- Price MN, Dehal PS & Arkin AP (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Providenti MA, Shaye RE, Lynes KD, *et al.* (2006) The locus coding for the 3-nitrobenzoate dioxygenase of *Comamonas* sp. strain JS46 is flanked by IS1071 elements and is subject to deletion and inversion events. *Appl Environ Microbiol* **72**: 2651-2660.

Puspita ID, Kamagata Y, Tanaka M, Asano K & Nakatsu CH (2012) Are uncultivated bacteria really uncultivable? *Microbes Environ* **27**: 356-366.

Quiroga MP, Andres P, Petroni A, Bistué AJS, Guerriero L, Vargas LJ & Galas M. (2007) Complex class 1 integrons with diverse variable regions, including *aac* (6')-Ib-cr, and a novel allele, *qnrB10*, associated with ISCR1 in clinical enterobacterial isolates from Argentina. *Antimicrobial agents and chemotherapy*, *51*(12), 4466-4470.

Quiroga MP, Arduino SM, Merkier AK, Quiroga C, Petroni A, Roy PH & Centron D (2013) "Distribution and functional identification of complex class 1 integrons". *Infect Genet Evol* **19**: 88-96.

Quiroga MP, Arduino SM, Merkier AK, *et al.* (2013) Distribution and functional identification of complex class 1 integrons. *Infection, Genetics and Evolution* **19**: 88-96.

Rahman NHA, Aziz SA & Hassan MA (2013) Production of ligninolytic enzymes by newly isolated bacteria from palm oil plantation soils. *BioResources*, *8*(4), 6136-6150.

Rajkumar B, Sharma GD & Paul AK (2012) Isolation and characterization of heavy metal resistant bacteria from Barak River contaminated with pulp paper mill effluent, South Assam. *Bull Environ Contam Toxicol* **89**: 263-268.

Rankin DJ, Rocha EP & Brown SP (2011) What traits are carried on mobile genetic elements, and why? *Heredity (Edinb)* **106**: 1-10.

Reniero D, Mozzon E, Galli E & Barbieri P (1998) Two aberrant mercury resistance transposons in the *Pseudomonas stutzeri* plasmid pPB. *Gene* **208**: 37-42.

Reuter JA, Spacek DV & Snyder MP (2015) High-throughput sequencing technologies. *Mol Cell* **58**: 586-597.

Roberts MC (2008) Update on macrolide–lincosamide–streptogramin, ketolide, and oxazolidinone resistance genes. *FEMS microbiology letters*, *282*(2), 147-159.

Robinson CJ, Bohannan BJ & Young VB (2010) From structure to function: the ecology of host-associated microbial communities. *Microbiol Mol Biol Rev* **74**: 453-476.

Romero D, Martínez-Salazar J, Ortiz E, Rodríguez C & Valencia-Morales E (1999) Repeated sequences in bacterial chromosomes and plasmids: a glimpse from sequenced genomes. *Research in microbiology* **150**: 735-743.

Rosenthal A, Coutelle O & Craxton M (1993) Large-scale production of DNA sequencing templates by microtitre format PCR. *Nucleic Acids Res* **21**: 173-174.

Rosewarne CP, Pettigrove V, Stokes HW & Parsons YM (2010) *Class 1 integrons in benthic bacterial communities: abundance, association with Tn402-like transposition modules and evidence for coselection with heavy-metal resistance.*

Roske I, Sabra W, Nacke H, Daniel R, Zeng AP, Antranikian G & Sahm K (2014) Microbial community composition and dynamics in high-temperature biogas reactors using industrial bioethanol waste as substrate. *Appl Microbiol Biotechnol* **98**: 9095-9106.

Ross EM, Petrovski S, Moate PJ & Hayes BJ (2013) Metagenomics of rumen bacteriophage from thirteen lactating dairy cattle. *BMC Microbiol* **13**: 242.

Rosser SJ & Young H-K (1999) Identification and characterization of class 1 integrons in bacteria from an aquatic environment. *Journal of Antimicrobial Chemotherapy* **44**: 11-18.

