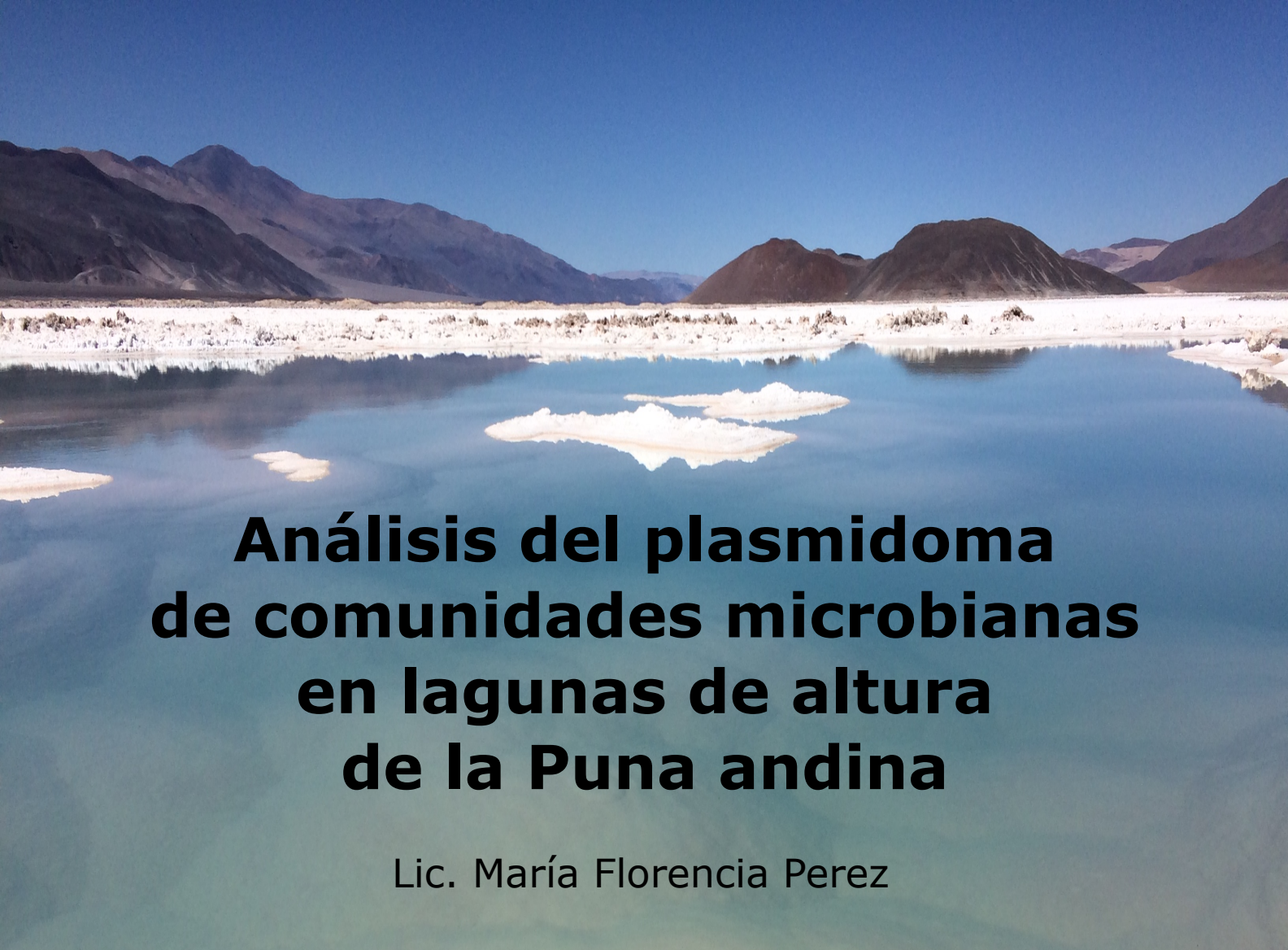




Análisis del plasmidoma de comunidades microbianas en lagunas de altura de la Puna andina

Lic. María Florencia Perez



UNIVERSIDAD NACIONAL DE TUCUMÁN
Facultad de Ciencias Naturales e
Instituto Miguel Lillo

Junio 2022

UNIVERSIDAD NACIONAL DE TUCUMÁN

**Facultad de Ciencias Naturales e
Instituto Miguel Lillo**



Tesis Doctoral

**Análisis del plasmidoma de comunidades microbianas
en lagunas de altura de la Puna andina**

**Lic. María Florencia Perez
Año 2022**



UNIVERSIDAD NACIONAL DE TUCUMÁN
Facultad de Ciencias Naturales e
Instituto Miguel Lillo

Miguel Lillo 205- Tel. 0381-4330633/4239456
CP: 4000 – San Miguel de Tucumán – República Argentina



HONORABLE CONSEJO DIRECTIVO

Dr. Eduardo Domínguez
Dra. Virginia Sara Luz Abdala
Dra. Iris Catiana Zampini
Dra. Lucia M. Ibáñez
Dra. Claudia Marcela Reina Herrera
Lic. Andrea Liliana Peralta
Srta. Claudia Melissa Carro
Srta. Constanza Anahí Belén Gómez
Sr. Marcos Darío Ponssa
Lic. Olga Myriam Sidán

DECANO

Dr. Hugo Rafael Fernández

VICE-DECANA

Dra. Norma Lidia Nasif

SECRETARIA ACADÉMICA

Dra. Alejandra María Kortsarz González

SECRETARIO DE POSGRADO

Dr. Eduardo Martín



UNIVERSIDAD NACIONAL DE TUCUMÁN
Facultad de Ciencias Naturales e
Instituto Miguel Lillo

Miguel Lillo 205- Tel. 0381-4330633/4239456
CP: 4000 – San Miguel de Tucumán – República Argentina



TRABAJO DE POSGRADO PARA LA OBTENCIÓN DEL GRADO
ACADÉMICO SUPERIOR DE DOCTORA EN CIENCIAS BIOLÓGICAS

CARRERA DE DOCTORADO EN CIENCIAS BIOLÓGICAS

Acreditado y Categorizado “A” ante la
Comisión Nacional de Acreditación Universitaria (CONEAU)
Res. 750/13
Reconocimiento Oficial y Validez Nacional
Res. 1396/13 - Ministerio de Educación Nacional

Director:

Dr. Carlos Javier Minahk

Vice-Directora:

Dra. María Pía Taranto

Comité Académico:

Dra. Aida Ben Altabef
Dra. Marta Eugenia Arias
Dra. María Paula Filippone
Dra. Fabiana del Carmen Cuezco

Suplentes:

Dra. María Teresa Vera
Dra. María Laura Ponssa
Dr. Pablo Edmundo Ortiz
Dra. Mirta María Rachid



UNIVERSIDAD NACIONAL DE TUCUMÁN
Facultad de Ciencias Naturales e
Instituto Miguel Lillo

Miguel Lillo 205- Tel. 0381-4330633/4239456
CP: 4000 – San Miguel de Tucumán – República Argentina



TRABAJO DE POSGRADO TITULADO

**Análisis del plasmidoma de comunidades microbianas
en lagunas de altura de la Puna andina**

TESISTA

Lic. María Florencia Perez

DIRECTOR

Dr. Julián Rafael Dib

CO-DIRECTORA

Dra. María Eugenia Farías

COMISIÓN DE SUPERVISIÓN

Dra. Marta Alejandra Polti
Dra. María Guadalupe Vizoso Pinto

Este trabajo de Tesis Doctoral se realizó en:

Planta Piloto de Procesos Industriales Microbiológicos

PROIMI – CONICET



Con el apoyo financiero de:

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Planta Piloto de Procesos Industriales Microbiológicos (PROIMI)

Laboratorio de Genómica de la Universidad de Göttingen
(Georg-August-Universität Göttingen, Alemania)

AGRADECIMIENTOS

En primer lugar, quisiera agradecer a las personas más importantes en mi vida, mis padres. Gracias por apoyarme y hacerme sentir acompañada en cada etapa, por su paciencia y sobre todo por tanto amor. Ojala fueran eternos.

A mis hermanas, Sofi y Emi, por estar conmigo cada día, por brindarme su apoyo y su cariño. Sé que siempre van a estar a mi lado. Gracias.

A mi director de tesis Julián, por confiar en mi capacidad y acompañar mi trabajo. A la Dra. María Eugenia Farías por abrirme las puertas para trabajar en su laboratorio.

A la Comisión de Supervisión, las doctoras Marta Polti y María Guadalupe Vizoso Pinto, por su valioso tiempo y por sus sugerencias durante el transcurso de este trabajo.

Al Dr. Daniel Kurth por su buena predisposición para ayudarme cada vez que necesité de sus conocimientos en bioinformática.

Al Prof. Dr. Rolf Daniel por brindarme la posibilidad de realizar una pasantía en el laboratorio de Genómica de la Universidad de Göttingen, Alemania. Gracias infinitas a Genis, por hacer que mi estancia allí sea productiva y por acompañarme, no solo desde lo académico sino también desde lo humano.

A mi querida amiga Martina, por lo que fue y es para mí. Por haber estado siempre, desde el inicio de nuestras carreras de grado hasta hoy que terminamos nuestros doctorados. Gracias por lo compartido durante todos estos años.

A Maru y Lucho, grandes amigos que conocí en PROIMI. Gracias por las risas compartidas en nuestras horas de trabajo, por ser mi sostén y escucharme cada día.

Al “cafetín”, por hacer que los días malos también tengan sus buenos momentos y transcurran con divertidas anécdotas. El trabajo siempre fue grato gracias a ustedes. Gracias también a todos mis compañeros del Laboratorio 27, en especial al Tano y Fede.

A todo el personal de PROIMI, en particular a Turi y Julia, por estar siempre dispuestos a colaborar con las necesidades de todos.

A mis amigas de siempre, Maris, Vichi, Agus, Anto, Naro, Coni y Sa. Gracias por su apoyo y por compartir conmigo tantos buenos momentos.

Por último, gracias al Consejo Nacional de Investigaciones Científicas y Técnicas por otorgarme la beca y el financiamiento para el desarrollo de esta tesis doctoral.

ÍNDICE

RESUMEN	1
Abstract	3
Abreviaturas	5
INTRODUCCIÓN GENERAL	7
1. Los plásmidos y sus características generales	8
2. Clasificación de plásmidos	10
3. Estructura de plásmidos	11
4. Replicación de plásmidos	12
4.1. Mecanismos de replicación	12
4.2. Control de la replicación plasmídica	14
5. Mantenimiento de plásmidos en el huésped	15
6. Movilización de plásmidos	16
7. Funciones accesorias en plásmidos	17
8. Estudio de plásmidos en procariotas	19
8.1. Estrategias de aislamiento de plásmidos a partir de muestras ambientales	20
8.2. Softwares para la reconstrucción <i>in silico</i> de plásmidos circulares	30
8.3. Plasmidoma y plásmidos lineales	31
9. Puna andina	32
9.1. Ecosistemas microbianos andinos	33
9.2. Estudios preliminares de plásmidos en ambientes andinos	36
Hipótesis	38
Objetivo	38
CAPÍTULO I: Análisis del plasmidoma de comunidades microbianas andinas	40
1. Introducción	41
2. Materiales y Métodos	43
2.1. Sitios de muestreo	43
2.2. Purificación de la biomasa	45
2.3. Extracciones de ADN	46
2.3.1. Extracción de ADN plasmídico	46
2.3.2. Extracción de ADN metagenómico	48
2.4. Eliminación del ADN cromosómico contaminante	48
2.5. Secuenciación, control de calidad y ensamblaje	50

2.6.	Análisis bioinformático	51
2.6.1.	Alineamiento plasmidoma-metagenoma	51
2.6.2.	Perfil funcional de los plasmidomas.....	51
2.6.2.1.	Funciones plasmídicas esenciales: replicación, movilización y mantención	52
2.6.2.2.	Funciones plasmídicas accesorias: resistencia a metales, factores de virulencia y resistencia antibiótica	52
2.6.3.	Elementos genéticos móviles: integrones, secuencias de inserción y transposasas ..	53
2.6.4.	Genes y secuencias de plásmidos conocidos.....	54
2.7.	Secuenciación de amplicones y análisis taxonómico	54
2.8.	Disponibilidad de datos	55
3.	Resultados y Discusión.....	57
3.1.	Aislamiento y análisis inicial de los plasmidomas	57
3.2.	Circularidad de secuencias en los plasmidomas.....	58
3.3.	Perfil funcional de los plasmidomas.....	58
3.4.	Comparación funcional entre plasmidomas	66
3.5.	Plasmidoma vs. metagenoma	69
3.6.	Funciones plasmídicas esenciales.....	73
3.6.1.	Replicación plasmídica.....	73
3.6.2.	Movilización plasmídica	76
3.6.3.	Mantención plasmídica.....	77
3.7.	Funciones plasmídicas accesorias	80
3.7.1.	Puquio de Campo Naranja.....	80
	Factores de virulencia.....	80
	Resistencia a metales.....	82
	Resistencia a los antibióticos.....	85
3.7.2.	Laguna Diamante	89
	Resistencia a los antibióticos y factores de virulencia	89
	Resistencia a metales y al arsénico.....	90
3.8.	Elementos genéticos móviles: integrones, secuencias de inserción y transposasas	94
3.9.	Genes y secuencias de plásmidos conocidos en los plasmidomas	96
3.9.1.	Base de datos NCBI	96
3.9.2.	Base de datos ACLAME	100
3.10.	Análisis taxonómico	101
4.	Conclusiones	107

CAPÍTULO II: Análisis y comparación de plásmidos lineales de cepas de <i>Micrococcus</i> de lagunas andinas	110
1. Introducción	111
2. Materiales y Métodos	115
2.1. Cepas y plásmidos	115
2.2. Análisis comparativo de megaplásmidos	116
3. Resultados y Discusión	116
3.1. Características generales	116
3.2. Comparación y funciones	117
3.3. Telómeros y proteínas terminales (<i>TPs</i>)	134
3.4. Replicación y mantención	135
3.5. Conjugación	138
4. Conclusiones	140
PRODUCCIÓN CIENTÍFICA	142
BIBLIOGRAFÍA	143
ANEXO	166

RESUMEN

Los plásmidos son elementos genéticos extra-cromosómicos auto-replicativos que se han encontrado en los tres dominios de la vida: Archaea, Bacteria y Eukarya. En principio, fueron considerados moléculas de ADN circular, pero años más tarde se descubrieron los primeros elementos lineales. Desde entonces, se han purificado, secuenciado y analizado una gran variedad de plásmidos a partir de microorganismos aislados. No obstante, hace algunos años, el avance en el campo de las tecnologías de secuenciación hizo posible el estudio de estos elementos en conjunto. Así surgió el término “plasmidoma” para referirse a la comunidad total de plásmidos de una muestra ambiental, proveniente tanto de organismos cultivables como no cultivables.

Los plásmidos, junto con otros elementos genéticos móviles, impulsan la evolución de los genomas a través de la transferencia horizontal de genes entre microorganismos. Generalmente llevan genes “accesorios” que codifican funciones que pueden ofrecer una ventaja para sus huéspedes. Por lo tanto, se espera que en un determinado nicho ecológico, los plásmidos se enriquezcan en funciones accesorias que son importantes para que sus huéspedes proliferen en él.

La Puna andina constituye un ambiente expuesto a múltiples condiciones extremas, tales como elevada radiación UV, alta sequedad, hipersalinidad, alcalinidad, altas concentraciones de metales pesados y metaloides como el arsénico, baja presión de oxígeno y grandes fluctuaciones térmicas. Los microorganismos que prosperan en este tipo de ambientes necesitan desarrollar mecanismos y estrategias eficientes para hacer frente a esas condiciones. Por lo tanto, presentan genes específicos que facilitan su supervivencia y que podrían, al menos en parte, estar codificados por plásmidos. Aunque algunos plásmidos de organismos cultivables de estos entornos ya se han analizado a nivel de secuencia, se desconoce el impacto de la información genética global transmitida por plásmidos en el contexto de las comunidades microbianas allí presentes.

Este trabajo de tesis doctoral se centró en caracterizar el plasmidoma de las comunidades microbianas presentes en dos lagunas de altura de la Puna andina (Puquio de Campo Naranja y Laguna Diamante), en predecir las funciones codificadas e identificar los posibles huéspedes. El análisis comprobó la existencia de funciones genéticas típicamente

codificadas por plásmidos. Por un lado, se detectaron funciones esenciales de replicación, movilización y mantenimiento plasmídico, y por otro, genes accesorios que permiten a los microorganismos que los hospedan sobrevivir bajo las duras condiciones ambientales. Se anotaron funciones relacionadas con la reparación del ADN que podrían ser un mecanismo adaptativo a la alta exposición a la radiación UV, y genes de resistencia a factores de estrés, incluidos aquellos relacionados con la respuesta al estrés oxidativo y osmótico. También se identificaron genes asociados con factores de virulencia y con la resistencia a metales, principalmente al cadmio, zinc, cobalto, cobre, mercurio, cromo y selenio. Además, la resistencia a los antibióticos estuvo ampliamente distribuida en uno de los plasmidomas, y la resistencia al arsénico en el otro. Adicionalmente, se observó una gran proporción de genes que codifican proteínas con función desconocida, lo que respalda la idea de que estos entornos prístinos son una fuente de nuevas biomoléculas.

Los resultados presentados en esta tesis suman evidencia sobre el papel clave del plasmidoma en la movilidad y la transferencia de genes dentro de las comunidades microbianas andinas: permite la diseminación de rasgos beneficiosos que brindan aptitud celular y la adaptación de poblaciones enteras a las condiciones extremas presentes en la Puna andina.

Abstract

Plasmids are self-replicating extra-chromosomal genetic elements that have been found in the three domains of life: Archaea, Bacteria, and Eukarya. At first, they were considered circular DNA molecules, but years later the first linear elements were discovered. Since then, a wide variety of plasmids have been purified, sequenced, and analyzed from isolated microorganisms. However, a few years ago, advances in the field of sequencing technologies made it possible to study these elements together. Thus, the term “plasmidome” was adopted to refer to the entire plasmid community in an environmental sample, coming from both cultivable and non-cultivable organisms.

Plasmids, along with other mobile genetic elements, drive the genomes evolution through horizontal gene transfer between microorganisms. They usually carry "accessory" genes encoding functions that may offer an advantage to their hosts. Therefore, it is expected that in a certain ecological niche, plasmids will be enriched in accessory functions that are important for their hosts to proliferate in it.

Andean Puna constitutes an environment exposed to multiple extreme conditions, such as high UV radiation, high dryness, hypersalinity, alkalinity, high concentrations of heavy metals and metalloids like arsenic, low oxygen pressure and large thermal fluctuations. Microorganisms that thrive in these types of environments need to develop efficient mechanisms and strategies to cope with these conditions. Therefore, they have specific genes that facilitate their survival and that could be encoded by plasmids. Although some plasmids from culturable organisms from these environments have already been analyzed at the sequence level, the impact of the global genetic information transmitted by plasmids in the context of the microbial communities present there is unknown.

This PhD thesis focused on characterizing the plasmidome of the microbial communities present in two high-altitude lagoons of the Andean Puna (Puquio de Campo Naranja and Laguna Diamante), predicting the encoded functions and identifying possible hosts. The analysis confirmed the existence of genetic functions typically encoded by plasmids. On the one hand, essential functions of plasmid replication, mobilization and maintenance were detected, and on the other, accessory genes that allow the microorganisms that host them to survive under harsh environmental conditions. Functions related to DNA repair that could be an adaptive mechanism to high UV radiation exposure and resistance

genes related to the oxidative and osmotic stress response were annotated. Genes associated with virulence factors and resistance to metals were also identified, mainly cadmium, zinc, cobalt, copper, mercury, chromium and selenium. Furthermore, antibiotic resistance was widespread in one of the plasmidomes, and arsenic resistance in the other. In addition, a large proportion of genes encoding proteins with unknown function were observed, supporting the idea that these pristine environments are a source of new biomolecules.

The results presented in this thesis add evidence on the key role of the plasmidome in the mobility and transfer of genes within the Andean microbial communities: it allows the dissemination of beneficial traits that provide cellular fitness and the adaptation of entire populations to the extreme conditions of the Andean Puna.

Abreviaturas

aa: aminoácido

ADN: ácido desoxirribonucleico

ADNasa: desoxirribonucleasa

ARG: *antibiotic resistance gene*, gen de resistencia antibiótica

ARN: ácido ribonucleico

ARNasa: ribonucleasa

ARNr: ácido ribonucleico ribosomal

ATP: adenosín trifosfato

ATPasa: adenosintrifosfatasa

c.s.p.: cantidad suficiente para

CIM: concentración inhibitoria mínima

ddH₂O: agua doblemente destilada

DMSO: dimetilsulfóxido

DNasas: desoxirribonucleasas

dNTP: deoxinucleósido trifosfato

EDTA: ácido etilendiaminotetraacético

EMAs: ecosistemas microbianos andinos

EPS: *extracellular polymeric substances*, sustancias poliméricas extracelulares

FPKM: fragmentos por kilobase por millón

HGT: *horizontal gene transfer*, transferencia horizontal de genes

HMM: *hidden Markov model*, modelo oculto de Markov

IS: *insertion sequence*, secuencia de inserción

kb: kilobase (mil pb)

LPS: lipopolisacárido

MGEs: *mobile genetic elements*, elementos genéticos móviles

m s.n.m.: metros sobre el nivel del mar

nt: nucleótidos

ORF: *open reading frame*, marco abierto de lectura

oriT: origen de transferencia plasmídica

oriV: origen de replicación plasmídica

pb: par de bases

PBS: *phosphate-buffered saline*, solución salina tamponada con fosfato

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

PFGE: *pulsed-field gel electrophoresis*, electroforesis en gel de campo pulsado

SDS: *sodium dodecyl sulfate*, dodecil sulfato sódico

TE: TRIS-EDTA

TIR: *terminal inverted repeat*, repetición invertida terminal

TP: *terminal protein*, proteína terminal

TRACA: *transposon-aided capture*, captura asistida por transposón

TRIS: tris (hidroximetil) aminometano

UV: ultravioleta

zOTU: *zero-radius operational taxonomic unit*, unidad taxonómica operativa de radio cero

INTRODUCCIÓN GENERAL

1. Los plásmidos y sus características generales

El término “plásmido” fue adoptado por Joshua Lederberg varias décadas atrás para referirse a cualquier determinante genético extra-cromosómico (Lederberg, 1952). A esta definición se agregó su capacidad de auto-replicarse, diferencia que mantienen con los replicones víricos que necesitan de la maquinaria celular del hospedador para iniciar su replicación. Sin embargo, existen algunos plásmidos denominados episomas que tienen la capacidad de insertarse en el cromosoma y, por lo tanto, también son reproducidos por la maquinaria de la célula (Jacob y Wollman, 1958).

Los plásmidos se han encontrado en especies de los tres dominios de la vida: Archaea, Bacteria y Eukarya. En principio, fueron considerados moléculas de ADN circular, sin embargo, años más tarde, se encontraron los primeros indicios de elementos extra-cromosómicos lineales en las mitocondrias del maíz (Pring y col., 1977), y luego se reportó el primer plásmido bacteriano lineal en *Streptomyces rochei* (pSLA2) (Hayakawa y col., 1979). Desde entonces, se han descubierto una gran cantidad de elementos lineales en una multitud de organismos, incluidos los procariotas y los eucariotas (Hishinuma y Hirai, 1991; Hinnebusch y Barbour, 1992; Meinhardt y col., 1997; Baker y col., 2007).

Las moléculas de ADN plasmídico generalmente tienen una conformación tipo doble hélice. A diferencia de los lineales, los plásmidos circulares adoptan varias topologías y por ello, a pesar de tener el mismo peso molecular, pueden migrar con velocidades distintas en un gel de electroforesis. Estas topologías son: circular abierto (corte en una sola hebra), lineal (corte en las dos hebras), circular relajado (sin cortes en las hebras pero de estructura “relajada” por la acción de enzimas), superenrollado (sin cortes en las hebras, totalmente compacto) y superenrollado desnaturalizado (sin cortes en las hebras pero menos compacto por una excesiva alcalinidad durante el proceso de purificación) (Figura 1).

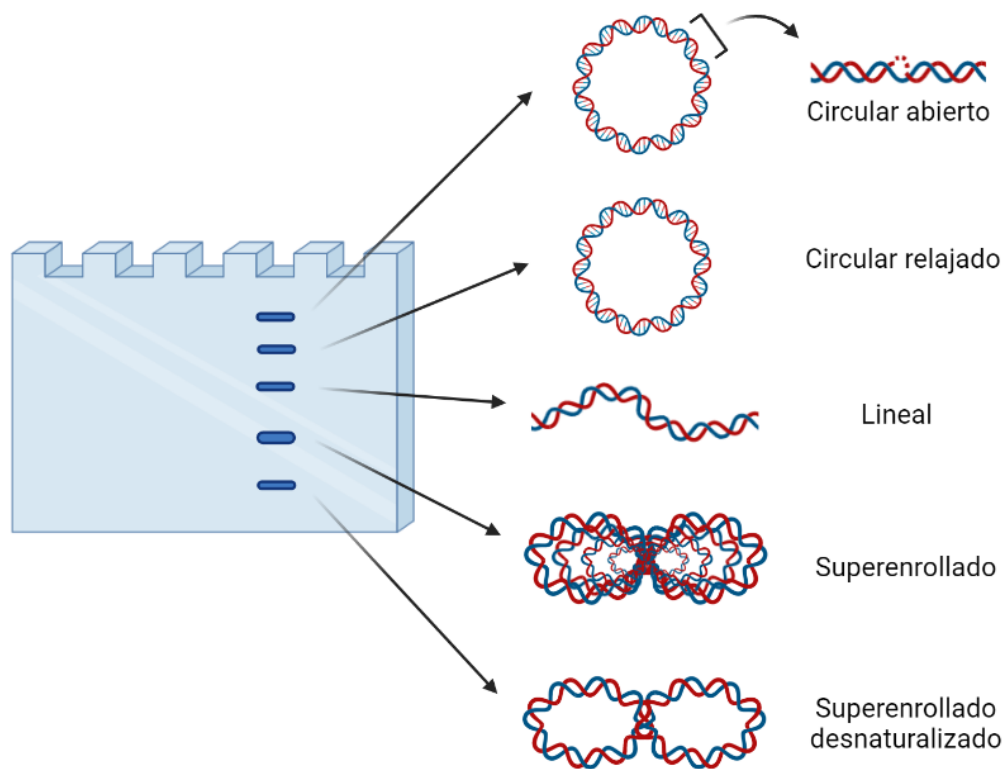


Figura 1. Patrón de movilidad electroforética de las distintas topologías que puede adoptar un plásmido circular.

El tamaño de los plásmidos es muy variable, existen algunos que solo contienen la información necesaria para su replicación, y otros que pueden alcanzar el orden de las megabases. Hasta la fecha, el plásmido más pequeño reportado es pRQ7 que fue aislado de la bacteria hipertermófila *Thermotoga* sp. RQ7, con 846 pb y un solo marco abierto de lectura (Harriott y col., 1994). Entre los plásmidos más grandes, a menudo llamados plásmidos gigantes o megaplásmidos, se encuentran pSCP1 de 356 kb (*Streptomyces coelicolor*; Bentley y col., 2004) o pRHL2 de 450 kb (*Rhodococcus jostii* RHA1; Shimizu y col., 2001). Uno de los representantes más grandes de megaplásmidos lineales es el pSCL4 de 1,8 Mb (*Streptomyces clavuligerus* ATCC 27064; Medema y col., 2010).

El número de copias de los plásmidos dentro de las células también varía ampliamente y está determinado por el mecanismo de control de replicación que posee cada uno (Del Solar y col., 1998). Además, generalmente, su contenido de GC es menor al del cromosoma de la célula huésped, aunque no difieren en más de un 10 % (Nishida, 2012). Probablemente, esto se deba a que la mayoría de las bacterias no pueden mantener y regular los plásmidos que muestran un contenido de GC muy diferente al suyo. Sin embargo, no está claro por qué no han adquirido ADN con un contenido de GC superior al de su propio cromosoma.

2. Clasificación de plásmidos

Existen diversos criterios para clasificar los plásmidos. En base a su habilidad para transferirse de un microorganismo a otro, estos elementos genéticos pueden ser conjugativos (codifican todos los componentes necesarios para transferirse), movilizables (codifican algunos de los componentes necesarios para transferirse, pero necesitan de otros que son aportados por un plásmido conjugativo denominado “*helper*”) o no movilizables (no codifican ninguno de los componentes necesarios para transferirse) (Garcillán-Barcia y de la Cruz, 2013).

De acuerdo a su función, pueden clasificarse de forma general en: plásmidos de resistencia, que otorgan una o varias resistencias a sus huéspedes (Poirel y col., 2016); metabólicos/degradativos, que hacen posible la digestión de ciertos compuestos (Williams y Murray, 1974); virulentos, que convierten a la célula huésped en un patógeno (Johnson y Nolan, 2009); y simbióticos, que codifican genes que hacen posible relaciones de simbiosis (Okazaki y col., 2015). Además, existen plásmidos que no tienen ninguna función conocida hasta ahora, los cuales se denominan crípticos.

De acuerdo a su capacidad de coexistir con otros plásmidos pueden ser compatibles o incompatibles. Cuando dos plásmidos no pueden replicarse y mantenerse de forma estable en una misma célula en ausencia de una presión selectiva permanente, se considera que son incompatibles. Esto se debe a que comparten sus sistemas de replicación y, por lo tanto, pertenecen al mismo grupo de incompatibilidad (Inc). Si los plásmidos coexisten, se consideran compatibles (Novick, 1987).

3. Estructura de plásmidos

El esqueleto molecular de los plásmidos cuenta con dos regiones principales. Por un lado, una región conservada que contiene genes esenciales para iniciar su replicación, asegurar su mantenimiento dentro del huésped y permitir su movilización hacia otro microorganismo. Y, por otro lado, una región variable con genes accesorios que codifican funciones que usualmente contribuyen a la evolución y que pueden conferirle una ventaja a quien los hospeda, como por ejemplo factores de virulencia, resistencia a antibióticos y metales pesados, producción de antimicrobianos, degradación de xenobióticos, entre otras (Figura 2) (Smalla y col., 2000b; Novick, 2003; Crossman, 2005; Gil y col., 2006; Heuer y col., 2009).

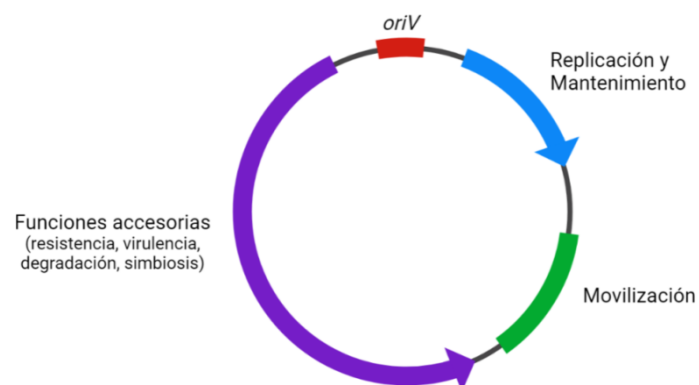


Figura 2. Estructura funcional de un plásmido. En rojo, azul y verde se observa la región conservada, y en violeta la región variable o accesorio. *oriV*: origen de replicación.

4. Replicación de plásmidos

4.1. Mecanismos de replicación

A pesar de su autonomía, los plásmidos utilizan ampliamente la maquinaria de replicación del huésped. Por lo tanto, existe una comunicación y cooperación entre los componentes de ambas partes. El inicio de la replicación es llevado a cabo por factores codificados por el plásmido y luego las proteínas celulares implicadas en el proceso, propias del huésped, continúan con la duplicación del material genético (Del Solar y col., 1998).

Todos los plásmidos tienen un sitio donde se inicia la replicación, denominado origen de replicación (*oriV*) (Figura 2). Entre sus características generales, se puede mencionar la presencia de una región adyacente rica en AT que contiene repeticiones de secuencia, donde se produce la apertura de las hebras. Además, en muchos casos existen repeticiones directas cortas en tándem (iterones), que son necesarias para el reconocimiento por parte de la proteína iniciadora Rep codificada por el propio plásmido. En otros casos, el inicio de la replicación es independiente de Rep, y en él intervienen ARNs antisentido, también codificados a nivel plasmídico (Del Solar y col., 1998; Konieczny y col., 2014).

Los mecanismos de replicación conocidos en plásmidos circulares son tres: mecanismo *theta*, desplazamiento de cadena y mecanismo de círculo rodante (Del Solar y col., 1998). Las características más importantes de cada uno de ellos se explican brevemente a continuación.

Mecanismo *theta*. La replicación del ADN plasmídico por este mecanismo es similar a la replicación del cromosoma ya que ocurre de manera simultánea en ambas cadenas, pero de forma continua en una (cadena líder) y discontinua en la otra (cadena retrasada) (Lilly y Camps, 2015). Luego de la apertura de dichas cadenas y de la síntesis de un cebador de ARN, comienza la síntesis de ADN por extensión del mismo, la cual es llevada a cabo por las ADN

polimerasas I y III del huésped (Lilly y Camps, 2015). Los intermediarios de replicación observados por microscopía electrónica presentan una forma característica que recuerda a la letra griega θ (*theta*), de ahí su nombre. El inicio de la replicación puede ser dependiente o independiente de Rep, en este último caso, como ya se mencionó antes, el ARN antisentido codificado por el plásmido actúa como pre-cebador (Del Solar y col., 1998).

Desplazamiento de cadena. En este mecanismo el inicio de la replicación depende de tres proteínas Rep codificadas por el plásmido (RepA, RepB, RepC). El reconocimiento de los iterones en el *oriV* es llevado a cabo por RepC, quien recluta a RepA que tiene actividad helicasa y permite la apertura local de las cadenas, exponiendo dos secuencias palindrómicas (*ssiA* y *ssiB*), una en cada hebra. A continuación RepB se une a dichas secuencias y sintetiza los cebadores que serán utilizados por las ADN polimerasas celulares para continuar la replicación. La síntesis de cada cadena se produce continuamente gracias al movimiento de RepA, y da como resultado el desplazamiento de la cadena complementaria (Del Solar y col., 1998).

Mecanismo de círculo rodante. A diferencia de lo que ocurre en el mecanismo *theta*, en este caso la síntesis de la cadena líder y de la cadena retrasada no están acopladas. Primero la proteína Rep codificada por el plásmido reconoce el *oriV* de doble cadena y realiza un corte en una de las hebras. Ese extremo 3' que se genera es utilizado como cebador por las ADN polimerasas celulares para la síntesis de la cadena líder. De esta forma se obtienen, por un lado, una molécula de ADN plasmídico constituida por la cadena parental molde y la cadena líder de síntesis reciente, y por el otro, un intermediario en forma de ADN circular monocatenario que corresponde a la cadena parental desplazada. La síntesis de la cadena retrasada se produce por acción de las proteínas de la célula que reconocen un origen de replicación de cadena sencilla en este intermediario. Por último, las dos moléculas de ADN

plasmídico producidas son superenrolladas por la ADN girasa del hospedador (Ruiz-Masó y col., 2015).

4.2. Control de la replicación plasmídica

El número de copias de los plásmidos dentro de las células huésped es controlado mediante la regulación de los mecanismos de replicación ya mencionados. Así, existen varias estrategias que explican cómo se mantiene este promedio de copias en una población en crecimiento y cómo se corrigen sus fluctuaciones. En todos los casos, se trata de una inhibición del inicio de la replicación que involucra elementos que son codificados por el mismo plásmido, como los ARNs antisentido, las Reps y los iterones presentes en los orígenes de replicación.

Los ARNs antisentido se sintetizan constitutivamente y son poco estables, por lo que un aumento en el número de copias del plásmido que los codifica genera inmediatamente un aumento en la concentración de estas moléculas. Los ARNs antisentido pueden inhibir la replicación mediante la atenuación de la transcripción de Reps o la inhibición de su traducción, o impidiendo la formación del complejo cebador-ADN molde para el caso de replications Reps-independientes (Brantl, 2014).

En los plásmidos con iterones pueden darse otros dos mecanismos de regulación, que consisten en la interacción de las proteínas Reps con estas secuencias repetidas directas o con otras invertidas. En el primero, cuando ocurre un aumento del número de copias del plásmido, las proteínas Reps pueden asociarse a dos de ellos a la vez, o bien a otras Reps que a su vez están unidas a otros plásmidos. Este fenómeno, conocido como “*handcuffing*”, genera complejos que resultan en un impedimento estérico para el desarrollo de la replicación. En el segundo mecanismo, las Reps forman dímeros que se asocian con las repeticiones invertidas río arriba del gen *rep* e impiden estéricamente la interacción de la ARN polimerasa del

huésped con su promotor. Este hecho se traduce en la inhibición de la transcripción de dichas proteínas (Del Solar y col., 1998; Konieczny y col., 2014).

5. Mantenimiento de plásmidos en el huésped

Durante la división celular puede ocurrir que las copias de los plásmidos no se segreguen de forma correcta en las células hijas. Para evitarlo, existen diversos sistemas que aseguran la estabilidad plasmídica en el huésped a través de las generaciones. Cada sistema corresponde a uno de tres tipos: resolución de multímeros, partición activa o muerte post-segregacional (toxina-antitoxina). Los plásmidos pueden tener codificados en su estructura molecular uno o varios de cada tipo (Sengupta y Austin, 2011).

La recombinación homóloga ocurre con frecuencia entre plásmidos, generando dímeros o multímeros que impiden la segregación de unidades plasmídicas independientes y aumentan así la probabilidad de obtener células hijas sin ellos. El sistema de resolución de multímeros consiste en la acción de recombinasas específicas de sitio, que se agrupan en dos familias (serina y tirosina recombinasas), y que reestablecen el número de unidades independientes (Crozat y col., 2014).

En el mecanismo de partición activa actúan dos proteínas principales. La primera, denominada CBP (del inglés *Centromere-Binding Protein*), se une al sitio *par* del plásmido y recluta a la segunda, una proteína motora con actividad NTPasa que, al polimerizar, moviliza a los plásmidos hacia los polos celulares asegurando su correcta segregación en la posterior división celular (Baxter y Funnell, 2015).

La muerte post-segregacional se da en el caso de que una célula hija no reciba ninguna copia del plásmido, y ocurre por la existencia de un sistema toxina-antitoxina (TA) codificado en él.

Las toxinas pueden actuar sobre diferentes objetivos, pero en todos los casos inhiben la viabilidad celular, mientras que las antitoxinas neutralizan dicha inhibición. Durante la división celular ambas moléculas, que se encuentran distribuidas en el citoplasma, son heredadas por las células hijas. Sin embargo, como la antitoxina (proteína de bajo peso molecular o ARN no codificante) es menos estable que la toxina (proteína), con el tiempo se genera un desequilibrio de concentraciones que tiene como resultado la activación de esta última. Así, las células que heredan una copia del plásmido con el sistema TA se mantienen, mientras que aquellas libres del mismo son eliminadas de la población (Hernández-Arriaga y col., 2015). Actualmente los sistemas TA se agrupan en seis clases, de acuerdo a la naturaleza molecular de la antitoxina y su modo de interacción con la toxina (Page y Peti, 2016).

6. Movilización de plásmidos

La transferencia horizontal de genes (*HGT*, *horizontal gene transfer*) contribuye significativamente tanto a la plasticidad como a la evolución de los genomas procariotas (Koonin y Wolf, 2008), y los elementos genéticos móviles actúan como conductores de este fenómeno (Frost y col., 2005). Entre tales elementos, los plásmidos son considerados vectores importantes en el intercambio genético entre bacterias (Halary y col., 2010).

El ADN plasmídico puede ingresar a la célula huésped por dos mecanismos bien estudiados: la transformación y la conjugación. Mientras que la transformación ocurre cuando una bacteria “competente” toma ADN exógeno libre del medio circundante, la conjugación requiere del contacto célula-célula y la formación de un *pili* sexual para la movilización del plásmido (Ou y Anderson, 1970). Para ello, en su esqueleto molecular, estos elementos conjugativos presentan genes que codifican dos sistemas: Dtr (del inglés *DNA Transfer and*

Replication) y Mpf (del inglés *Mating Pair Formation*). En el primero, la enzima relaxasa es la más importante y la encargada del reconocimiento del origen de transferencia (*oriT*) y el procesamiento del ADN a transferir (Lanka y Wilkins, 1995). El segundo es el encargado del proceso de translocación hacia la célula receptora y pertenece a la familia de sistemas de secreción de tipo IV (T4SS) (Schröder y Lanka, 2005).

7. Funciones accesorias en plásmidos

Como se ha mencionado anteriormente, los plásmidos llevan con frecuencia un conjunto de genes "accesorios" que contribuyen con rasgos adaptativos a los huéspedes y, potencialmente, a otros procariotas residentes dentro de nichos ecológicos específicos (Norman y col., 2009). Por tal motivo, se considera que estos elementos móviles no son indispensables para las células, es decir que no cumplen funciones del metabolismo central, esenciales para la vida de los microorganismos. Además, debido al incremento en la aptitud celular, se espera que las funciones accesorias varíen en los distintos plásmidos de acuerdo al entorno ambiental ocupado por su huésped. En general, se pueden agrupar en las siguientes categorías (Figura 3):

Funciones de resistencia: permiten al huésped resistir en ambientes con antibióticos o metales pesados. Los microorganismos pueden resistir a estos compuestos por diversos mecanismos, por lo que existe una gran variedad de genes codificados por los plásmidos de resistencia. Asimismo, existen plásmidos con genes de resistencia a múltiples antimicrobianos, o con integrones, que son estructuras especializadas que contienen varios casetes de genes de resistencia bajo el control de un solo promotor (Smalla y col., 2006; Carattoli, 2013; Chen y col., 2014; Szmolka y col., 2015).

Funciones metabólicas/degradativas: permiten al huésped utilizar recursos alternativos como fuente de carbono y energía. Los plásmidos con estas funciones accesorias suelen ser de alto peso molecular debido a que codifican rutas metabólicas de degradación de distintos compuestos orgánicos, y han sido aislados a partir de microorganismos seleccionados de ambientes expuestos a ellos. Su potencial biotecnológico como agentes de biorremediación de xenobióticos en ambientes contaminados ha sido ampliamente estudiado (Basta y col., 2004; Ghosal y col., 2016).

Funciones de virulencia: permiten al huésped actuar como un patógeno otorgándole la capacidad de invadir un tejido o producir un compuesto tóxico, y de esta forma dañar a los organismos superiores blancos de ataque, mientras benefician su persistencia y multiplicación (Hale, 1991; Wijetunge y col., 2014).

Funciones de simbiosis: permiten al huésped desarrollar una relación simbiótica. Entre los plásmidos con estas funciones accesorias, se encuentran los de rizobios simbioses de leguminosas. A partir de los genes codificados en ellos, estos microorganismos portadores son capaces de formar nódulos en las raíces y fijar nitrógeno atmosférico en forma de sustratos asimilables por la planta, a cambio de protección y nutrientes (Bánfalvi y col., 1981; Kondorosi y col., 1982).

Es importante destacar que los plásmidos pueden perder, adquirir o modificar su información accesoria por recombinación homóloga o por la presencia de otros elementos móviles en su estructura, como las secuencias de inserción o los transposones (Cohen, 1976). Asimismo, la posibilidad de transferirse de un microorganismo a otro, los convierte en elementos claves en los procesos de adaptación de poblaciones microbianas a nuevos ambientes (Norman y col., 2009).

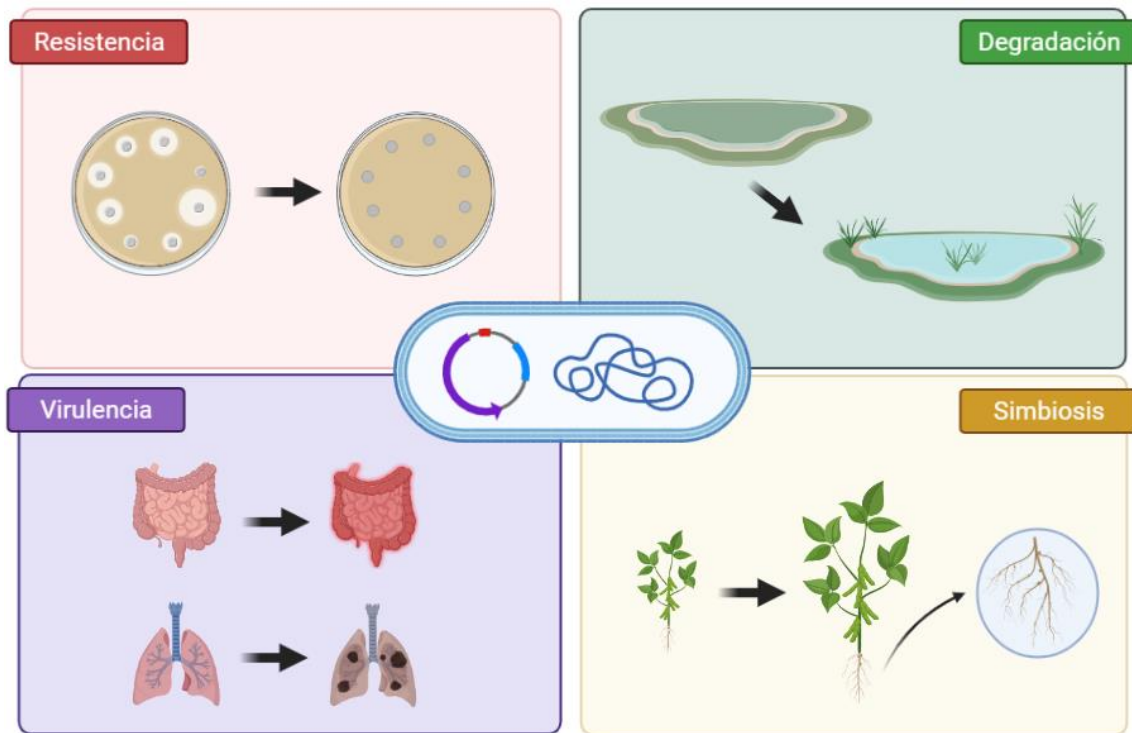


Figura 3. Principales categorías de funciones accesorias en plásmidos.

8. Estudio de plásmidos en procariotas

Por lo general, el estudio de los plásmidos ha estado orientado a aquellos obtenidos a partir de microorganismos individuales y cultivables. Sin embargo, en los últimos años, el gran avance de la tecnología de secuenciación ha permitido el análisis de comunidades microbianas completas. La importancia de un enfoque metagenómico surge, principalmente, de la necesidad de eliminar el sesgo producido por la dependencia de cultivo, ya que la mayoría de los microorganismos dentro de una comunidad se consideran no cultivables en condiciones de laboratorio (Schloss y Handelsman, 2005). Además, la metagenómica, no solo constituye una poderosa herramienta para el análisis genético de los problemas ambientales como el incremento del número de bacterias resistentes a antibióticos o la intensa contaminación presente en suelos, ríos y mares, sino que también resulta útil para encontrar nuevas

biomoléculas de interés biotecnológico a partir de organismos hasta ahora desconocidos (Daniel, 2004; Hemme y col., 2010; Schmieder y Edwards, 2012; Tripathi y col., 2018).

Los datos de secuencia obtenidos a partir de un enfoque metagenómico requieren de un gran esfuerzo de análisis, el cual incluye, en general, el filtrado de las lecturas de calidad y su ensamblaje en *contigs*, la predicción de genes y la determinación de la diversidad taxonómica. De hecho, la gran cantidad de información biológica generada, en ocasiones, puede causar confusión, ya que la muestra ambiental analizada consiste en una mezcla compleja de material genético perteneciente a diversos organismos (McHardy y Rigoutsos, 2007). Además, las muestras metagenómicas no solo incluyen ADN cromosómico, sino también todo tipo de elementos genéticos accesorios. Entre ellos, los plásmidos han recibido atención especial debido a su papel destacado como actores evolutivos. Sin embargo, estos elementos han sido subestimados en los estudios metagenómicos convencionales, ya que la relación de ambos permanece enormemente a favor de los cromosomas por la baja tasa de incidencia que presentan los plásmidos en aislamientos bacterianos o en colecciones de cepas (en algunos casos menor al 10 %) (Heuer y col., 2008; Li y col., 2012). De esta forma, se hace evidente la necesidad de protocolos mejorados y más específicos que permitan separarlos *a priori* del ADN cromosómico, y así caracterizarlos en conjunto para poder comprender a fondo sus funciones como así también su significancia e impacto ecológico en los ecosistemas microbianos.

8.1. Estrategias de aislamiento de plásmidos a partir de muestras ambientales

Dependiendo del objetivo del análisis, el estudio de los elementos genéticos extra-cromosómicos de poblaciones de microorganismos de diversos entornos puede incluir enfoques dependientes o independientes del cultivo.

Entre los primeros trabajos dependientes de cultivo reportados para el estudio global de plásmidos a partir de muestras ambientales, se destacan aquellos realizados a partir de plantas de tratamiento de aguas residuales, por considerarse reservorios de bacterias con resistencia antibiótica (Schlüter y col., 2008; Szczepanowski y col., 2008). En estos casos, la estrategia consistió en exponer las bacterias del lodo activado a condiciones selectivas, utilizando diferentes antibióticos. A partir de los organismos viables, es decir, aquellos cultivables en dichas condiciones, se purificó el ADN plasmídico por medio de un método de lisis alcalina seguido de una ultracentrifugación en gradiente de densidad de cloruro de cesio. La desventaja de este tipo de enfoques es que los datos obtenidos dependen en gran medida del procedimiento de extracción de plásmidos, ya que se excluyen los elementos de organismos que son, en general, no cultivables o que no prosperan en las condiciones elegidas.

En contraste, varios métodos independientes de cultivo han demostrado facilitar la detección y posterior caracterización de nuevos elementos genéticos accesorios a partir de poblaciones bacterianas de diversos entornos. Entre ellos, se encuentra el aislamiento exógeno de plásmidos, que consiste en la captura de estos elementos por conjugaciones bi- o tri-parentales usando una bacteria receptora (Bale y col., 1988). Los apareamientos bi-parentales utilizan a toda la comunidad microbiana de la muestra como donantes de plásmidos con el rasgo de selección deseado (por ej. una resistencia), y una cepa con otro marcador de selección como receptora. Los transconjugantes se obtienen a partir de la viabilidad en un medio de cultivo que contiene tanto el marcador del receptor como el codificado por los plásmidos autotransferibles capturados a partir de la muestra ambiental. Mientras que este enfoque bi-parental se basa en la expresión de marcadores seleccionables, el aislamiento de plásmidos en apareamientos tri-parentales se basa únicamente en su capacidad para movilizar (a un nuevo receptor) pequeños plásmidos movilizables que llevan marcadores seleccionables. Por lo

tanto, la detección de transconjugantes está basada en el rasgo adquirido por el receptor debido a la captación del plásmido movilizable del tercer socio del apareamiento (Smalla y Sobecky, 2002). A pesar de que estos procedimientos se aplicaron con bastante éxito para aislar plásmidos que confieren resistencia a antibióticos o metales pesados (Hill y col., 1992; Lilley y col., 1996; Dahlberg y col., 1997; Smalla y col., 2000a; Binh y col., 2008), tienen limitaciones intrínsecas: los elementos a capturar tienen que ser conjugativos y deben replicarse de manera estable. Además, la conjugación es específica del huésped (excepto para plásmidos de amplio rango de huéspedes) y requiere la presencia de marcadores de selección plasmídicos.

Por su parte, Jones y Marchesi (2007) desarrollaron un método de captura asistida por transposón (TRACA, *transposon-aided capture*) para estudiar los plásmidos residentes dentro de la comunidad bacteriana del intestino humano, el cual también es independiente de cultivo pero no se basa en ningún rasgo codificado por los mismos. La técnica consiste en la extracción del ADN total de la muestra y el tratamiento con una ADNasa (*plasmid-safe DNase*, Epicentre, EE.UU.) que elimina selectivamente el ADN lineal, dejando intacto el ADN plasmídico circular. Este se somete luego a una reacción de transposición *in vitro* para insertar el transposón EZ-Tn5 que posee un origen de replicación de *Escherichia coli* y un marcador seleccionable (*OriV/Kan2*). Posteriormente, se transforma un huésped sustituto (*E. coli*) con los elementos híbridos circulares resultantes, se realiza el aislamiento del ADN plasmídico y se secuencia. Aplicando esta metodología, Warburton y col. (2011) fueron capaces de aislar treinta y dos plásmidos nuevos presentes en la comunidad microbiana de la cavidad bucal humana. Sin embargo, a pesar de su comprobada eficacia, el método tiene sus limitaciones, como la inactivación de genes por el transposón y la captura de plásmidos principalmente pequeños (3 a 10 kb). De hecho, los plásmidos grandes difícilmente puedan

transformar a *E. coli* (Szostková y Horáková, 1998). Más aún, aquellos elementos que sean inestables en este huésped o intratables por transposición, así como los presentes en un bajo número de copias, difícilmente puedan capturarse. La eficacia de este método se ve influenciada también por la calidad del ADN y la integridad de los plásmidos aislados.

Zhang y col. (2011) también emplearon el sistema TRACA para aislar nuevos plásmidos de lodos activados de una planta de tratamiento de aguas residuales en Hong Kong, China. Sin embargo, su enfoque fue algo más amplio: aislaron el ADN total del lodo utilizando un kit de extracción (*FastDNA Soil Kit*, MP Biomedicals, California, EE.UU.) y, a partir de este, purificaron el ADN plasmídico con otro kit específico para enriquecer los elementos extra-cromosómicos (*QIAGEN Plasmid Mini Kit*, QIAGEN SCIENCES INC, Maryland, EE.UU.); a continuación eliminaron el ADN cromosómico contaminante aplicando una ADNasa (*plasmid-safe DNase*, Epicentre, EE.UU.), y utilizaron esas muestras para la secuenciación de alto rendimiento por Illumina o para la captura de plásmidos novedosos por TRACA. El análisis de datos de secuencia reveló altos niveles de genes de resistencia antibiótica, así como elementos genéticos móviles (*MGEs*, *mobile genetic elements*), incluidos integrones, transposones y plásmidos. De esta forma, el enfoque metagenómico propuesto por los autores permitió demostrar el papel clave que desempeñan este tipo de comunidades, caracterizadas por una alta densidad y diversidad microbiana, en la transferencia horizontal de genes de resistencia antibiótica a través de *MGEs*.

Es aquí cuando el término “plasmidoma”, también conocido como metagenoma plasmídico, empieza a tomar fuerza para referirse al ADN plasmídico total de una muestra ambiental, obtenido a partir de metodologías independientes de cultivo. En consecuencia, en esta tesis nos referiremos al plasmidoma como la comunidad total de plásmidos de un entorno dado, proveniente tanto de organismos cultivables como no cultivables, que se estudia mediante

enfoques metagenómicos usando tecnologías de secuenciación masiva o de nueva generación (Figura 4).

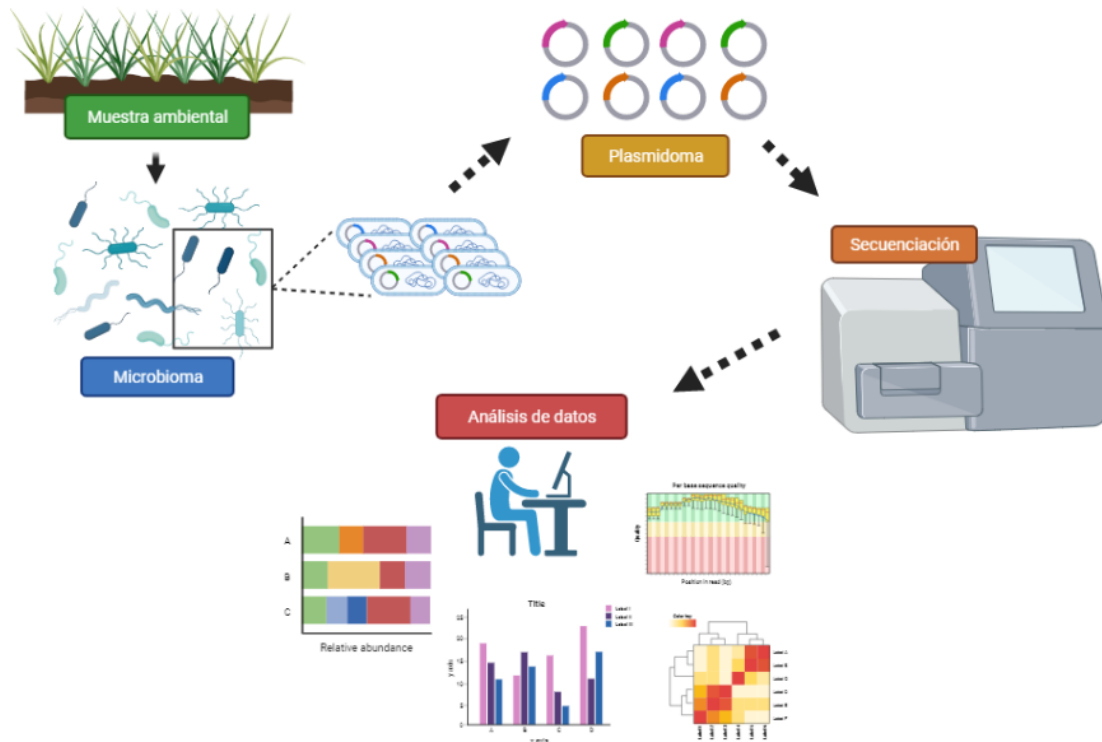


Figura 4. Contextualización del término “plasmidoma” como enfoque metagenómico. El plasmidoma se refiere a la comunidad total de plásmidos de una muestra ambiental, proveniente tanto de organismos cultivables como no cultivables.

Aunque es indudable la relevancia ecológica de este tipo de estudios, hasta el momento son escasas las investigaciones referidas al plasmidoma de ecosistemas microbianos que emplean métodos independientes de cultivo. Posiblemente, esto se deba a la complejidad del procesamiento de las muestras y a la necesidad de herramientas bioinformáticas a medida para un análisis adecuado. Entre los pioneros en este campo de estudio se pueden mencionar a Kav y col. (2012), quienes estudiaron la composición de plásmidos en la comunidad microbiana asociada al rumen vacuno. En este caso, se incluyeron algunas modificaciones para mejorar

tanto la calidad como la cantidad del ADN plasmídico total a secuenciar. A diferencia del enfoque anterior (Zhang y col., 2011), las extracciones de ADN se realizaron a partir de suspensiones concentradas de células bacterianas (previamente separadas del resto de las muestras) mediante tres métodos diferentes para maximizar y asegurar la lisis de diversas bacterias (Kav y col., 2013). Una vez más, el ADN cromosómico contaminante se eliminó utilizando la ADNasa que degrada el ADN lineal (*plasmid-safe DNase*, Epicentre, EE.UU.) y, además, el ADN circular se amplificó utilizando la ADN polimerasa $\Phi 29$ (New England Biolabs, Massachusetts, EE.UU.) con el objetivo de aumentar los elementos de bajo número de copias y garantizar una cantidad de plásmidos que permita la secuenciación (Illumina). A partir de la comparación de los datos plasmídicos obtenidos con estudios metagenómicos sobre el rumen, los autores determinaron la importancia de estos elementos extra-cromosómicos, ya que se hizo evidente que el plasmidoma ruminal codifica funciones que se enriquecen en ese nicho ecológico y que podrían conferir ventajas a sus huéspedes. Además, un gran número de *contigs* del plasmidoma alineó con plásmidos de bacterias ruminales aisladas previamente, lo que sugiere que no solo los taxones bacterianos, sino también sus plásmidos y perfiles funcionales, están asociados o definidos por el entorno.

En los estudios posteriores sobre plasmidomas, se utilizaron distintas combinaciones de las técnicas ya descritas (Figura 5, Tabla 1). De esta forma, Li y col. (2012) encontraron principalmente genes que codifican funciones plasmídicas esenciales, como la estabilidad y la conjugación, en el plasmidoma de una muestra de aguas residuales de una planta de tratamiento en Dinamarca. Además, el análisis les permitió recuperar varios cientos de replicasas plasmídicas putativas nuevas, indicando claramente que los plasmidomas ambientales permanecen en gran parte sin explotar y que se pueden realizar muchos descubrimientos mediante la realización de estudios directamente centrados en ellos. Por su

parte, Sentchilo y col. (2013) estudiaron el plasmidoma de comunidades microbianas en dos sistemas de lodos activados de plantas de tratamiento de aguas residuales suizas, una de las cuales recibe principalmente aguas residuales domésticas y la otra de la industria química. En ellos encontraron una gran abundancia de genes para el metabolismo de carbohidratos, así como también genes de resistencia a antibióticos y metales pesados, demostrando nuevamente que las funciones enriquecidas en los plasmidomas están determinadas por las condiciones locales.

Jørgensen y col. (2014) propusieron un procedimiento robusto *in silico* para identificar pequeños plásmidos completos en conjuntos de datos metagenómicos. A partir del plasmidoma del ciego de una rata, identificaron 616 secuencias circulares, de las cuales 160 portaban un gen con un dominio de replicación plasmídica. El análisis de homología indicó, además, que la mayoría de estas secuencias plasmídicas eran nuevas. Por lo tanto, su trabajo brinda la posibilidad de descubrir nuevos grupos de plásmidos completos en conjuntos de datos metagenómicos existentes y así expandir las bases de datos actuales de estos elementos. Basándose en este procedimiento, Kothari y col. (2019) presentaron su propia *pipeline* para detectar circularidad en el plasmidoma de muestras de agua subterránea, normalmente caracterizada por una baja densidad celular. En este estudio se identificó una gran diversidad de plásmidos circulares (incluidos algunos grandes), representados por la presencia de siete grupos de incompatibilidad y altamente enriquecidos en resistencias a metales.

Recientemente, Kirstahler y col. (2021) realizaron el primer estudio a escala global a partir del análisis del plasmidoma de 24 muestras de aguas residuales de 22 países distintos, distribuidos en 5 continentes. Combinando las tecnologías de secuenciación de Oxford Nanopore e Illumina NextSeq fueron capaces de ensamblar 159.322 *contigs* circulares, de los cuales aproximadamente el 90 % no tenían ninguna similitud con plásmidos conocidos. En

ellos detectaron una gran abundancia de genes de resistencia a los antimicrobianos y una extensa variedad de funciones, incluida la replicación, el mantenimiento, la movilización y la conjugación plasmídica.

Teniendo en cuenta las limitaciones del sistema TRACA, los enfoques subsiguientes utilizados para estudiar plásmidos de comunidades microbianas demostraron ser más eficientes, revelando un gran aumento de la diversidad y la cantidad de plásmidos. Además del alto rendimiento, brindan amplia información sobre los elementos extra-cromosómicos en conjunto, tales como los potenciales huéspedes y los genes de resistencia (resistoma); y pueden ser comparados fácilmente con los datos metagenómicos del mismo entorno o de entornos similares. Sin embargo, la amplificación del ADN circular por desplazamiento de cadena con la ADN polimerasa $\Phi 29$, aplicada en la mayoría de las estrategias, favorece principalmente a los plásmidos pequeños (< 10 kb), llevando a que estos elementos estén representados en gran abundancia en los plasmidomas (Norman y col., 2014). En consecuencia, estos autores propusieron un protocolo mejorado que utiliza un paso de electroelución adicional, previo a la amplificación, para eliminar dicho sesgo y permitir una captura más eficiente de plásmidos de un rango de tamaño superior (> 10 kb).

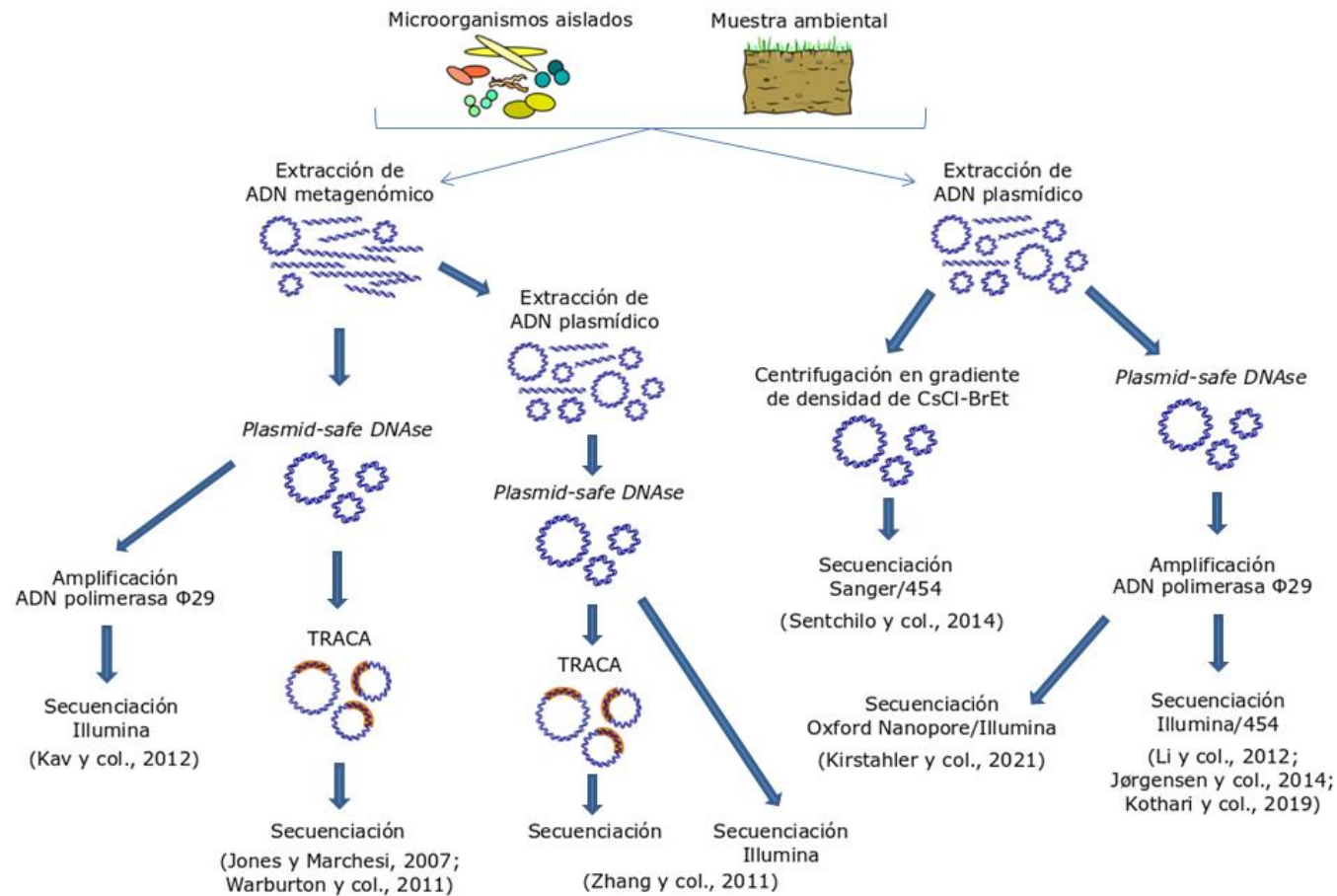


Figura 5. Representación esquemática de las estrategias de aislamiento de plásmidos en estudios de plasmidomas independientes de cultivo. Imagen modificada a partir de Dib y col. (2015b).

Tabla 1. Metodologías de aislamiento de plásmidos en estudios de plasmidomas independientes de cultivo.

Muestra	Extracción de ADN total/plasmídico	Purificación de ADN plasmídico	Amplificación	Secuenciación	Referencia
Intestino humano	<i>Genomix Cell and Tissue kit</i> , Talent srl; Lisis alcalina; <i>QIAquick plasmid mini prep kit</i> , QIAGEN	<i>Plasmid-safe DNase</i> , Epicentre; TRACA	NA	NE	Jones y Marchesi, 2007
Cavidad bucal humana	CTAB+Lisozima	<i>Plasmid-safe DNase</i> , Epicentre; TRACA	NA	NE	Warburton y col., 2011
Planta de tratamiento de aguas residuales (Hong Kong, China)	<i>FastDNA Soil Kit</i> , MP Biomedicals; <i>QIAGEN Plasmid Mini Kit</i> , QIAGEN SCIENCES INC	<i>Plasmid-safe DNase</i> , Epicentre; TRACA	NA	Illumina Hiseq 2000	Zhang y col., 2011
Rumen bovino	Tres métodos de lisis distintos	<i>Plasmid-safe DNase</i> , Epicentre	ADN polimerasa $\Phi 29$	Illumina GAIIX	Kav y col., 2012
Planta de tratamiento de aguas residuales (Dinamarca)	<i>Plasmid mini AX kit</i> , A&A Biotechnology	<i>Plasmid-safe DNase</i> , Epicentre	ADN polimerasa $\Phi 29$	454 FLX Titanium	Li y col., 2012
Plantas de tratamiento de aguas residuales (Suiza)	Dos métodos de lisis alcalina	Centrifugación en gradiente de densidad de CsCl-BrEt	NA	Sanger; 454 FLX Titanium	Sentchilo y col., 2013
Ciego de rata	<i>Plasmid mini AX kit</i> , A&A Biotechnology	<i>Plasmid-safe DNase</i> , Epicentre	ADN polimerasa $\Phi 29$	Illumina Hiseq 2000; 454 FLX Titanium	Jørgensen y col., 2014
Aguas subterráneas (Tennessee, EE.UU.)	Hidrólisis alcalina	<i>Plasmid-safe DNase</i> , Epicentre	ADN polimerasa $\Phi 29$	Illumina Miseq	Kothari y col., 2019
Aguas residuales globales (22 países)	<i>QIAGEN Plasmid Mini Kit</i> , QIAGEN	<i>Plasmid-safe DNase</i> , Epicentre	ADN polimerasa $\Phi 29$	Oxford Nanopore; Illumina NextSeq	Kirstahler y col., 2021

TRACA: método de captura asistida por transposón

NA: no aplica

NE: no especificada

CTAB: Bromuro de cetiltrimetilamonio

8.2. Softwares para la reconstrucción *in silico* de plásmidos circulares

Como ya se mencionó antes, los avances en el campo de la secuenciación han generado gran cantidad de datos que requieren de un constante esfuerzo de análisis. En cualquier estrategia que implique secuenciación de alto rendimiento de lectura corta, el ensamblaje de *contigs* es un desafío y a veces resulta difícil obtener una secuencia plasmídica completa a partir de datos metagenómicos tan fragmentados, especialmente para elementos grandes o con numerosas secuencias repetitivas (Arredondo-Alonso y col., 2017). Se han desarrollado varias herramientas para permitir la predicción y reconstrucción automatizada *in silico* de plásmidos a partir de datos de secuenciación de genomas. No obstante, la mayoría de ellas no pueden ser aplicadas a conjuntos de datos de metagenomas o plasmidomas. Recycler constituye una de las excepciones, ya que es capaz de extraer con éxito secuencias completas de plásmidos circulares a partir de los gráficos de ensamblaje proporcionados por un ensamblador *de novo* convencional. Sin embargo, como la mayoría de los plásmidos grandes tienden a ensamblarse en varios *contigs*, la herramienta se encuentra limitada a la detección de plásmidos o elementos circulares pequeños (Rozov y col., 2016). Por otra parte, anteriormente se mencionó que Kothari y col. (2019) desarrollaron recientemente una *pipeline* para identificar aquellos *contigs* que son circulares, permitiendo de esta forma obtener plásmidos de mayor tamaño. Sin embargo, la predicción correcta, utilizando únicamente datos de secuenciación de lectura corta, sigue siendo un desafío debido a la abundante presencia de secuencias repetidas que dificultan el proceso de ensamblaje. Aunque una posible solución podría ser complementar esta información con datos de secuenciación de lectura larga, los elevados costos han restringido su uso.

8.3. Plasmidoma y plásmidos lineales

A pesar de la demostrada presencia en diversos ambientes microbianos y de su indudable importancia ecológica (Meinhardt y col., 1990, 1997; Ravel y col., 2000), los elementos genéticos lineales se ignoran en gran medida en los estudios de plasmidomas actuales (Dib y col., 2015b). El uso de ADNasa (*plasmid-safe DNase*) en los protocolos experimentales aplicados elimina cualquier tipo de ADN lineal, es decir que no solo degrada los fragmentos cromosómicos, sino también los plásmidos lineales. De hecho, como estos elementos suelen ser relativamente grandes, su aislamiento, purificación y caracterización requieren procedimientos específicos (Dib y col., 2010a, 2010b). Además, sus características estructurales pueden interferir con el registro de sus secuencias completas (Parschat y col., 2007; Fan y col., 2012), e incluso, las excesivas repeticiones de secuencias internas y el alto sesgo de GC, pueden requerir la combinación de secuenciación de alto rendimiento y el método convencional de Sanger para obtener una cobertura confiable y luego un ensamblaje de lecturas adecuado (Wagenknecht y col., 2010).

Por lo tanto, a pesar de que numerosos plásmidos lineales de microorganismos aislados de diversos ambientes han sido completamente secuenciados y caracterizados genéticamente (Le Dantec y col., 2001; Stecker y col., 2003; Bentley y col., 2004; Parschat y col., 2007; Dib y col., 2013b, 2013c; Miller y col., 2013), el impacto ecológico de tales elementos en su conjunto sigue siendo en gran parte desconocido e intrigante, debido a la falta de protocolos experimentales adecuados para el análisis del plasmidoma “lineal” y a las características antes mencionadas. En consecuencia, los elementos lineales, por ahora, se siguen estudiando de forma individual.

9. Puna andina

Los Andes centrales de América del Sur comprenden el área que se extiende desde el suroeste de Bolivia hasta el norte de Chile y noroeste de Argentina (Figura 6). Esta región de gran altitud (3.000 - 6.000 m s.n.m.), también conocida como Altiplano (en Bolivia), Desierto de Atacama (en Chile) y Puna andina (en Argentina), muestra una geografía única de cuencas endorreicas donde existe una gran reserva de lagos andinos, salares y volcanes activos. Se trata de ambientes expuestos a múltiples condiciones extremas, tales como elevada radiación UV, alta sequedad, hipersalinidad, alcalinidad, altas concentraciones de metales pesados y metaloides como el arsénico, baja presión de oxígeno y grandes fluctuaciones térmicas (Farías, 2020; Vignale y col., 2021).



Figura 6. Ubicación geográfica de los Andes centrales de América del Sur (recuadro en color rojo). Fuente: Google Maps.

9.1. Ecosistemas microbianos andinos

Los Andes centrales constituyen un gran reservorio de ecosistemas microbianos andinos (EMAs), que incluye biopelículas (o *biofilms*), tapetes microbianos, microbialitos y endoevaporitas (Vignale y col., 2021).

Las biopelículas son comunidades de microorganismos que se asocian a una gran variedad de superficies sólidas. La colonización microbiana de esta superficie depende de la formación de una matriz de sustancias poliméricas extracelulares (*EPS*), que pueden atrapar compuestos inorgánicos y abióticos, además de inmovilizar el agua (Marshall y col., 2012). En cuanto a su composición taxonómica, las biopelículas pueden variar desde una sola especie hasta comunidades complejas.

Los tapetes microbianos son biopelículas distribuidas en capas que se definen por la disponibilidad de luz y oxígeno, y por otros requerimientos fisicoquímicos. Los microorganismos que construyen el tapete escapan del entierro por sedimentos recién depositados migrando verticalmente hacia arriba, donde pueden colonizar nuevamente las superficies recién depositadas (Prieto-Barajas y col., 2018). Algunas de estas comunidades pueden influir en la precipitación de carbonatos dando lugar a estructuras denominadas microbialitos. Estos se definen por lo tanto, como depósitos órgano-sedimentarios acrecentados por la captura, la unión y/o la precipitación *in situ* de sedimentos debido al crecimiento, las actividades metabólicas y la matriz de *EPS* producida por los microorganismos del tapete microbiano (Burne y Moore, 1987). Los registros geológicos indican que aparecieron por primera vez hace 3.500 millones de años y fueron la principal evidencia de vida en la Tierra durante los siguientes 2.000 millones de años (Allwood y col., 2006; Schopf, 2006). Además, al ser las primeras comunidades en realizar una fotosíntesis

oxigénica significativa, se cree que desempeñaron un papel importante en la oxigenación de la atmósfera de la Tierra (Kasting, 1991).

Estas comunidades complejas se desarrollan en una amplia variedad de ambientes acuáticos poco profundos, que van desde agua dulce hasta agua hipersalina, e incluso dentro de una costra de halita o yeso (Stivaletta y Barbieri, 2009). De hecho, las endoevaporitas se han descrito como evaporitas con una comunidad microbiana endolítica laminada asociada (Rothschild y col., 1994; Sahl y col., 2008). Las evaporitas son estructuras que se originan a partir de la evaporación del agua salina en las áreas poco profundas de los lagos hipersalinos. Como las condiciones ambientales presentes en los Andes centrales se asemejan bastante a las condiciones de la Tierra primitiva, el estudio de los EMAs podría aportar información valiosa sobre la evolución temprana de la vida en la Tierra. En la última década, se ha reportado la presencia de numerosos ecosistemas microbianos en estos ambientes. En la Tabla 2, se resumen aquellos encontrados en lagunas de Argentina (Puna andina).

Tabla 2. EMAs encontrados en lagunas de altura de la Puna andina (Argentina).

Provincia	Laguna andina	Coordenadas	Altitud (m s.n.m.)	EMAs	Referencia
Jujuy	Laguna Vilama	22° 35' S 66° 55' O	4.660	Tapete microbiano	Farías y col., 2009
Salta	Laguna Socompa	24° 28' S 68° 17' O	3.570	Microbialitos, Endoevaporitas	Farías y col., 2013; Kurth y col., 2017; Toneatti y col., 2017
	Ojos de Mar de Tolar Grande	24° 37' 23.51" S 67° 22' 14.35" O	3.510	Endoevaporitas	Farías y col., 2011
Catamarca	Laguna Verde y Bosques del Salar	25° 28' 45.73" S 67° 33' 17.72" O	3.343	Endoevaporitas, Tapete microbiano	Farías y col., 2020
	Laguna Pozo Bravo	25° 30' 48.55" S 67° 34' 37.18" O	3.333	Microbialitos, Tapete microbiano	Farías y col., 2020
	Puquio de Campo Naranja	25° 36' 38.05" S 67° 40' 20.95" O	3.340	Microbialitos, Tapete microbiano	Farías y col., 2020
	Las Quinoas	25° 52' 15.50" S 67° 54' 25.56" O	3.338	Microbialitos, Tapete microbiano	Villafañe y col., 2021a
	Laguna El Peinado	26° 30' 46.16" S 68° 05' 50.09" O	3.748	Microbialitos, Tapete microbiano	Valero-Garcés y col., 2001; Villafañe y col., 2021b
	Laguna Carachipampa	26° 27' 02.29" S 67° 30' 39.38" O	3.018	Microbialitos, Tapete microbiano	Farías, 2018; Farías y col., 2020
	Laguna Diamante	26° 00' 51.04" S 67° 01' 46.42" O	4.589	Microbialitos, <i>Biofilms</i>	Rascovan y col., 2016; Saona y col., 2019; Sancho-Tomás y col., 2020; Stepanenko y col., 2020
	Laguna Negra	27° 38' 34.82" S 68° 33' 53.00" O	4.200	Microbialitos, Tapete microbiano	Gomez y col., 2014, 2020; Beeler y col., 2020; Boidi y col., 2020

9.2. Estudios preliminares de plásmidos en ambientes andinos

En trabajos previos, se demostró que en estos ambientes reside una amplia diversidad de microorganismos extremos que necesitan desarrollar mecanismos eficientes y estrategias para sobrevivir bajo las condiciones prevalecientes (Dib y col., 2009; Ordoñez y col., 2009; Farías y col., 2013, 2014; Rasuk y col., 2014, 2016; Albarracín y col., 2015, 2016). Entre las bacterias aisladas, se encontraron proteobacterias, firmicutes y actinobacterias resistentes a una elevada radiación UV, a arsénico y a múltiples antibióticos (Dib y col., 2008, 2009; Flores y col., 2009; Ordoñez y col., 2009). Además, se observó que algunas de estas cepas, particularmente las de actinobacterias, llevan plásmidos de gran tamaño, entre 75 y 245 kb aproximadamente. Mediante ensayos experimentales, se demostró que tales elementos son de naturaleza linear (Dib y col., 2010a, 2010b). Las secuencias de nucleótidos de algunos de estos plásmidos revelaron genes asociados a dichas resistencias, lo que apoya la teoría de que muchos factores adaptativos están codificados a nivel extra-cromosómico (Dib y col., 2013b, 2013c, 2015a). Por ejemplo, en la anotación del plásmido pAP13 (89 kb) de *Brevibacterium* sp. Ap13, se encontraron genes involucrados en la reparación del daño del ADN inducido por rayos UV (*umuC* y *umuD*), lo que potencialmente mejora la supervivencia celular en ambientes altamente irradiados (Dib y col., 2013b). Más detalles sobre estos trabajos serán abordados en el capítulo II de esta tesis.

En el caso de comunidades microbianas complejas, tales como microbialitos, tapetes microbianos, endoevaporitas y biopelículas asociadas con los humedales de la Puna andina, aún no se han reportado estudios sobre los elementos móviles en su conjunto. Al ser los plásmidos responsables de albergar y transferir genes importantes para la supervivencia y la aptitud de sus huéspedes en un determinado nicho ecológico, los perfiles funcionales de estos elementos suelen estar asociados con las condiciones presentes en su entorno. Por lo tanto, un

enfoque metagenómico dirigido exclusivamente a la población plasmídica podría revelar adaptaciones únicas requeridas por estas comunidades microbianas extremas, permitiendo así conocer su papel e impacto ecológico en los ambientes andinos.

Hipótesis

El plasmidoma (población plasmídica total) de comunidades microbianas de lagunas de altura de la Puna andina contiene genes accesorios que permiten a los microorganismos que los hospedan sobrevivir y adaptarse a las condiciones extremas de estos ambientes.

El plasmidoma es un factor clave en la evolución y flexibilidad de los genomas en los organismos de ambientes extremos.

Objetivo general

El objetivo general de esta tesis doctoral es la caracterización del plasmidoma de comunidades microbianas de lagunas de altura de la Puna andina, analizando sus funciones y significancia ecológica, utilizando tecnología de secuenciación masiva. Se busca identificar determinantes genéticos de naturaleza plasmídica que permitan a sus huéspedes sobrevivir en las condiciones extremas y adversas presentes en estos ambientes.

Objetivos específicos

- Extracción de ADN plasmídico de comunidades microbianas de lagunas de altura de la Puna andina, utilizando métodos independientes de cultivo.
- Secuenciación del ADN plasmídico. Prefiltrado y ensamblaje de las secuencias generadas.
- Anotación de las regiones codificantes en los *contigs* ensamblados.
- Análisis comparativo de los datos obtenidos con el metagenoma de estas comunidades, con bases de datos de plásmidos y con plasmidomas de otros ambientes.
- Análisis taxonómico de las secuencias plasmídicas obtenidas.
- Identificación y anotación funcional de genes de interés, especialmente aquellos relacionados a la supervivencia en condiciones extremas.
- Análisis comparativo de megaplásmidos lineales de microorganismos aislados previamente de estas lagunas.

En el capítulo I, se describen los plasmidomas obtenidos a partir de las comunidades microbianas de dos lagunas de altura de la Puna andina.

En el capítulo II, se analizan y comparan plásmidos lineales de microorganismos aislados de este tipo de ambientes que, como ya se explicó, a pesar de su frecuente aparición, son excluidos en las concepciones actuales de plasmidomas debido a las estrategias de aislamiento de plásmidos que se utilizan.

CAPÍTULO I

Análisis del plasmidoma de comunidades microbianas andinas

1. Introducción

Las comunidades microbianas que se enfrentan a un entorno como la Puna andina deben desarrollar mecanismos de adaptación adecuados que garanticen su supervivencia. Los estudios metagenómicos preliminares realizados a partir de estas comunidades andinas demuestran la existencia de una gran diversidad microbiana y riqueza genética, lo que lleva a constituir las como fuentes de reservas genómicas y proteómicas que resultan de interés para futuras aplicaciones biotecnológicas (Rascovan y col., 2016; Kurth y col., 2017). Entre las reservas genéticas mencionadas, los plásmidos pueden jugar un rol significativo en la ecología de los microorganismos asociados a aquellos ambientes, ya que con bastante frecuencia llevan los genes que les permiten a sus huéspedes resistir las severas condiciones circundantes (Wagenknecht y col., 2020). Sin embargo, en los estudios metagenómicos convencionales mencionados, posiblemente se subestime el impacto de los plásmidos, ya que es difícil discriminarlos del ADN cromosómico por su bajo número de copias y su relativa baja proporción.

Puquios de Campo está localizado en el Salar de Antofalla, a 3.340 m s.n.m., en el noroeste de la provincia de Catamarca, Argentina. Entre ocho pequeñas piletas naturales (alrededor de 40 m de diámetro), Puquio de Campo Naranja se destaca debido a su intenso color (Figura I.1). Además, el análisis de suelo de este sitio reveló alta salinidad y concentración de metales (Anexo, Tabla A1).

Laguna Diamante se encuentra dentro de la caldera del Volcán Galán (40 km de diámetro), a una altitud de 4.589 m s.n.m., también en la provincia de Catamarca, Argentina (Figura I.1). Constituye la laguna andina con mayor concentración de arsénico reportada hasta el momento (hasta 347 mg L⁻¹ en el verano) (Saona Acuña y col., 2020). En ella, Rascovan y col. (2016) describieron una biopelícula roja que consta de un 94 % de arqueas, principalmente con

miembros pertenecientes a la clase haloarchaea (anteriormente denominada halobacteria). En los estudios metagenómicos sobre esta comunidad microbiana (Rascovan y col., 2016; Saona y col., 2019), se detectó una abundancia de genes relacionados con la respiración anaeróbica del arsenato (*arr*) y la oxidación del arsenito (*aio*), lo que apoya firmemente la suposición de que el arsénico se usa en procesos bioenergéticos. Además, una investigación adicional demostró la presencia de estos genes en 11 haloarcheas aisladas de la misma biopelícula, y su expresión en dos de ellas (Ordoñez y col., 2018). También se encontraron genes relacionados con los mecanismos desintoxicantes para eliminar el arsénico intracelular, como el gen *acr3* y el operón *arsABCRD* (Rascovan y col., 2016; Saona y col., 2019).

Hasta ahora, no se conocían antecedentes con respecto al contenido global de los plásmidos ni sobre su impacto en estos nichos microbianos andinos. La implementación de un enfoque metagenómico direccionado exclusivamente a estos elementos permitió el estudio de los mismos a partir de microorganismos cultivables y no cultivables, superando cualquier dependencia del huésped. Así, el objetivo de este capítulo fue caracterizar el plasmidoma de las comunidades microbianas presentes en Puquio de Campo Naranja y Laguna Diamante, e identificar los potenciales huéspedes y las funciones codificadas. Además, ambos se compararon con el plasmidoma de una planta de tratamiento de aguas residuales que contiene efluentes de la industria química (Visp, Suiza), ya que también se trata de un entorno con altas concentraciones de metales (Sentchilo y col., 2013).