Sakai Y, Ogawa N, Shimomura Y & Fujii T (2014) A 2, 4-dichlorophenoxyacetic acid degradation plasmid pM7012 discloses distribution of an unclassified megaplasmid group across bacterial species. *Microbiology*, **160**(3), 525-536.

Sambrook J, Fritsch EF & Maniatis T (1989) *Molecular cloning : a laboratory manual*.

Sanger F & Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94**: 441-448.

Saunders AM, Albertsen M, Vollertsen J & Nielsen PH (2015) The activated sludge ecosystem contains a core community of abundant organisms. *ISME J*.

Schafer A, Tauch A, Jager W, Kalinowski J, Thierbach G & Puhler A (1994) Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene* **145**: 69-73.

Schlüter A, Krause L, Szczepanowski R, Goesmann A & Pühler A (2008) Genetic diversity and composition of a plasmid metagenome from a wastewater treatment plant. *Journal of biotechnology*, **136**(1), 65-76.

Schmieder R & Edwards R (2012) Insights into antibiotic resistance through metagenomic approaches. *Future Microbiol* **7**: 73-89.

Sebastianelli A & Bruce, IJ (2007) Tn5530 from Burkholderia cepacia strain 2a encodes a chloride channel protein essential for the catabolism of 2, 4-dichlorophenoxyacetic acid. *Environmental microbiology*, **9**(1), 256-265.

Segura A, Molina L & Ramos JL (2014) Plasmid-Mediated Tolerance Toward Environmental Pollutants. *Microbiol Spectr* **2**.

Selbitschka W, Dresing U, Hagen M, Niemann S & Pühler A (1995) A biological containment system for Rhizobium meliloti based on the use of recombination-deficient (recA-) strains. *FEMS Microbiology Ecology* **16**: 223-232.

Sentchilo V, Mayer AP, Guy L, *et al.* (2013) Community-wide plasmid gene mobilization and selection. *ISME J* **7**: 1173-1186.

Serban DE (2015) Microbiota in Inflammatory Bowel Disease Pathogenesis and Therapy: Is It All About Diet? *Nutr Clin Pract*.

Shahcheraghi F, Rahmati Ghezalgeh F, Nobari S, Torabi E, Mousavi SF, Aslani MM & Shahcheraghi F (2014) Identification and characterization of class 1 integrons among atypical enteropathogenic *Escherichia coli* isolated from children under 5 years of age. *Iranian Journal of Microbiology* **6**: 156-162.

Sharma A & Thakur IS (2009) Identification and characterization of integron mediated antibiotic resistance in pentachlorophenol degrading bacterium isolated from the chemostat. *J Environ Sci (China)* **21**: 858-864.

Sheldon AT, Jr. (2005) Antibiotic resistance: a survival strategy. *Clin Lab Sci* **18**: 170-180.

Shin DH, Kim DU, Seong CN, Song HG & Ka JO (2012) Genetic and phenotypic diversity of carbofuran-degrading bacteria isolated from agricultural soils. *J Microbiol Biotechnol* **22**: 448-456.

Siguier P, Perochon J, Lestrade L, Mahillon J & Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**: D32-36.

Simon R, Priefer U & Puhler A (1983) A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Nat Biotech* **1**: 784-791.

Smalla K, Heuer H, Gotz A, Niemeyer D, Krogerrecklenfort E & Tietze E (2000) Exogenous isolation of antibiotic resistance plasmids from piggery manure slurries reveals a high prevalence and diversity of IncQ-like plasmids. *Appl Environ Microbiol* **66**: 4854-4862.

Smalla K, Heuer H, Götz A, Niemeyer D, Krögerrecklenfort E & Tietze E (2000) Exogenous Isolation of Antibiotic Resistance Plasmids from Piggery Manure Slurries Reveals a High Prevalence and Diversity of IncQ-Like Plasmids. *Applied and Environmental Microbiology* **66**: 4854-4862.