2. Materiales y Métodos

2.1. Sitios de muestreo

Las muestras para el análisis de los plasmidomas se tomaron a partir de comunidades microbianas de dos lagunas de altura de la Puna andina: Puquio de Campo Naranja y Laguna Diamante, ambas ubicadas en la provincia de Catamarca, Argentina (Figura I.1). Los permisos para la colección de muestras fueron otorgados por la Secretaría de Medio Ambiente de esa provincia (N° 22935/2016).

Puquio de Campo Naranja

Las muestras fueron tomadas a partir de microbialitos de Puquio de Campo Naranja (25° 36' 38.05" S, 67° 40' 20.95" O) en Mayo de 2017 (Figura I.1). Los microbialitos se encontraron a una distancia de 10 cm de la orilla del lago. Se tomaron *cores* por triplicado (3 cm² cada uno) a una profundidad de 3 cm, incluyendo varias capas, y se agruparon para obtener muestras representativas. Las muestras se disgregaron manualmente en frascos de plástico estériles, se almacenaron a 4 °C y se procesaron dentro de la misma semana.

Laguna Diamante

Las muestras se tomaron asépticamente a partir de las biopelículas rojas adheridas a cristales de gaylussita en la parte inferior de microbialitos sumergidos en Laguna Diamante (26° 00' 51.04" S, 67° 01' 46.42" O), en septiembre de 2019 (Figura I.1). Los microbialitos se encontraron a una distancia de 2 m de la orilla del lago. Se tomaron muestras de tres sitios elegidos al azar y se combinaron para asegurar la representatividad. Las muestras agrupadas se almacenaron en frascos de plástico estériles a 4 °C y se procesaron dentro de la misma semana.

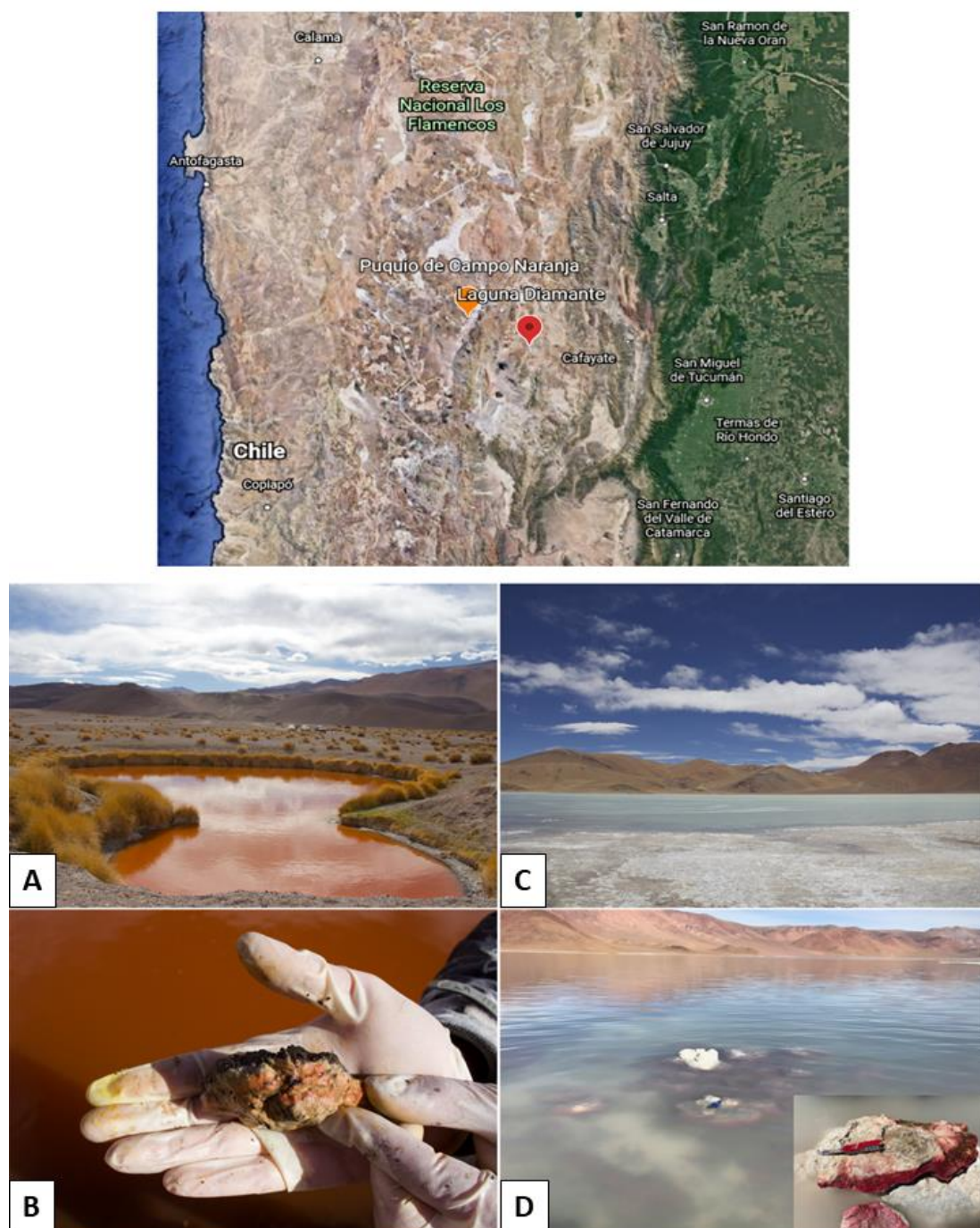


Figura I.1. Ubicación geográfica de las lagunas de altura de la Puna andina, en la provincia de Catamarca, Argentina. Fuente: Google Maps. (A) Puquío de Campo Naranja. (B) Microbialito de Puquío de Campo Naranja. (C) Laguna Diamante. (D) Biopelículas rojas adheridas a cristales de gaylussita en la parte inferior de microbialitos sumergidos en Laguna Diamante.

2.2. Purificación de la biomasa

Los microorganismos fueron separados de las muestras usando el siguiente protocolo: 40 mL de NaCl al 10 % (p/v) fueron agregados a 15 g de muestra e incubados a 40 °C durante 15 min para disolver las sustancias poliméricas extracelulares (Farías y col., 2013). Después se centrifugó a 4 °C y $7.100 \times g$ durante 15 min (UNIVERSAL 320 R, Hettich, Alemania) y el pellet se resuspendió en 15 mL de buffer PBS 4X (solución *stock* 10X: NaCl 80 g L⁻¹; KCl 2 g L⁻¹; Na₂HPO₄ 14,4 g L⁻¹; KH₂PO₄ 2,4 g L⁻¹; pH 7,4). A continuación, la muestra se sometió a un baño de ultrasonido durante 15 min (TB 010 TA, Testlab, Argentina), y se recogieron las células mediante centrifugación suave a 4 °C y $700 \times g$ durante 1 min. Este último procedimiento con PBS 4X fue repetido tres veces y los sobrenadantes resultantes se agruparon y centrifugaron a 4 °C y $7.100 \times g$ durante 15 min (Figura I.2). La purificación de biomasa se realizó dos veces a partir de la muestra de Puquio de Campo Naranja, y una vez a partir de la muestra de Laguna Diamante. Los tres pellets se mantuvieron a -20 °C hasta la extracción del ADN plasmídico.

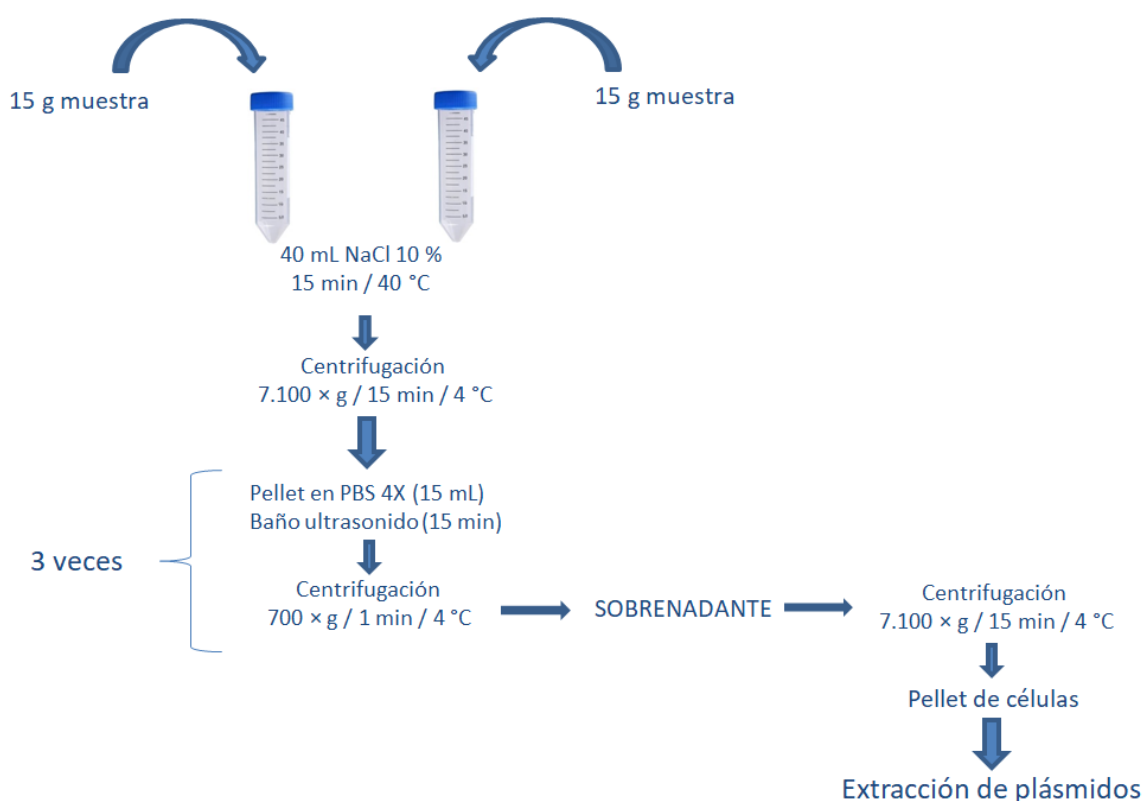


Figura I.2. Representación esquemática del proceso de purificación de biomasa a partir de las muestras de las comunidades microbianas.

2.3. Extracciones de ADN

2.3.1. Extracción de ADN plasmídico

Puquio de Campo Naranja

El ADN plasmídico total fue extraído por dos métodos diferentes con el objetivo de maximizar el rendimiento obtenido. El primer protocolo se realizó según Kado y Liu (1981), el pellet de células fue resuspendido en 35 mL de buffer de lisis (SDS 3 % p/v; TRIS 50 mM; pH 12,6) e incubado a 60 °C durante 1 h. Después, se extrajo dos veces con un volumen de solución fenol/cloroformo (1:1, v/v) y la fase acuosa fue separada por centrifugación a 4 °C y $16.000 \times g$ durante 15 min (Sorvall RC-5C, SS-34 rotor, DuPont, EE.UU.). Luego de dos

extracciones con un volumen de cloroformo y las mismas condiciones de centrifugación anteriores, la fase acuosa fue transferida a un tubo nuevo y el ADN se precipitó con la adición de 0,6 volúmenes de isopropanol y 0,1 volúmenes de acetato de sodio 3 M (pH 5,2), seguido de una incubación *overnight* a 4 °C. El pellet de ADN plasmídico fue obtenido mediante una centrifugación a 4 °C y $23.600 \times g$ durante 30 min, y luego lavado dos veces con etanol 70 %. Finalmente, el ADN se secó a temperatura ambiente durante 10 min y se resuspendió en 400 μ L de agua destilada libre de ADNasas. El segundo procedimiento de extracción fue realizado siguiendo el protocolo de Hansen y Olsen (1978). Las células fueron suspendidas en 16 mL de sucrosa al 25 % (p/v) y TRIS 50 mM (pH 8), y luego se agregaron 120 μ L de lisozima (50 mg mL⁻¹; USB Corporation, Ohio, EE.UU.). Las células suspendidas fueron mezcladas por inversión e incubadas en hielo durante 5 min. A continuación, se agregaron 3 mL de EDTA 250 mM (pH 8) y nuevamente se mezcló por inversión e incubó en hielo durante 5 min. 3 mL de una solución de SDS al 20 % en buffer TE (TRIS 10 mM; EDTA 1 mM; pH 8) fueron también agregados y mezclados por inversión. Se llevaron a cabo ocho ciclos de calor y mezclado (un ciclo incluye 15 s a 55 °C y cinco inversiones durante 15 s a temperatura ambiente). Para desnaturalizar el ADN, se agregaron 3 mL de NaOH 3 M. Después de 3 min de inversión, la mezcla se neutralizó con 6 mL de TRIS 2 M (pH 7). Se añadieron 4 mL de una solución de SDS al 20 % en buffer TE, e inmediatamente después 7,5 mL de NaCl 5 M, se mezcló y se enfrió en hielo. Luego de una incubación *overnight* a 4 °C y una centrifugación a $9.200 \times g$ durante 30 min a la misma temperatura, el sobrenadante conteniendo el ADN plasmídico fue precipitado y lavado como se describió en el primer protocolo.

Laguna Diamante

El ADN plasmídico total se aisló utilizando el kit QIAGEN Large-Construct (QIAGEN, Hilden, Alemania) según las recomendaciones del fabricante.

2.3.2. Extracción de ADN metagenómico

La extracción de ADN metagenómico de muestras de microbialitos de Puquio de Campo Naranja y de la biopelícula roja de Laguna Diamante se realizó usando el kit para suelo FastDNA™ SPIN (MP Biomedicals, California, EE.UU.), siguiendo las recomendaciones del fabricante. En ambos casos, el ADN metagenómico extraído sirvió como molde para la secuenciación del amplicón del gen ARNr 16S (apartado 2.7.).

El conjunto de datos generados a partir de la secuenciación del metagenoma de Puquio de Campo Naranja fue usado para comparar el perfil funcional con los datos del plasmidoma correspondiente. En el caso de Laguna Diamante, para esta comparación, se utilizó el metagenoma previamente secuenciado por Saona y col. (2019).

2.4. Eliminación del ADN cromosómico contaminante

El ADN plasmídico obtenido a partir de ambas muestras fue sometido a una digestión *overnight* a 37 °C con exonucleasa V Rec BCD (New England Biolabs, Massachusetts, EE.UU.). Esta enzima, dependiente de ATP, digiere el ADN lineal de doble cadena en las direcciones 3' - 5' y 5' - 3', asegurando la eliminación del ADN cromosómico. A continuación, se realizaron pruebas de *PCR* usando cebadores universales que abarcan la región V4 del gen ARNr 16S para comprobar la existencia de contaminación con ADN cromosómico (Mastercycler Nexus Gradient, Eppendorf, Alemania). Los cebadores usados para bacterias fueron 341F (5'-CCTACGGGNGGCWGCAG-3') y 785R (5'-GGATTAGATACCCBDGTAGTC-3') (Klindworth y col., 2013); y para arqueas fueron 340F (5'-CCCTAYGGGGYGCASCAG-3') y 915R (5'-ATTAGAKACCCSNGTAGTCC-3')

(Gantner y col., 2011). Las concentraciones y programas empleados se detallan en las Tablas I.1 y I.2, respectivamente.

El ADN plasmídico fue purificado con el kit SureClean Plus (BIOLINE, Londres, Reino Unido) siguiendo las recomendaciones del fabricante. La concentración se midió por cuantificación fluorométrica con Qubit (Invitrogen, Thermo Fisher Scientific, California, EE.UU.).

Tabla I.1. Concentraciones utilizadas en las *PCRs* para determinar la presencia de ADN perteneciente al gen ARNr 16S de bacterias y arqueas.

Concentraciones de PCR		
Componente	Bacterias	Arqueas
5X <i>Phusion HF Buffer</i>	5 µL	5 µL
DMSO	1,25 µL	1,25 µL
Cebador <i>forward</i> (10 µM)	1 µL	0,5 µL
Cebador <i>reverse</i> (10 µM)	1 µL	0,5 µL
MgCl ₂ (50 mM)	0,75 µL	0,1 µL
dNTPs (10 mM)	0,5 µL	0,5 µL
<i>Phusion DNA Polymerase</i> (New England Biolabs; 2 U µL ⁻¹)	0,25 µL	0,25 µL
Templado (ADN 100 ng µL ⁻¹)	0,5 µL	0,5 µL
ddH ₂ O	c.s.p. 25 µL	c.s.p. 25 µL

Tabla I.2. Programas de *PCR* utilizados para determinar la presencia de ADN perteneciente al gen ARNr 16S de bacterias y arqueas.

Programas de PCR				
Etapa	Bacterias		Arqueas	
	Temperatura	Tiempo	Temperatura	Tiempo
Desnaturalización inicial	98 °C	30 s	98 °C	1 min
Desnaturalización	98 °C	10 s	98 °C	45 s
Hibridación	60 °C	30 s	63(53) °C *	45 s
Enlongación	72 °C	30 s	72 °C	30 s
Enlongación final	72 °C	5 min	72 °C	5 min
Conservación	4 °C	∞	4 °C	∞

*se realizaron 10 ciclos con una temperatura de hibridación de 63 °C y 15 ciclos con 53 °C, sumando un total de 25 ciclos

2.5. Secuenciación, control de calidad y ensamblaje

Debido al gran avance de las tecnologías de secuenciación, en la estrategia utilizada en este caso, se eliminó el paso de amplificación con la ADN polimerasa $\Phi 29$ (y por consiguiente el sesgo hacia plásmidos pequeños) comúnmente utilizado en los estudios previos, ya que la cantidad de ADN requerida para secuenciar actualmente es menor.

La secuenciación de todas las muestras se realizó en el Laboratorio de Genómica de la Universidad de Göttingen (Georg-August-Universität Göttingen, Alemania) mediante la tecnología Illumina, que está basada en la fragmentación del ADN, la amplificación en puente para obtener *clusters* de clones del fragmento inicial y la secuenciación propiamente dicha con nucleótidos terminadores, marcados cada uno con un fluorocromo distinto.

A partir del ADN plasmídico se generaron bibliotecas de secuenciación *paired-end* utilizando el sistema MiSeq y el kit de reactivos MiSeq versión 3 (600 ciclos), según las indicaciones del fabricante (Illumina, California, EE.UU.).

La secuenciación *paired-end* del metagenoma de Puquio de Campo Naranja se realizó mediante el sistema HiSeq 2500 (2x250) usando el kit SBS v2 según lo recomendado por el fabricante (Illumina, California, EE.UU.).

El control de calidad de los datos de secuencia en bruto se llevó a cabo mediante la herramienta FastQC (versión 0.11.5 y 0.11.9) (Andrews, 2010) y las lecturas fueron filtradas usando Trimmomatic (versión 0.36 y 0.38) (Bolger y col., 2014). A continuación, las lecturas obtenidas se ensamblaron *de novo* con el software metaSPAdes (versión 3.11.0) (Nurk y col., 2017). A partir de los gráficos de ensamblaje proporcionados por este software, se utilizó el algoritmo Recycler para obtener secuencias cíclicas que probablemente sean plásmidos, fagos y otros elementos circulares (Rozov y col., 2016). La *pipeline* descrita por Kothari y col. (2019) también se utilizó para identificar *contigs* circulares cerrados completos. Los ciclos

obtenidos en ambos casos se compararon con DoriC 10.0, una base de datos de orígenes de replicación en genomas procarióticos que incluyen cromosomas y plásmidos (Luo y Gao, 2019).

La Tabla I.3, al final de esta sección, resume los softwares, las herramientas bioinformáticas y las bases de datos utilizadas para los diferentes objetivos de análisis que se detallan a continuación.

2.6. Análisis bioinformático

2.6.1. Alineamiento plasmidoma-metagenoma

Las lecturas generadas por la secuenciación del ADN plasmídico fueron alineadas con los *contigs* del metagenoma correspondiente a cada comunidad usando la herramienta Bowtie2 (Langmead y Salzberg, 2012), con el objetivo de evaluar que porcentaje de la información genética presente en el plasmidoma se encuentra representada en el metagenoma.

2.6.2. Perfil funcional de los plasmidomas

La anotación de todas las características genómicas relevantes sobre los *contigs* de ambos plasmidomas fueron hechas con el software Prokka (versión 1.11 y 1.14.5) (Seemann, 2014). Además, cada conjunto de datos plasmídicos ensamblados fue cargado en el servidor MG-RAST (Meyer y col., 2008) para un análisis funcional y taxonómico. Se comparó por similitud con la base de datos de subsistemas SEED ($E\text{-value} \leq 10^{-5}$).

El perfil funcional de cada plasmidoma fue comparado con el perfil funcional del metagenoma correspondiente, también obtenido a partir del servidor MG-RAST. Las comparaciones se realizaron mediante el software estadístico STAMP (*Statistical Analysis of*

Metagenomic Profiles) (Parks y Beiko, 2010). Además, los perfiles funcionales de ambos plasmidomas fueron comparados entre sí y con el derivado del plasmidoma de una planta de tratamiento de aguas residuales publicado previamente (Sentchilo y col., 2013).

2.6.2.1. Funciones plasmídicas esenciales: replicación, movilización y mantención

A partir de las secuencias de proteínas identificadas con Prokka, se buscaron dominios funcionales relacionados a la replicación y movilización plasmídica utilizando HMMER 3.3 ($E\text{-value} \leq 10^{-3}$) (Mistry y col., 2013) y las bases de datos Pfam (versión 32.0) (El-Gebali y col., 2019) y MOBFamDB (Garcillán-Barcia y col., 2020). Las familias de Pfam que cubren dominios de replicones plasmídicos conocidos se seleccionaron a partir de Schlüter y col. (2008), Ma y col. (2012), Sentchilo y col. (2013), Jørgensen y col. (2014), Shi y col. (2018), y adiciones propias. Además, los *contigs* de ambos plasmidomas se compararon con la base de datos TADB 2.0 (Xie y col., 2018) mediante la herramienta *blastn* (Altschul y col., 1990) para identificar sistemas de toxina-antitoxina (TA).

2.6.2.2. Funciones plasmídicas accesorias: resistencia a metales, factores de virulencia y resistencia antibiótica

Los genes que codifican resistencia a metales y factores de virulencia se buscaron mediante comparaciones entre los plasmidomas y las bases de datos BacMet (Pal y col., 2014) y VFDB (Chen y col., 2005), respectivamente. Para este propósito, se utilizaron los archivos de anotación de Prokka y la herramienta *crb-blast* (*Conditional Reciprocal Best BLAST*) con un $E\text{-value} \leq 10^{-3}$ (Aubry y col., 2014). Los *hits* con una identidad ≥ 70 % y una cobertura de alineamiento ≥ 90 % fueron seleccionados y sus abundancias fueron calculadas en FPKM

(Fragmentos Por Kilobase por Millón). Los FPKM son una estimación normalizada de la abundancia de genes basada en datos de secuencia. Se calculan a partir del número de lecturas que se asignaron a cada secuencia de gen en particular, teniendo en cuenta la longitud del gen (se espera que se produzcan más lecturas a partir de genes más largos) y la profundidad de secuenciación (se espera que se produzcan más lecturas de la muestra que ha sido secuenciada a una mayor profundidad).

El software RGI (*Resistance Gene Identifier*) fue empleado para la predicción de genes de resistencia antibiótica a partir de la base de datos CARD (*Comprehensive Antibiotic Resistance Database*) (Jia y col., 2017).

Resistencia al arsénico en Laguna Diamante

Debido a la alta concentración de arsénico presente en Laguna Diamante (Saona Acuña y col., 2020), las proteínas relacionadas con la resistencia a este metaloide se buscaron de forma manual en el plasmidoma. Para este propósito, las secuencias de aminoácidos se descargaron de Uniprot (UniProt Consortium, 2019) y se sometieron a PSI-BLAST (*Position-Specific Iterative BLAST*) (Altschul y col., 1997). Para crear conjuntos de datos no redundantes se utilizó CD-HIT (versión 4.8.1) (Li y col., 2002) y para alinear las secuencias Clustal Omega (versión 1.2.4) (Sievers y col., 2011). Los perfiles de los modelos ocultos de Markov (*HMMs*) se construyeron y se buscaron en las secuencias aminoacídicas correspondientes a los genes identificados por Prokka utilizando HMMER 3.3 ($E\text{-value} \leq 10^{-3}$) (Mistry y col., 2013).

2.6.3. Elementos genéticos móviles: integrones, secuencias de inserción y transposasas

Se descargaron todas las secuencias nucleotídicas de integrones de la base de datos INTEGRALL (Moura y col., 2009) y se compararon contra los plasmidomas mediante las

herramientas *blastn* y Bowtie2. El software ISEScan fue utilizado para la búsqueda de otros elementos móviles, tales como secuencias de inserción (Xie y Tang, 2017). La predicción de transposasas procariotas se realizó con HMMER 3.3 y los perfiles *HMMs* descargados de la web TnpPred ($E\text{-value} \leq 10^{-3}$) (Riadi y col., 2012).

2.6.4. Genes y secuencias de plásmidos conocidos

Una base de datos de plásmidos fue creada a partir de todas las secuencias plasmídicas descargadas del NCBI (*National Center for Biotechnology Information*) y comparada con los *contigs* de los plasmidomas mediante la herramienta *blastn*. Los mejores *hits* fueron seleccionados en base a los siguientes parámetros: identidad $\geq 90\%$ y cobertura del *query* $\geq 90\%$ (*hit* de al menos 100 pb de longitud). Además, los archivos de genes anotados por Prokka de ambos plasmidomas se compararon con las secuencias de genes plasmídicos de la base de datos ACLAME (*A Classification of Mobile genetic Elements*) (Leplae y col., 2010) mediante la herramienta *crb-blast* ($E\text{-value} \leq 10^{-3}$). Se seleccionaron los *hits* con una identidad $\geq 70\%$ y una cobertura de alineamiento $\geq 90\%$.

2.7. Secuenciación de amplicones y análisis taxonómico

La secuenciación de amplicones del ADN metagenómico extraído de ambas comunidades se realizó mediante el sistema MiSeq y el kit de reactivos MiSeq versión 3 (600 ciclos), según lo recomendado por el fabricante (Illumina, California, EE. UU.). Se utilizaron los programas, las concentraciones y los cebadores universales que cubren la región V4 del gen ARNr 16S descriptos previamente (apartado 2.4.). Para el control de calidad y el análisis se empleó QIIME 1.9 (Caporaso y col., 2010). Primero, las lecturas *paired-end* se unieron con PEAR

(pear-0.9.10-bin-64) (Zhang y col., 2014). Después se realizó el filtrado de calidad utilizando el script *split_libraries_fastq.py* y se removieron los cebadores usando *cutadapt* (Martin, 2011). Los *zOTUs* (*zero-radius Operational Taxonomic Units*) fueron determinados con USEARCH v11 (Edgar, 2010) y la taxonomía se asignó por comparación con la base de datos Silva 128 para Puquio de Campo Naranja y Silva 132 para Laguna Diamante (Yilmaz y col., 2014).

2.8. Disponibilidad de datos

Los datos de secuencia del plasmidoma, del amplicón del gen ARNr 16S y del metagenoma de Puquio de Campo Naranja se depositaron en el ENA (*European Nucleotide Archive*) con los números de acceso PRJEB34475 y PRJEB30901. Mientras que los datos del plasmidoma y del amplicón del gen ARNr 16S de Laguna Diamante se depositaron en el NCBI (*National Center for Biotechnology Information*) con los números de acceso SRR13795604, SRR13795605 y SRR13795606.

Tabla I.3. Resumen de softwares, herramientas bioinformáticas y bases de datos utilizadas.

Objetivo de análisis	Software/Herramienta	Base de datos	Referencias
Control de calidad de lecturas	FastQC	-	Andrews, 2010
Filtrado de lecturas	Trimmomatic	-	Bolger y col., 2014
Ensamblaje	MetaSPAdes	-	Nurk y col., 2017
Ciclos y <i>contigs</i> circulares	Recycler	-	Rozov y col., 2016
	<i>Pipeline</i> de circularidad	-	Kothari y col., 2019
Orígenes de replicación plasmídica (<i>oriV</i>)	DoriC (web)	DoriC 10.0	Luo y Gao, 2019
Alineamiento plasmidoma-metagenoma	Bowtie2	-	Langmead y Salzberg, 2012
Anotación funcional	Prokka	-	Seemann, 2014
Análisis funcional y taxonómico	MG-RAST	Subsistemas SEED RefSeq	Meyer y col., 2008
Comparación estadística	STAMP	-	Parks y Beiko, 2010
Funciones de replicación y movilización plasmídica	HMMER 3.3	Pfam MOBfamDB	Mistry y col., 2013; El-Gebali y col., 2019 Garcillán-Barcia y col., 2020
Funciones de mantención: sistemas toxina-antitoxina	Blastn	TADB 2.0	Altschul y col., 1990; Xie y col., 2018
Resistencia a metales	crb-blast	BacMet	Aubry y col., 2014; Pal y col., 2014
Factores de virulencia	crb-blast	VFDB	Chen y col., 2005; Aubry y col., 2014
Resistencia antibiótica	RGI	CARD	Jia y col., 2017
Resistencia al arsénico	PSI-BLAST	Uniprot	Altschul y col., 1997; UniProt Consortium, 2019
	CD-HIT		Li y col., 2002
	Clustal Omega		Sievers y col., 2011
	HMMER 3.3		Mistry y col., 2013
Integrones	blastn	INTEGRALL	Altschul y col., 1990; Moura y col., 2009
	Bowtie2		Langmead y Salzberg, 2012
Secuencias de inserción	ISEScan	-	Xie y Tang, 2017
Transposasas	HMMER 3.3	TnpPred	Riadi y col., 2012; Mistry y col., 2013
Genes y secuencias de plásmidos conocidos	blastn	NCBI	Altschul y col., 1990
	crb-blast	ACLAME	Lepplae y col., 2010; Aubry y col., 2014
Análisis taxonómico de amplicones del gen ARNr 16S	QIIME 1.9	Silva 128, Silva 132	Caporaso y col., 2010; Yilmaz y col., 2014

3. Resultados y Discusión

3.1. Aislamiento y análisis inicial de los plasmidomas

El ADN plasmídico de la comunidad microbiana presente en Puquio de Campo Naranja se aisló utilizando dos métodos diferentes. Ambos fueron efectivos y permitieron obtener suficiente ADN para estudios posteriores. Después de su agrupamiento, del tratamiento con la exonucleasa y de la purificación, la concentración final de ADN fue de $58,6 \mu\text{g mL}^{-1}$. La secuenciación produjo 7.075.125 de lecturas de alta calidad, que resultó en el ensamblaje de 135.813 *contigs* ($> 500 \text{ pb}$) con una longitud de *contig* total de aproximadamente 127,9 Mb. Por su parte, la concentración final de ADN plasmídico obtenido a partir de la comunidad de Laguna Diamante fue de $40 \mu\text{g mL}^{-1}$. La secuenciación generó 1.071.941 de lecturas *paired-end*, de las cuales 941.587 pasaron el filtrado de calidad. En este caso, el ensamblador metaSPAdes produjo 13.492 *contigs* ($> 500 \text{ pb}$) correspondientes a aproximadamente 16,9 Mb, lo cual resulta bastante menor a lo obtenido para el plasmidoma de Puquio de Campo Naranja (Tabla I.4).

Tabla I.4. Datos de ensamblaje de los plasmidomas de Puquio de Campo Naranja y Laguna Diamante.

	Puquio de Campo Naranja		Laguna Diamante	
Tamaño del <i>contig</i>	N° de <i>contigs</i>	Longitud total (pb)	N° de <i>contigs</i>	Longitud total (pb)
$\geq 0 \text{ pb}$	404.273	228.249.002	47.872	24.663.296
$\geq 500 \text{ pb}$	135.813	127.870.923	13.492	16.950.743
$\geq 1.000 \text{ pb}$	30.545	58.962.838	5.008	11.238.923
$\geq 5.000 \text{ pb}$	1.100	9.097.016	292	2.160.508
$\geq 10.000 \text{ pb}$	223	3.305.940	37	462.897
$\geq 25.000 \text{ pb}$	15	453.749	0	0
<i>Contig</i> de mayor longitud	37.006 pb		20.415 pb	

3.2. Circularidad de secuencias en los plasmidomas

A pesar de la baja probabilidad de obtener una secuencia plasmídica completa a partir de datos metagenómicos tan fragmentados, se intentó reconstruir plásmidos circulares *in silico*. El software Recycler predijo 39 replicones cerrados en cada plasmidoma (> 1.000 pb), en Puquio de Campo Naranja el más grande fue de 3.341 pb y en Laguna Diamante de 3.313 pb. Sin embargo, ninguno mostró un origen de replicación plasmídica (*oriV*) conocido en comparación con la base de datos DoriC 10.0. Por otra parte, la *pipeline* utilizada para detectar circularidad predijo solo 3 *contigs* circulares pequeños en el plasmidoma de Puquio de Campo Naranja (2.119, 2.330 y 2.583 pb). Aunque en ninguno se encontró un *oriV* conocido, en uno de ellos se pudo anotar una proteína de la familia II/X, la cual incluye proteínas de replicación de fagos y plásmidos. En el plasmidoma de Laguna Diamante, se identificaron 20 *contigs* circulares, el más grande de 8.295 pb y el más pequeño de 2.025 pb. Sólo uno de ellos mostró un *oriV* conocido (87 % de similitud, longitud de alineación de 70 nt), correspondiente al plásmido pSN encontrado en *Haloterrigena thermotolerans* H13 (DoriC ID: pORI00000477). Solo se pudieron anotar proteínas hipotéticas en los marcos abiertos de lectura de todos los ciclos predichos.

3.3. Perfil funcional de los plasmidomas

El análisis de MG-RAST reveló que los plasmidomas contienen una gran fracción de ADN desconocido. En Puquio de Campo Naranja, un total de 248.147 secuencias codificantes (61 %) correspondieron a proteínas predichas con función conocida, mientras que 155.205 (39 %) codificaron proteínas putativas sin función asignada. En Laguna Diamante, los números fueron 18.314 (41,5 %) y 25.800 (58,5 %) para funciones conocidas y desconocidas,

respectivamente. De esta forma, los plasmidomas aparecen como una fuente de nuevos genes, confirmando que los enfoques de secuenciación masiva son adecuados para detectar proteínas con funciones hasta ahora desconocidas.

Por otra parte, la diferencia entre los dos entornos que muestran características medioambientales bastante similares posiblemente se deba a la gran proporción de arqueas presentes en la comunidad de Laguna Diamante. Los genomas de las arqueas contienen típicamente una mayor fracción de "materia oscura" en comparación con los genomas bacterianos, ya que el aislamiento y el cultivo de la mayoría de ellas y, en consecuencia, la caracterización experimental de sus productos génicos, es bastante complicado (Makarova y col., 2019). Asimismo, Sentschilo y col. (2013) reportaron 52 y 66 % de secuencias codificantes sin función asignada en dos plasmidomas de plantas de tratamiento de aguas residuales, mientras Kav y col. (2012) encontraron cerca del 65 % en el plasmidoma de rumen bovino. Sin embargo, en tales casos, posiblemente se deba a la poca información disponible en ese momento; desde entonces, las bases de datos han crecido considerablemente.

Puquio de Campo Naranja

La asignación funcional del plasmidoma de Puquio de Campo Naranja a partir de la base de datos SEED, con 153.038 anotaciones, reveló que los subsistemas más representados son "*Clustering-based subsystems*" (12,8 %) y "*Carbohydrates*" (11,5 %), seguidos de "*Protein Metabolism*" (9,3 %) y "*Amino Acids and Derivatives*" (8,7 %). Los siguientes subsistemas más abundantes fueron "*Miscellaneous*" (6,6 %) y "*DNA Metabolism*" (5,6 %) (Figura I.3A). Estas categorías funcionales también fueron las más representadas en el plasmidoma de agua subterránea reportado por Kothari y col. (2019).

En el subsistema "*DNA Metabolism*", aproximadamente el 47 % de las proteínas estuvieron relacionadas con las funciones de reparación del ADN y el 21 % con las funciones de replicación del ADN. Este último incluía proteínas específicas para la replicación (RepA, RepB y RepC) y la partición (ParA y ParB) de plásmidos.

La gran proporción de las funciones de reparación del ADN podría explicarse a través de la alta exposición de los microorganismos a la radiación UV en la Puna andina, que puede afectar directa e indirectamente la integridad del ADN. Las vías de reparación, además de proteger las células contra los efectos del daño endógeno (Dalhus y col., 2009), confieren resistencia o son un mecanismo de adaptación al entorno extremo. En estudios previos, se aisló e identificó un conjunto de casi cien cepas resistentes a los rayos UV en otros lagos andinos de gran altitud (Fernández Zenoff y col., 2006; Zenoff y col., 2006; Dib y col., 2008; Flores y col., 2009; Ordoñez y col., 2009; Bequer Urbano y col., 2013). Además, a partir de datos genómicos y proteómicos, Kurth y col. (2015) definieron el "resistoma UV" de *Acinetobacter* sp. Ver 3, el cual abarca una variedad de genes que apoyan la excelente resistencia a los rayos UV de esta cepa aislada de Laguna Verde (3.343 m s.n.m.). Asimismo, Alonso-Reyes y col. (2021) determinaron dicho resistoma para *Nesterenkonia* sp. Act20, aislada de Laguna Socompa (3.570 m s.n.m.), mediante el análisis de su genoma y la caracterización fisiológica. En línea con esto, Alonso-Reyes y col. (2020) realizaron un estudio metagenómico de comunidades microbianas acuáticas de diferentes regiones geográficas de todo el mundo (incluidas dos lagunas andinas) considerando a la radiación UV como una variable ecológica. Los datos metagenómicos combinados con mediciones de radiación solar georreferenciadas indicaron que cuanto mayor es el nivel de la misma en el entorno del microbioma, mayor es la abundancia de genes que codifican fotoliasas (enzimas que reparan el daño en el ADN causado por los rayos UV) y menor la diversidad microbiana.

Es decir, existe una concordancia entre la abundancia de estos genes y la intensidad de la radiación. Además, observaron que es más probable que dichos genes estén presentes en los taxones dominantes de los microbiomas altamente irradiados, lo que sugiere una fuerza evolutiva para la supervivencia y el dominio bajo una exposición solar extrema (Alonso-Reyes y col., 2020).

Aunque menos abundantes, los subsistemas "*Stress Response*" y "*Virulence, Disease and Defense*" también estuvieron representados en el plasmidoma. En este último, aproximadamente el 83 % de las proteínas pertenecían a la categoría "*Resistance to antibiotics and toxic compounds*", mientras que en el primero las proteínas relacionadas a las respuestas al estrés oxidativo y osmótico fueron las principales (Figura I.3B). Como ya se mencionó, la radiación UV también actúa indirectamente al producir compuestos foto-oxidantes y especies reactivas de oxígeno, como los radicales peróxidos de hidrógeno, superóxidos e hidroxilos, los cuales dañan el ADN, las proteínas y los lípidos (Demple y Harrison, 1994; Wilson y col., 2004). Por lo tanto, esto indica que la adaptación a este entorno requiere también del desarrollo de otros mecanismos que brinden tolerancia a la radiación UV como los sistemas de respuesta al estrés oxidativo.

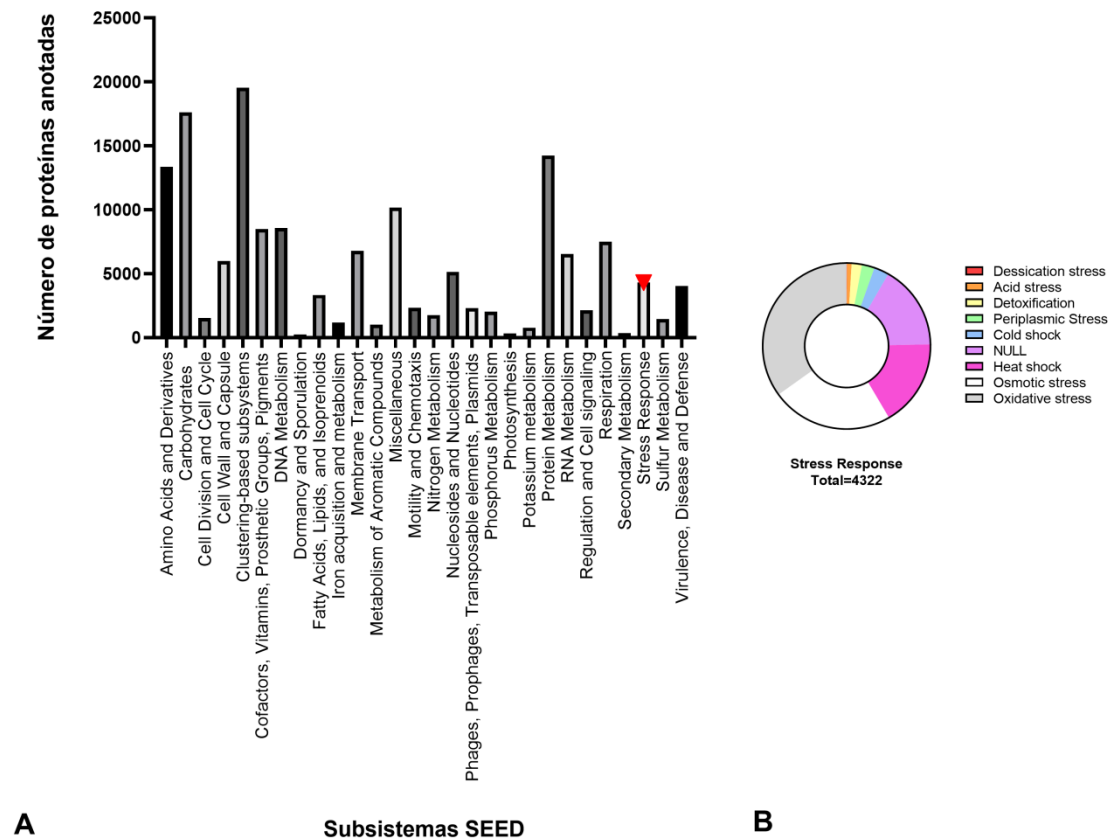


Figura I.3. Perfil funcional del plasmidoma de Puquio de Campo Naranja. **(A)** Número de proteínas anotadas en cada subsistema funcional SEED ($E\text{-value} \leq 10^{-5}$). Los subsistemas SEED con menos del 1 % de los *hits* no son mostrados. **(B)** Clasificación por categorías dentro del subsistema "Stress Response".

Laguna Diamante

En el plasmidoma de Laguna Diamante sólo 5.196 proteínas predichas (28,4 %) se asignaron a subsistemas funcionales SEED, la mayoría de las cuales cubrieron funciones metabólicas básicas, como el metabolismo del ADN, ARN y proteínas (Figura I.4A). Entre las proteínas implicadas en el metabolismo del ADN, aquellas relacionadas con la reparación del ADN fueron bastante diversas (Figura I.4B). Estos hallazgos no resultan inesperados, ya que, como se mencionó en el apartado anterior, la reparación del ADN juega un papel clave como

mecanismo de adaptación para resistir la alta radiación UV presente en la Puna andina (Fernández Zenoff y col., 2006; Flores y col., 2009; Ordoñez y col., 2009; Albarracín y col., 2012; Kurth y col., 2015; Alonso-Reyes y col., 2020, 2021). Además, en este plasmidoma, las proteínas implicadas en la respuesta al estrés oxidativo también se encontraron con mayor frecuencia (57,4 %) dentro del subsistema “*Stress Response*” (Figura I.5). Al igual que los mecanismos de reparación del ADN, los sistemas de respuesta al estrés oxidativo contribuyen a proteger a los microorganismos del daño causado por los rayos UV.

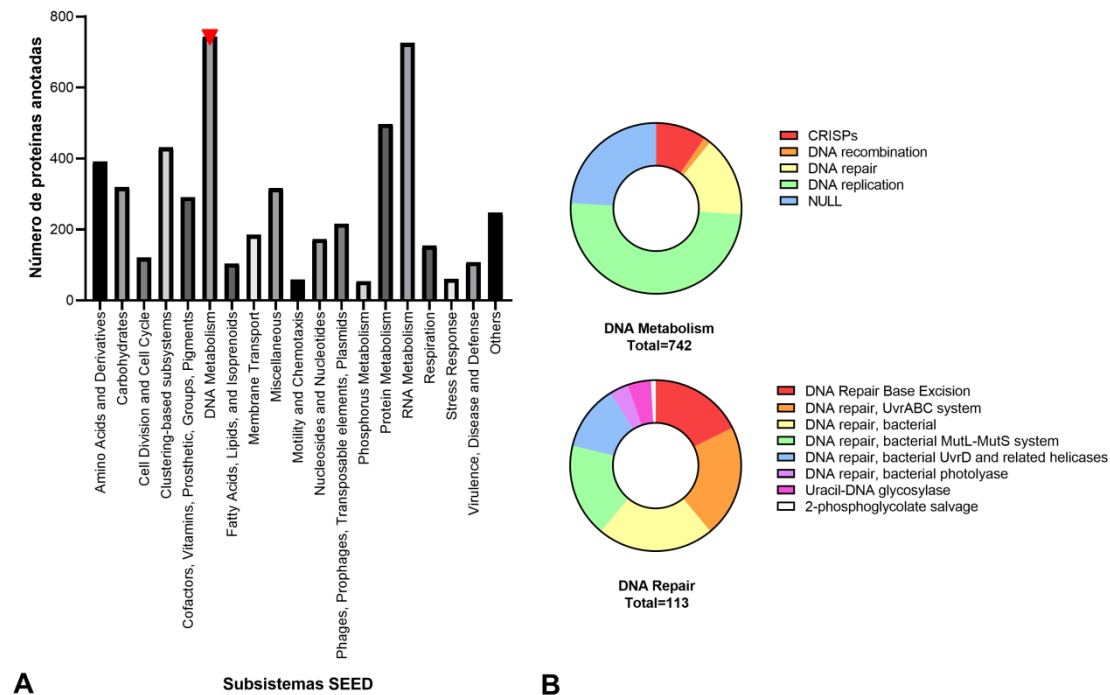


Figura I.4. Perfil funcional del plasmidoma de Laguna Diamante. **(A)** Número de proteínas anotadas en cada subsistema funcional SEED ($E\text{-value} \leq 10^{-5}$). Los subsistemas SEED con menos de 50 *hits* fueron agrupados en “Others”. **(B)** Arriba, clasificación por categorías dentro del subsistema “DNA Metabolism”. Abajo, clasificación dentro de la categoría “DNA Repair” ($E\text{-value} \leq 10^{-5}$).

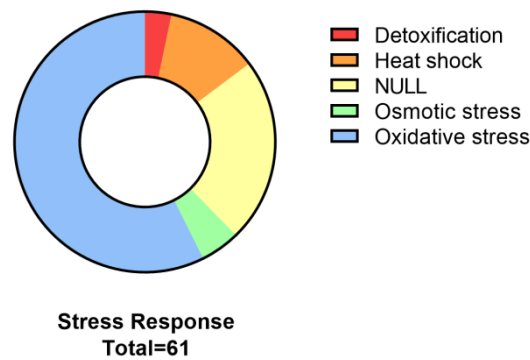


Figura I.5. Perfil funcional del plasmidoma de Laguna Diamante. Clasificación por categorías dentro del subsistema “*Stress Response*” ($E\text{-value} \leq 10^{-5}$).

El 82,4 % de las asignaciones del subsistema “*Virulence, Disease and Defense*” correspondió a la categoría “*Resistance to antibiotics and toxic compounds*”, predominando los sistemas de resistencia al arsénico (67,4 %). Con respecto a este último, las proteínas predichas incluían una arsenato reductasa (ArsC), una ATPasa que impulsa una bomba que conduce el arsenito fuera de la célula (ArsA), un represor de acción *trans* del operón de resistencia al arsénico (ArsD) y una proteína de resistencia al arsénico ACR3 (Figura I.6).

Las características únicas de este entorno, como su ubicación a gran altura, la exposición a condiciones extremas y las peculiaridades de su composición microbiana (gran proporción de arqueas) podrían explicar las relativamente pocas anotaciones funcionales a subsistemas SEED.

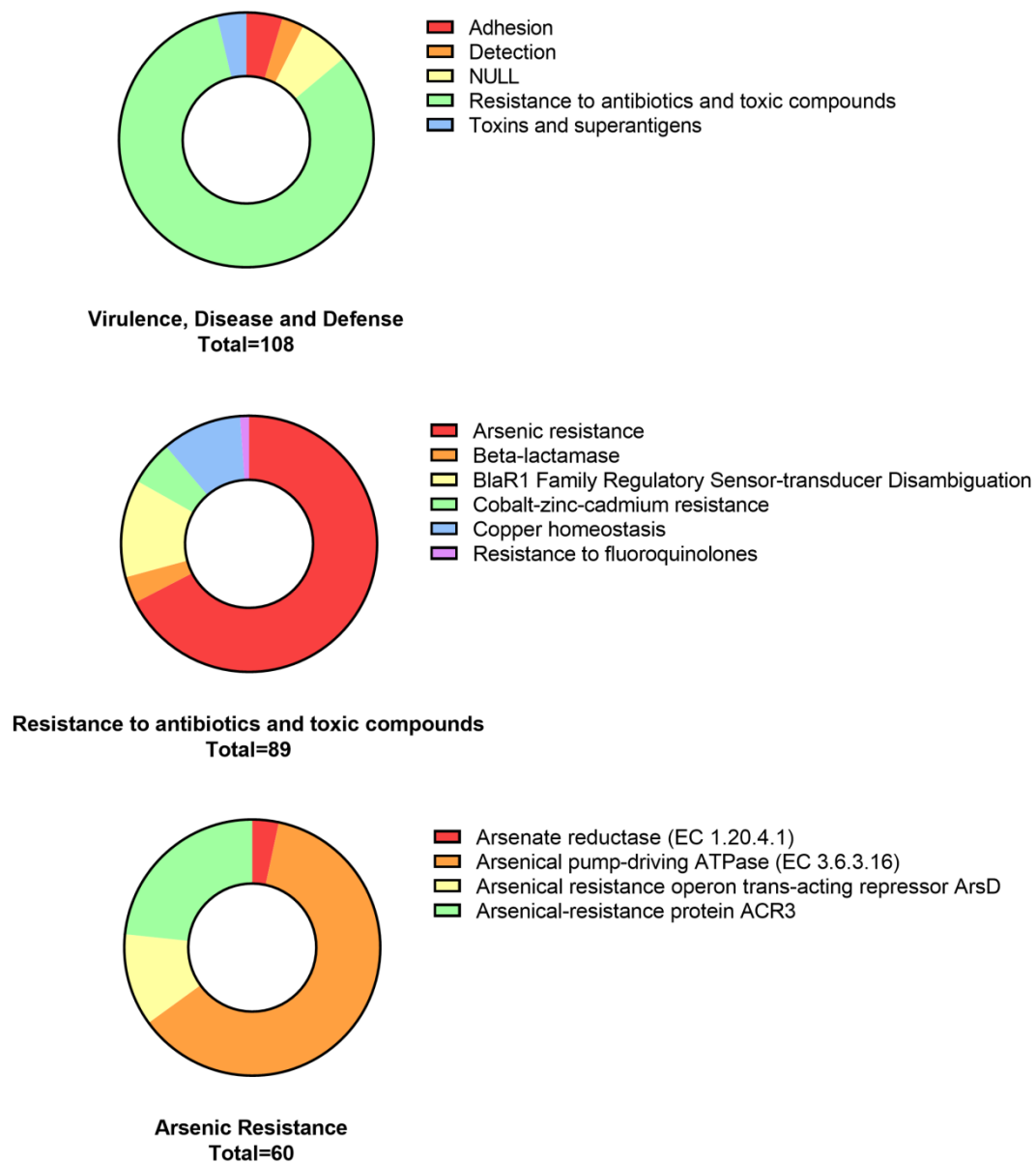


Figura I.6. Perfil funcional del plasmidoma de Laguna Diamante. Arriba, clasificación por categorías dentro del subsistema “*Virulence, Disease and Defense*”. En el medio, clasificación dentro de la categoría “*Resistance to antibiotics and toxic compounds*”. Abajo, clasificación dentro de la categoría “*Arsenic resistance*” ($E\text{-value} \leq 10^{-5}$).

3.4. Comparación funcional entre plasmidomas

A partir de la comparación de los perfiles funcionales de ambos plasmidomas, se observó que los subsistemas “*RNA Metabolism*”, “*DNA Metabolism*”, “*Phages, Prophages, Transposable elements, Plasmids*” y “*Cell Division and Cell Cycle*” están representados en mayor proporción en Laguna Diamante, mientras que “*Carbohydrates*”, “*Cell Wall and Capsule*”, “*Clustering-based subsystems*”, “*Stress Response*” y “*Respiration*” son más abundantes en Puquio de Campo Naranja. No se observaron diferencias significativas en la abundancia de otros subsistemas, lo que sugiere un cierto grado de similitud entre los perfiles funcionales predichos para ambos EMAs (Figura I.7).

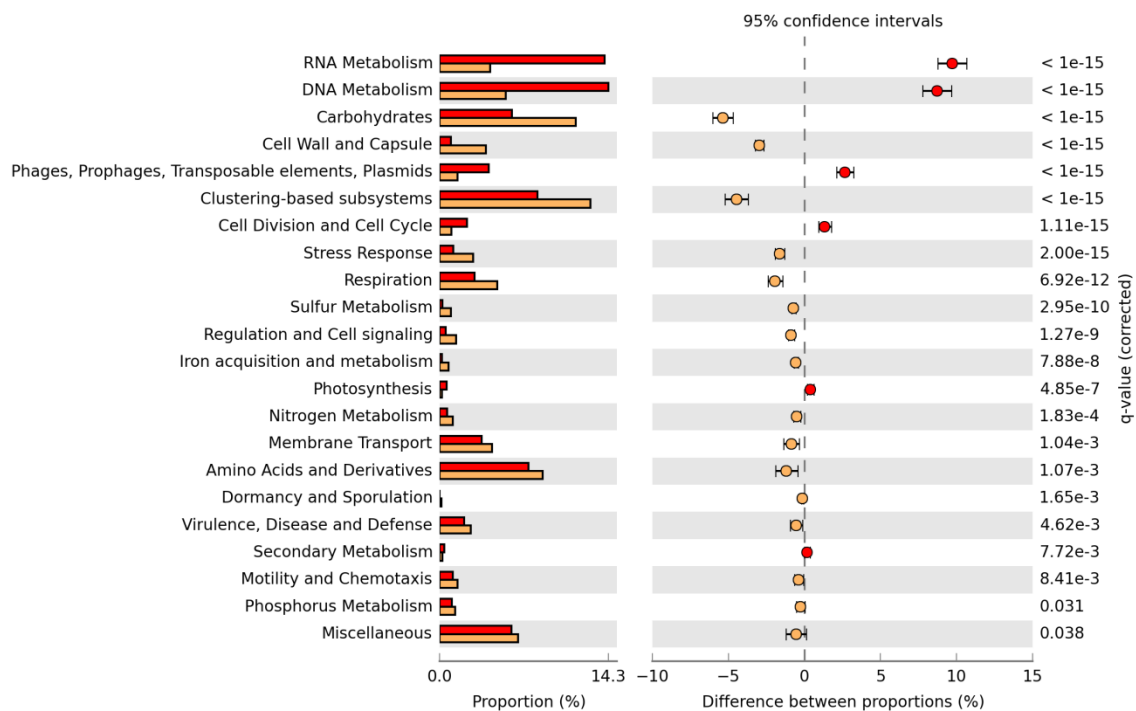


Figura I.7. Comparación de perfiles funcionales entre el plasmidoma de Puquio de Campo Naranja (barras naranjas) y el plasmidoma de Laguna Diamante (barras rojas). Se muestra el porcentaje de genes codificantes de proteínas que pertenecen a cada subsistema SEED y la diferencia de proporciones por pares calculadas con el software STAMP. Se utilizó el test exacto de Fisher y se calcularon los *P-value* corregidos empleando el método Storey's FDR. Solo se muestran los subsistemas estadísticamente significativos (*q-value* < 0,05).

El plasmidoma de Puquio de Campo Naranja se comparó con el plasmidoma de una planta de tratamiento de aguas residuales ubicada en Visp, Suiza, que procesa una mezcla de aguas residuales domésticas y efluentes de la industria química (Sentchilo y col., 2013). Cuando se analizaron las resistencias a antibióticos y a compuestos tóxicos, los genes que codifican bombas de eflujo de resistencia a múltiples fármacos, resistencia al zinc, resistencia al cobalto-zinc-cadmio, así como resistencia a las fluoroquinolonas y antibióticos β -lactámicos mostraron una mayor abundancia relativa en el plasmidoma de Puquio de Campo Naranja que en el plasmidoma de la planta de tratamiento de aguas residuales (Figura I.8). Los genes que codifican resistencia al arsénico se encontraron en mayor abundancia en el plasmidoma de esta última, pero también estuvieron bien representados en el plasmidoma de la comunidad extrema, ubicada en un sitio donde se encontró una alta concentración de arsénico (82,2–109 mg Kg⁻¹; Anexo, Tabla A1). No se detectaron diferencias significativas en la abundancia de otros genes de resistencia, lo que sugiere un cierto grado de similitud entre ambos plasmidomas, a pesar de tratarse de un entorno prístino y un ambiente con impacto antropogénico (Figura I.8).

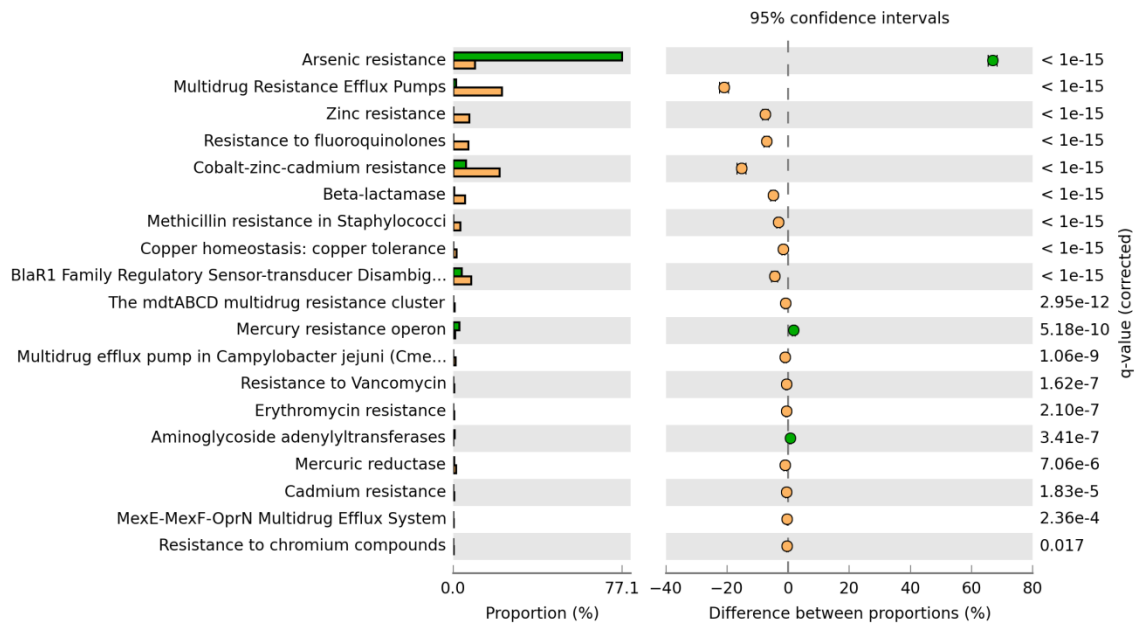


Figura I.8. Comparación de perfiles funcionales entre el plasmidoma de Puquio de Campo Naranja (barras naranjas) y el plasmidoma de la planta de tratamiento de aguas residuales en Visp, Suiza (barras verdes). Se muestran las proporciones de genes codificantes de proteínas dentro de la categoría "*Resistance to antibiotics and toxic compounds*", que pertenece al subsistema SEED "*Virulence, Disease and Defense*"; y la diferencia de proporciones por pares calculadas usando el software STAMP. Se utilizó el test exacto de Fisher y se calcularon los *P-value* corregidos empleando el método Benjamini-Hochberg's FDR. Solo se muestran las categorías estadísticamente significativas (*q-value* < 0,05).

Por otra parte, en la comparación entre el plasmidoma de Laguna Diamante y el de Puquio de Campo Naranja, las diferencias de proporciones entre los genes que codifican proteínas categorizadas dentro de "*Resistance to antibiotics and toxic compounds*", pertenecientes al subsistema SEED "*Virulence, Disease and Defense*", mostraron que la resistencia al arsénico es más abundante en Laguna Diamante que en Puquio de Campo Naranja (Figura I.9). Además, su comparación con la planta de tratamiento de aguas residuales, que contenía efluentes de la industria química (Visp, Suiza) (Sentchilo y col., 2013), reveló que los genes de resistencia al arsénico son también más abundantes en el plasmidoma de Laguna Diamante

(Figura I.9). Estos resultados son de esperarse ya que Laguna Diamante se encuentra entre los ambientes acuáticos con mayor concentración de arsénico del mundo (Rascovan y col., 2016; Sancho-Tomás y col., 2020; Saona Acuña y col., 2020).

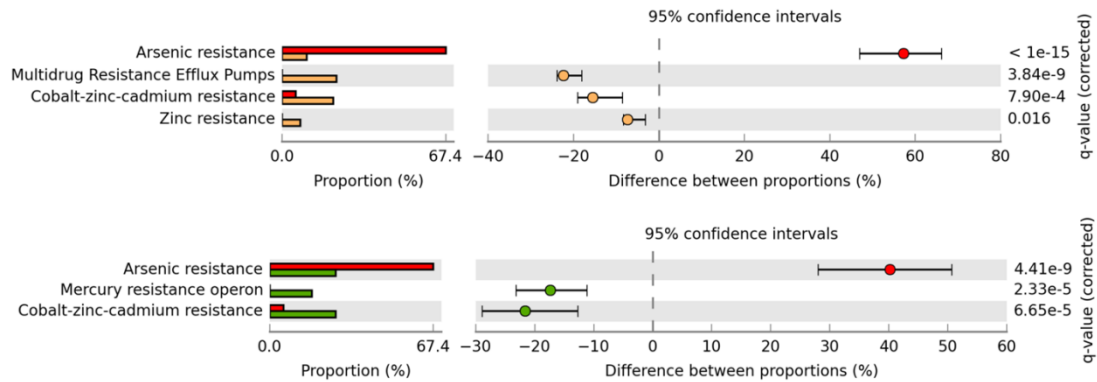


Figura I.9. Comparación de perfiles funcionales entre el plasmidoma de Laguna Diamante (barras rojas), el de Puquio de Campo Naranja (barras naranjas) y el plasmidoma de la planta de tratamiento de aguas residuales en Visp, Suiza (barras verdes). Se muestran las proporciones de genes codificantes de proteínas dentro de la categoría "*Resistance to antibiotics and toxic compounds*", que pertenece al subsistema SEED "*Virulence, Disease and Defense*"; y la diferencia de proporciones por pares calculadas usando el software STAMP. Se utilizó el test exacto de Fisher y se calcularon los *P-value* corregidos empleando el método Benjamini-Hochberg's FDR. Solo se muestran las categorías estadísticamente significativas (*q-value* < 0,05).

3.5. Plasmidoma vs. metagenoma

La incidencia de plásmidos en aislamientos bacterianos de diversos hábitats o en colecciones de cepas es muy variable, incluso menor al 10 % en algunos casos (Heuer y col., 2008). Por lo tanto, debido a su baja tasa de aparición y/o número de copias, los plásmidos generalmente representan solo una pequeña fracción del ADN total de una comunidad microbiana (Li y col., 2012). De esta forma, se podría decir que las secuencias de plásmidos están subrepresentadas

y de alguna manera ocultas en el ADN metagenómico. En consecuencia, aunque se registran casualmente mediante métodos convencionales de metagenómica, la purificación experimental de plásmidos antes de la secuenciación permite un análisis dirigido específicamente a las poblaciones plasmídicas de un modo independiente de cultivo, en el mejor de los casos sin perder información. Además, actualmente las herramientas bioinformáticas para el análisis de plásmidos a partir del ADN metagenómico son limitadas (Arredondo-Alonso y col., 2017), y la posibilidad de secuenciar solo el ADN extra-cromosómico, en lugar del metagenoma, debería aumentar la posibilidad de obtener la secuencia completa de plásmidos particulares. Sin embargo, cabe destacar que esto sigue siendo un desafío cuando se utilizan solamente datos de secuenciación de lectura corta. Esta dificultad ha sido superada en gran medida por la aplicación de la tecnología Oxford Nanopore (Kirstahler y col., 2021), que en combinación con Illumina, permite ampliar la identificación de moléculas circulares completas.

El ADN metagenómico de la muestra de Puquio de Campo Naranja fue extraído y analizado para permitir la comparación con el plasmidoma correspondiente. Solo el 30 % de las lecturas del plasmidoma se alinearon con los *contigs* derivados del metagenoma. Lo mismo ocurrió con la comunidad de Laguna Diamante, en la que el 52 % de las lecturas del plasmidoma se alinearon con los *contigs* del metagenoma descrito anteriormente por Saona y col. (2019). Por lo tanto, estos resultados apoyan la idea de que la purificación de plásmidos previa a la secuenciación cumple de manera más satisfactoria los requisitos para juzgar exhaustivamente la importancia ecológica de las secuencias transmitidas por plásmidos.

La distribución funcional de los plasmidomas se comparó con la de los metagenomas utilizando los conjuntos de datos derivados de MG-RAST y STAMP para el análisis estadístico. En Puquio de Campo Naranja, se encontró que el plasmidoma está

significativamente enriquecido en genes asociados con la respiración, el transporte de membrana, la regulación y la señalización celular, y el metabolismo de proteínas (Figura I.10). Los genes que comúnmente codifican los plásmidos, pertenecientes a los subsistemas "Phages, Prophages, Transposable elements, Plasmids" y "Stress Response" también fueron más abundantes en el plasmidoma que en el metagenoma. Lo mismo ocurrió con los genes asociados al metabolismo del nitrógeno, entre ellos aquellos relacionados con la fijación de nitrógeno (Figura I.10). Si bien el subsistema "Virulence, Disease and Defense" se mostró más enriquecido en el metagenoma, también estuvo bien representado en el plasmidoma (Figura I.3A y I.10). En cambio, en la categoría "DNA Metabolism", que contiene genes asociados a la reparación del ADN, no se detectaron diferencias significativas (Figura I.10).

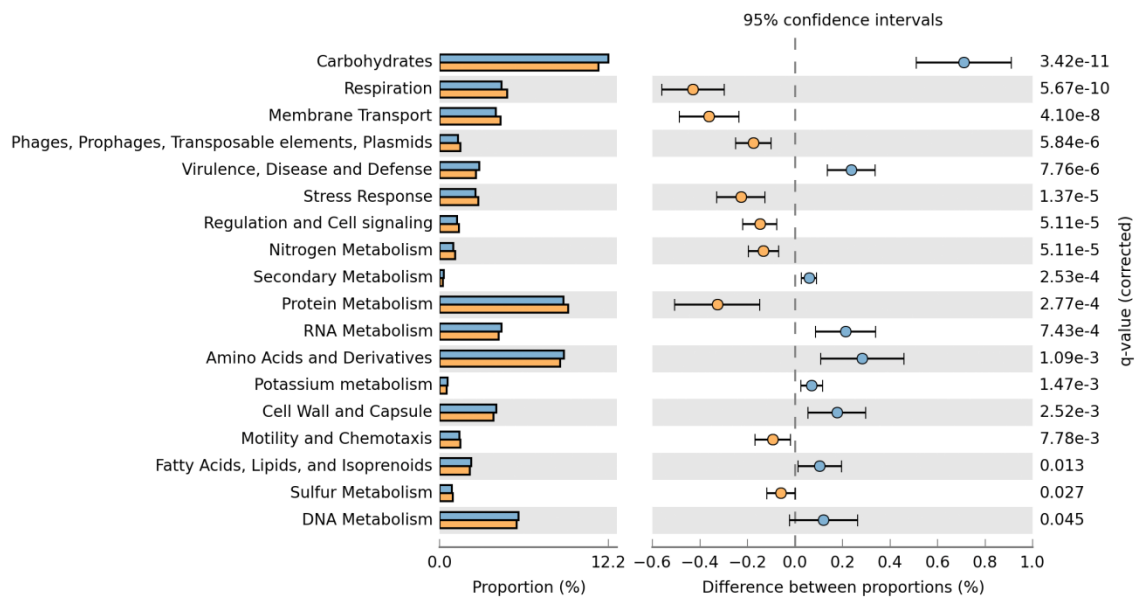


Figura I.10. Comparación de perfiles funcionales entre el plasmidoma (barras naranjas) y el metagenoma (barras azules) de Puquio de Campo Naranja. Se muestran las proporciones de genes codificantes de proteínas categorizables de SEED y la diferencia de proporciones por pares calculadas usando el software STAMP. Se utilizó el test exacto de Fisher y se calcularon los *P-value* corregidos empleando el método Storey's FDR. Solo se muestran los subsistemas SEED estadísticamente significativos ($q\text{-value} < 0,05$).

En Laguna Diamante, la comparación por pares con el objetivo de distinguir el acervo génico del plasmidoma y del metagenoma, reveló que los subsistemas “*Phages, Prophages, Transposable elements, Plasmids*” y “*Membrane Transport*” están representados con mayor frecuencia en el plasmidoma (Figura I.11).

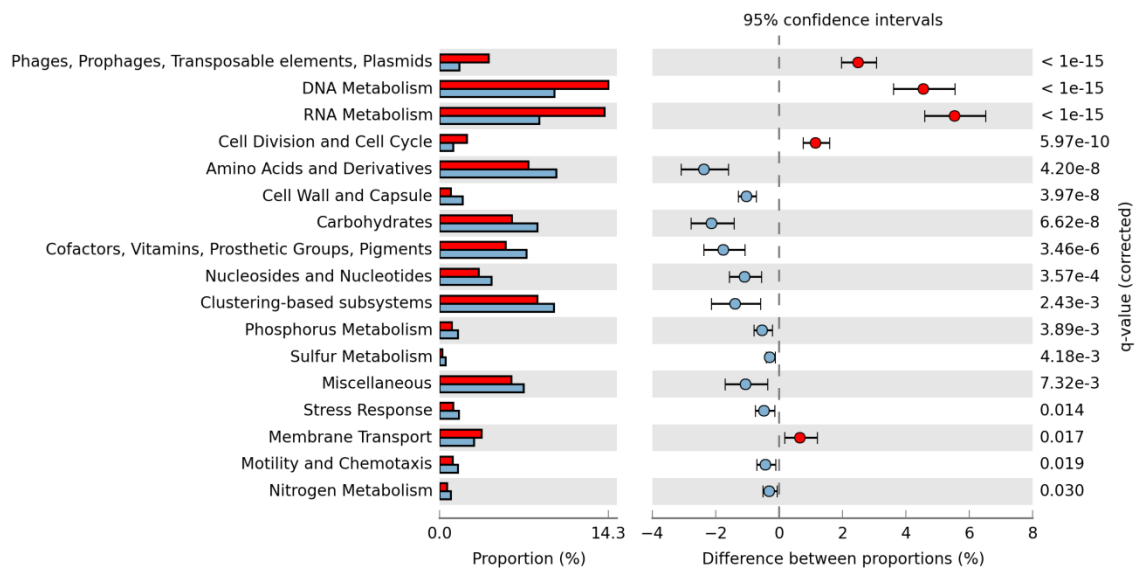


Figura I.11. Comparación de perfiles funcionales entre el plasmidoma (barras rojas) y el metagenoma (barras azules) de Laguna Diamante. Se muestran las proporciones de genes codificantes de proteínas categorizables de SEED y la diferencia de proporciones por pares calculadas usando el software STAMP. Se utilizó el G-test (w/Yates') y se calcularon los *P-value* corregidos empleando el método Benjamini-Hochberg's FDR. Solo se muestran los subsistemas SEED estadísticamente significativos (*q-value* < 0,05).

3.6. Funciones plasmídicas esenciales

Con el fin de identificar rasgos plasmídicos esenciales en los plasmidomas, se buscaron dominios Pfam relacionados con la replicación de plásmidos, familias de relaxasas de tipo MOB que están implicadas en su movilización y sistemas toxina-antitoxina (TA) relacionados con la mantención plasmídica.

3.6.1. Replicación plasmídica

Las familias de proteínas que cubren los productos génicos putativos implicados en la replicación de plásmidos y su regulación se enumeran en la Tabla I.5. En Puquio de Campo Naranja, se detectaron 21 de las 32 familias de proteínas conocidas, donde RHH_1 (PF01402.21), RepL (PF05732.11) y SSB (PF00436.25) fueron las más abundantes. La familia RHH_1 contiene el represor transcripcional CopG codificado por plásmidos, anteriormente denominado RepA (Del Solar y Espinosa, 1992; Del Solar y col., 1995). Este es el represor transcripcional más corto reportado hasta la fecha y el primero aislado de un plásmido. CopG reprime su propia síntesis y la de la proteína iniciadora de replicación RepB. La familia RepL cubre las proteínas de inicio de replicación RepL, que están involucradas en la replicación de plásmidos en Firmicutes, mientras que SSB incluye proteínas de unión "monocatenarias" y también la proteína de replicación N (PriB). En el plasmidoma de Laguna Diamante, RHH_1 y RepL también estuvieron entre los principales dominios Pfam de replicación plasmídica encontrados, pero en este caso acompañados de DUF1424, una familia de varias proteínas arqueales que parecen encontrarse exclusivamente en las especies de *Halobacterium* y *Haloferax* (Tabla I.5). Aunque se desconoce la función de esta familia, es probable que sus miembros sean proteínas Rep debido a la presencia de motivos funcionales conservados (Bateman y col., 2010; El-Gebali y col., 2019).

Rep_1 y Rep_3 son los dominios pertenecientes a las principales familias de proteínas de iniciación de la replicación. Se han reportado entre los más abundantes en plasmidomas de plantas de tratamiento de aguas residuales y de un ciego de rata (Schlüter y col., 2008; Sentchilo y col., 2013; Jørgensen y col., 2014; Shi y col., 2018). Aunque con una tasa de aciertos menor, estos dominios también se encontraron en los plasmidomas andinos; mientras que otros dominios de familias de proteínas conocidas no fueron detectados (Tabla I.5). Posiblemente, existan sistemas de replicación para los cuales los detalles moleculares y los mecanismos se desconocen en la actualidad, principalmente porque la mayoría de los microorganismos contribuyentes no fueron cultivados y, en el caso particular de Laguna Diamante, debido a la composición taxonómica inusual dominada por taxones específicos (Rascovan y col., 2016). De todas formas, la diversidad de familias de proteínas de replicación sugiere que los plasmidomas albergan tipos de plásmidos muy diferentes.

Tabla I.5. Pfam relacionadas a la replicación plasmídica en los plasmidomas de Puquio de Campo Naranja (PCN) y Laguna Diamante (LD).

Nombre Pfam	Código Pfam	Descripción	Hits PCN	Hits LD
Rep_1	PF01446.17	<i>Replication protein</i>	14	45
Rep_2	PF01719.17	<i>Plasmid replication protein</i>	0	1
Rep_3	PF01051.21	<i>Initiator Replication protein</i>	15	5
RepL	PF05732.11	<i>Firmicute plasmid replication protein (RepL)</i>	59	85
TrfA	PF07042.11	<i>TrfA protein</i>	6	3
RepA_C	PF04796.12	<i>Plasmid encoded RepA protein</i>	6	2
Rep_trans	PF02486.19	<i>Replication initiation factor</i>	5	0
RHH_1	PF01402.21	<i>Ribbon-helix-helix protein, copG family</i>	293	344
Rop	PF01815.16	<i>Rop protein</i>	0	0
RP-C	PF03428.13	<i>Replication protein C N-terminal domain</i>	13	32
RPA	PF10134.9	<i>Replication initiator protein A</i>	12	2
RepA_N	PF06970.11	<i>Replication initiator protein A (RepA) N-terminus</i>	0	5
RepC	PF06504.11	<i>Replication protein C (RepC)</i>	4	4
Replicase	PF03090.17	<i>Replicase family</i>	0	5
IncFII_repA	PF02387.15	<i>IncFII RepA protein family</i>	0	0
PriCT_1	PF08708.11	<i>Primase C terminal 1 (PriCT-1)</i>	3	4
DUF1424	PF07232.11	<i>Putative rep protein (DUF1424)</i>	1	179
Phage_CRI	PF05144.14	<i>Phage replication protein CRI</i>	4	0
PriCT_2	PF08707.11	<i>Primase C terminal 2 (PriCT-2)</i>	0	1
Prim-Pol	PF09250.11	<i>Bifunctional DNA primase/polymerase, N-terminal</i>	0	12
SSB	PF00436.25	<i>Single-strand binding protein family</i>	45	0
RepA1_leader	PF08048.12	<i>Tap RepA1 leader peptide</i>	0	0
DUF1738	PF08401.11	<i>Domain of unknown function (DUF1738)</i>	3	0
UvrD-helicase	PF00580.21	<i>UvrD/REP helicase N-terminal domain</i>	29	12
RepB-RCR_reg	PF10723.9	<i>Replication regulatory protein RepB</i>	10	30
Replic_Relax	PF13814.6	<i>Replication-relaxation</i>	24	73
KorB_C	PF06613.11	<i>KorB C-terminal beta-barrel domain</i>	0	0
KorB	PF08535.10	<i>KorB domain</i>	39	6
Activator-TraM	PF11657.8	<i>Transcriptional activator TraM</i>	2	1
pRN1_helical	PF13010.6	<i>Primase helical domain</i>	0	0
KORA	PF16509.5	<i>TrfB plasmid transcriptional repressor</i>	19	17
Rol_Rep_N	PF18106.1	<i>Rolling Circle replication initiation protein N-terminal domain</i>	0	0

3.6.2. Movilización plasmídica

Se buscaron proteínas con dominios relacionados a la movilización de plásmidos, en particular relaxasas de tipo MOB. En el plasmidoma de Puquio de Campo Naranja, se detectó al menos un representante de cada una de las nueve familias diferentes de relaxasas (Tabla I.6). La familia más abundante fue MOB_T, para la cual el perfil *HMM* (profile_MOBT) se construyó recientemente a partir de una alineación de 498 relaxasas (Garcillán-Barcia y col., 2020). Por su parte, en el plasmidoma de Laguna Diamante, las familias de relaxasas más abundantes fueron MOB_C y MOB_P, para las cuales se contaron 41 y 18 coincidencias de dominios proteicos, respectivamente. También estuvieron presentes MOB_T, MOB_V y MOB_M (3, 2 y 1 coincidencia de dominios de proteína, respectivamente) (Tabla I.6).

Aunque los elementos de movilización ya se han informado en plasmidomas anteriores (Schlüter y col., 2008; Kav y col., 2012; Sentchilo y col., 2013; Jørgensen y col., 2014), la clasificación en familias de relaxasas MOB propuesta por Garcillán-Barcia y col. (2009, 2020) no se realizó en la mayoría de ellos; excepto por Kothari y col. (2019), quienes informaron que las familias MOB_Q y MOB_P son las más abundantes en plásmidos circulares de plasmidomas de agua subterránea.

Tabla I.6. Familias de relaxasas MOB en los plasmidomas de Puquio de Campo Naranja (PCN) y Laguna Diamante (LD).

Familia MOB	Perfil HMM	Hits PCN	Hits LD
MOB _B	T4SS_MOBB	1	0
MOB _C	T4SS_MOBC	4	41
MOB _F	profile_MOBF	3	0
MOB _H	T4SS_MOBH	4	0
MOB _P	T4SS_MOBP1	5	18
	T4SS_MOBP2	0	0
	T4SS_MOBP3	1	0
MOB _Q	T4SS_MOBQ	5	0
MOB _T	profile_MOBT	29	3
MOB _V	T4SS_MOBV	3	2
MOB _M	profile_MOBM	6	1

3.6.3. Mantención plasmídica

En ambos plasmidomas se detectaron secuencias que albergan genes fundamentales para el mantenimiento de plásmidos, como aquellos correspondientes a los sistemas de toxina-antitoxina (TA). Todos los genes identificados, con una identidad y cobertura de al menos 85 %, pertenecen a los sistemas TA de tipo II (Tabla I.7). Solo se pudo anotar un único sistema completo -una toxina con su respectiva antitoxina- en cada plasmidoma. En Puquio de Campo Naranja, se presentó en el *contig* 5461 y corresponde al sistema T3767-AT3767, previamente encontrado en la cepa *Rhodobacter sphaeroides* ATCC 17025 (NC_009428.1). La toxina (102 aa) pertenece a la superfamilia RelE/ParE, la cual incluye toxinas que escinden las transcripciones asociadas a los ribosomas y otras que funcionan como inhibidores de la ADN girasa. Los factores de transcripción (antitoxinas) que las acompañan pueden pertenecer a superfamilias no relacionadas o relacionadas lejanamente (Anantharaman y Aravind, 2003). En Laguna Diamante, el sistema TA se encontró en el *contig* 2880 (T2787-AT2787) y

corresponde a la familia VapBC, donde la toxina es una ribonucleasa de dominio PIN (145 aa) y la antitoxina es un factor de transcripción (98 aa) (Arcus y col., 2011). Resulta interesante que este sistema TA se conoce a partir de *Haloquadratum walsbyi* DSM 16790 (NC_008212.1), una arquea halófila que fue aislada de una salina solar en Brac del Port (Alicante, España) y que domina la mayoría de los ambientes saturados de NaCl talásico (Bolhuis y col., 2004), por tanto podría tratarse de un sistema común en haloarchaeas.

En los estudios de plasmidoma anteriores, no se tuvieron en cuenta los sistemas TA. Solo Kothari y col. (2019) informaron los sistemas TA de tipo II YoeB-YefM y RelE/StbE-RelB/StbD en algunos plásmidos circulares de plasmidomas de agua subterránea. YoeB y RelE son toxinas ARNasa dependientes de ribosomas que se unen directamente al sitio A del ribosoma, donde escinden el ARNm asociado al ribosoma (Page y Peti, 2016).

En general, no existe una gran cantidad de secuencias estrechamente relacionadas con las entradas de las bases de datos actuales. Es probable que los plasmidomas codifiquen sistemas de replicación, movilización y sistemas TA aún desconocidos, particularmente porque la mayoría de los microorganismos contribuyentes no son cultivables y derivan de ambientes poco estudiados, a lo cual se suma la gran proporción de arqueas presentes en la comunidad de Laguna Diamante.

Tabla I.7. Genes putativos de sistemas toxina-antitoxina (TA) en los plasmidomas de Puquio de Campo Naranja y Laguna Diamante.

Toxina o Antitoxina	ID Contig	Identidad (%)	Longitud (pb)	Cobertura (%)	Huésped	NCBI	Tipo
Puquio de Campo Naranja							
TADB AT2883	NODE_18150	90,535	243	100	<i>Alkalilimnicola ehrlichii</i> MLHE-1	NC_008340.1	II_AT
TADB AT3158	NODE_3568	89,562	297	97	<i>Shewanella</i> sp. ANA-3	NC_008577.1	II_AT
TADB AT3539	NODE_5461	89,412	255	100	<i>Rhodobacter sphaeroides</i> ATCC 17029	NC_009050.1	II_AT
TADB AT3767	NODE_5461	88,976	254	99	<i>R. sphaeroides</i> ATCC 17025	NC_009428.1	II_AT
TADB AT4728	NODE_92792	92,661	327	100	<i>Dinoroseobacter shibae</i> DFL 12, plásmido pDSHI01	NC_009955.1	II_AT
TADB AT4730	NODE_92792	93,578	327	100	<i>D. shibae</i> DFL 12, plásmido pDSHI03	NC_009957.1	II_AT
TADB T3767	NODE_5461	87,055	309	99	<i>R. sphaeroides</i> ATCC 17025	NC_009428.1	II_T
TADB T4686	NODE_166621	92,832	279	100	<i>Desulfococcus oleovorans</i> Hxd3	NC_009943.1	II_T
Laguna Diamante							
TADB AT2015	NODE_13167	87,824	386	89	<i>Natronomonas pharaonis</i> DSM 2160, plásmido PL131	NC_007427.1	II_AT
TADB AT2778	NODE_9276	89,717	389	99	<i>Haloquadratum walsbyi</i> DSM 16790	NC_008212.1	II_AT
TADB AT2787	NODE_2880	94,613	297	100	<i>H. walsbyi</i> DSM 16790	NC_008212.1	II_AT
TADB AT5082	NODE_5968	91,667	216	86	<i>Halobacterium salinarum</i> R1, plásmido PHS3	NC_010368.1	II_AT
TADB T2011	NODE_25283	86,463	229	93	<i>N. pharaonis</i> DSM 2160	NC_007426.1	II_T
TADB T2017	NODE_25283	86,463	229	93	<i>N. pharaonis</i> DSM 2160, plásmido PL23	NC_007428.1	II_T
TADB T2787	NODE_2880	85,492	386	88	<i>H. walsbyi</i> DSM 16790	NC_008212.1	II_T

ID: identificación

AT: antitoxina

T: toxina

3.7. Funciones plasmídicas accesorias

3.7.1. Puquio de Campo Naranja

Factores de virulencia

Las bacterias patógenas infectan a los huéspedes empleando factores de virulencia. En algunos casos, estos factores están codificados a nivel extra-cromosómico. Para evaluarlos, se analizó el plasmidoma de Puquio de Campo Naranja a través de la base de datos VFDB, la cual es una fuente de referencia integrada y completa para factores de virulencia de patógenos bacterianos (Chen y col., 2005). En la Figura I.12, se muestran los factores de virulencia presentes en dicho plasmidoma. Los genes que codifican a KatAB y al antígeno O fueron los más abundantes. El gen *katB* codifica una peroxidasa catalasa (KatB), la cual es importante para la supervivencia intracelular y la transmisión del patógeno *Legionella pneumophila* (Bandyopadhyay y col., 2003). El factor de virulencia del antígeno O incluye el gen *gmd* que codifica la 4,6-deshidratasa GDP-manosa, presente en el patógeno *Yersinia enterocolitica*. Esta proteína está involucrada en la ruta de biosíntesis del antígeno O LPS, que es parte de la biogénesis bacteriana de la membrana externa. La ausencia de antígeno O en la membrana externa afecta la capacidad de colonizar el huésped y la expresión de otros factores de virulencia de *Yersinia* (Bengoechea y col., 2004).

El factor de virulencia LPS, implicado en la supervivencia y replicación de *Brucella melitensis* en sus huéspedes (Lapaque y col., 2005), también exhibió una abundancia relativa alta. Los siguientes genes putativos más abundantes codifican al pili tipo IV en *Pseudomonas aeruginosa*. Estos genes incluyen el gen *pilT* que codifica la proteína de motilidad de contracción PilT, la cual se une a las células huésped. Los pili de tipo IV también están involucrados en la formación de biopelículas. Además, Carter y col. (2010) mostraron la

importancia de un pilus tipo IV para la transferencia horizontal de genes de la isla de patogenicidad PAPI-1 de *P. aeruginosa*, que se transfiere mediante un mecanismo conjugativo. PAPI-1 está codificada por un grupo que comprende 10 genes, que son similares a los genes correspondientes del plásmido conjugativo enterobacteriano R64 (Kim y Komano, 1997).