Smalla K, Haines AS, Jones K, Krogerrecklenfort E, Heuer H, Schloter M & Thomas CM (2006) Increased abundance of IncP-1beta plasmids and mercury resistance genes in mercury-polluted river sediments: first discovery of IncP-1beta plasmids with a complex mer transposon as the sole accessory element. *Appl Environ Microbiol* **72**: 7253-7259.

Smillie C, Garcillan-Barcia MP, Francia MV, Rocha EP & de la Cruz F (2010) Mobility of plasmids. *Microbiol Mol Biol Rev* **74**: 434-452.

Smit E, Wolters A & van Elsas JD (1998) Self-Transmissible Mercury Resistance Plasmids with Gene-Mobilizing Capacity in Soil Bacterial Populations: Influence of Wheat Roots and Mercury Addition. *Applied and Environmental Microbiology* **64**: 1210-1219.

Snell-Castro R, Godon JJ, Delgenes JP & Dabert P (2005) Characterisation of the microbial diversity in a pig manure storage pit using small subunit rDNA sequence analysis. *FEMS Microbiol Ecol* **52**: 229-242.

Sniegowski K & Springael D (2015) Establishment of multiple pesticide biodegradation capacities from pesticide-primed materials in on-farm biopurification system microcosms treating complex pesticide-contaminated wastewater. *Pest Manag Sci* **71**: 986-995.

- Soden DM & Dobson AD (2001) Differential regulation of laccase gene expression in *Pleurotus sajor-caju*. *Microbiology* **147**: 1755-1763.
- Stokes HW, Holmes AJ, Nield BS, Holley M P, Nevalainen KH, Mabbutt BC & Gillings MR (2001) Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Applied and environmental microbiology*, *67*(11), 5240-5246.
- Soler Bistue AJ, Birshan D, Tomaras AP, *et al.* (2008) *Klebsiella pneumoniae* multiresistance plasmid pMET1: similarity with the *Yersinia pestis* plasmid pCRY and integrative conjugative elements. *PLoS ONE* **3**: e1800.
- Song X, Sun J, Mikalsen T, Roberts AP & Sundsfjord A (2013) Characterisation of the plasmidome within *Enterococcus faecalis* isolated from marginal periodontitis patients in Norway. *PLoS ONE* **8**: e62248.
- Soto MJ, Zorzano A, Olivares J & Toro N (1992) Sequence of ISRm4 from *Rhizobium meliloti* strain GR4. *Gene* **120**: 125-126.
- Stahl DA, Lane DJ, Olsen GJ & Pace NR (1984) Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. *Science* **224**: 409-411.
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-2690.
- Stokes HW & Gillings MR (2011) Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS microbiology reviews* **35**: 790-819.
- Stokes HW & Gillings MR (2011) Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiol Rev* **35**: 790-819.
- Stokes HW, Nesbø CL, Holley M, Bahl MI, Gillings MR & Boucher Y (2006) Class 1 Integrons Potentially Predating the Association with Tn402-Like Transposition Genes Are Present in a Sediment Microbial Community. *Journal of Bacteriology* **188**: 5722-5730.
- Stolze Y, Zakrzewski M, Maus I, *et al.* (2015) Comparative metagenomics of biogas-producing microbial communities from production-scale biogas plants operating under wet or dry fermentation conditions. *Biotechnol Biofuels* **8**: 14.
- Suresh PS, Kumar A, Kumar R & Singh VP (2008) An in silico [correction of insilico] approach to bioremediation: laccase as a case study. *J Mol Graph Model* **26**: 845-849.
- Szczepanowski R, Bekel T, Goesmann A, *et al.* (2008) Insight into the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to antimicrobial drugs analysed by the 454-pyrosequencing technology. *J Biotechnol* **136**: 54-64.
- Tabata M, Endo R, Ito M, Ohtsubo Y, Kumar A, Tsuda M & Nagata Y (2011) The *lin* genes for gamma-hexachlorocyclohexane degradation in *Sphingomonas* sp. MM-1 proved to be dispersed across multiple plasmids. *Biosci Biotechnol Biochem* **75**: 466-472.