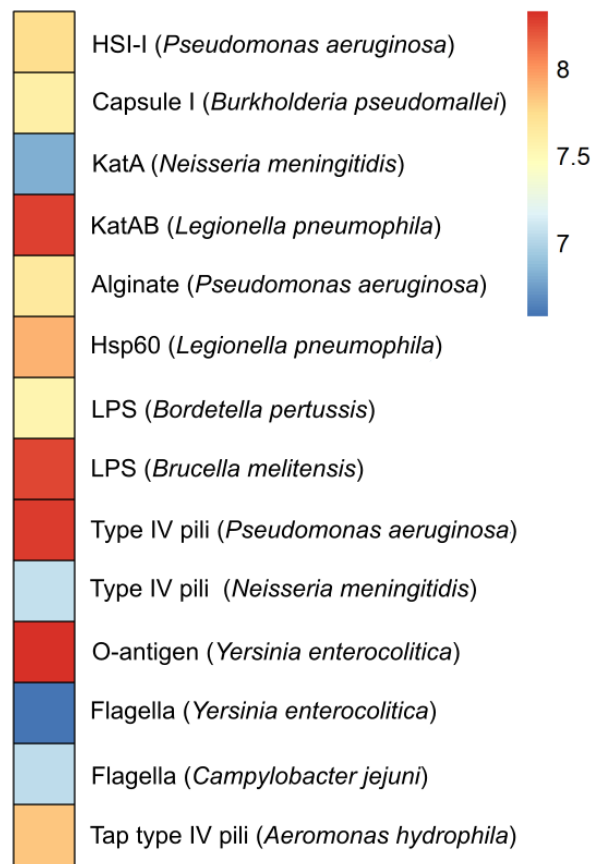


Figura I.12. Factores de virulencia en el plasmidoma de Puquio de Campo Naranja. La escala de colores indica la abundancia de cada factor de virulencia en FPKM (Fragmentos Por Kilobase por Millón). Los valores se trazan en una escala logarítmica. Las bacterias patógenas se especifican entre paréntesis.

Dado que la diversidad eucariota es limitada en este entorno extremo, no es probable que el ADN de bacterias patógenas asociadas con grandes organismos multicelulares esté

representado en el plasmidoma. Es por ello que la aparición y distribución de genes de virulencia bacteriana clínicamente relevantes en estos entornos naturales (no humanos) parece no ser significativa ni estar bien estudiada. Sin embargo, Søborg y col. (2016) también informaron la aparición de homólogos a genes de virulencia bacteriana humana en conjuntos de datos metagenómicos de una variedad de nichos ecológicos, incluida una mata hipersalina de Guerrero Negro (México), microbialitos vivos de Cuatro Ciénegas (México) y una mata microbiana de aguas termales del Parque Nacional de Yellowstone (Estados Unidos). Además, los análisis filogenéticos sugirieron que representan linajes antiguos en la filogenia de los genes clínicos (Søborg y col., 2016). En consecuencia, estos enfoques permitirían comprender mejor el papel de los entornos naturales en la evolución de la virulencia bacteriana.

Resistencia a metales

Algunos genes asociados con la resistencia a metales también estuvieron presentes en el plasmidoma de Puquio de Campo Naranja (Figura I.13). Los genes *copB* y *sodB* fueron los más abundantes, seguidos por *ruvB* y *merA*. El gen *copB* se describió en *Archaeoglobus fulgidus* y codifica una ATPasa de tipo P transportadora de cobre que está involucrada en el flujo del mismo (Mana-Capelli y col., 2003). Aunque los genes *sodB* y *ruvB* generalmente están relacionados con la resistencia al estrés oxidativo y la reparación del ADN, también se ha informado de su participación en la resistencia a los metales. El gen *sodB* codifica una superóxido dismutasa dependiente de Fe que reduce los radicales aniones superóxido que se producen dentro de las células y que son tóxicos para los sistemas biológicos. El gen también está asociado con la resistencia al selenio, ya que la actividad superóxido dismutasa es esencial para la defensa celular contra las sales de selenio (Bébién y col., 2002). El gen *ruvB*

codifica una ADN helicasa dependiente de ATP que participa en la reparación del daño en el ADN causado por el cromato o sus derivados. Además, confiere resistencia a telurito y selenito (Miranda y col., 2005). El gen *merA* se describió en el plásmido pPB (80 kb) de una cepa de *Pseudomonas stutzeri* que degrada el O-xileno (Baggi y col., 1987; Barbieri y col., 1989). Además, este gen también fue codificado por el plásmido p5343 (8 kb), identificado en un plasmidoma de agua subterránea por Kothari y col. (2019). MerA es una reductasa mercúrica, que cataliza la reducción de Hg^{2+} a mercurio metálico volátil Hg^0 (Reniero y col., 1995).

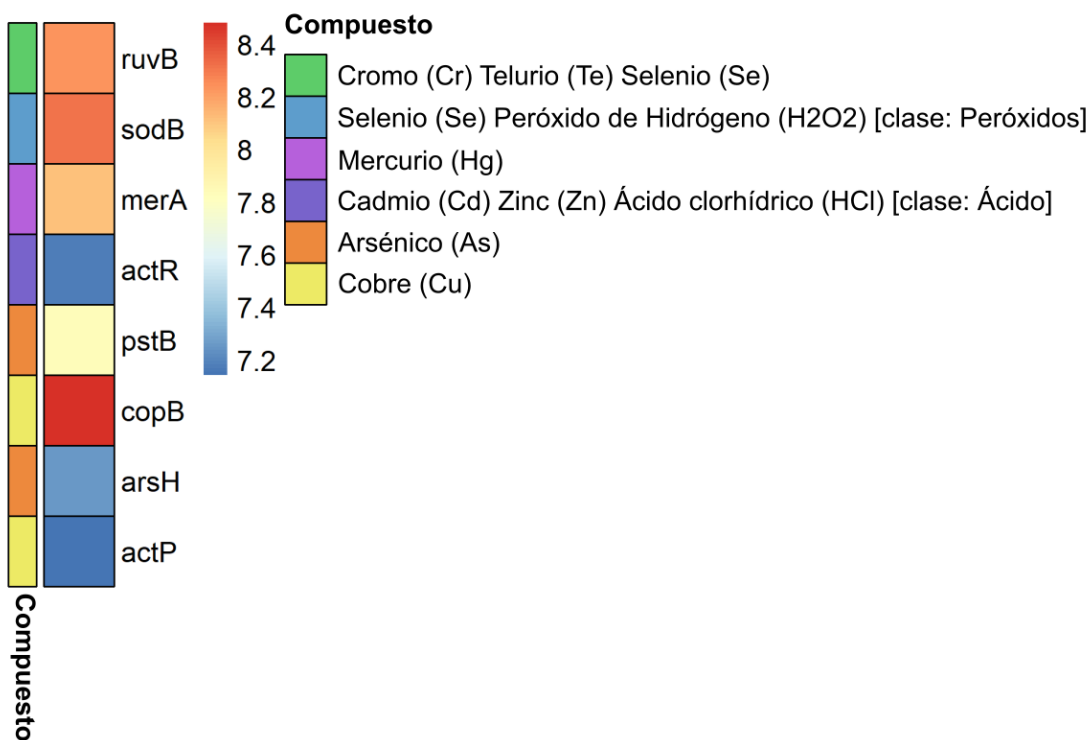


Figura I.13. Genes de resistencia a metales en el plasmidoma de Puquio de Campo Naranja. La escala de colores indica la abundancia de cada gen en FPKM (Fragmentos Por Kilobase por Millón). Los valores se trazan en una escala logarítmica. Los otros colores indican el metal al que confieren resistencia.

No es la primera vez que se detectan genes de resistencia a metales en comunidades complejas derivadas de ambientes extremos. En estudios metagenómicos de microbialitos de Pavilion Lake (Canadá), las funciones de SEED y las vías de MetaCyc relacionadas con la desintoxicación de metales pesados fueron abundantes. Se encontró la presencia de proteínas de resistencia a cobalto-zinc-cadmio, proteínas de bombas de eflujo, degradación del acetato de fenilmercurio y resistencia al cromato (White y col., 2016). Posteriormente, el análisis del genoma de dos bacterias aisladas de estos microbialitos confirmó la presencia de genes relacionados con estas vías a nivel cromosómico y plasmídico (White y col., 2018, 2019). El análisis metagenómico de matas microbianas hipersalinas y estromatolitos de Shark Bay (Australia) también reveló anotaciones funcionales para una variedad de vías involucradas en la resistencia a metales pesados -incluidas las de zinc, cromo, mercurio, cobre y cadmio- y al metaloide arsénico (Ruvindy y col., 2016). Los genes del metabolismo del arsénico son de particular interés en estas comunidades microbianas debido a un estudio que proporciona evidencia de ciclos del arsénico en estromatolitos fósiles (Sforna y col., 2014). Dichos genes también estuvieron presentes en los microbialitos de Pavilion Lake (White y col., 2016). Asimismo, el metagenoma de un estromatolito del lago andino Socompa (Argentina) reveló un potencial metabólico sorprendentemente diverso que comprende todos los tipos conocidos de resistencia al arsénico y vías generadoras de energía (Kurth y col., 2017). Además, estudios geoquímicos de una mata microbiana de Laguna Brava (Chile) (Sancho-Tomás y col., 2018) y de microbialitos de Laguna Diamante (Sancho-Tomás y col., 2020) determinaron la presencia simultánea de dos estados de oxidación, As(III) y As(V), evidenciando ciclos del arsénico mediados por microbios en esas comunidades.

El análisis funcional del plasmidoma de Puquio de Campo Naranja sugiere la presencia de diversos grupos de genes de resistencia a metales en la comunidad estudiada. La existencia de

genes que codifican resistencia al cromo, cadmio, arsénico y cobre podría explicarse por las altas concentraciones de estos metales en el medio ambiente (Anexo, Tabla A1). Sin embargo, también se detectaron genes que codifican resistencia al mercurio a pesar de su baja concentración ($< 0,1 \text{ mg Kg}^{-1}$). Ruvindy y col. (2016) y White y col. (2016) informaron hallazgos similares en Shark Bay y Pavilion Lake, respectivamente. La idea de que la resistencia a los metales tóxicos surgió durante la vida temprana en la Tierra está en línea con estos resultados, ya que el medio ambiente estaba contaminado por actividades volcánicas y otras actividades geológicas (Silver y Phung, 1996). Sin embargo, la razón por la cual los mecanismos de resistencia se mantienen en estos microbialitos aún no está clara. Una hipótesis es que la presión antibiótica ayuda a mantener la presencia de resistencia metálica ya que ambos están físicamente unidos en los plásmidos (Foster, 1983). Además, White y col. (2018) han propuesto la hipótesis alternativa de que inicialmente estos genes pueden haber conferido resistencia a metales pesados, pero ahora están operando bajo otros factores estresantes.

Resistencia a los antibióticos

Se identificó un total de 123 genes de resistencia a antibióticos (*ARGs*) en el plasmidoma de Puquio de Campo Naranja, los cuales confieren resistencia a 26 tipos diferentes de ellos. Principalmente, se detectaron genes relacionados con la resistencia a penam, cefalosporina y macrólidos. Además, también estuvieron presentes genes de resistencia a los aminoglucósidos, carbapenémicos, tetraciclinas, fluoroquinolonas y fenicoles (Figura I.14). Coincidentemente, Zhang y col. (2011) revelaron altos niveles de resistencias antibióticas conocidas (35 *ARGs*) en el plasmidoma de lodo activado de la planta de tratamiento de aguas residuales de Shatin (Hong Kong, China). Los *ARGs* que codifican resistencia a tetraciclina

(27,2 %), a macrólidos (25,0 %) y a múltiples fármacos (24,9 %) se mostraron muy enriquecidos, lo cual es de esperarse en este entorno afectado por el ser humano. Del mismo modo, Sentchilo y col. (2013) estudiaron el metagenoma plasmídico de comunidades microbianas en dos sistemas de lodos activados, uno de los cuales recibe principalmente aguas residuales domésticas y el otro aguas residuales de la industria química. Ambos plasmidomas contenían una amplia variedad de genes que codifican resistencia a antibióticos. En contraste, hay poca información sobre las resistencias a antibióticos en entornos remotos y prístinos. Van Goethem y col. (2018) recientemente describieron la presencia de 177 ARGs en 17 metagenomas del suelo remoto y no contaminado de la superficie antártica dentro de la región del Glaciar Mackay. La mayoría codificó bombas de flujo de salida de uno o múltiples fármacos, y los autores sugirieron que probablemente representan genes históricos funcionales eficientes que se han heredado verticalmente durante generaciones. En ese sentido, en este estudio también se confirmó la presencia de ARGs en un entorno sin impacto antropogénico, además de que previamente se había informado la resistencia a varios antibióticos en bacterias aisladas de otros lagos de gran altitud de la Puna andina (Dib y col., 2008, 2009). En dichas cepas se encontró un predominio de resistencias a azitromicina, eritromicina, claritromicina, roxitromicina, estreptomicina, cloranfenicol, gentamicina, kanamicina, tetraciclina y ampicilina, con concentraciones inhibitorias mínimas inesperadamente altas (CIMs $> 2 \text{ mg mL}^{-1}$) para macrólidos (Dib y col., 2008). Además, se detectó evidencia experimental de una resistencia codificada a nivel plasmídico en una cepa de actinobacteria aislada (Dib y col., 2010a) y, posteriormente, el análisis de secuencia de plásmidos lineales confirmó la presencia de genes de resistencia (Dib y col., 2018). Asimismo, el análisis de secuencia del plasmidoma también mostró que las resistencias a los antibióticos están muy extendidas en el ambiente prístino estudiado, en el que se supone que

la presión selectiva está ausente. Sin embargo, se sabe que la presión de metales puede ayudar a retener los genes de resistencia a los antibióticos en las bacterias (Baker-Austin y col., 2006), ya que estos genes generalmente están vinculados en los plásmidos (Foster, 1983). De esta forma, los genes de resistencia a los antibióticos codificados por el plasmidoma, junto con las resistencias metálicas, podrían transferirse horizontalmente entre las bacterias de la comunidad. Además, los metales pesados han demostrado la capacidad de aumentar la resistencia a los antibióticos en bacterias, independientemente de cualquier exposición a antibióticos (Nisanian y col., 2014).

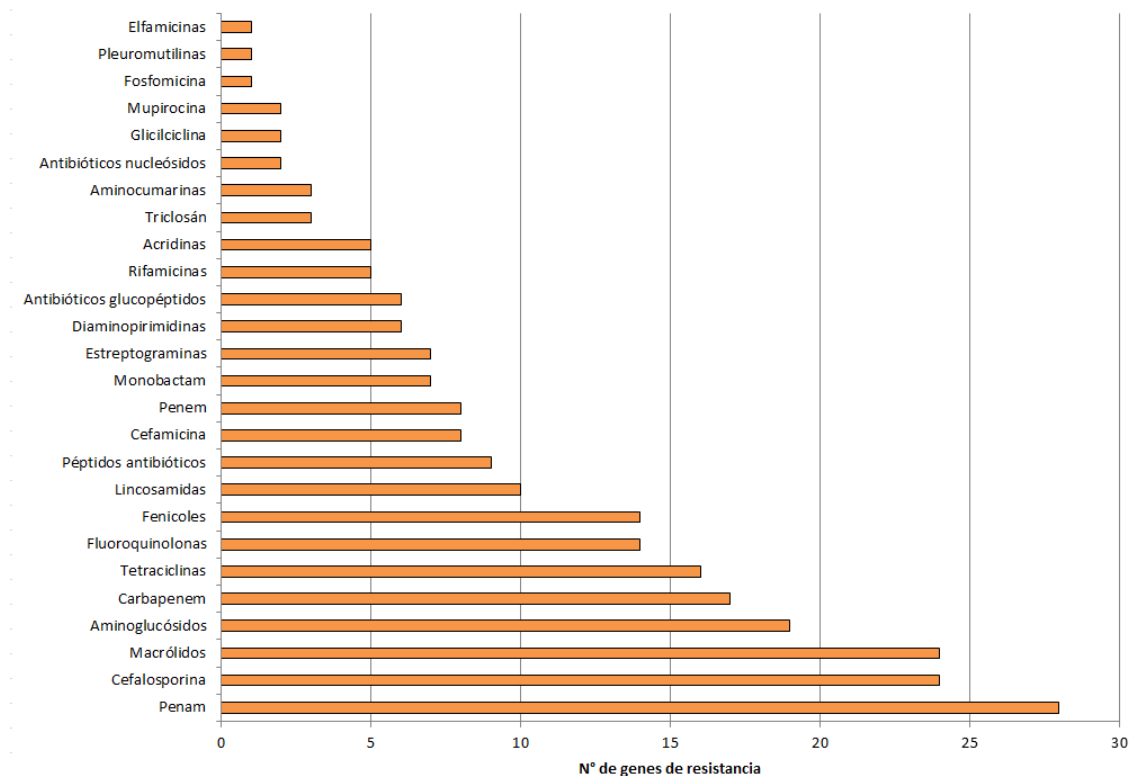


Figura I.14. Número de genes que confieren resistencia a diferentes tipos de antibióticos en el plasmidoma de Puquio de Campo Naranja. Los ARGs putativos fueron identificados por el software RGI utilizando la base de datos CARD.

Otros estudios metagenómicos de comunidades microbianas complejas también informaron genes relacionados con la resistencia a los antibióticos. El metagenoma de microbialito de Pavilion Lake incluía vías de resistencia a los antibióticos betalactámicos (White y col., 2016). Asimismo, se identificaron sistemas de resistencia como bombas de eflujo de múltiples fármacos, betalactamasas, resistencia a fluoroquinolonas, meticilina y eritromicina en metagenomas de matas y estromatolitos de Shark Bay (Ruvindy, 2014). Cabe destacar que estos resultados coinciden con los obtenidos en el plasmidoma de microbialito de Puquio de Campo Naranja aquí analizado (Figura I.8). Además, Peimbert y col. (2012) encontraron vías para producir antibióticos como tetraciclina, penicilina, estreptomicina, novomicina, ansamicina y vancomicina, así como genes que codifican betalactamasas y metalolactamasas, en una mata microbiana de Cuatro Ciénegas (México).

Como estos entornos están expuestos a poca o ninguna influencia humana, es posible que la presencia de mecanismos de resistencia a los antibióticos se deba a factores ecológicos más que a factores antropogénicos. De hecho, se han descripto otras funciones en las que también participan los antibióticos. Pueden servir para competir con otros organismos por la supervivencia o como fuente de nutrientes (Dantas y col., 2008), y funcionar como moléculas de señalización a bajas concentraciones (Yim y col., 2007). Además, D'costa y col. (2011) han demostrado que la resistencia a los antibióticos es un fenómeno natural que precede a la presión selectiva del uso clínico de antibióticos. Por estas razones, los ambientes extremos podrían constituir un reservorio de genes de resistencia, y los estudios adicionales en comunidades ambientales complejas podrían ser clínicamente significativos en un enfoque epidemiológico integrado. Sin duda, para afrontar los retos relacionados con la resistencia antibiótica en salud pública, es necesario comprender los procesos de adquisición, estabilización y propagación de estos genes. En este sentido, está claro que los estudios de

plasmidomas son un enfoque amplio y eficiente para enriquecer la anotación de *ARGs* (presentes con frecuencia en plásmidos) y permitir un análisis centrado en aquellos potencialmente transferibles por *HGT*.

3.7.2. Laguna Diamante

Resistencia a los antibióticos y factores de virulencia

A diferencia de lo ocurrido con Puquio de Campo Naranja, en el análisis de secuencia del plasmidoma de Laguna Diamante solo se pudieron clasificar 8 *ARGs*, los cuales transmiten resistencia a 10 clases de drogas, entre ellas, macrólidos, carbapenémicos, cefalosporinas y penams. Esta notable diferencia con respecto al número de *ARGs* encontrados en ambientes extremos tan similares probablemente se deba a las distinciones de las comunidades microbianas presentes, ya que como se mencionó antes, la biopelícula roja de Laguna Diamante se encuentra dominada por arqueas. Los procesos metabólicos y las paredes celulares de bacterias y arqueas presentan diferencias significativas, lo que explica que varios antibióticos que son eficaces contra los primeros no amenacen a los segundos (Khelaifia y Drancourt, 2012). Además, el estudio de los mecanismos de resistencia a los antibióticos es, debido a la relevancia clínica, de mayor necesidad en el dominio bacteriano, ya que aún no se han identificado arqueas patógenas (Cavicchioli y col., 2003). Solamente se ha reportado, de manera exclusiva, sobre una relación entre la gravedad de la enfermedad periodontal y la abundancia relativa de *Methanobrevibacter oralis* (Lepp y col., 2004). En consecuencia, todos los intentos de buscar factores de virulencia en el plasmidoma de Laguna Diamante también fueron en vano.

El sesgo introducido por las bases de datos existentes que están desarrolladas a partir de la información disponible actualmente, así como el predominio de las arqueas en la comunidad

microbiana en cuestión, condicionan el análisis del resistoma codificado en el plasmidoma de Laguna Diamante. Por lo tanto, no debe descartarse que sea la carencia de estudios en arqueas lo que eclipse el número de ARGs y factores de virulencia identificados.

Resistencia a metales y al arsénico

El análisis del plasmidoma de Laguna Diamante no reveló genes relacionados con la resistencia a metales, posiblemente esto se deba a la razón antes mencionada, ya que la base de datos BacMet consta únicamente de entradas bacterianas. Sin embargo, la anotación manual permitió identificar genes de resistencia al arsénico.

El arsénico es un metaloide que puede afectar el desarrollo de varios microorganismos, sin embargo, algunas bacterias y arqueas poseen sistemas de desintoxicación que permiten el crecimiento incluso en altas concentraciones del mismo (Andres y Bertin, 2016). El sistema de resistencia más común está codificado por el operón *ars*, para el cual se describieron diferentes organizaciones genéticas entre diversos procariotas (Ben Fekih y col., 2018). Además de los sistemas de extrusión, compuestos por los productos génicos codificados por *arsA*, *arsB*, *arsC*, *arsD*, *arsR* y *acr3*, se reportó otro mecanismo que involucra una arsenito-metiltransferasa (ArsM) en *Halobacterium* sp. NRC-1 (Wang y col., 2004) (Figura I.15). En el plasmidoma de esta comunidad microbiana se identificaron 28 proteínas posiblemente relacionadas con la resistencia al arsénico, diez de las cuales habían sido anotadas automáticamente como *hypothetical protein* por Prokka (Tabla I.8). Por lo tanto, la anotación automática, aunque es relativamente rápida, asume el riesgo de ofrecer resultados menos precisos o la divulgación de un número menor de genes de los que realmente existen. Cabe destacar que los genes que permiten a los microorganismos realizar la respiración anaeróbica del arsenato (*arr*) y la oxidación del arsenito (*aio*), presentes en los estudios metagenómicos

previos de esta comunidad (Rascovan y col., 2016; Saona y col., 2019) y respaldados por evidencia geoquímica (Sancho-Tomás y col., 2020), no fueron detectados en el plasmidoma, asumiendo su ubicación cromosómica.

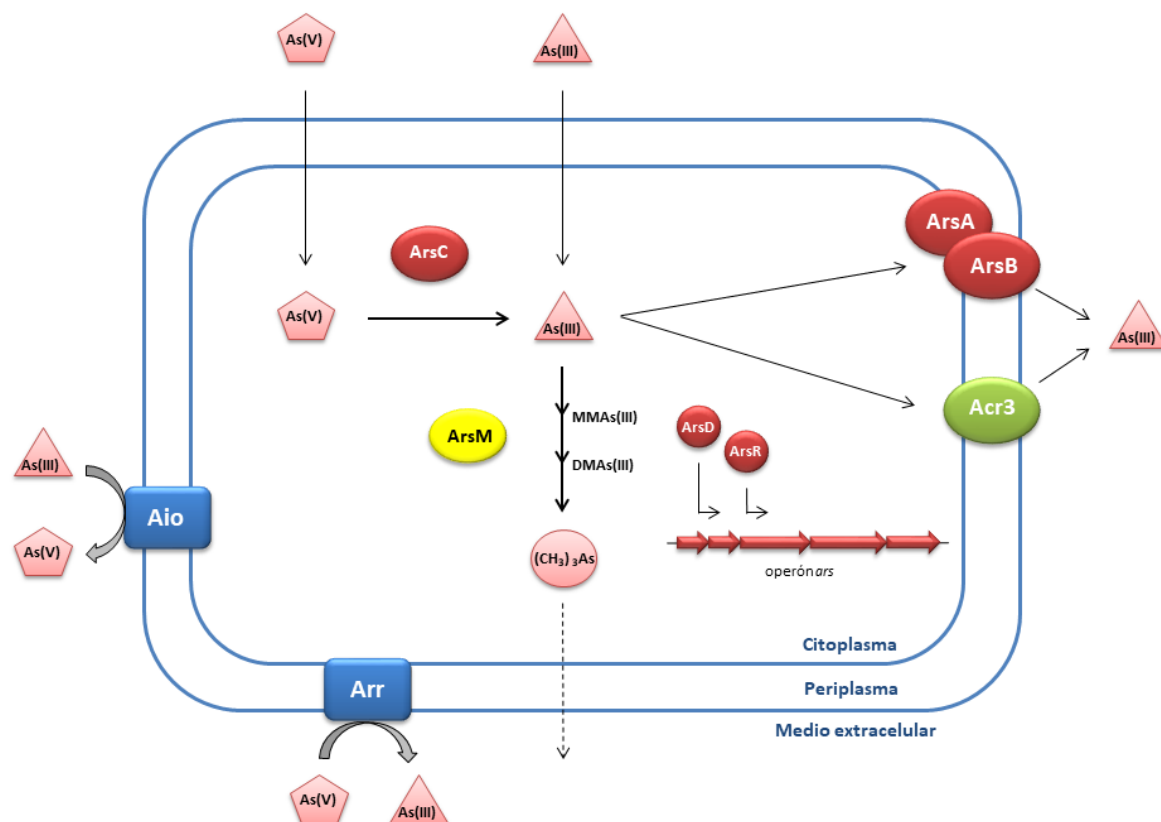


Figura I.15. Esquema general del metabolismo del arsénico (As). El metaloide ingresa a la célula a través de distintos transportadores inespecíficos. ArsC (arsenato reductasa) reduce el arsenato a arsenito que luego puede ser expulsado de la célula por ArsAB o Acr3. ArsD y ArsR actúan como reguladores del operón *ars*. Alternativamente, el arsenito puede ser metilado por ArsM (arsenito metiltransferasa), formando los intermediarios metilados MMAs(III) (monometil arsenito) y DMAs(III) (dimetil arsenito) que pueden ser expulsados de la célula por un gradiente de concentración, y (CH₃)₃As (trimetilarsina) que, al ser volátil, difunde hacia el exterior. El arsénico fuera del citoplasma se puede utilizar para obtener energía mediante la oxidación del arsenito por Aio (arsenito oxidasa) o la respiración anaeróbica del arsenato por Arr (arsenato reductasa respiratoria).

Tabla I.8. Proteínas relacionadas a la resistencia al arsénico en el plasmidoma de Laguna Diamante.

Proteína	ID Prokka	Anotación de Prokka	Longitud (aa)	E-value	Score	ID Contig
ArsA	JDIPpm3_10560	<i>Putative arsenical pump-driving ATPase</i>	641	8,2 e ⁻²⁸²	936,5	NODE_116
	JDIPpm3_27810	<i>Putative arsenical pump-driving ATPase</i>	314	9,2 e ⁻⁶²	209,7	NODE_444
	JDIPpm3_23530	<i>Putative arsenical pump-driving ATPase</i>	326	5,7 e ⁻⁵⁸	197,1	NODE_349
	JDIPpm3_196340	<i>Putative arsenical pump-driving ATPase</i>	125	8,0 e ⁻³⁵	120,7	NODE_20952
	JDIPpm3_89920	<i>Putative arsenical pump-driving ATPase</i>	167	2,0 e ⁻¹⁰	40	NODE_2989
ArsB	JDIPpm3_27270	<i>hypothetical protein</i>	403	1,3 e ⁻¹²	47,3	NODE_429
ArsC	JDIPpm3_18280	<i>Putative arsenate reductase</i>	141	3,1 e ⁻⁷⁶	256	NODE_249
	JDIPpm3_127220	<i>Putative arsenate reductase</i>	168	3,9 e ⁻⁷¹	239,2	NODE_6178
	JDIPpm3_19440	<i>Putative arsenate reductase</i>	121	5,8 e ⁻⁶⁴	215,8	NODE_272
	JDIPpm3_186670	<i>Putative arsenate reductase</i>	54	6,1 e ⁻²³	81,6	NODE_17695
ArsD	JDIPpm3_10570	<i>Putative arsenical resistance operon repressor ArsD</i>	109	1,3 e ⁻²⁸	99,6	NODE_116
	JDIPpm3_106070	<i>Putative arsenical resistance operon repressor ArsD</i>	109	3,3 e ⁻²⁶	91,8	NODE_4121
ArsR	JDIPpm3_106080	<i>Putative arsenical resistance operon repressor ArsR</i>	133	4,9 e ⁻⁶⁷	227,1	NODE_4121
	JDIPpm3_102030	<i>Putative arsenical resistance operon repressor ArsR</i>	133	2,6 e ⁻⁶⁶	224,7	NODE_3804
	JDIPpm3_110750	<i>Putative arsenical resistance operon repressor ArsR</i>	136	4,8 e ⁻⁶⁶	223,8	NODE_4519
	JDIPpm3_10590	<i>Putative arsenical resistance operon repressor ArsR</i>	131	6,6 e ⁻⁶⁵	220,1	NODE_116
	JDIPpm3_172300	<i>Putative arsenical resistance operon repressor ArsR</i>	104	8,4 e ⁻⁵⁵	186,9	NODE_13710
	JDIPpm3_152720	<i>Putative arsenical resistance operon repressor ArsR2</i>	119	1,3 e ⁻⁵¹	176,4	NODE_9685
	JDIPpm3_37070	<i>hypothetical protein</i>	130	1,6 e ⁻³²	113,9	NODE_680
	JDIPpm3_148600	<i>hypothetical protein</i>	81	1,5 e ⁻²⁷	97,5	NODE_8999
	JDIPpm3_109740	<i>hypothetical protein</i>	132	4,1 e ⁻²⁷	96,1	NODE_4429
Acr3	JDIPpm3_18260	<i>hypothetical protein</i>	326	1,1 e ⁻¹²⁸	429,3	NODE_249
	JDIPpm3_19450	<i>hypothetical protein</i>	325	6,7 e ⁻¹²⁸	426,8	NODE_272
	JDIPpm3_19180	<i>hypothetical protein</i>	331	3,8 e ⁻⁴³	148	NODE_267
	JDIPpm3_109760	<i>hypothetical protein</i>	231	4,6 e ⁻³⁰	104,9	NODE_4432
	JDIPpm3_13280	<i>hypothetical protein</i>	332	1,5 e ⁻¹⁷	63,8	NODE_161
	JDIPpm3_192940	<i>hypothetical protein</i>	67	5,9 e ⁻¹⁴	51,9	NODE_19718
ArsM	JDIPpm3_82590	<i>Arsenite methyltransferase</i>	190	2,6 e ⁻¹¹⁰	370	NODE_2535

*Los colores resaltan proteínas codificadas en el mismo contig

ID: identificación, aa: aminoácidos

Como se mencionó antes, la organización genética de los operones *ars* puede variar entre los diversos microorganismos. En el plasmidoma de Laguna Diamante, los genes *arsC* (arsenato reductasa) y *acr3* (transportador de eflujo de arsenito) se encontraron dos veces muy cerca, en el mismo *contig* (Tabla I.8). Sin embargo, el grupo de genes *arsADR* del *contig* 116 fue el más relevante, ya que su disposición genética no solo concuerda con la descrita para pHLAC01 y pNRC100 de *Halorubrum lacusprofundi* ATCC 49239 y *Halobacterium* sp. NRC-1, respectivamente, sino también con la encontrada en el genoma de *Halorubrum* sp. DM2, un aislado extremadamente halófilo de la biopelícula roja de Laguna Diamante (Ordoñez y col., 2018) (Figura I.16). En todos los casos, los genes *arsDA* y *arsR* se transcriben en direcciones opuestas y se nota la ausencia del gen que codifica el transportador de arsenito ArsB. Esta estructura de operón, por lo demás muy inusual, es aparentemente característica de la clase haloarchaea (Wang y col., 2004; Ben Fekih y col., 2018).

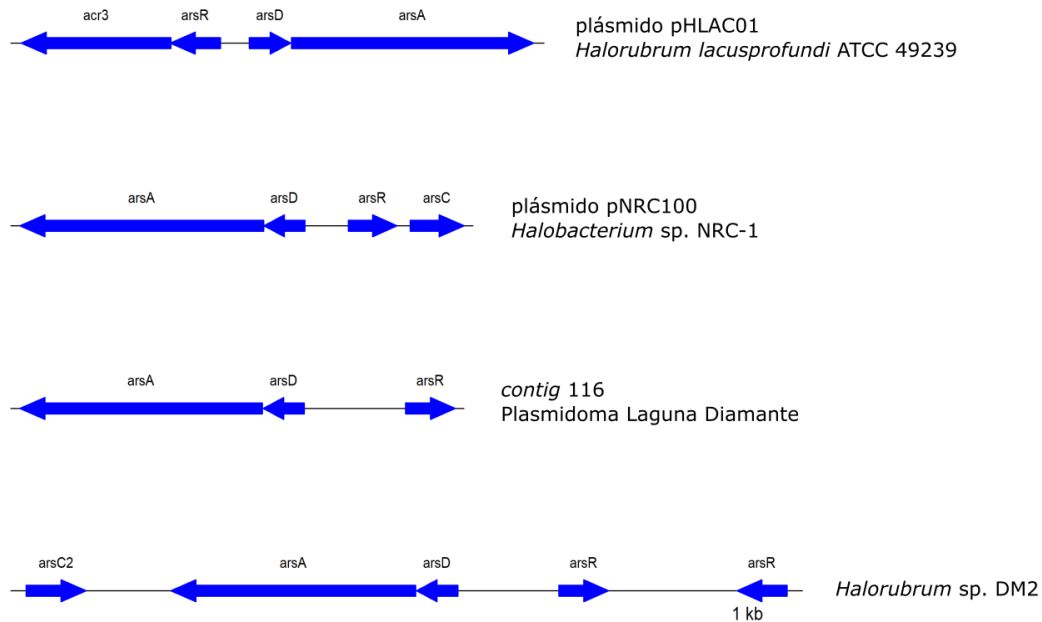


Figura I.16. Organización genética del operón *ars* en el *contig* 116 del plasmidoma de Laguna Diamante, en el genoma de un aislado de la misma comunidad (*Halorubrum* sp. DM2) y en plásmidos de cepas de haloarchaeas. Los mapas genéticos fueron creados usando el paquete genoplots de R (Guy y col., 2010).

3.8. Elementos genéticos móviles: integrones, secuencias de inserción y transposasas

Los integrones son posibles vehículos para la distribución de genes de resistencia a antibióticos. Probablemente no son móviles por sí mismos, pero a menudo se encuentran en transposones, que a su vez se encuentran en plásmidos conjugativos involucrados en la transferencia horizontal (Sabaté y Prats, 2002). Se investigó la presencia de integrones en los plasmidomas mediante la comparación con la base de datos INTEGRALL, pero no se detectó ninguno.

A partir del software ISEScan, un total de 6.569 secuencias de inserción (IS) fueron determinadas en el plasmidoma de Puquio de Campo Naranja. Los elementos pertenecientes a las familias IS3, IS21, IS66 e IS110 fueron los más frecuentes, seguidos por los miembros de las familias IS1634 e IS4 (Tabla I.9). En el plasmidoma de Laguna Diamante, se identificó un número algo menor, 533 IS, siendo IS200/IS605 e IS5 las familias más representadas, seguidas por IS4, IS6, IS630 e ISH3. Las primeras cinco son las principales familias de IS que se propagan en las halobacterias, precisamente la clase dominante en la comunidad estudiada (Filée y col., 2007). La mayoría de las IS de arqueas pertenecen a familias que se encuentran también en bacterias, pero otras están restringidas al dominio Archaea, como los miembros de la familia ISH3 (Craig, 2020). En este caso, también se clasificaron dos nuevas IS potenciales no atribuibles a ninguna de las familias conocidas (Nº de acceso GenBank: OK172335, OK172336) (Tabla I.9).

Los transposones de la familia Tn3 son algunos de los transposones más abundantes en los genomas bacterianos y se transponen preferentemente en moléculas de plásmidos (Szuplewska y col., 2015). En el plasmidoma de Puquio de Campo Naranja, se predijeron 10 transposasas pertenecientes a la familia Tn3. Sin embargo, en el plasmidoma de Laguna Diamante no se identificaron secuencias asociadas con estos transposones. Entre los reportes

sobre plasmidomas descritos hasta la fecha, solo unos pocos estudiaron la presencia de estos elementos (Zhang y col., 2011; Jørgensen y col., 2014). Zhang y col. (2011) informaron la presencia de dos IS asociadas con Tn3 (ISSod9 e ISXc5) en el plasmidoma de una planta de tratamiento de aguas residuales, aunque cabe aclarar que mencionaron una importante contaminación con ADN cromosómico. La ausencia de este tipo de elementos y la menor variedad de IS en el plasmidoma de Laguna Diamante con respecto al de Puquio de Campo Naranja, podría explicarse por las características únicas de la comunidad microbiana en cuestión, que incluye una gran proporción de arqueas y que, por tanto, podría contener elementos móviles aún no caracterizados.

En general, la presencia de tantos elementos IS apoya la noción de que el plasmidoma contribuye de forma sustancial tanto a la evolución de los genomas como a los procesos de adaptación, al facilitar la adquisición de genes novedosos y rasgos beneficiosos (Frost y col., 2005; Hülter y col., 2017).

Tabla I.9. Secuencias de inserción (IS) presentes en los plasmidomas de Puquio de Campo Naranja (PCN) y Laguna Diamante (LD).

Familia IS	N° de copias PCN	N° de copias LD
IS200/IS605	315	151
ISNCY	238	29
IS5	345	107
IS4	410	48
IS6	18	45
IS30	84	1
IS701	95	14
IS1595	124	15
IS110	517	9
ISH3	8	42
IS21	703	3
IS1182	163	2
ISAS1	198	1
IS630	350	43
IS66	563	20
IS481	116	1
IS982	31	0
ISAZO13	74	0
IS256	321	0
ISL3	245	0
IS1380	203	0
IS1634	419	0
IS3	740	0
IS91	207	0
ISKRA4	71	0
IS1	9	0
IS607	2	0
Nueva	0	2
Total	6.569	533

3.9. Genes y secuencias de plásmidos conocidos en los plasmidomas

3.9.1. Base de datos NCBI

Se buscaron secuencias pertenecientes a plásmidos de la base de datos del NCBI en ambos plasmidomas. En el de Puquio de Campo Naranja, las coincidencias (*matches*) se adjudicaron

a los plásmidos pDSHI01, pDSHI03 y pDSHI04 de *Dinoroseobacter shibae* DFL 12, que se aisló de cultivos de dinoflagelados marinos derivados de la Bahía de Tokio, Japón (Biebl y col., 2005) (Tabla I.10). Esta bacteria marina es miembro del grupo *Roseobacter* y representa un organismo modelo para el estudio de la vida de dicho grupo en el océano. La cepa DFL 12 alberga cinco plásmidos que transportan 628 kb de información extra-cromosómica, incluyendo un grupo de genes *vir*, genes relacionados con la degradación de compuestos aromáticos, genes *cox* que contribuyen a la producción de energía, y una gran cantidad de genes para el metabolismo de carbohidratos (Wagner-Döbler y col., 2010). Además, cuatro *contigs* de este plasmidoma coincidieron con las secuencias del plásmido pP73B de *Celeribacter indicus* P73^T, una bacteria que degrada hidrocarburos aromáticos policíclicos, aislada de sedimentos de aguas profundas del Océano Índico (Lai y col., 2014). Kothari y col. (2019) identificaron un plásmido que revela homología con otro plásmido de dicha cepa (pP73C, 122 kb) en un plasmidoma de agua subterránea. También se encontraron coincidencias con los plásmidos pNS6001, pNS6002 y pNS6003 de *Confluentimicrobium* sp. EMB200-NS6. Esta alfa-proteobacteria pertenece a la familia Rhodobacteraceae y se aisló de una planicie de marea contaminada con aceite (Tully y col., 2018) (Tabla I.10).

En el plasmidoma de Laguna Diamante, se encontraron secuencias pertenecientes a 24 megaplásmidos descritos en 13 cepas de arqueas halófilas aisladas de diferentes ambientes salinos (Tabla I.11). La mayoría de las coincidencias se observaron con el plásmido de *Halobacterium* sp. DL1 (315 kb), que se aisló de un estanque de agua dulce y cuya secuencia fue recientemente depositada en la base de datos del NCBI (NZ_CP007061.1). También se detectaron catorce coincidencias con secuencias del plásmido pHLAC01 (431 kb) de *Halorubrum lacusprofundi* ATCC 49239, que se aisló de *Deep Lake*, un lago antártico hipersalino. Cabe destacar que una de estas coincidencias (identidad 92 %, longitud 4.123 pb)

corresponde al *contig* 116 del plasmidoma, el cual alberga los genes *arsDA* (Figura I.17). Otras trece secuencias alinearon con el plásmido pHTIA (330 kb) de *Halorhabdus tiamatea* SARL4B, que se aisló del lago anóxico hipersalino de aguas profundas de Shaban en el Mar Rojo (Werner y col., 2014) (Tabla I.11).

Aunque todos los plásmidos mencionados anteriormente se aislaron de microorganismos que prosperan en ambientes salinos, no pueden considerarse comparables debido a que la salinidad en el entorno bajo estudio es considerablemente mayor que en los otros ambientes. Sin embargo, los resultados indican la existencia de secuencias plasmídicas conservadas en entornos diferentes.

Tabla I.10. Plásmidos de la base de datos del NCBI presentes parcialmente en el plasmidoma de Puquio de Campo Naranja.

Plásmido	Huésped	Tamaño (kb)	Fuente de aislamiento	Matches	NCBI	Referencia
pDSHI01	<i>Dinoroseobacter shibae</i> DFL 12	191	Dinoflagelado marino derivado de la Bahía de Tokio, Japón	3	NC_009955.1	Wagner-Döbler y col., 2010
pDSHI03		126		1	NC_009955.7.1	
pDSHI04		86		1	NC_009955.8.1	
pP73B	<i>Celeribacter indicus</i> P73 ^T	136	Sedimento de aguas profundas del Océano Índico	4	NZ_CP004395.1	Cao y col., 2015
pNS6001	<i>Confluentimicrobium</i> sp. EMB200-NS6	184	Planicie de marea contaminada con aceite	3	NZ_CP010870.1	ND
pNS6002		157		2	NZ_CP010871.1	
pNS6003		156		1	NZ_CP010872.1	

ND: no disponible

Tabla I.11. Plásmidos de la base de datos del NCBI presentes parcialmente en el plasmidoma de Laguna Diamante.

Plásmido	Huésped	Tamaño (kb)	Fuente de aislamiento	Matches	NCBI	Referencia
pNRC100	<i>Halobacterium</i> sp. NRC-1	191	Alta concentración	6	NC_001869.1	Ng y col., 2000
pNRC200		365	de sales (no especificada)	8	NC_002608.1	
pNG500	<i>Haloarcula marismortui</i>	132	Mar Muerto	1	NC_006393.1	Baliga y col., 2004
pNG600	ATCC 43049	155		7	NC_006394.1	
PHS1	<i>Halobacterium salinarum</i> R1	147	Alta concentración	3	NC_010366.1	Pfeiffer y col., 2008
PHS4		41	de sales (no especificada)	1	NC_010367.1	
PHS3		284		4	NC_010368.1	
PHS2		195		2	NC_010369.1	
pHLAC01	<i>Halorubrum lacusprofundi</i> ATCC 49239	431	Deep Lake, Antártida	14	NC_012030.1	ND
pHTUR01	<i>Haloterrigena turkmenica</i> DSM 5511	698	Suelo salino en Turkmenistán	1	NC_013744.1	Saunders y col., 2010
pHV3	<i>Haloferax volcanii</i> DS2	438	Sedimentos	3	NC_013964.1	Hartman y col., 2010
pHV4		635	del fondo del	4	NC_013966.1	
pHV1		85	Mar Muerto	1	NC_013968.1	
p1	<i>Halalkalicoccus jeotgali</i> B3	406	Mariscos fermentados en sal, Corea del Sur	2	NC_014298.1	Roh y col., 2010
p2		363		1	NC_014299.1	
pHBOR02	<i>Halogeometricum borinquense</i>	339	Salinas solares de Cabo Rojo, Puerto Rico	1	NC_014731.1	Malfatti y col., 2009
pHBOR01	DSM 11551	362		1	NC_014735.1	
pHBOR03		210		1	NC_014736.1	
pHALXA01	<i>Halopiger xanaduensis</i> SH-6	436	Sedimento salino del lago Shangmata, China	2	NC_015658.1	Anderson y col., 2012
pHALXA02		181		1	NC_015667.1	
pHTIA	<i>Halorhabdus tiamatea</i> SARL4B	330	Lago anóxico hipersalino de aguas profundas de Shaban en el Mar Rojo	13	NC_021913.1	Werner y col., 2014
pHH126	<i>Haloarcula hispanica</i> N601	125	Salina solar en Alicante, España	1	NC_023011.1	Ding y col., 2014
pXH-48	<i>Halostagnicola larsenii</i> XH-48	131	Sedimento salino del lago Xilinhoh, China	1	NZ_CP007057.1	Castillo y col., 2006
ND	<i>Halobacterium</i> sp. DL1	315	Agua dulce, estanques	23	NZ_CP007061.1	ND

ND: no disponible

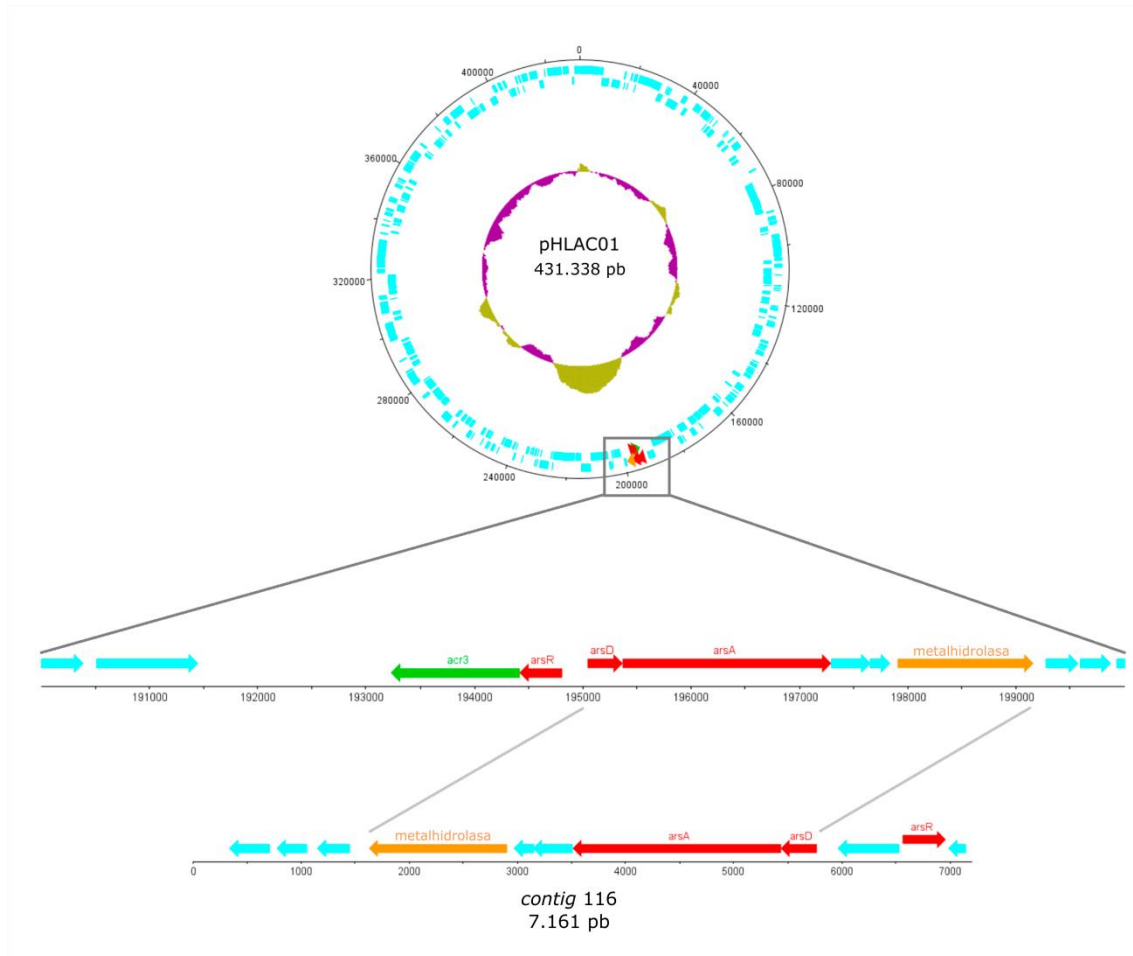


Figura I.17. Match entre el plásmido pHLAC01 de *Halorubrum lacusprofundi* ATCC 49239 y el contig 116 del plasmidoma de Laguna Diamante. Los mapas fueron generados con el software DNAPlotter (Carver y col., 2009).

3.9.2. Base de datos ACLAME

Al comparar las anotaciones funcionales de Prokka con la base de datos ACLAME, se encontraron 313 genes de origen plasmídico en el plasmidoma de Puquio de Campo Naranja (Anexo, Tabla A2). Estos genes pertenecen a 61 plásmidos aislados de 43 cepas distintas, incluyendo dos arqueas y algunas bacterias patógenas tales como *Burkholderia vietnamiensis*, *Klebsiella pneumoniae*, *Ochrobactrum anthropi*, *Pseudomonas aeruginosa*, *Ralstonia*

solanacearum, *Salmonella enterica* subsp. *enterica* serovar, *Serratia marcescens* y *Yersinia ruckeri*. Entre ellos, no se detectó ningún gen determinante de virulencia, pero si de respuesta al daño oxidativo y de resistencia al mercurio, además de aquellos propios del metabolismo y el transporte, algunos relacionados a la partición de plásmidos, y otros sin función asignada. En el caso de Laguna Diamante, las 125 coincidencias correspondieron a genes previamente encontrados en 11 megaplásmidos de cinco arqueas halófilas diferentes y una cepa de actinobacteria (*Rhodococcus* sp. RHA1) (Anexo, Tabla A3). La mayoría de ellos están relacionados con el metabolismo, la transposición y la recombinación del ADN. También se identificaron genes relacionados con la resistencia al arsénico, la reparación del ADN y la partición de plásmidos. A partir de esto, fue posible atribuir funciones a muchas de las proteínas hipotéticas identificadas por Prokka, sin embargo, no a todas. De todos modos, la comparación proporcionó una justificación adicional de la práctica, ya que se demostró la presencia de genes asociados a plásmidos conocidos en los conjuntos de datos de ambos plasmidomas.

3.10. Análisis taxonómico

Aunque los plásmidos se pueden transferir entre diferentes microorganismos, la asignación taxonómica de los *contigs* plasmídicos permite la estimación de los potenciales huéspedes.

Puquio de Campo Naranja

En el plasmidoma de Puquio de Campo Naranja, aproximadamente el 95 % de los *contigs* fueron asignados al dominio Bacteria, un poco más del 4 % a Archaea y el resto a Eukarya. En el dominio bacteriano, la mayoría de los *contigs* correspondieron a Proteobacterias (54 %), Bacteroidetes (10,8 %) y Firmicutes (10,7 %). Esta distribución de filos bacterianos se

comparó con el análisis taxonómico del ADN metagenómico de la comunidad microbiana, obtenido a partir de la secuenciación del amplicón del gen ARNr 16S. En dicho caso, los filos con mayor abundancia relativa de *zOTUs* fueron Proteobacterias (20 %), Cianobacterias (15 %) y Bacteroidetes (14 %) (Figura I.18). La distribución por clase bacteriana según cada análisis se muestra en la Figura I.19. Con respecto al dominio Archaea, el filo Euryarchaeota fue el de mayor abundancia relativa tanto en el plasmidoma como en el metagenoma (94,4 y 71,5 %, respectivamente; Figura I.20).

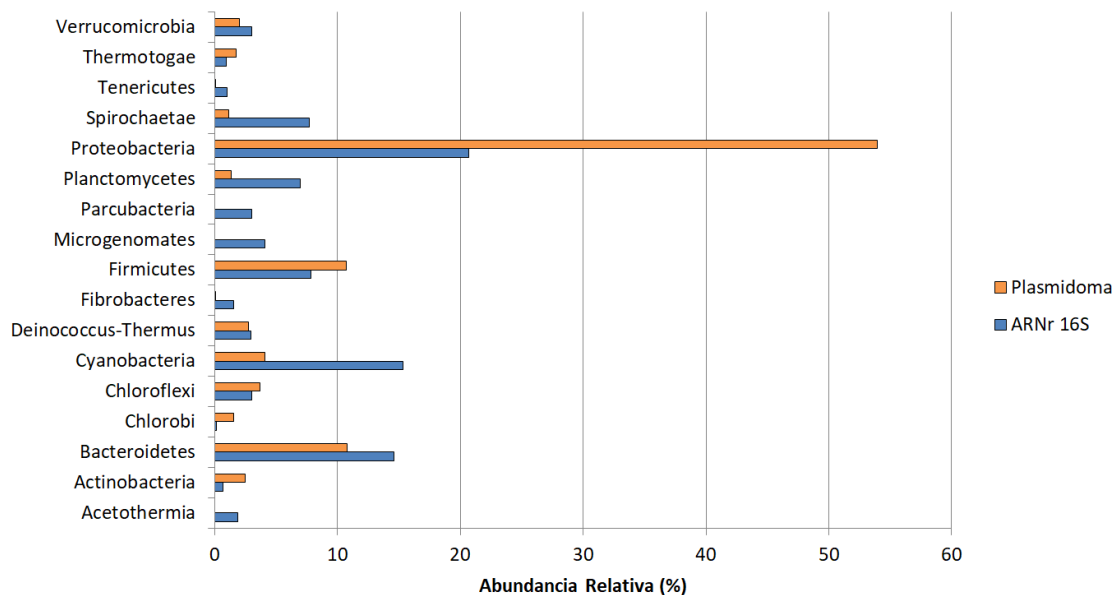


Figura I.18. Análisis taxonómico del plasmidoma de Puquio de Campo Naranja a nivel de filo bacteriano. Las barras naranjas muestran la abundancia relativa de las asignaciones de *contigs* a cada filo bacteriano a partir del análisis de MG-RAST, por similitud con la base de datos RefSeq ($E\text{-value} \leq 10^{-5}$). Las barras azules muestran la abundancia relativa de *zOTUs* de cada filo bacteriano a partir del análisis de ADN metagenómico, utilizando secuenciación de amplicones del gen ARNr 16S. Los filos que fueron menos abundantes que el 1 % en ambos conjuntos de datos no fueron incluidos.

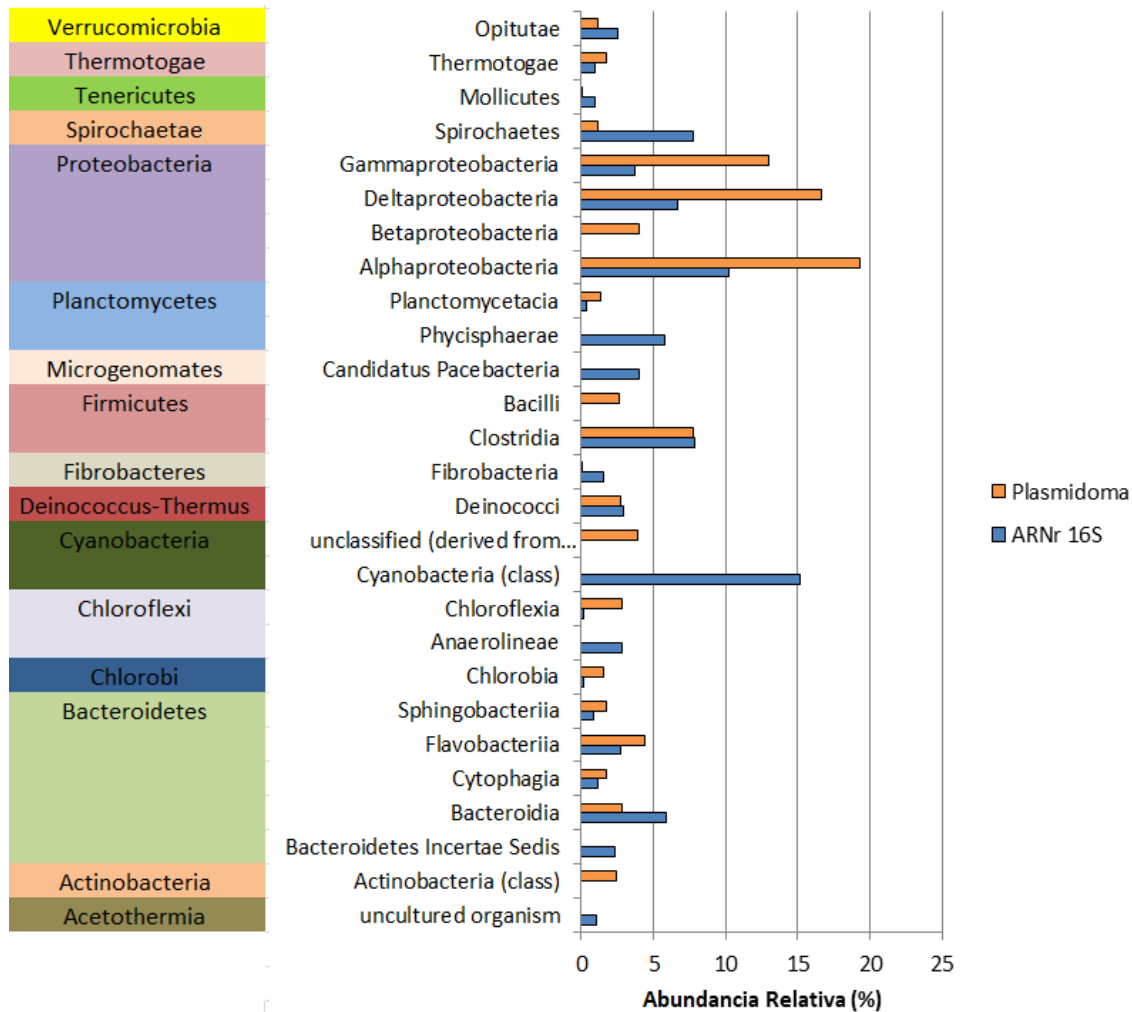


Figura I.19. Análisis taxonómico del plasmidoma de Puquio de Campo Naranja a nivel de clase bacteriana. Las barras naranjas muestran la abundancia relativa de las asignaciones de *contigs* a cada clase bacteriana a partir del análisis de MG-RAST, por similitud con la base de datos RefSeq ($E\text{-value} \leq 10^{-5}$). Las barras azules muestran la abundancia relativa de *zOTUs* de cada clase bacteriana a partir del análisis de ADN metagenómico, utilizando secuenciación de amplicones del gen ARNr 16S. Las clases que fueron menos abundantes que el 1 % en ambos conjuntos de datos no fueron incluidas. Hacia la izquierda, con diferentes colores, se muestran los filos bacterianos a los que pertenece cada clase bacteriana.

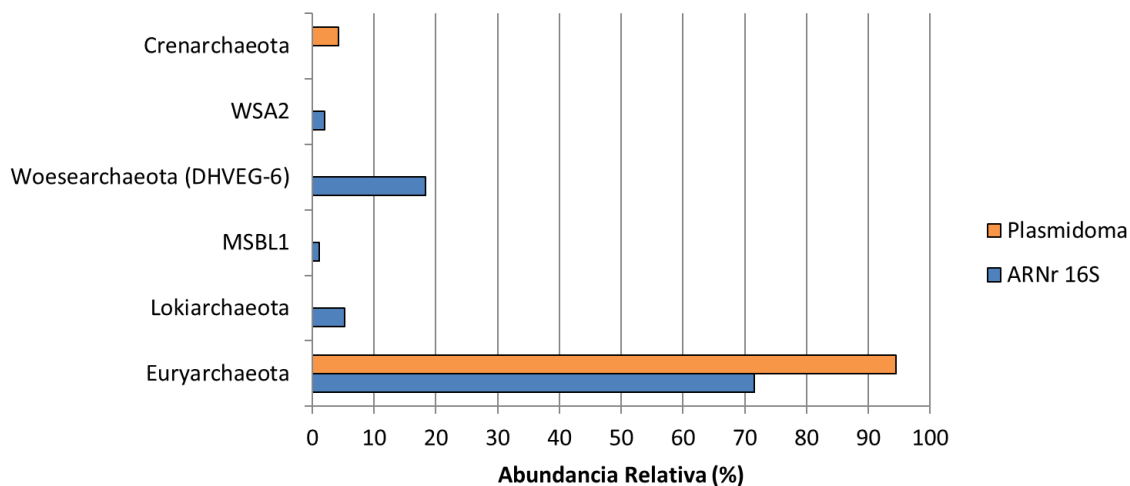


Figura I.20. Análisis taxonómico del plasmidoma de Puquio de Campo Naranja a nivel de filo arqueano. Las barras naranjas muestran la abundancia relativa de las asignaciones de *contigs* a cada filo arqueano a partir del análisis de MG-RAST, por similitud con la base de datos RefSeq ($E\text{-value} \leq 10^{-5}$). Las barras azules muestran la abundancia relativa de *zOTUs* de cada filo arqueano a partir del análisis de ADN metagenómico, utilizando secuenciación de amplicones del gen ARNr 16S. Los filos que fueron menos abundantes que el 1 % en ambos conjuntos de datos no fueron incluidos.

En general, la diferencia en la distribución taxonómica observada a partir de ambos análisis podría deberse a que ciertos filos albergan más plásmidos que otros, al sesgo introducido por las bases de datos, o al sesgo creado por los métodos de extracción de plásmidos utilizados en este estudio, que podrían favorecer la lisis de cierto grupo de microorganismos con respecto a otros.

Laguna Diamante

En el plasmidoma de Laguna Diamante, el 88 % de los *contigs* se asignaron al dominio Archaea, mientras que el 11 % se asignó a Bacteria y el 1 % restante a Eukarya. Entre las Archaea, dominó el filo Euryarchaeota (99,85 %). Cuando se comparó la distribución de filos del plasmidoma con los datos de la secuenciación del amplicón del ARNr 16S de la muestra

de ADN metagenómico correspondiente, el filo Euryarchaeota se volvió a destacar con la mayor abundancia relativa de *zOTUs* (Figura I.21). En ambos casos, dominó la clase halobacteria.

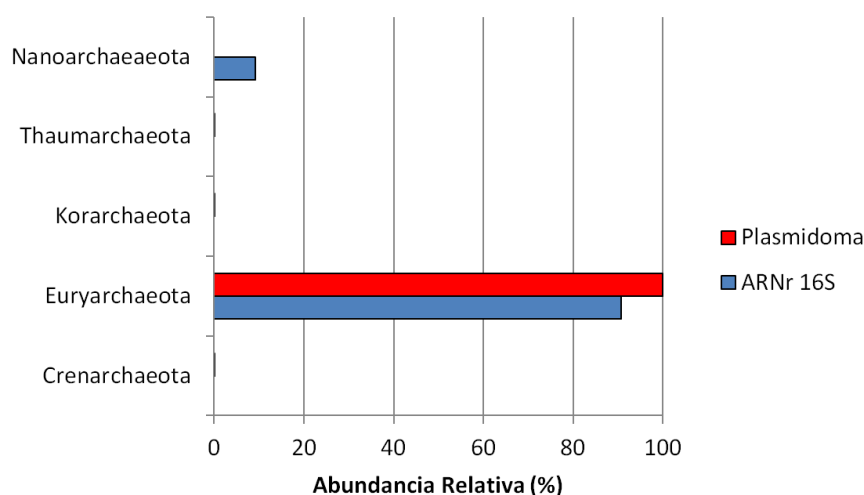


Figura I.21. Análisis taxonómico del plasmidoma de Laguna Diamante a nivel de filo arqueano. Las barras rojas muestran la abundancia relativa de las asignaciones de *contigs* a cada filo arqueano a partir del análisis de MG-RAST, por similitud con la base de datos RefSeq ($E\text{-value} \leq 10^{-5}$). Las barras azules muestran la abundancia relativa de *zOTUs* de cada filo arqueano a partir del análisis de ADN metagenómico, utilizando secuenciación de amplicones del gen ARNr 16S. Los fillos que fueron menos abundantes que el 1 % en ambos conjuntos de datos no fueron incluidos.

Con respecto al dominio Bacteria, las Proteobacterias (36 %), los Firmicutes (22,3 %) y las Actinobacterias (10,5 %) agruparon la mayoría de los *contigs* asignados en el plasmidoma. En el análisis del ARNr 16S correspondiente, las Proteobacterias (66,9 %) y los Firmicutes (17,3 %) también representaron los fillos con la mayor abundancia relativa de *zOTUs*, pero en este caso los Bacteroidetes ocuparon el tercer lugar (10,9 %). En particular, se obtuvieron *contigs* plasmídicos de Actinobacteria, Chloroflexi y Deinococcus-Thermus, mientras que el análisis de ARNr 16S no reveló miembros de estos fillos (Figura I.22). Una posible explicación para

esto podría ser la transferencia horizontal de genes transmitidos por plásmidos entre filos bacterianos, o la existencia de plásmidos con un amplio rango de huéspedes. Por otro lado, el análisis de ARNr 16S indicó la presencia de miembros del filo Halanaerobiaeota y, sin embargo, ningún *contig* del plasmidoma fue asignado a dicho grupo (Figura I.22). Esta situación, posiblemente, se deba a un sesgo en las bases de datos de plásmidos (cuando un filo no está bien representado) o tal vez, simplemente, a la ausencia de plásmidos dentro del taxón.

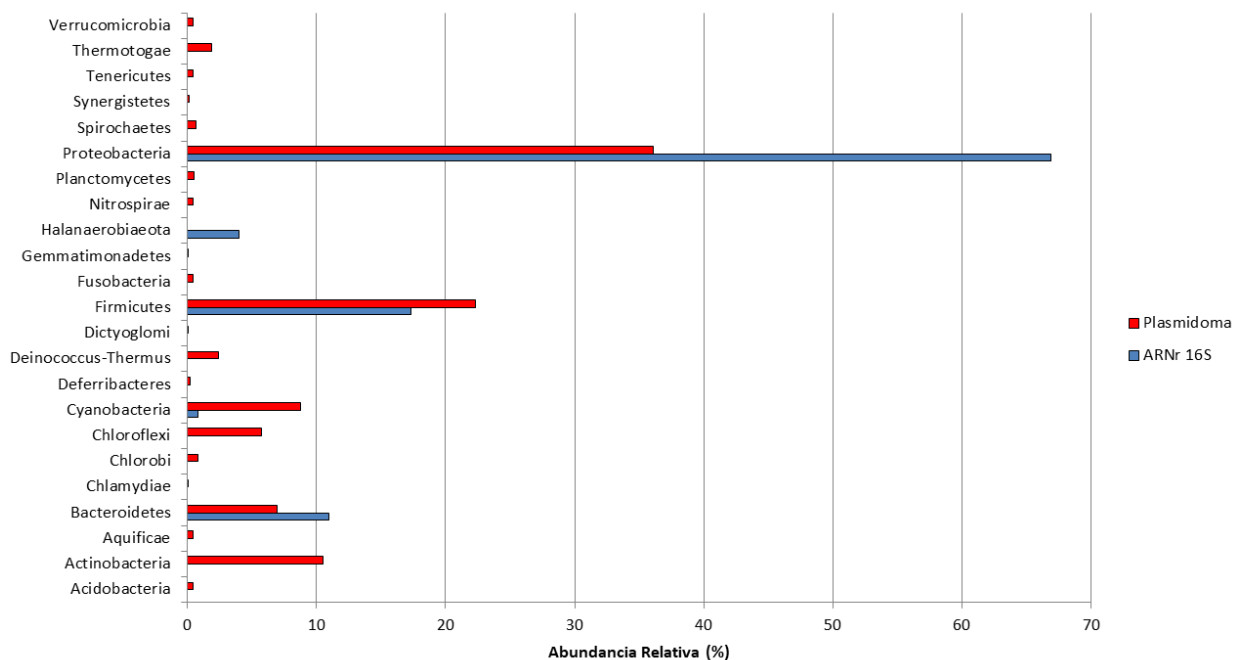


Figura I.22. Análisis taxonómico del plasmidoma de Laguna Diamante a nivel de filo bacteriano. Las barras rojas muestran la abundancia relativa de las asignaciones de *contigs* a cada filo bacteriano a partir del análisis de MG-RAST, por similitud con la base de datos RefSeq ($E\text{-value} \leq 10^{-5}$). Las barras azules muestran la abundancia relativa de *zOTUs* de cada filo bacteriano a partir del análisis de ADN metagenómico, utilizando secuenciación de amplicones del gen ARNr 16S. Los filos que fueron menos abundantes que el 1 % en ambos conjuntos de datos no fueron incluidos.

4. Conclusiones

En este capítulo se describieron las estructuras y las potenciales funciones de los plasmidomas de dos EMAs, un microbialito de Puquio de Campo Naranja y la biopelícula roja de Laguna Diamante, ambos ubicados en la región andina de la provincia de Catamarca, Argentina.

Aunque actualmente es posible estudiar los elementos plasmídicos en el curso de un análisis metagenómico convencional, este enfoque, que apunta específicamente a poblaciones de plásmidos, permite superar las limitaciones inherentes de las herramientas bioinformáticas aplicadas para el análisis de plásmidos a partir del ADN total de una comunidad. Bajo esta perspectiva, el estudio de ambos plasmidomas demostró que gran parte de la información de plásmidos no se registra si no se lleva a cabo la purificación experimental de los mismos antes de la secuenciación. Por lo tanto, se puede decir que un plasmidoma proporciona información nueva, la cual aparece oculta en los estudios metagenómicos tradicionales.

El análisis de la población total de plásmidos de estas comunidades microbianas permitió la identificación de una gran proporción de genes que codifican proteínas con función desconocida, lo que respalda la idea de que estos entornos prístinos son una fuente de nuevas biomoléculas. Sin embargo, también se encontraron características genéticas típicamente codificadas por plásmidos, entre ellas, los genes que permiten a los microorganismos que los hospedan sobrevivir bajo las duras condiciones ambientales. La abundante presencia de genes que confieren competencia en condiciones de estrés podría deberse al hecho de que una gran cantidad de copias de dichos genes proporciona un fenotipo más intenso. Además, su ubicación plasmídica permite la movilidad y la transferencia dentro de la comunidad microbiana.

En ambos plasmidomas, se anotaron funciones relacionadas con la reparación del ADN. Estas funciones podrían conferir resistencia o ser un mecanismo adaptativo a la alta exposición a la

radiación UV existente en la Puna andina. Los genes de resistencia a los factores de estrés también estuvieron bien representados, incluidos aquellos relacionados con el estrés oxidativo y osmótico. La radiación UV también actúa indirectamente al producir radicales libres que dañan el ADN, las proteínas y los lípidos. Por lo tanto, se requieren sistemas eficaces de respuesta al estrés oxidativo que brinden tolerancia a los rayos UV.

En el plasmidoma de Puquio de Campo Naranja, se identificaron genes asociados con factores de virulencia y con la resistencia a metales, principalmente al cadmio, zinc, cobalto, cobre, arsénico, mercurio, cromo y selenio. Además, las resistencias a los antibióticos también estuvieron ampliamente distribuidas en este ambiente prístino en el que se supone que la presión selectiva está ausente. A pesar de las diferencias que existen entre un entorno no modificado por el hombre como la Puna andina y una planta de tratamiento de aguas residuales con impacto antropogénico, se registró cierto grado de similitud con respecto a la proporción de genes de resistencia a antibióticos y metales entre sus plasmidomas.

La comparación del plasmidoma de Laguna Diamante con el de Puquio de Campo Naranja, también reveló un cierto grado de similitud entre los perfiles funcionales predichos de ambos EMAs. No obstante, se registraron diferencias notables con respecto a la resistencia a los antibióticos y al arsénico. Las secuencias relacionadas con la resistencia a este último fueron más abundantes en el plasmidoma de Diamante que en el de Puquio, e incluso que en el plasmidoma derivado de la planta de tratamiento de aguas residuales con efluentes de la industria química. Estos hallazgos reflejan la gran cantidad de arsénico presente en Laguna Diamante.

Por otra parte, y en contraste a lo encontrado en Puquio de Campo Naranja, la presencia de *ARGs* en el plasmidoma de Laguna Diamante fue escasa. Este hecho, sumado a las relativamente pocas anotaciones funcionales en general, posiblemente se deba a las

peculiaridades de la composición microbiana de la comunidad bajo estudio, la cual incluye una proporción predominante de arqueas; además, el sesgo introducido por las bases de datos existentes indudablemente contribuye al resultado.

Cabe destacar que, en ambos plasmidomas, también se detectaron rasgos que se esperaba encontrar en un pool de plásmidos, como dominios Pfam relacionados con la replicación plasmídica, familias de relaxasas de tipo MOB implicadas en la movilización de plásmidos y genes pertenecientes a sistemas toxina-antitoxina de tipo II que están relacionados con el mantenimiento de los mismos. Además, se encontraron secuencias conocidas de megaplásmidos de microorganismos aislados de diferentes entornos salinos, lo que proporciona pruebas adicionales y convincentes de genes asociados a plásmidos conocidos en los conjuntos de datos obtenidos.

Los resultados presentados aquí, junto con la detección de numerosos elementos IS, suman evidencia sobre el papel clave que desarrolla el plasmidoma en la movilidad y la transferencia de genes dentro de estas comunidades microbianas. Esto permite la diseminación de rasgos beneficiosos que brindan aptitud celular y la adaptación de poblaciones enteras a las condiciones extremas presentes en la Puna andina.

CAPÍTULO II

Análisis y comparación de plásmidos lineales de cepas de Micrococcus de lagunas andinas

1. Introducción

Las actinobacterias representan una de las unidades taxonómicas más grandes dentro del dominio de las bacterias, ya que existen numerosas especies identificadas y una gran variedad de ellas. Comprenden bacterias Gram-positivas con un alto contenido de GC, el cual oscila entre el 50 % a más del 70 % (Ventura y col., 2007). Entre ellas se destacan varios géneros, tales como *Micrococcus*, *Brevibacterium*, *Arthrobacter*, *Streptomyces*, *Mycobacterium* y *Rhodococcus*.

Las actinobacterias presentan asimismo una amplia variedad de propiedades fisiológicas y rasgos metabólicos. Algunas producen enzimas extracelulares de importancia biotecnológica (Maruyama y col., 2005; Manikandan y col., 2011) y metabolitos secundarios que actúan como potentes antibióticos (Demain y Blander, 1999), lo que explica su importancia farmacéutica y económica; mientras que otras se pueden utilizar para procesos de biodegradación y biorremediación (Doddamani y Ninnekar, 2001; Li y col., 2005). Además, estas bacterias se destacan por su gran diversidad morfológica y por encontrarse en una amplia variedad de hábitats, tanto terrestres como acuáticos (Manivasagan y col., 2014; Rangseekaew y Pathom-Aree, 2019), adoptando una multitud de estilos de vida diferentes: comensales de plantas, patógenos, simbioses fijadores de nitrógeno, habitantes del tracto gastrointestinal y, principalmente, habitantes del suelo (Ventura y col., 2007).

En los últimos años se ha reportado la presencia de actinobacterias en ambientes extremos de la Puna andina, entre los que destacan miembros del género *Micrococcus*: *Micrococcus* sp. A1 y A7 aislados de Laguna Azul, *Micrococcus* sp. H5 aislado de Laguna Huaca Huasi, *Micrococcus* sp. V7 aislado de Laguna Vilama y *Micrococcus* sp. D12 aislado de Laguna Diamante (Dib y col., 2008). Laguna Azul (27° 34' S, 68° 32' O) y Laguna Diamante (26° 00' S, 67° 01' O) se encuentran al noroeste de Argentina, en la

provincia de Catamarca, a 4.560 y 4.589 m s.n.m., respectivamente. Laguna Vilama (22° 35' S, 66° 55' O) se ubica en Jujuy a 4.660 m s.n.m. y Laguna Huaca Huasi a 4.250 m s.n.m en las Cumbres Calchaquíes (26° 39' S, 65° 43' O), al norte de la provincia de Tucumán. Estos entornos prístinos comparten las características típicas de los ambientes andinos ya mencionadas anteriormente. Como podría esperarse, las cepas de *Micrococcus* aisladas mostraron resistencia a niveles altos de radiación UV y arsénico; y, además, presentaron múltiples resistencias a los antibióticos (particularmente a los macrólidos) (Dib y col., 2008, 2009; Ordoñez y col., 2009).

Los estudios posteriores revelaron la presencia de plásmidos de gran tamaño (megaplásmidos) en dichas cepas: pLMA1 (110 kb) en *Micrococcus* sp. A1, pLMA7 (80 kb) en *Micrococcus* sp. A7, pLMH5 (110 kb) en *Micrococcus* sp. H5, pLMV7 (90 kb) en *Micrococcus* sp. V7 y pJD12 (75 kb) en *Micrococcus* sp. D12 (Figura II.1). Como la migración de los plásmidos en electroforesis en gel de campo pulsado (PFGE), utilizando diferentes condiciones de corrida (tiempo de pulso), se mantuvo constante en relación con el marcador, se consideró que los plásmidos eran lineales (Figura II.2). Además, mediante tratamientos con exonucleasa III y exonucleasa λ que hidrolizan el ADN doble cadena en la dirección 3'-5' y 5'-3' respectivamente, se demostró que estos replicones lineales poseen proteínas terminales (TPs) unidas covalentemente a sus extremos 5' (Dib y col., 2010, 2013a), lo cual es una característica típica de los elementos lineales del tipo invertron (Sakaguchi, 1990). Por otra parte, un análisis por la técnica de Southern blot y los patrones de digestión con endonucleasas de restricción, demostraron que pLMA1 y pLMH5 eran plásmidos indistinguibles (Dib y col., 2010a). Desde un punto de vista ecológico, el hecho de que estos plásmidos se hayan encontrado en cepas diferentes, aisladas de lagunas separadas por más de 300 km, apunta a la difusión de los elementos por transferencia horizontal, y puede tomarse

como una indicación de que la información genética codificada por ellos es beneficiosa para la supervivencia en esas condiciones ambientales.

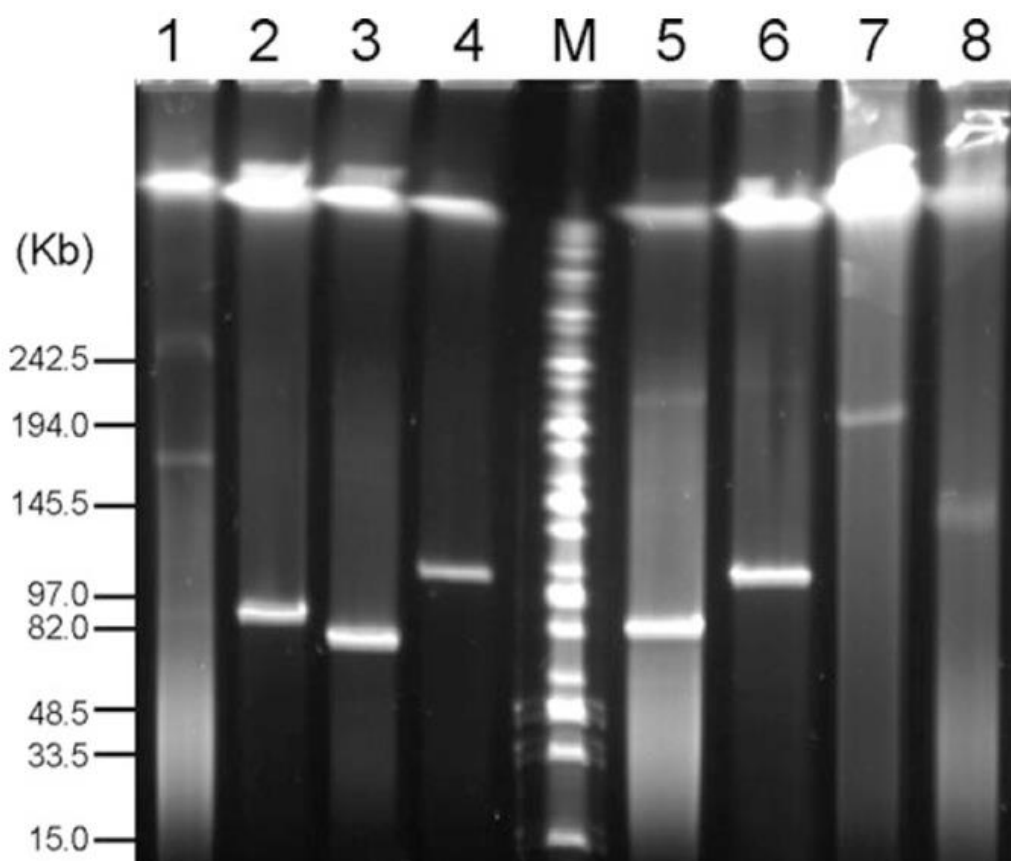


Figura II.1. Electroforesis en gel de campo pulsado (PFGE) de ADN de actinobacterias aisladas de lagos andinos de gran altitud. Calle 1: *Nocardia* sp. A5; calle 2: *Micrococcus* sp. V7; calle 3: *Micrococcus* sp. A7; calle 4: *Micrococcus* sp. A1; calle 5: *Brevibacterium* sp. Ap13; calle 6: *Micrococcus* sp. H5; calle 7: *Rhodococcus* sp. Ch13; calle 8: *Kokuria* sp. N33; calle M: Marcador *MidRange I PFG Marker* (New England Biolabs, Massachusetts, EE.UU.). Condiciones de corrida: 1-25 s durante 24 h (6 V cm^{-1} , 14°C). Imagen tomada de Wagenknecht y col. (2020).

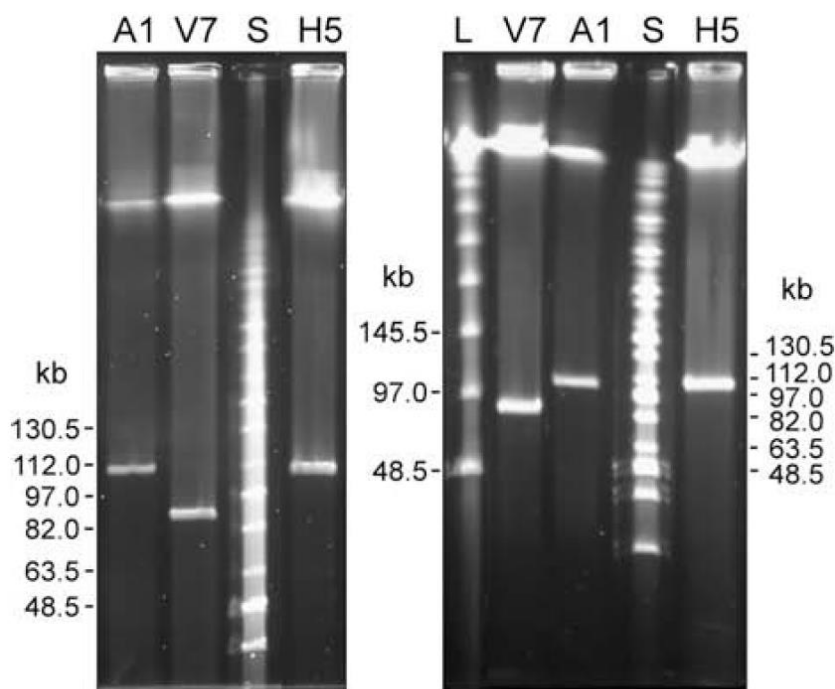


Figura II.2. PFGE del ADN total de tres cepas de *Micrococcus* en diferentes condiciones de corrida. Los tiempos de pulso fueron de 10-30 s (izquierda) y de 1-25 s (derecha), durante 24 h (6 V cm^{-1} , $15 \text{ }^{\circ}\text{C}$). A1: *Micrococcus* sp. A1; H5: *Micrococcus* sp. H5; V7: *Micrococcus* sp. V7; S: Marcador *MidRange I PFG Marker* (New England Biolabs, Massachusetts, EE. UU.); L: Marcador *Lambda Ladder PFG Marker* (New England Biolabs, Massachusetts, EE.UU.). Imagen tomada de Dib y col. (2010a).

Estos elementos extra-cromosómicos fueron secuenciados y anotados ya que, como se explicó anteriormente, la adaptación de sus huéspedes a las condiciones ambientales podría ser, al menos parcialmente, conferida por ellos (Dib y col., 2013c, 2015a, 2018). De hecho, en un trabajo previo, se obtuvo evidencia de que la resistencia a eritromicina está codificada por el plásmido pLMA1 al verificar el perfil de resistencia de un derivado de *Micrococcus* sp. A1 curado de dicho plásmido (Dib y col., 2010a).

En este capítulo, se realiza un análisis comparativo de las secuencias de los megaplásmidos lineales antes mencionados, con el objetivo de identificar regiones accesorias y conservadas, y estudiar posibles relaciones de transferencia horizontal de

genes (*HGT*), fenómeno significativo en la adaptación de las actinobacterias a estos ambientes extremos. Las condiciones allí presentes podrían actuar como presión de selección hacia mutaciones genéticas y *HGT*, permitiendo así la diseminación de propiedades fenotípicas generalizadas en la población, como la resistencia a antibióticos y metales.

El estudio de los plásmidos lineales individuales en esta tesis brinda indicios sobre el papel y el impacto de este tipo de elementos en ambientes de la Puna andina. Como se mencionó previamente, no han podido ser estudiados globalmente debido a las limitaciones técnicas de los protocolos hasta ahora aplicados para los plasmidomas.

2. Materiales y Métodos

2.1. Cepas y plásmidos

Los plásmidos analizados y comparados en este capítulo fueron purificados previamente a partir de cepas de *Micrococcus* aisladas de muestras de agua de lagunas andinas (Tabla II.1) (Dib y col., 2008). Los plásmidos se aislaron mediante electroforesis en gel de campo pulsado y se purificaron del gel mediante electroelución. El plásmido pLMA7 se secuenció utilizando el método convencional de Sanger, mientras que para los plásmidos pLMA1, pLMV7 y pJD12 se utilizó una combinación de Sanger y pirosecuenciación 454. Esta última estrategia de secuenciación se aplicó luego de comprobar que el ensamblaje de lecturas solamente a partir de la tecnología 454 resultaba imposible debido a la presencia de numerosas repeticiones de secuencia de longitud variada en pLMA1 (Wagenknecht y col., 2010).

Tabla II.1. Cepas de *Micrococcus* aisladas de lagunas andinas con plásmidos lineales.