Taghavi S, Mergeay M & van der Lelie D (1997) Genetic and physical maps of the *Alcaligenes eutrophus* CH34 megaplasmid pMOL28 and its derivative pMOL50 obtained after temperature-induced mutagenesis and mortality. *Plasmid* **37**: 22-34.

Tan HM (1999) Bacterial catabolic transposons. *Appl Microbiol Biotechnol* **51**: 1-12.

Tennstedt T, Szczepanowski R, Krahn I, Puhler A & Schluter A (2005) Sequence of the 68,869 bp IncP-1alpha plasmid pTB11 from a waste-water treatment plant reveals a highly conserved backbone, a Tn402-like integron and other transposable elements. *Plasmid* **53**: 218-238.

Thurlow LR, Thomas VC & Hancock LE (2009) Capsular polysaccharide production in *Enterococcus faecalis* and contribution of CpsF to capsule serospecificity. *J Bacteriol* **191**: 6203-6210.

Top EM, Springael D & Boon N (2002) Catabolic mobile genetic elements and their potential use in bioaugmentation of polluted soils and waters. *FEMS Microbiology Ecology*, *42*(2), 199-208.

Torstensson L & Castillo MDP (1997) Use of biobeds in Sweden to minimize environmental spillages from agricultural spraying equipment *Pest Outlook* **8**: 24-24.

Tortella GR, Mella-Herrera RA, Sousa DZ, Rubilar O, Acuna JJ, Briceno G & Diez MC (2013) Atrazine dissipation and its impact on the microbial communities and community level physiological profiles in a microcosm simulating the biomixture of on-farm biopurification system. *J Hazard Mater* **260**: 459-467.

Traglia GM, Almuzara M, Merzier AK, *et al.* (2012) *Achromobacter xylosoxidans*: an emerging pathogen carrying different elements involved in horizontal genetic transfer. *Current microbiology* **65**: 673-678.

Trefault N, De la Iglesia R, Molina AM, *et al.* (2004) Genetic organization of the catabolic plasmid pJP4 from *Ralstonia eutropha* JMP134 (pJP4) reveals mechanisms of adaptation to chloroaromatic pollutants and evolution of specialized chloroaromatic degradation pathways. *Environ Microbiol* **6**: 655-668.

Troster M, Felisberto-Rodrigues C, Christie PJ & Waksman G (2014) Recent advances in the structural and molecular biology of type IV secretion systems. *Current opinion in structural biology* **27**: 16-23.

Tung WL & Chow KC (1995) A modified medium for efficient electrotransformation of *E. coli*. *Trends Genet* **11**: 128-129.

Van de Peer Y, Jansen J, De Rijk P & De Wachter R (1997) Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Res* **25**: 111-116.

Vedler E, Koiv V & Heinaru A (2000) Analysis of the 2,4-dichlorophenoxyacetic acid-degradative plasmid pEST4011 of *Achromobacter xylosoxidans* subspecies *denitrificans* strain EST4002. *Gene* **255**: 281-288.

Verhagen P, Destino C, Boon N & De Gelder L (2015) Spatial heterogeneity in degradation characteristics and microbial community composition of pesticide biopurification systems. *J Appl Microbiol* **118**: 368-378.

Vishnivetskaya TA, Hamilton-Brehm SD, Podar M, *et al.* (2015) Community analysis of plant biomass-degrading microorganisms from Obsidian Pool, Yellowstone National Park. *Microb Ecol* **69**: 333-345.

Vlassak KM, Vanderleyden J & Graham PH (1997) Factors Influencing Nodule Occupancy by Inoculant Rhizobia. *Critical Reviews in Plant Sciences* **16**: 163-229.

Walker A (2012) Welcome to the plasmidome. *Nat Rev Microbiol* **10**: 379.

Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ & Scott KP (2015) 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome* **3**: 26.

Wang WL, Xu SY, Ren ZG, Tao L, Jiang JW & Zheng SS (2015) Application of metagenomics in the human gut microbiome. *World J Gastroenterol* **21**: 803-814.

Warburton PJ, Allan E, Hunter S, Ward J, Booth V, Wade WG & Mullany P (2011) Isolation of bacterial extrachromosomal DNA from human dental plaque associated with periodontal disease, using transposon-aided capture (TRACA). *FEMS Microbiol Ecol* **78**: 349-354.

Ward DM, Weller R & Bateson MM (1990) 16S rRNA sequences reveal uncultured inhabitants of a well-studied thermal community. *FEMS Microbiol Rev* **6**: 105-115.

Weidner S, Arnold W & Puhler A (1996) Diversity of uncultured microorganisms associated with the seagrass *Halophila stipulacea* estimated by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. *Appl Environ Microbiol* **62**: 766-771.

Whiteley AS, Jenkins S, Waite I, *et al.* (2012) Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *J Microbiol Methods* **91**: 80-88.

Whitman WB, Coleman DC & Wiebe WJ (1998) Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* **95**: 6578-6583.

Wiedenbeck J & Cohan FM (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* **35**: 957-976.

Wilkins D, Lauro FM, Williams TJ, *et al.* (2013) Biogeographic partitioning of Southern Ocean microorganisms revealed by metagenomics. *Environ Microbiol* **15**: 1318-1333.

Woese CR (1987) Bacterial evolution. *Microbiol Rev* **51**: 221-271.

Wu J, Gao W, Johnson RH, Zhang W & Meldrum DR (2013) Integrated metagenomic and metatranscriptomic analyses of microbial communities in the meso- and bathypelagic realm of north pacific ocean. *Mar Drugs* **11**: 3777-3801.

Wu S, Dalsgaard A, Hammerum AM, Porsbo LJ & Jensen LB (2010) Prevalence and characterization of plasmids carrying sulfonamide resistance genes among *Escherichia coli* from pigs, pig carcasses and human. *Acta Vet Scand* **52**: 47.

Yang Y, Li B, Zou S, Fang HH & Zhang T (2014) Fate of antibiotic resistance genes in sewage treatment plant revealed by metagenomic approach. *Water Res* **62**: 97-106.

Yang L, Wenfeng Z, Yadong L & Xingguo W (2012) Cloning of multicopper oxidase gene from *Ochrobactrum* sp. 531 and characterization of its alkaline laccase activity towards phenolic substrates. *Advances in Biological Chemistry*, 2012.

Ye M, Li G, Liang WQ & Liu YH (2010) Molecular cloning and characterization of a novel metagenome-derived multicopper oxidase with alkaline laccase activity and highly soluble expression. *Applied microbiology and biotechnology* **87**: 1023-1031.

Yen KM & Serdar CM (1988) Genetics of naphthalene catabolism in pseudomonads. *Crit Rev Microbiol* **15**: 247-268.

Yousefi S, Nahaei MR, Farajnia S, *et al.* (2010) Class 1 integron and Imipenem Resistance in Clinical Isolates of *Pseudomonas aeruginosa*: Prevalence and Antibiotic Susceptibility. *Iranian Journal of Microbiology* **2**: 115-121.

Zhang T, Zhang XX & Ye L (2011) Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. *PLoS ONE* **6**: e26041.

Zhang Y, Jiang Z, Cao B, Hu M, Wang Z & Dong X (2011) Chemotaxis to atrazine and detection of a xenobiotic catabolic plasmid in *Arthrobacter* sp. DNS10. *Environ Sci Pollut Res Int* **19**: 2951-2958.

Zhou JS, Chen JT, Wu XQ & Jin DD (2002) [Application of BACTEC MGIT 960 system and molecular identification of mycobacteria in the diagnosis of spinal tuberculosis]. *Di Yi Jun Yi Da Xue Xue Bao* **22**: 830-832.