Cepa	Fuente de aislamiento	Coordenadas	Plásmido	Tamaño aprox. (kb)	Nº de acceso GenBank	Referencias
<i>Micrococcus</i> sp. A1	L. Azul (Catamarca)	27° 34' S 68° 32' O	pLMA1	110	LK056645.1	Dib y col., 2010a, 2018; Wagenknecht y col., 2010
<i>Micrococcus</i> sp. A7	L. Azul (Catamarca)	27° 34' S 68° 32' O	pLMA7	80	KJ599675.1	Dib y col., 2018
<i>Micrococcus</i> sp. V7	L. Vilama (Jujuy)	22° 35' S 66° 55' O	pLMV7	90	KF577591.1	Dib y col., 2010a, 2013c
<i>Micrococcus</i> sp. D12	L. Diamante (Catamarca)	26° 00' S 67° 01' O	pJD12	75	KR152226.1	Dib y col., 2015a

L.: Laguna

2.2. Análisis comparativo de megaplásmidos

Las secuencias anotadas de los plásmidos lineales pLMA1, pLMA7, pLMV7 y pJD12 fueron descargadas de la base de datos NCBI (Nº de acceso en Tabla II.1). Para obtener información más actualizada, las secuencias aminoacídicas de las proteínas codificadas fueron nuevamente comparadas con la base de datos NCBI mediante BLAST (Altschul y col., 1990). Se utilizó DNAPlotter para hacer el mapa genético de cada megaplásmido (Carver y col., 2009) y la herramienta ACT de Artemis para comparar la similitud de sus secuencias (Carver y col., 2005).

3. Resultados y Discusión

3.1. Características generales

La secuencia de nucleótidos de pLMA1, con un contenido promedio de GC del 68,40 %, consta de 109.112 pb y comprende 125 genes que codifican proteínas predichas, de las cuales 62 se asignaron a funciones conocidas. El plásmido pLMA7 comprende 82.075 pb (69,56 % GC) y alberga 93 genes que codifican proteínas predichas, con 33

de ellas asignadas a funciones conocidas. Por su parte, el plásmido pLMV7 presenta una longitud de 92.815 pb (69,40 % GC) y 114 marcos abiertos de lectura (*ORFs*), de los cuales solo 28 codifican proteínas con función conocida. El plásmido pJD12 consta de ADN lineal que abarca 75.989 pb con un contenido medio de GC del 68,86 % y su anotación revela la presencia de 80 *ORFs* (27 con función conocida) (Tabla II.2).

Tabla II.2. Características de secuencia de los plásmidos analizados.

Plásmido	Longitud (pb)	Contenido GC (%)	Nº <i>ORFs</i>	Función conocida
pLMA1	109.112	68,40	125	62
pLMA7	82.075	69,56	93	33
pLMV7	92.815	69,40	114	28
pJD12	75.989	68,86	80	27

3.2. Comparación y funciones

La comparación de secuencias con la herramienta ACT de Artemis, reveló que el plásmido pLMV7 no presenta regiones con similitud a las secuencias de nucleótidos de los otros plásmidos (identidad > 89 %), a pesar de que la cepa huésped (*Micrococcus* sp. V7) proviene de un entorno similar y pertenece al mismo género que las demás. Por lo tanto, se continuó el análisis con los plásmidos pLMA1, pLMA7 y pJD12. En ellos se encontraron 6 bloques de ADN altamente homólogos, con una identidad de secuencia que varía entre el 92 y 98 % (Figura II.3). Esos bloques de ADN contienen mayormente genes que codifican para proteínas de función desconocida, pero también se observaron genes de funciones conservadas implicadas en la replicación, mantención y conjugación plasmídica. Además, comparando el contenido de GC (%) para cada segmento de secuencia de cada megaplásmido, se pudo evidenciar que la información accesoria de

las regiones variables, tiene un contenido de GC distinto al del esqueleto molecular común (compuesto por los 6 bloques homólogos) (Figura II.4).

Los genes anotados en cada megaplásmido se clasificaron en nueve grupos de acuerdo a la función que les fue asignada. La organización genética de cada uno puede verse en los mapas de las Figuras II.5, II.6 y II.7, y la función correspondiente a cada *ORF* en las Tablas II.3, II.4 y II.5.

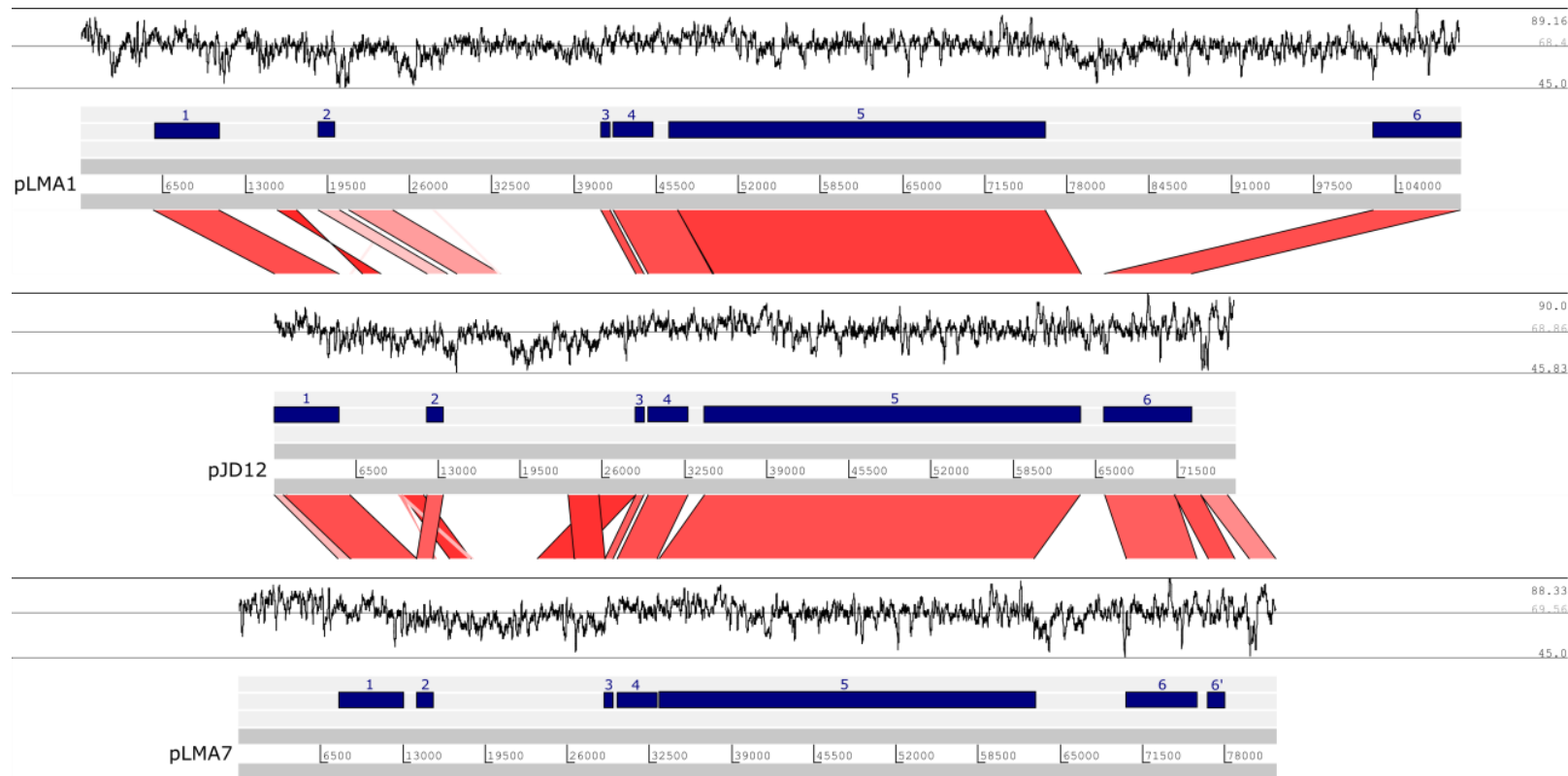


Figura II.3. Comparación de secuencia de los plásmidos lineales pLMA1 (109.112 pb), pJD12 (75.989 pb) y pLMA7 (82.075 pb). Las regiones conservadas entre cada par de plásmidos se muestran en rojo, las distintas intensidades se refieren a diferentes valores de similitud, a mayor intensidad mayor identidad de secuencia. Se estableció como punto de corte una identidad mayor al 89 % y un mínimo de 157 pares de bases. Las regiones conservadas en los tres plásmidos se muestran en bloques de color azul (1-6). Por encima de la secuencia de cada plásmido se grafica el contenido de GC, que varía del 45 al 89,16 % con un promedio del 68,40 % para pLMA1, del 45,83 al 90 % con un promedio del 68,86 % para pJD12, y del 45 al 88,33 % con un promedio del 69,56 % para pLMA7.

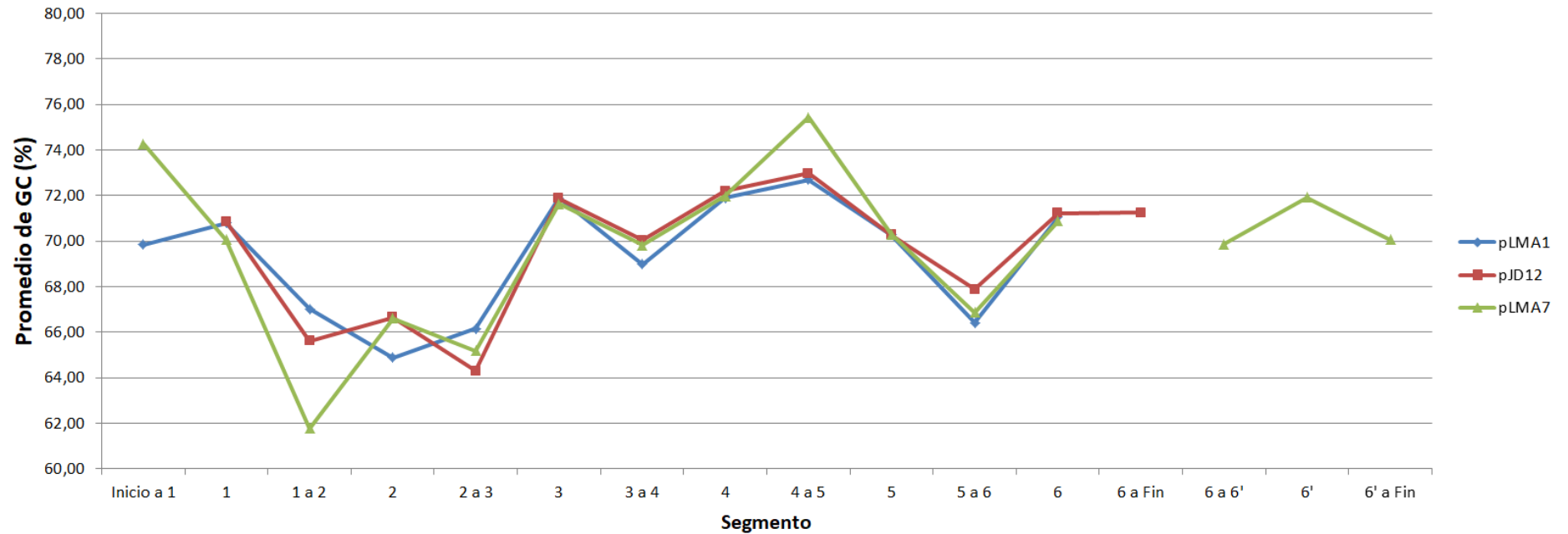


Figura II.4. Promedio del contenido de GC (%) por segmento de secuencia para cada megaplásmido. En todos los segmentos homólogos, con excepción del N°2 en el que existe una leve variación para pLMA1, se puede observar la similitud en el contenido de GC al tratarse de regiones conservadas. Los segmentos no homólogos, donde se encuentran las funciones accesorias, presentan mayores desviaciones en el contenido de GC, tanto entre plásmidos como con respecto al esqueleto molecular común.

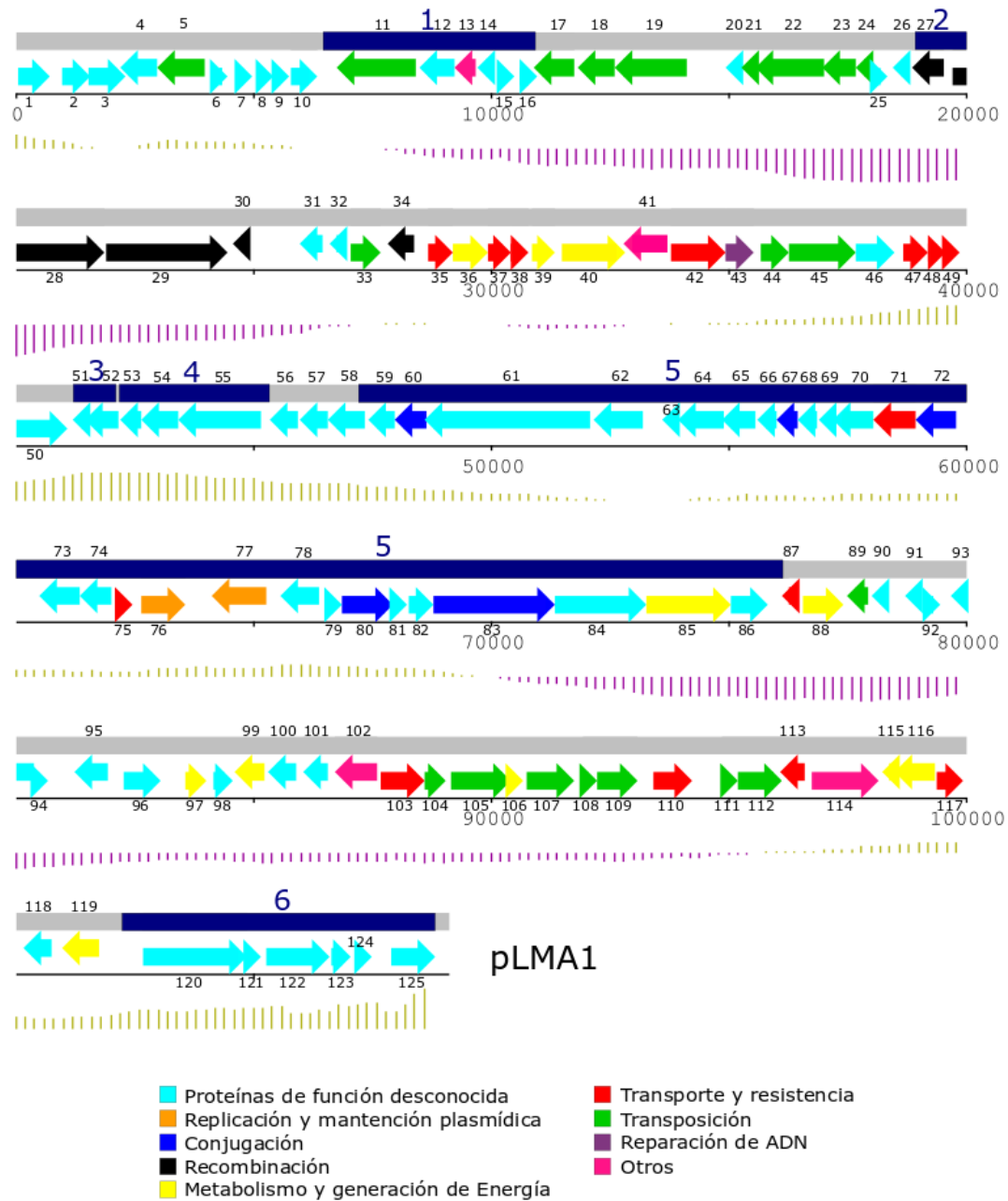


Figura II.5. Mapa genético del plásmido lineal pLMA1. Las flechas indican los genes predichos y su orientación indica el sentido de la transcripción (los *ORFs* se identifican con números secuenciales). El color de cada *ORF* indica el grupo funcional al cual fue asignado el gen, de acuerdo con el código en la parte inferior. Los bloques de color azul oscuro (1-6) se refieren a las secuencias conservadas en los tres plásmidos (pLMA1, pJD12 y pLMA7). La variación del contenido de GC a lo largo de la secuencia se grafica con barras verticales de color verde (por encima del promedio de GC) y violeta (por debajo del promedio de GC).

Tabla II.3. Función asignada a cada *ORF* en el plásmido pLMA1.

<i>ORF</i>	Posición de inicio	Posición de término	Longitud (nt)	Cadena	Nombre de la Proteína / Función	Región conservada
pLMA1						
1	60	740	680	+	<i>Hypothetical protein</i>	
2	997	1560	563	+	<i>Hypothetical protein</i>	
3	1545	2324	779	+	<i>Hypothetical protein</i>	
4	2222	2965	743	-	<i>Hypothetical protein</i>	
5	2975	3985	1010	-	<i>IS481 family transposase</i>	
6	4350	4451	101	+	<i>Hypothetical protein</i>	
7	4661	4990	329	+	<i>Hypothetical protein</i>	
8	5151	5420	269	+	<i>Hypothetical protein</i>	
9	5500	5769	269	+	<i>Hypothetical protein</i>	
10	5798	6349	551	+	<i>Hypothetical protein</i>	
11	6773	8434	1661	-	<i>Putative transposase</i>	1
12	8517	9254	737	-	<i>Hypothetical protein</i>	
13	9251	9700	449	-	<i>Helix-turn-helix transcriptional regulator</i>	
14	9704	10003	299	-	<i>Hypothetical protein</i>	
15	10314	10490	176	+	<i>Hypothetical protein</i>	
16	10628	10984	356	+	<i>Hypothetical protein</i>	
17	10905	11774	869	-	<i>IS21 putative ATP-binding protein IstB</i>	
18	11842	12627	785	-	<i>IS21-like element helper ATPase IstB</i>	
19	12624	14147	1523	-	<i>IS21 family transposase</i>	
20	14933	15133	200	-	<i>Hypothetical protein</i>	
21	15275	15592	317	-	<i>Integrase core domain-containing protein</i>	
22	15573	17015	1442	-	<i>Transposase InsF for insertion sequence IS3</i>	
23	16981	17691	710	-	<i>IS3 family transposase</i>	
24	17688	18017	329	-	<i>Transposase</i>	
25	18211	18345	134	+	<i>Hypothetical protein</i>	
26	18452	18784	332	-	<i>Hypothetical protein</i>	
27	18881	19546	665	-	<i>HNH endonuclease</i>	2
28	19710	21884	2174	+	<i>Site-specific DNA-methyltransferase</i>	
29	21890	24475	2585	+	<i>Type III restriction enzyme, res subunit</i>	
30	24562	24879	317	-	<i>Resolvase, N-terminal domain protein</i>	
31	25999	26475	476	-	<i>Hypothetical protein</i>	
32	26594	26749	155	-	<i>Hypothetical protein</i>	
33	27071	27697	626	+	<i>Transposon Tn21 resolvase</i>	
34	27824	28393	569	-	<i>Recombinase family protein</i>	
35	28694	29200	506	+	<i>Universal stress protein</i>	
36	29197	29943	746	+	<i>2,3-diphosphoglycerate-dependent phosphoglycerate mutase</i>	
37	29940	30443	503	+	<i>Putative fluoride ion transporter</i>	

CrcB 1					
38	30440	30811	371	+	CrcB family protein
39	30859	31344	485	+	Inorganic diphosphatase
40	31487	32821	1334	+	Phosphopyruvate hydratase
41	32806	33732	926	-	LysR family transcriptional regulator
42	33803	34960	1157	+	Tellurite-Resistance/Dicarboxylate Transporter (TDT) family
43	34957	35541	584	+	Uracil-DNA glycosylase
44	35682	36284	602	+	Transposase for transposon Tn552
45	36291	37685	1394	+	DDE-type integrase/transposase/recombinase
46	37682	38494	812	+	Hypothetical protein
47	38686	39216	530	+	TetR family transcriptional regulator
48	39213	39557	344	+	Multidrug resistance protein, SMR family
49	39557	39886	329	+	QacE family quaternary ammonium compound efflux SMR transporter
50	40023	41078	1055	+	Hypothetical protein
51	41199	41471	272	-	Hypothetical protein
52	41468	42181	713	-	Hypothetical protein
53	42191	42652	461	-	Hypothetical protein
54	42643	43422	779	-	Hypothetical protein
55	43415	45184	1769	-	Hypothetical protein
56	45339	45953	614	-	Hypothetical protein
57	45997	46584	587	-	Hypothetical protein
58	46574	47344	770	-	Hypothetical protein
59	47439	47993	554	-	Hypothetical protein
60	47998	48633	635	-	Prepilin peptidase
61	48614	52075	3461	-	Hypothetical protein
62	52170	53222	1052	-	Hypothetical protein
63	53621	53788	167	-	Hypothetical protein
64	53952	54890	938	-	Hypothetical protein
65	54884	55585	701	-	Hypothetical protein
66	55601	56017	416	-	Hypothetical protein
67	56020	56469	449	-	Pilus assembly protein
68	56466	56876	410	-	Hypothetical protein
69	56920	57102	182	-	Hypothetical protein
70	57202	58074	872	-	Hypothetical protein
71	58071	58937	866	-	Type II secretion system F family protein
72	58934	59800	866	-	Pilus assembly protein, ATPase of CpaF family
73	60515	61339	824	-	Hypothetical protein
74	61336	62013	677	-	Hypothetical protein
75	62185	62481	296	+	MarR family transcriptional regulator
76	62666	63589	923	+	ParA family protein
77	64112	65263	1151	-	Plasmid encoded RepA protein
78	65572	66378	806	-	Hypothetical protein

79	66641	66868	227	+	<i>Hypothetical protein</i>	
80	66865	67959	1094	+	<i>Conjugal transfer protein</i>	
81	67986	68255	269	+	<i>Hypothetical protein</i>	
82	68273	68788	515	+	<i>Hypothetical protein</i>	
83	68785	71340	2555	+	<i>IV secretion system protein VirB4</i>	
84	71337	73283	1946	+	<i>Hypothetical protein</i>	
85	73280	75046	1766	+	<i>Peptidoglycan DD-metalloendopeptidase family protein</i>	
86	75043	75828	785	+	<i>Hypothetical protein</i>	
87	76145	76291	146	-	<i>Fluoroacetate dehalogenase</i>	
88	76563	77429	866	+	<i>Alpha/beta hydrolase</i>	
89	77489	77935	446	-	<i>IS5 family transposase</i>	
90	78018	78368	350	-	<i>Hypothetical protein</i>	
91	78704	79042	338	-	<i>Hypothetical protein</i>	
92	79320	79460	140	+	<i>Hypothetical protein</i>	
93	79692	80405	713	-	<i>Hypothetical protein</i>	
94	80434	80679	245	+	<i>Hypothetical protein</i>	
95	81250	81948	698	-	<i>Hypothetical protein</i>	
96	82284	83057	773	+	<i>Hypothetical protein</i>	
97	83556	84032	476	+	<i>ADP-ribose pyrophosphatase</i>	
98	84160	84591	431	+	<i>Hypothetical protein</i>	
99	84617	85237	620	-	<i>SAM-dependent methyltransferase</i>	
100	85332	85901	569	-	<i>Hypothetical protein</i>	
101	86064	86591	527	-	<i>Hypothetical protein</i>	
102	86727	87620	893	-	<i>LysR family transcriptional regulator</i>	
103	87700	88599	899	+	<i>EamA family transporter</i>	
104	88596	89060	464	+	<i>Transposase for insertion sequence element IS1081</i>	
105	89171	90415	1244	+	<i>IS256 family transposase</i>	
106	90421	90672	251	+	<i>NAD dependent epimerase/dehydratase family protein</i>	
107	90759	91745	986	+	<i>IS481 family transposase</i>	
108	91932	92225	293	+	<i>IS3 family transposase</i>	
109	92225	93106	881	+	<i>IS3 family transposase</i>	
110	93411	94229	818	+	<i>23S ribosomal RNA methyltransferase Erm</i>	
111	94872	95201	329	+	<i>Transposase for insertion sequence element IS904</i>	
112	95198	96121	923	+	<i>IS3 family transposase</i>	
113	96095	96622	527	-	<i>DMT family transporter</i>	
114	96762	98174	1412	+	<i>Transcriptional regulator, GntR family</i>	
115	98224	98400	176	-	<i>Putative short-chain type dehydrogenase/reductase VdIC</i>	
116	98518	99348	830	-	<i>SDR family NAD(P)-dependent oxidoreductase</i>	
117	99402	99929	527	+	<i>TetR family transcriptional regulator</i>	
118	100166	100741	575	-	<i>Hypothetical protein</i>	
119	100998	101771	773	-	<i>Prolyl oligopeptidase family protein</i>	

120	102668	104869	2201	+	Hypothetical protein	6
121	104871	105176	305	+	Hypothetical protein	
122	105280	106623	1343	+	Hypothetical protein	
123	106657	107052	395	+	Hypothetical protein	
124	107256	107486	230	+	Hypothetical protein	
125	107889	108845	956	+	Hypothetical protein	

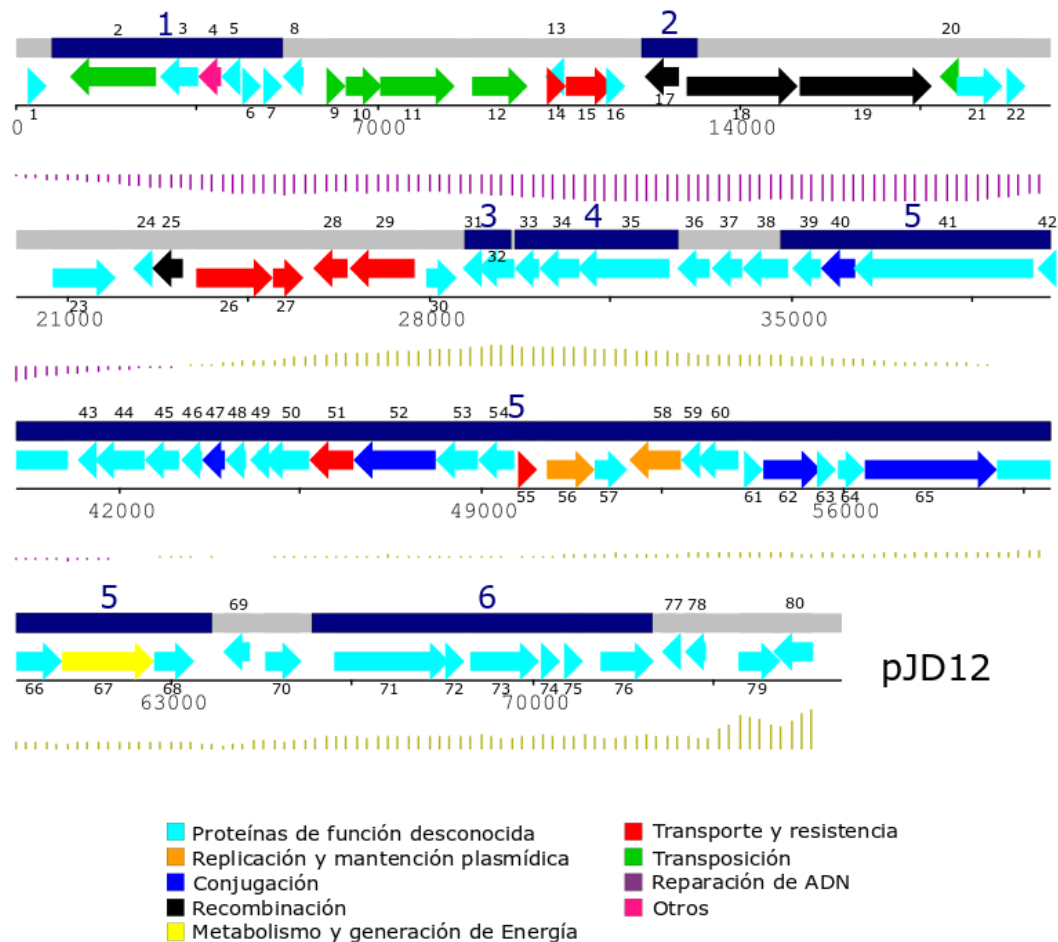


Figura II.6. Mapa genético del plásmido lineal pJD12. Las flechas indican los genes predichos y su orientación indica el sentido de la transcripción (los ORFs se identifican con números secuenciales). El color de cada ORF indica el grupo funcional al cual fue asignado el gen, de acuerdo con el código en la parte inferior. Los bloques de color azul oscuro (1-6) se refieren a las secuencias conservadas en los tres plásmidos (pLMA1, pJD12 y pLMA7). La variación del contenido de GC a lo largo de la secuencia se grafica con barras verticales de color verde (por encima del promedio de GC) y violeta (por debajo del promedio de GC).

Tabla II.4. Función asignada a cada *ORF* en el plásmido pJD12.

<i>ORF</i>	Posición de inicio	Posición de término	Longitud (nt)	Cadena	Nombre de la Proteína / Función	Región conservada
pJD12						
1	277	624	347	+	<i>Hypothetical protein</i>	1
2	1048	2709	1661	-	<i>Transposase</i>	
3	2792	3529	737	-	<i>Hypothetical protein</i>	
4	3526	3975	449	-	<i>Helix-turn-helix transcriptional regulator</i>	
5	3979	4278	299	-	<i>Hypothetical protein</i>	
6	4590	4766	176	+	<i>Hypothetical protein</i>	
7	4919	5179	260	+	<i>Hypothetical protein</i>	
8	5176	5556	380	-	<i>Hypothetical protein</i>	2
9	6055	6372	317	+	<i>Transposase</i>	
10	6375	7076	701	+	<i>IS3 family transposase</i>	
11	7042	8484	1442	+	<i>IS30 family transposase</i>	
12	8838	9908	1070	+	<i>IS110 family transposase</i>	
13	10242	10424	182	-	<i>Hypothetical protein</i>	
14	10425	10649	224	+	<i>Bacterial regulatory protein, arsR family</i>	
15	10649	11581	932	+	<i>Co/Zn/Cd efflux system component protein</i>	
16	11553	11804	251	+	<i>Hypothetical protein</i>	
17	12163	12828	665	-	<i>HNH endonuclease</i>	
18	12992	15145	2153	+	<i>Site-specific DNA-methyltransferase</i>	
19	15151	17736	2585	+	<i>Type III restriction enzyme, res subunit</i>	
20	17873	18163	290	-	<i>Transposase</i>	3
21	18205	19080	875	+	<i>Hypothetical protein</i>	
22	19282	19542	260	+	<i>Hypothetical protein</i>	
23	20709	21950	1241	+	<i>Hypothetical protein</i>	
24	22289	22645	356	-	<i>Hypothetical protein</i>	
25	22656	23225	569	-	<i>Site-specific DNA recombinase</i>	
26	23505	24971	1466	+	<i>MFS transporter</i>	
27	24989	25558	569	+	<i>Transcriptional regulator, TetR family</i>	
28	25764	26444	680	-	<i>Iron (metal) dependent repressor, DtxR family</i>	
29	26461	27735	1274	-	<i>Mn²⁺ and Fe²⁺ transporters of the NRAMP family</i>	
30	27926	28522	596	+	<i>Hypothetical protein</i>	4
31	28660	28932	272	-	<i>Hypothetical protein</i>	
32	28929	29645	716	-	<i>Hypothetical protein</i>	
33	29656	30123	467	-	<i>Hypothetical protein</i>	
34	30114	30893	779	-	<i>Hypothetical protein</i>	
35	30886	32655	1769	-	<i>Hypothetical protein</i>	
36	32810	33424	614	-	<i>Hypothetical protein</i>	

37	33468	34055	587	-	<i>Hypothetical protein</i>	5
38	34058	34933	875	-	<i>Hypothetical protein</i>	
39	35030	35584	554	-	<i>Hypothetical protein</i>	
40	35589	36224	635	-	<i>Prepilin peptidase</i>	
41	36205	39672	3467	-	<i>Hypothetical protein</i>	
42	39768	41030	1262	-	<i>Hypothetical protein</i>	
43	41219	41461	242	-	<i>Hypothetical protein</i>	
44	41550	42488	938	-	<i>Hypothetical protein</i>	
45	42482	43183	701	-	<i>Hypothetical protein</i>	
46	43199	43615	416	-	<i>Hypothetical protein</i>	
47	43618	44067	449	-	<i>Pilus assembly protein</i>	
48	44064	44474	410	-	<i>Hypothetical protein</i>	
49	44519	44701	182	-	<i>Hypothetical protein</i>	
50	44801	45673	872	-	<i>Hypothetical protein</i>	
51	45683	46537	854	-	<i>Type II secretion system F family protein</i>	
52	46534	48117	1583	-	<i>Pilus assembly protein, ATPase of CpaF family</i>	
53	48114	48938	824	-	<i>Hypothetical protein</i>	
54	48935	49639	704	-	<i>Hypothetical protein</i>	
55	49811	50080	269	+	<i>MarR family transcriptional regulator</i>	
56	50265	51188	923	+	<i>ParA family protein</i>	
57	51190	51816	626	+	<i>Hypothetical protein</i>	
58	51869	52888	1019	-	<i>Plasmid encoded RepA protein</i>	
59	52857	53132	275	-	<i>Hypothetical protein</i>	
60	53173	53979	806	-	<i>Hypothetical protein</i>	
61	54242	54469	227	+	<i>Hypothetical protein</i>	
62	54466	55560	1094	+	<i>Conjugal transfer protein</i>	
63	55592	55885	293	+	<i>Hypothetical protein</i>	
64	55903	56418	515	+	<i>Hypothetical protein</i>	
65	56415	58970	2555	+	<i>IV secretion system protein VirB4</i>	
66	58967	60913	1946	+	<i>Hypothetical protein</i>	
67	60910	62676	1766	+	<i>Peptidoglycan DD-metalloendopeptidase family protein</i>	
68	62673	63458	785	+	<i>Hypothetical protein</i>	
69	64012	64545	533	-	<i>Hypothetical protein</i>	6
70	64835	65536	701	+	<i>Hypothetical protein</i>	
71	66177	68390	2213	+	<i>Hypothetical protein</i>	
72	68392	68697	305	+	<i>Hypothetical protein</i>	
73	68801	70117	1316	+	<i>Hypothetical protein</i>	
74	70151	70546	395	+	<i>Hypothetical protein</i>	
75	70750	70980	230	+	<i>Hypothetical protein</i>	
76	71322	72353	1031	+	<i>Hypothetical protein</i>	
77	72502	72741	239	-	<i>Hypothetical protein</i>	
78	72956	73342	386	-	<i>Hypothetical protein</i>	
79	73975	74802	827	+	<i>Hypothetical protein</i>	

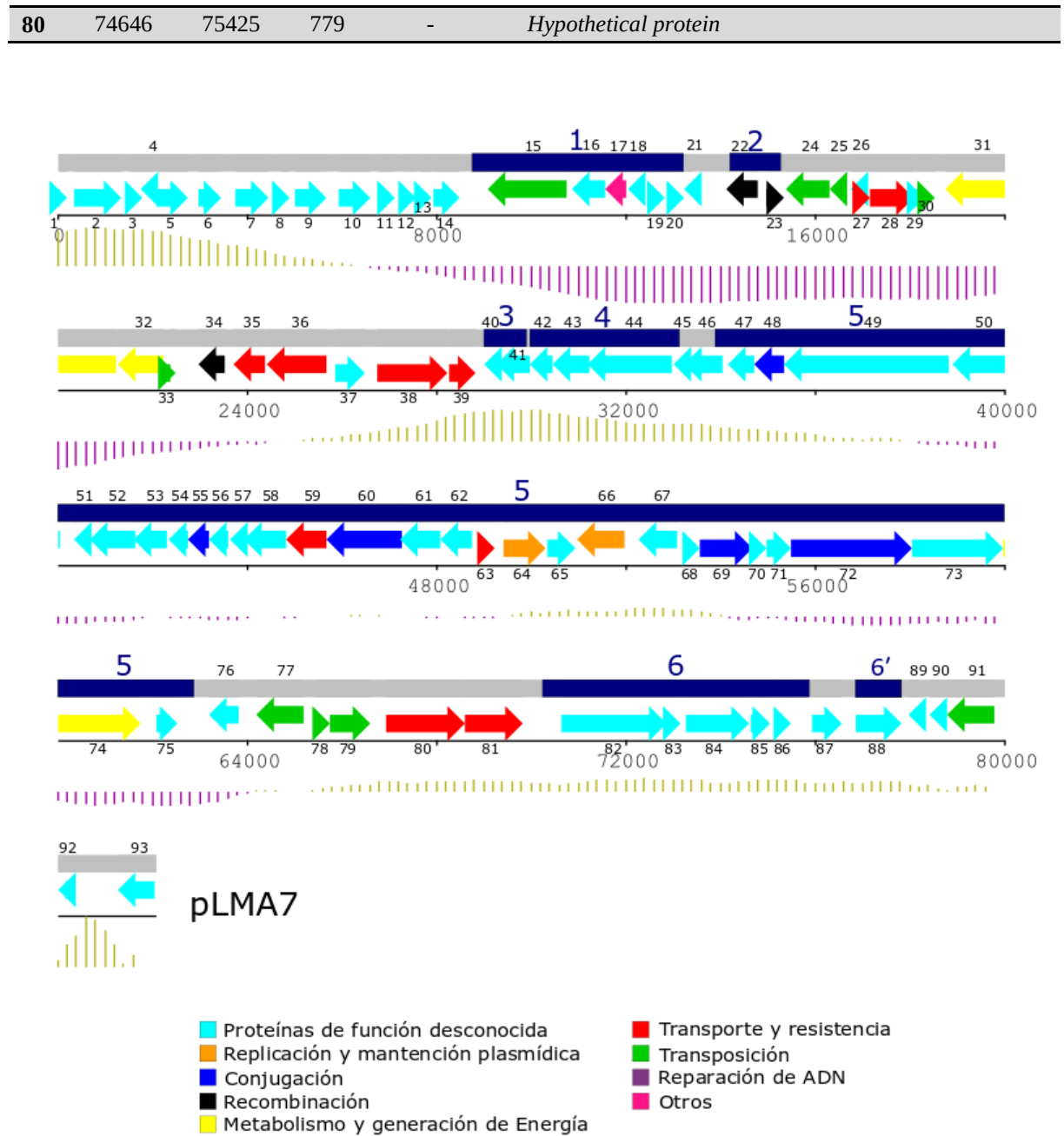


Figura II.7. Mapa genético del plásmido lineal pLMA7. Las flechas indican los genes predichos y su orientación indica el sentido de la transcripción (los *ORFs* se identifican con números secuenciales). El color de cada *ORF* indica el grupo funcional al cual fue asignado el gen, de acuerdo con el código en la parte inferior. Los bloques de color azul oscuro (1-6) se refieren a las secuencias conservadas en los tres plásmidos (pLMA1, pJD12 y pLMA7). La variación del contenido de GC a lo largo de la secuencia se grafica con barras verticales de color verde (por encima del promedio de GC) y violeta (por debajo del promedio de GC).

Tabla II.5. Función asignada a cada *ORF* en el plásmido pLMA7.

<i>ORF</i>	Posición de inicio	Posición de término	Longitud (nt)	Cadena	Nombre de la Proteína / Función	Región conservada
pLMA7						
1	41	211	170	+	<i>Hypothetical protein</i>	
2	366	1343	977	+	<i>Hypothetical protein</i>	
3	1540	1797	257	+	<i>Hypothetical protein</i>	
4	1761	2057	296	-	<i>Hypothetical protein</i>	
5	2094	2774	680	+	<i>Hypothetical protein</i>	
6	2994	3479	485	+	<i>Hypothetical protein</i>	
7	3748	4461	713	+	<i>Hypothetical protein</i>	
8	4725	4916	191	+	<i>Hypothetical protein</i>	
9	5033	5683	650	+	<i>Hypothetical protein</i>	
10	5935	6600	665	+	<i>Hypothetical protein</i>	
11	6813	7142	329	+	<i>Hypothetical protein</i>	
12	7303	7572	269	+	<i>Hypothetical protein</i>	
13	7652	7921	269	+	<i>Hypothetical protein</i>	
14	7951	8502	551	+	<i>Hypothetical protein</i>	
15	9108	10769	1661	-	<i>Transposase</i>	1
16	10852	11589	737	-	<i>Hypothetical protein</i>	
17	11586	12035	449	-	<i>Helix-turn-helix transcriptional regulator</i>	
18	12039	12338	299	-	<i>Hypothetical protein</i>	
19	12648	12824	176	+	<i>Hypothetical protein</i>	
20	12968	13237	269	+	<i>Hypothetical protein</i>	
21	13234	13413	179	-	<i>Hypothetical protein</i>	
22	14147	14812	665	-	<i>HNH endonuclease</i>	2
23	14976	15362	386	+	<i>Site-specific DNA-methyltransferase (type III restriction-modification system methylation subunit)</i>	
24	15387	16301	914	-	<i>IS3 family transposase</i>	
25	16298	16618	320	-	<i>Transposase</i>	
26	16744	16959	215	-	<i>Hypothetical protein</i>	
27	16927	17151	224	+	<i>Regulatory protein ArsR</i>	
28	17151	18083	932	+	<i>Cobalt-zinc-cadmium resistance protein CzcD</i>	
29	18157	18306	149	+	<i>Hypothetical protein</i>	
30	18341	18538	197	+	<i>Transposase</i>	
31	18748	21225	2477	-	<i>S8 family serine peptidase</i>	
32	21261	22238	977	-	<i>ATP-binding protein</i>	
33	22388	22513	125	+	<i>Transposase</i>	
34	22988	23542	554	-	<i>Recombinase family protein</i>	
35	23725	24405	680	-	<i>Mn-dependent transcriptional regulator MntR</i>	
36	24422	25696	1274	-	<i>Manganese transport protein MntH</i>	
37	25887	26483	596	+	<i>Hypothetical protein</i>	
38	26776	28242	1466	+	<i>MFS transporter</i>	

39	28260	28829	569	+	Transcriptional regulator, TetR family	3
40	29001	29273	272	-	Hypothetical protein	
41	29270	29983	713	-	Hypothetical protein	
42	29993	30454	461	-	Hypothetical protein	4
43	30445	31224	779	-	Hypothetical protein	
44	31217	32986	1769	-	Hypothetical protein	
45	33032	33235	203	-	Hypothetical protein	5
46	33225	34073	848	-	Hypothetical protein	
47	34170	34724	554	-	Hypothetical protein	
48	34729	35364	635	-	Prepilin peptidase	
49	35345	38818	3473	-	Hypothetical protein	
50	38913	40040	1127	-	Hypothetical protein	
51	40364	40531	167	-	Hypothetical protein	
52	40695	41633	938	-	Hypothetical protein	
53	41627	42328	701	-	Hypothetical protein	
54	42344	42760	416	-	Hypothetical protein	
55	42763	43212	449	-	Pilus assembly protein	
56	43209	43619	410	-	Hypothetical protein	
57	43664	43846	182	-	Hypothetical protein	
58	43946	44818	872	-	Hypothetical protein	
59	44815	45681	866	-	Type II secretion system F family protein	
60	45678	47261	1583	-	Pilus assembly protein, ATPase of CpaF family	
61	47258	48082	824	-	Hypothetical protein	
62	48079	48756	677	-	Hypothetical protein	
63	48928	49224	296	+	MarR family transcriptional regulator	
64	49409	50332	923	+	ParA family protein	
65	50334	50960	626	+	Hypothetical protein	
66	50979	51980	1001	-	Plasmid encoded RepA protein	
67	52283	53089	806	-	Hypothetical protein	
68	53352	53579	227	+	Hypothetical protein	
69	53576	54670	1094	+	Conjugal transfer protein	
70	54697	54966	269	+	Hypothetical protein	
71	54984	55499	515	+	Hypothetical protein	
72	55496	58051	2555	+	IV secretion system protein VirB4	
73	58048	59994	1946	+	Hypothetical protein	
74	59991	61757	1766	+	Peptidoglycan DD-metalloendopeptidase family protein	
75	62089	62538	449	+	Hypothetical protein	
76	63205	63831	626	-	Hypothetical protein	
77	64200	65207	1007	-	IS481 family transposase	
78	65445	65747	302	+	Transposase	
79	65744	66619	875	+	IS3 family transposase	
80	66962	68614	1652	+	Putative ABC transporter permease protein	

81	68611	69843	1232	+	<i>Tetracycline-resistance determinant TetV</i>	
82	70656	72869	2213	+	<i>Hypothetical protein</i>	6
83	72871	73176	305	+	<i>Hypothetical protein</i>	
84	73280	74623	1343	+	<i>Hypothetical protein</i>	
85	74657	75052	395	+	<i>Hypothetical protein</i>	
86	75256	75486	230	+	<i>Hypothetical protein</i>	
87	75952	76569	617	+	<i>Hypothetical protein</i>	
88	76862	77818	956	+	<i>Hypothetical protein</i>	6`
89	77966	78205	239	-	<i>Hypothetical protein</i>	
90	78420	78767	347	-	<i>Hypothetical protein</i>	
91	78794	79804	1010	-	<i>IS481 family transposase</i>	
92	80004	80183	179	-	<i>Hypothetical protein</i>	
93	81293	82072	779	-	<i>Hypothetical protein</i>	

Además de los genes en común necesarios para las funciones plasmídicas, estos tres elementos albergan, principalmente en sus regiones variables, genes accesorios relacionados con la resistencia a las condiciones ambientales extremas que enfrentan sus huéspedes. En el caso de pLMA1, se encontró que el *ORF* 35 codifica una proteína de estrés universal, miembro de una superfamilia que abarca un grupo antiguo y conservado de proteínas que se encuentran ampliamente distribuidas entre los organismos. Aunque las funciones exactas de estas proteínas siguen siendo un enigma y en algunos casos parecen estar vinculadas a la resistencia a agentes que dañan el ADN y a desacopladores respiratorios (Kvint y col., 2003), Havis y col. (2019), recientemente han reportado que una proteína de este tipo (UspA616) es fundamental para la supervivencia y la regulación del fenotipo de latencia de *Micrococcus luteus* en condiciones de inanición e hipoxia. También se anotaron genes involucrados en la reducción de la concentración de fluoruro dentro la célula (*ORF* 37, 38 y 87), y en el transporte y resistencia a otros compuestos (*ORF* 42, 47, 48, 49, 75, 103, 113 y 117). El análisis de la secuencia de pLMA1 reveló, además, un gen potencial de resistencia a la eritromicina correspondiente al *ORF* 110 (23S ribosomal RNA methyltransferase *Erm*). De

hecho, Dib y col. (2010) obtuvieron previamente evidencia experimental de dicha resistencia codificada por el plásmido a partir del curado del mismo en *Micrococcus* sp. A1.

El plásmido pLMA7 también contiene información genética que puede ser favorable para la supervivencia y el crecimiento del huésped en el entorno hostil. En su secuencia se encontró un regulador transcripcional de la familia ArsR y un sistema de salida de cobalto-zinc-cadmio (*ORF* 27 y 28) y otro de manganeso (*ORF* 35 y 36), potencialmente involucrados en hacer frente a la intoxicación por metales pesados. Además, se identificó un gen relacionado con la resistencia a tetraciclina (*ORF* 81). Asimismo, el plásmido pJD12 también codifica un regulador transcripcional de la familia ArsR y un sistema de salida de cobalto-zinc-cadmio putativo (*ORF* 14 y 15). Es más, estos *ORFs* se encuentran comprendidos en una región que es compartida por ambos plásmidos con una similitud de secuencia del 99 % (Figura II.3).

Al comparar la similitud de secuencia entre pLMA1 y pLMA7, se encontraron 12 regiones conservadas (89 – 99 % de identidad), es decir el doble de aquellas compartidas por los tres plásmidos (Figura II.8). Este hecho no resulta extraño debido a que, además de compartir el mismo huésped, se trata de plásmidos aislados de la misma laguna.

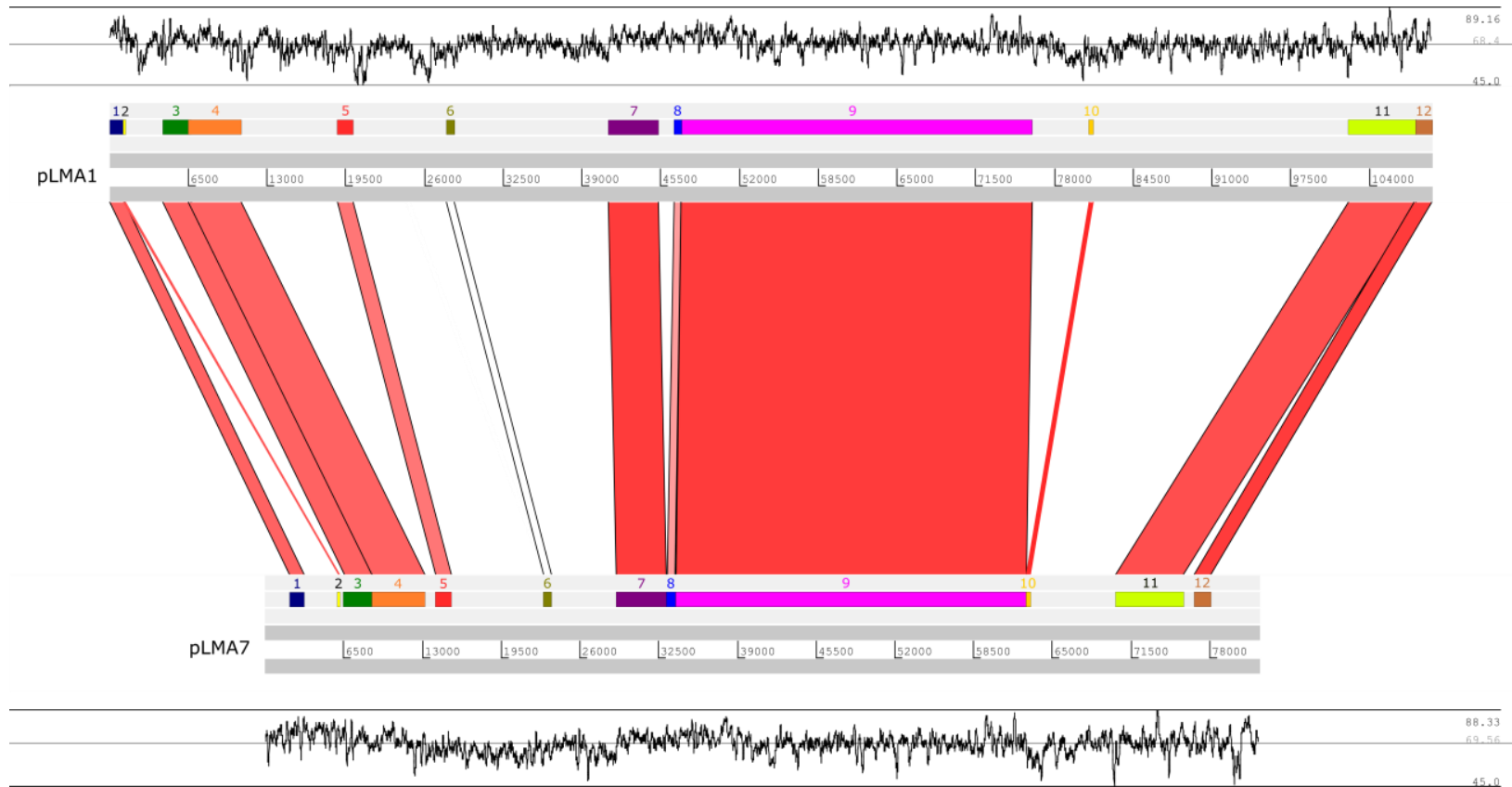


Figura II.8. Comparación de secuencia de los plásmidos lineales pLMA1 (109.112 pb) y pLMA7 (82.075 pb). Las regiones homólogas se muestran en rojo, las distintas intensidades se refieren a diferentes valores de similitud, a mayor intensidad mayor identidad. Se estableció como punto de corte una identidad mayor al 89 % y un mínimo de 157 pares de bases. Las regiones conservadas en ambos plásmidos se grafican en bloques de distintos colores (1-12). Por encima y por debajo de la secuencia de cada plásmido se grafica el contenido de GC, que varía del 45 al 89,16 % con un promedio del 68,40 % para pLMA1 y del 45 al 88,33 % con un promedio del 69,56 % para pLMA7.

3.3. Telómeros y proteínas terminales (TPs)

Existen dos tipos de plásmidos lineales que presentan diferencias sustanciales en sus estructuras, los llamados elementos “*hairpin*”, que terminan en horquillas por la unión covalente de las dos hebras simples de ADN en sus extremos (Kikuchi y col., 1985); y los que entran dentro del grupo de los invertrones, que se caracterizan por tener proteínas terminales (TPs) unidas covalentemente al extremo 5' (Sakaguchi, 1990). Ambos presentan repeticiones invertidas terminales (TIRs) (Figura II.9).

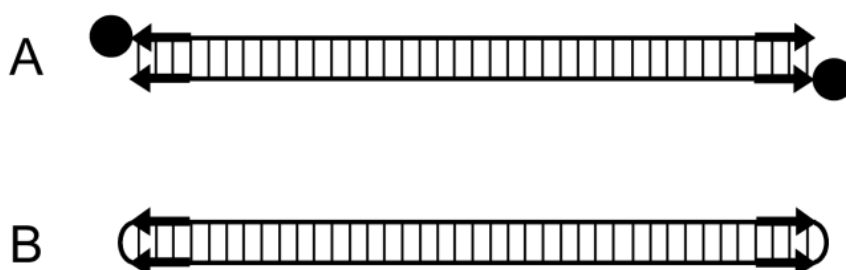


Figura II.9. Estructura de elementos lineales con proteínas terminales (TPs) unidas covalentemente al extremo 5' (A) y elementos “*hairpin*” (B). Las TPs y TIRs están representadas por los círculos negros y las flechas, respectivamente. Imagen tomada de Wagenknecht y col. (2020).

Los plásmidos lineales descriptos en actinobacterias son del tipo invertron. Si bien el género *Micrococcus* aún no ha sido muy estudiado, las primeras TPs descriptas en *Streptomyces* indican tamaño (184-185 aminoácidos) y secuencia muy conservada (Bao y Cohen, 2001). Los genes que codifican estas proteínas (*tpg*, *terminal protein gen*) suelen estar localizados en las regiones teloméricas, las cuales no fueron secuenciadas para pLMA1 y pLMA7 (Dib y col., 2018). En el caso de pJD12, un protocolo de restricción, auto-ligación y amplificación por PCR permitió obtener la secuencia de los telómeros plasmídicos (Dib y col., 2015). Sin embargo, tampoco se encontraron ORFs con homología a genes *tpg*, lo cual podría deberse a

la falta de información sobre *TPs* de replicones lineales en *Micrococcus*. Asimismo, se sabe que los plásmidos lineales más pequeños tienden a carecer de un homólogo de *tpg*, y pueden estar cubiertos por las *TPs* codificadas por plásmidos más grandes o por el cromosoma del huésped (Bao y Cohen, 2001).

3.4. Replicación y mantención

En la mayoría de los replicones lineales del tipo invertron, la replicación consiste en la síntesis continua de la molécula hija por desplazamiento de cadena, en el cual las *TPs* actúan como cebadores. Este mecanismo fue descrito en bacteriófagos y adenovirus (Salas, 1991), y en elementos extra-cromosómicos presentes en hongos filamentosos y levaduras (Klassen y Meinhardt, 2007). En actinobacterias, se reportó un mecanismo diferente para elementos lineales del género *Streptomyces*, el cual consiste en una replicación bidireccional con un origen de replicación interno (Chen, 2007). En *Micrococcus*, el mecanismo aún se desconoce, sin embargo, en los tres megaplásmidos analizados, se detectaron *ORFs* con homología a genes *rep* e iterones putativos a gran distancia (Tabla II.6). Mientras que los genes *rep* se ubicaron en el bloque conservado N° 5 en los tres plásmidos, los iterones lo hicieron en el bloque N° 1. Esto difiere en gran medida de lo reportado para los plásmidos lineales de *Streptomyces* más estudiados (pSCL1, pSLA2, SCP1, SLP2), donde los genes *rep* se encuentran adyacentes a las secuencias de iterones (Wu y Roy, 1993; Chang y col., 1996; Redenbach y col., 1999; Huang y col., 2003). Incluso Zhang y col. (2009) advirtieron la existencia de una variedad inesperada de componentes, posiciones y combinaciones de *loci* de replicación entre plásmidos lineales de *Streptomyces*, como los plásmidos pRL4 y pFRL2 donde los genes *rep* se encuentran adyacentes a secuencias no codificantes fundamentales para la replicación; o el plásmido pRL2 que contiene un *loci* discontinuo, con genes no

esenciales entre los iterones y los dos genes implicados en la replicación. Sin embargo, esta es la primera vez que se observan distancias tan grandes (mayores a 40 kb) entre los componentes implicados en la replicación de un plásmido. Además, como en algunos casos se encontró más de un *loci* de replicación en el mismo plásmido (pSLA2-L, Zhang y col. 2009), se necesitan ensayos adicionales para confirmar la funcionalidad de estos elementos en el proceso. De todos modos, el hallazgo podría contribuir como una primera aproximación a la comprensión del mecanismo de replicación de plásmidos lineales en el género *Micrococcus*; y sugiere, además, que dichos plásmidos comparten mecanismos de replicación y que sus componentes son elementos individuales que pueden separarse o combinarse durante la evolución.

El mantenimiento estable de los plásmidos en las bacterias es impulsado activamente por mecanismos de partición. Estos sistemas son responsables del posicionamiento de los plásmidos dentro de la célula para garantizar una segregación adecuada después de la replicación. En general, la partición es el determinante más importante en la estabilidad de los plásmidos de bajo número de copias, mientras que en los de alto número de copias no hay sistemas de partición codificados típicamente porque la segregación aleatoria es suficiente para mantenerlos (Baxter y Funnell, 2015).

Aunque poco se conoce sobre la partición de plásmidos lineales en actinobacterias, se sabe que los sistemas en *Streptomyces* cuentan con dos genes adyacentes, *parA* y *parB*. Mientras que ParB se encarga del reconocimiento de una secuencia de ADN palindrómica específica (*parS*) en los replicones y se une formando un complejo de partición; ParA, con actividad ATPasa, forma estructuras parecidas a un citoesqueleto y es responsable de segregar los dos replicones hijos (Chen, 2007). En los tres megaplásmidos de *Micrococcus* aquí analizados se encontraron *ORFs* con homología a genes *parA*, en cercanía a los genes *rep* (Tabla II.3, II.4 y

II.5; Figura II.5, II.6 y II.7). Sin embargo, no se detectaron secuencias homólogas a genes *parB* en ningún caso. Es de resaltar que también se encontraron genes *parA* solitarios en plásmidos lineales de otros huéspedes actinomicetos, como por ejemplo pCLP en *Mycobacterium celatum* (Le Dantec y col., 2001), pRHL3 en *Rhodococcus* sp. RHA1 (Warren y col., 2004), pBD2 en *Rhodococcus erythropolis* BD2 (Stecker y col., 2003) y pREL1 en *R. erythropolis* PR4 (Sekine y col., 2006).

Tabla II.6. Ubicación de genes de replicación plasmídica e iterones putativos en pLMA1, pJD12 y pLMA7.

pLMA1					
Gen rep		Iterones			
Nº ORF	Posición	Nº	Longitud (pb)	Posición	Secuencia
77	64112..65263	1	29	6499..6527	5'-GGAAGCTCCGCCC G CGCAGGGATGAGCCA-3' *
		2	29	6560..6588	5'-GGAAGCCCCGCGCACGCAGGGATGAGCCC-3'
		3	29	6621..6649	5'-GGAAGCCCCGCGCACGCAGGGATGAGCCC-3'
		4	29	6682..6710	5'-GGAAGCCCCGCGCACGCAGGGATGAGCCC-3'
		5	29	6743..6771	5'-GGAAGCCCCGCGCACGCAGGGATGAGCCC-3'
pJD12					
Gen rep		Iterones			
Nº ORF	Posición	Nº	Longitud (pb)	Posición	Secuencia
58	51869..52888	1	29	774..802	5'-GGAAGCCCCGCGCACGCAGGGATGAGCCA-3' *
		2	29	835..863	5'-GGAAGCCCCGCGCACGCAGGGATGAGCCC-3' *
		3	29	896..924	5'-GGAGGCCCCGCGCACGCAGGGATGAGCCC-3'
		4	29	957..985	5'-GGAGGCCCCGCGCACGCAGGGATGAGCCC-3'
		5	29	1018..1046	5'-GGAGGCCCCGCGCACGCAGGGATGAGCCC-3'
pLMA7					
Gen rep		Iterones			
Nº ORF	Posición	Nº	Longitud (pb)	Posición	Secuencia
66	50979..51980	1	29	8651..8679	5'-GGAAGCTCCGCCC G CGCAGGGATGAGCCA-3' *
		2	29	8712..8740	5'-GGAAGCCCCGCGCACGCAGGGATGAGCCC-3'
		3	29	8773..8801	5'-GGAAGCCCCGCGCACGCAGGGATGAGCCA-3' *
		4	29	8834..8862	5'-GGAAGCCCCGCGCACGCAGGGATGAGCCC-3'
		5	29	8895..8923	5'-GGAAGCCCCGCGCACGCAGGGATGAGCCC-3'
		6	29	8956..8984	5'-GGAAGCCCCGCGCACGCAGGGATGAGCCC-3'
		7	29	9017..9045	5'-GGAAGCTCCGCGCACGCAGGGATGAGCCC-3' *
		8	29	9078..9106	5'-GGAAGCCCCGCGCACGCAGGGATGAGCCC-3'

*Las bases resaltadas en negrita corresponden a cambios respecto a la secuencia consenso

3.5. Conjugación

En los plásmidos circulares, el mecanismo de conjugación se encuentra ampliamente estudiado. La transferencia conyugal se inicia en un sitio específico de la secuencia de ADN plasmídico (*oriT*), donde la enzima relaxasa realiza un corte mono-hebra y permanece unida covalentemente al extremo 5' de la hebra mellada (Lanka y Wilkins, 1995). Esto inicia una replicación en círculo rodante alrededor de todo el plásmido y la hebra desplazada es conducida y translocada hacia la célula receptora por medio de un sistema de secreción de tipo IV (*T4SS*) (Schröder y Lanka, 2005). En pLMA1, pJD12 y pLMA7 se encontraron *ORFs* con homología a proteínas involucradas en la formación del pili y el sistema *T4SS* en la región conservada N° 5 (Tabla II.3, II.4 y II.5). Sin embargo, si el mismo mecanismo se iniciara en un *oriT* interno de un plásmido lineal, la síntesis de ADN procedería unidireccionalmente y solo se replicaría una parte de la molécula plasmídica. Es por eso que resta saber si estos genes son funcionales o si solo se trata de replicones con un antecesor circular común; o bien si existe un mecanismo, aún desconocido, por el cual el ADN plasmídico lineal es transferido por tal sistema (*T4SS*) pero en una forma de doble hebra. Aunque no hay evidencia experimental de este tipo de transferencia para plásmidos lineales de actinobacterias, el análisis de secuencia del plásmido pAL1 de *Arthrobacter nitroguajacolicus* Rü61a también reveló la presencia de genes putativos pertenecientes al sistema de secreción tipo IV (Parschat y col., 2007). En el género *Streptomyces* se ha descripto la transferencia de ADN doble hebra, pero para plásmidos circulares y a través de un sistema distinto que consiste en una única proteína traslocadora (TraB) y una secuencia plasmídica no codificante (Reuther y col., 2006; Thoma y Muth, 2012).

Por otra parte, en los tres plásmidos bajo estudio, se encontraron *ORFs* con homología a una proteína de transferencia conyugal multiespecie que podría estar implicada en el proceso de

conjugación (*ORF* 80 en pLMA1, 62 en pJD12 y 69 en pLMA7). Este tipo de proteínas también fueron encontradas en los plásmidos lineales pAL1 (113 kb) y pAP13 (89 kb) de *Arthrobacter nitroguajacolicus* R61a y *Brevibacterium* sp. Ap13, ambos géneros pertenecientes al orden de los Micrococcales (Parschat y col., 2007; Dib y col., 2013b). Otra alternativa sería que pLMA1, pJD12 y pLMA7 se transfieran mediante el modelo “*end first*” (“extremo primero”) descrito por Chen (1996) para replicones lineales en *Streptomyces*, donde la transferencia se inicia en un telómero y la síntesis del ADN lineal es cebada por una *TP*. Al igual que en el modelo de círculo rodante, la hebra desplazada es conducida hacia el receptor por una proteína unida covalentemente al extremo 5', pero en este caso se trata de la *TP* parental (Figura II.10). A pesar de todo lo expuesto, no hay que descartar la posibilidad de que estos plásmidos sean replicones compuestos y tengan más de un mecanismo de conjugación, como ocurre con el plásmido lineal SLP2 (50 kb) de *Streptomyces lividans* (Huang y col., 2003). Es así que el mecanismo de conjugación en el género *Micrococcus* aún está lejos de ser elucidado y se requieren ensayos experimentales que soporten las distintas hipótesis.

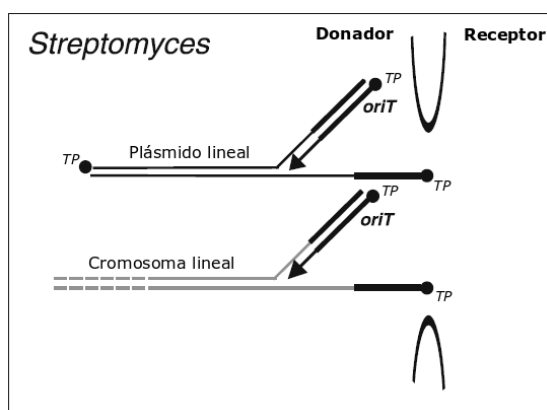


Figura II.10. Modelo “*end first*” de transferencia conjugugal de replicones lineales en *Streptomyces*. El telómero puede actuar como el origen de transferencia (*oriT*), la replicación se inicia cebada por una *TP* y la hebra 5' cubierta con la *TP* parental se desplaza y se transfiere. Imagen adaptada de Chen (1996).

4. Conclusiones

La gran identidad de secuencia de más de 30 kb en los tres megaplásmidos sugiere un vínculo evolutivo entre estos elementos. Si bien los plásmidos se identificaron en cepas hospedadoras diferentes, el género *Micrococcus* es el huésped común. En el caso de pLMA1 y pLMA7, las cepas se aislaron de la misma laguna, mientras que pJD12 se encontró en una cepa aislada de una laguna bastante distante (aprox. a 230 km). El hecho de encontrar una estructura plasmídica similar en ambientes distintos confirma el éxito de estos elementos para persistir y diseminarse, incluso en zonas alejadas. Al parecer tienen, por un lado, mecanismos en común que les permiten replicarse, mantenerse y conjugarse de forma eficiente, y por otro, una gran flexibilidad genética que se evidencia por la diversidad de sus regiones accesorias. De hecho, en las regiones conservadas de estos elementos se encontraron principalmente genes relacionados a las funciones plasmídicas esenciales: replicación, mantención y conjugación. Esta última está presente en la mayoría en los plásmidos lineales de actinobacterias y permite que dichos elementos se diseminen dentro de la misma especie y hacia organismos estrechamente relacionados. De esta forma, los actinomicetos poseen el potencial de compartir y beneficiarse de los atributos transmitidos por sus plásmidos.

Además de genes con funciones poco claras o desconocidas que podrían ser responsables de fenotipos importantes en su entorno natural, se encontraron genes relacionados con la resistencia a metales y antibióticos. Sin embargo, se necesitan más investigaciones para aclarar su papel en la fisiología de los plásmidos.

Por otra parte, la gran cantidad de *ORFs* relacionados a funciones de transposición y recombinación soporta la idea de que las secuencias de inserción y los transposones han jugado un papel central en la historia evolutiva de los tres plásmidos lineales. Además, las desviaciones del promedio del contenido de GC, a lo largo de la secuencia, evidencian la

naturaleza de mosaico de estos elementos y una notable plasticidad genética entre plásmidos lineales bacterianos de estos ambientes. Por lo tanto, se puede esperar que, en el entorno particular de las lagunas andinas, la *HGT* ocurra con frecuencia y proporcione a la comunidad microbiana nativa un acervo genético amplio y beneficioso que aumente la capacidad para resistir las condiciones ambientales extremas.

Aunque se han descrito plásmidos grandes en distintas actinobacterias, se sabe muy poco sobre la incidencia y diversidad de megaplásmidos lineales en actinobacterias de ambientes extremos. El análisis realizado en este capítulo permitió encontrar, de algún modo, un ancestro común de plásmidos lineales de *Micrococcus*. Por lo tanto, sería interesante estudiar otros plásmidos lineales de cepas del género aisladas de ambientes semejantes, y comprobar en que medida este ancestro común se encuentra diseminado.

PRODUCCIÓN CIENTÍFICA

Publicaciones

- **Perez MF**, Saona LA, Farías ME, Poehlein A, Meinhardt F, Daniel R, Dib JR. **2021**. Assessment of the plasmidome of an extremophilic microbial community from the Diamante Lake, Argentina. *Scientific Reports*, 11: 21459. doi: 10.1038/s41598-021-00753-1
- Noor ZZ, Rabiou Z, Sani MHM, Samad AFA, Kamaroddin MFA, **Perez MF**, Dib JR... & Zakaria ZA. **2021**. A Review of Bacterial Antibiotic Resistance Genes and Their Removal Strategies from Wastewater. *Current Pollution Reports*, 1-16. doi: 10.1007/s40726-021-00198-0
- **Perez MF**, Kurth D, Farías ME, Soria MN, Castillo Villamizar GA, Poehlein A, Daniel R, Dib JR. **2020**. First report on the plasmidome from a high-altitude lake of the Andean Puna. *Frontiers in Microbiology*, 11: 1343. doi: 10.3389/fmicb.2020.01343
- Wagenknecht M, **Perez MF**, Dib JR. **2020**. Linear Megaplasmsids Spreading in the Andean Resistome. In: Farías ME (Ed) Microbial Ecosystems in Central Andes Extreme Environments: Biofilms, Microbial Mats, Microbialites and Endoevaporites. *Springer Nature*. doi: 10.1007/978-3-030-36192-1
- Dib JR, **Perez MF**, Schuldes J, Poehlein A, Wagenknecht M, Farías ME, Meinhardt F, Daniel R. **2018**. Complete Genome Sequences of pLMA1 and pLMA7, Two Large Linear Plasmids of Micrococcus Strains Isolated from a High-Altitude Lake in Argentina. *Genome Announcements*, 6(8), e00010-18. doi: 10.1128/genomeA.00010-18

Presentaciones a Congresos

- **Perez MF**, Kurth D, Castillo Villamizar GA, Daniel R, Farías ME, Dib JR. “Análisis del plasmidoma de Puquio de Campo Naranja, una laguna de altura de la Puna Andina”. PRIMER CONGRESO ISME AMERICA LATINA. Universidad Técnica Federico Santa María. Septiembre, 2019. Valparaíso, Chile.
- **Perez MF**, Rasuk MC, Farías ME, Dib JR. “Haloarqueas aisladas de ecosistemas microbianos de ambientes extremos: resistencia al arsénico y perfil plasmídico”. JORNADAS DE JÓVENES INVESTIGADORES 2016. Secretaría de Ciencia, Arte e Innovación Tecnológica. Universidad Nacional de Tucumán. Tucumán, Argentina.

BIBLIOGRAFÍA

- Albarracín, V. H., Gärtner, W., y Farias, M. E. (2016). Forged under the Sun: Life and Art of Extremophiles from Andean Lakes. *Photochem. Photobiol.* 92, 14-28. doi:10.1111/php.12555.
- Albarracín, V. H., Kurth, D., Ordoñez, O. F., Belfiore, C., Luccini, E., ... Farías, M. E. (2015). High-up: A remote reservoir of microbial extremophiles in central Andean Wetlands. *Front. Microbiol.* 6, 1404. doi:10.3389/FMICB.2015.01404/FULL.
- Albarracín, V. H., Pathak, G. P., Douki, T., Cadet, J., Borsarelli, C. D., ... Farias, M. E. (2012). Extremophilic *Acinetobacter* strains from high-altitude lakes in Argentinean Puna: remarkable UV-B resistance and efficient DNA damage repair. *Orig. Life Evol. Biosph.* 42, 201-221. doi:10.1007/s11084-012-9276-3.
- Allwood, A. C., Walter, M. R., Kamber, B. S., Marshall, C. P., y Burch, I. W. (2006). Stromatolite reef from the Early Archaean era of Australia. *Nature* 441, 714-718. doi:10.1038/nature04764.
- Alonso-Reyes, D. G., Farias, M. E., y Albarracín, V. H. (2020). Uncovering cryptochrome/photolyase gene diversity in aquatic microbiomes exposed to diverse UV-B regimes. *Aquat. Microb. Ecol.* 85, 141-154. doi: 10.3354/ame01947.
- Alonso-Reyes, D. G., Galván, F. S., Portero, L. R., Alvarado, N. N., Farías, M. E., ... Albarracín, V. H. (2021). Genomic insights into an andean multiresistant soil actinobacterium of biotechnological interest. *World J. Microbiol. Biotechnol.* 37, 1-16. doi: 10.1007/s11274-021-03129-9.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., y Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410. doi:10.1016/S0022-2836(05)80360-2.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., ... Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402. doi:10.1093/nar/25.17.3389.
- Anantharaman, V., y Aravind, L. (2003). New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. *Genome Biol.* 4, R81. doi:10.1186/GB-2003-4-12-R81.
- Anderson, I., Tindall, B. J., Rohde, M., Lucas, S., Han, J., ... Ivanova, N. (2012). Complete genome sequence of *Halopiger xanaduensis* type strain (SH-6 T). *Stand. Genomic Sci.* 6, 31-42. doi:10.4056/sigs.2505605.
- Andres, J., y Bertin, P. N. (2016). The microbial genomics of arsenic. *FEMS Microbiol. Rev.* 40, 299-322. doi:10.1093/femsre/fuv050.
- Andrews, S. (2010). FastQC - A quality control tool for high throughput sequence data. *Babraham Bioinforma.*, <http://www.bioinformatics.babraham.ac.uk/projects/>.

- Arcus, V. L., McKenzie, J. L., Robson, J., y Cook, G. M. (2011). The PIN-domain ribonucleases and the prokaryotic VapBC toxin–antitoxin array. *Protein Eng. Des. Sel.* 24, 33-40. doi:10.1093/PROTEIN/GZQ081.
- Arredondo-Alonso, S., Willems, R. J., van Schaik, W., y Schürch, A. C. (2017). On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb. Genomics* 3. doi:10.1099/mgen.0.000128.
- Aubry, S., Kelly, S., Kümpers, B. M. C., Smith-Unna, R. D., y Hibberd, J. M. (2014). Deep Evolutionary Comparison of Gene Expression Identifies Parallel Recruitment of Trans-Factors in Two Independent Origins of C4 Photosynthesis. *PLOS Genet.* 10, e1004365. doi:10.1371/JOURNAL.PGEN.1004365.
- Baggi, G., Barbieri, P., Galli, E., y Tollari, S. (1987). Isolation of a *Pseudomonas stutzeri* strain that degrades o-xylene. *Appl. Environ. Microbiol.* 53, 2129-2132. doi:10.1128/aem.53.9.2129-2132.1987.
- Baker-Austin, C., Wright, M. S., Stepanauskas, R., y McArthur, J. V. (2006). Co-selection of antibiotic and metal resistance. *Trends Microbiol.* 14, 176-182. doi:10.1016/j.tim.2006.02.006.
- Baker, S., Hardy, J., Sanderson, K. E., Quail, M., Goodhead, I., ... Dougan, G. (2007). A novel linear plasmid mediates flagellar variation in *Salmonella* Typhi. *PLoS Pathog.* 3, e59. doi:10.1371/journal.ppat.0030059.
- Bale, M. J., Day, M. J., y Fry, J. C. (1988). Novel method for studying plasmid transfer in undisturbed river epilithon. *Appl. Environ. Microbiol.* 54, 2756-2758. doi:10.1128/AEM.54.11.2756-2758.1988.
- Baliga, N. S., Bonneau, R., Facciotti, M. T., Pan, M., Glusman, G., ... Ng, W. V. (2004). Genome sequence of *Haloarcula marismortui*: A halophilic archaeon from the Dead Sea. *Genome Res.* 14, 2221-2234. doi:10.1101/gr.2700304.
- Bandyopadhyay, P., Byrne, B., Chan, Y., Swanson, M. S., y Steinman, H. M. (2003). *Legionella pneumophila* catalase-peroxidases are required for proper trafficking and growth in primary macrophages. *Infect. Immun.* 71, 4526-4535. doi:10.1128/IAI.71.8.4526-4535.2003.
- Bánfalvi, Z., Sakanyan, V., Koncz, C., Kiss, A., Dusha, I., y Kondorosi, A. (1981). Location of nodulation and nitrogen fixation genes on a high molecular weight plasmid of *R. meliloti*. *MGG Mol. Gen. Genet.* 184, 318-325. doi:10.1007/BF00272925.
- Bao, K., y Cohen, S. N. (2001). Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. *Genes Dev.* 15, 1518-1527. doi:10.1101/gad.896201.
- Barbieri, P., Galassi, G., y Galli, E. (1989). Plasmid-encoded mercury resistance in a *Pseudomonas stutzeri* strain that degrades o-xylene. *FEMS Microbiol. Lett.* 62, 375-383. doi:10.1111/j.1574-6968.1989.tb03393.x.

- Basta, T., Keck, A., Klein, J., y Stolz, A. (2004). Detection and characterization of conjugative degradative plasmids in xenobiotic-degrading *Sphingomonas* strains. *J. Bacteriol.* 186, 3862-3872. doi:10.1128/JB.186.12.3862-3872.2004.
- Bateman, A., Coghill, P., y Finn, R. D. (2010). DUFs: families in search of function. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 66, 1148-1152. doi:10.1107/S1744309110001685.
- Baxter, J. C., y Funnell, B. E. (2015). «Plasmid Partition Mechanisms», en *Plasmids: Biology and Impact in Biotechnology and Discovery*, eds. M. E. Tolmasky y J. C. Alonso (ASM Press), 133-155. doi:10.1128/9781555818982.CH8.
- Bébian, M., Lagniel, G., Garin, J., Touati, D., Verméglio, A., y Labarre, J. (2002). Involvement of superoxide dismutases in the response of *Escherichia coli* to selenium oxides. *J. Bacteriol.* 184, 1556-1564. doi:10.1128/JB.184.6.1556-1564.2002.
- Beeler, S. R., Gomez, F. J., y Bradley, A. S. (2020). Controls of extreme isotopic enrichment in modern microbialites and associated abiogenic carbonates. *Geochim. Cosmochim. Acta* 269, 136-149. doi:10.1016/J.GCA.2019.10.022.
- Ben Fekih, I., Zhang, C., Li, Y. P., Zhao, Y., Alwathnani, H. A., ... Cervantes, C. (2018). Distribution of arsenic resistance genes in prokaryotes. *Front. Microbiol.* 9, 2473. doi:10.3389/fmicb.2018.02473.
- Bengoechea, J. A., Najdenski, H., y Skurnik, M. (2004). Lipopolysaccharide O antigen status of *Yersinia enterocolitica* O:8 is essential for virulence and absence of O antigen affects the expression of other *Yersinia* virulence factors. *Mol. Microbiol.* 52, 451-469. doi:10.1111/j.1365-2958.2004.03987.x.
- Bentley, S. D., Brown, S., Murphy, L. D., Harris, D. E., Quail, M. A., ... Chater, K. F. (2004). SCP1, a 356 023 bp linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* 51, 1615-1628. doi:10.1111/j.1365-2958.2003.03949.x.
- Bequer Urbano, S., Albarracín, V. H., Ordoñez, O. F., Farías, M. E., y Alvarez, H. M. (2013). Lipid storage in high-altitude Andean Lakes extremophiles and its mobilization under stress conditions in *Rhodococcus* sp. A5, a UV-resistant actinobacterium. *Extremophiles* 17, 217-227. doi:10.1007/s00792-012-0508-2.
- Biebl, H., Allgaier, M., Tindall, B. J., Koblizek, M., Lünsdorf, H., ... Wagner-Döbler, I. (2005). *Dinoroseobacter shibae* gen. nov., sp. nov., a new aerobic phototrophic bacterium isolated from dinoflagellates. *Int. J. Syst. Evol. Microbiol.* 55, 1089-1096. doi:10.1099/ij.s.0.63511-0.
- Binh, C., Heuer, H., Kaupenjohann, M., y Smalla, K. (2008). Piggery manure used for soil fertilization is a reservoir for transferable antibiotic resistance plasmids. *FEMS Microbiol. Ecol.* 66, 25-37. doi:10.1111/j.1574-6941.2008.00526.x.

- Boidi, F. J., Mlewski, E. C., Gomez, F. J., y Gérard, E. (2020). «Characterization of Microbialites and Microbial Mats of the Laguna Negra Hypersaline Lake (Puna of Catamarca, Argentina)», en *Microbial Ecosystems in Central Andes Extreme Environments*, ed. M. E. Farías (Springer International Publishing), 183-203. doi:10.1007/978-3-030-36192-1_13.
- Bolger, A. M., Lohse, M., y Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120. doi:10.1093/bioinformatics/btu170.
- Bolhuis, H., Te Poele, E. M., y Rodriguez-Valera, F. (2004). Isolation and cultivation of Walsby's square archaeon. *Environ. Microbiol.* 6, 1287-1291. doi:10.1111/j.1462-2920.2004.00692.x.
- Brantl, S. (2014). Plasmid replication control by antisense RNAs. *Microbiol. Spectr.* 2. doi:10.1128/microbiolspec.plas-0001-2013.
- Burne, R. V., y Moore, L. S. (1987). Microbialites: Organosedimentary Deposits of Benthic Microbial Communities. *Palaos* 2, 241-254. doi:10.2307/3514674.
- Cao, J., Lai, Q., Yuan, J., y Shao, Z. (2015). Genomic and metabolic analysis of fluoranthene degradation pathway in *Celeribacter indicus* P73 T. *Sci. Rep.* 5, 1-12. doi:10.1038/srep07741.
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., ... Knight, R. (2010). QIIME allows analysis of high- throughput community sequencing data. *Nat. Publ. Gr.* 7, 335-336. doi:10.1038/nmeth0510-335.
- Carattoli, A. (2013). Plasmids and the spread of resistance. *Int. J. Med. Microbiol.* 303, 298-304. doi:10.1016/j.ijmm.2013.02.001.
- Carter, M. Q., Chen, J., y Lory, S. (2010). The *Pseudomonas aeruginosa* pathogenicity island PAPI-1 is transferred via a novel type IV pilus. *J. Bacteriol.* 192, 3249-3258. doi:10.1128/JB.00041-10.
- Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., y Parkhill, J. (2005). ACT: The Artemis comparison tool. *Bioinformatics* 21, 3422-3423. doi:10.1093/bioinformatics/bti553.
- Carver, T., Thomson, N., Bleasby, A., Berriman, M., y Parkhill, J. (2009). DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25, 119-120. doi:10.1093/bioinformatics/btn578.
- Castillo, A. M., Gutiérrez, M. C., Kamekura, M., Xue, Y., Ma, Y., ... Ventosa, A. (2006). *Halostagnicola larsenii* gen. nov., sp. nov., an extremely halophilic archaeon from a saline lake in Inner Mongolia, China. *Int. J. Syst. Evol. Microbiol.* 56, 1519-1524. doi:10.1099/ijls.0.64286-0.
- Cavicchioli, R., Curmi, P. M. G., Saunders, N., y Thomas, T. (2003). Pathogenic archaea: do they exist? *BioEssays* 25, 1119-1128. doi:10.1002/bies.10354.
- Chang, P.-C., Kim, E.-S., y Cohen, S. N. (1996). *Streptomyces* linear plasmids that contain a phage-like, centrally located, replication origin. *Mol. Microbiol.* 22, 789-800. doi:10.1046/J.1365-

- 2958.1996.01526.X.
- Chen, C. (1996). Complications and implications of linear bacterial chromosomes. *Trends Genet.* 12, 192-196. doi:10.1016/0168-9525(96)30014-0.
- Chen, C. W. (2007). «*Streptomyces* Linear Plasmids: Replication and Telomeres», en *Microbial Linear Plasmids*, eds. R. Meinhardt y F. Klassen (Springer Berlin Heidelberg), 33-61. doi:10.1007/7171_2007_090.
- Chen, L., Hu, H., Chavda, K. D., Zhao, S., Liu, R., ... Kreiswirth, B. N. (2014). Complete sequence of a KPC-producing IncN multidrug-resistant plasmid from an epidemic *Escherichia coli* sequence type 131 strain in China. *Antimicrob. Agents Chemother.* 58, 2422-2425. doi:10.1128/AAC.02587-13.
- Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., ... Jin, Q. (2005). VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Res.* 33, D325-D328. doi:10.1093/nar/gki008.
- Cohen, S. N. (1976). Transposable genetic elements and plasmid evolution. *Nature* 263, 731-738. doi:10.1038/263731a0.
- Craig, N. (2020). *Mobile DNA III*. John Wiley & Sons, Ltd.
- Crossman, L. C. (2005). Plasmid replicons of *Rhizobium*. *Biochem. Soc. Trans.* 33, 157–158. doi:10.1042/bst0330157.
- Crozat, E., Fournes, F., Cornet, F., Hallet, B., y Rousseau, P. (2014). Resolution of Multimeric Forms of Circular Plasmids and Chromosomes. *Microbiol. Spectr.* 2. doi:10.1128/microbiolspec.PLAS-0025-2014.
- Dahlberg, C., Linberg, C., Torsvik, V. L., y 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. doi:10.1128/AEM.63.12.4692-4697.1997.
- Dalhus, B., Laerdahl, J. K., Backe, P. H., y Bjørås, M. (2009). DNA base repair - Recognition and initiation of catalysis. *FEMS Microbiol. Rev.* 33, 1044-1078. doi:10.1111/j.1574-6976.2009.00188.x.
- Daniel, R. (2004). The soil metagenome - A rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol.* 15, 199-204. doi:10.1016/j.copbio.2004.04.005.
- Dantas, G., Sommer, M. O. A., Oluwasegun, R. D., y Church, G. M. (2008). Bacteria subsisting on antibiotics. *Science* 320, 100-103. doi:10.1126/science.1155157.
- D’costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W. L., ... Wright, G. D. (2011). Antibiotic resistance is ancient. *Nature* 477, 457-461. doi:10.1038/nature10388.
- Del Solar, G., Acebo, P., y Espinosa, M. (1995). Replication control of plasmid pLS1: Efficient regulation of plasmid copy number is exerted by the combined action of two plasmid

- components, CopG and RNA II. *Mol. Microbiol.* 18, 913-924. doi:10.1111/j.1365-2958.1995.18050913.x.
- Del Solar, G., y Espinosa, M. (1992). The copy number of plasmid pLS1 is regulated by two trans-acting plasmid products: the antisense RNA II and the repressor protein, RepA. *Mol. Microbiol.* 6, 83-94. doi:10.1111/j.1365-2958.1992.tb00840.x.
- Del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M., y Díaz-Orejas, R. (1998). Replication and Control of Circular Bacterial Plasmids. *Microbiol. Mol. Biol. Rev.* 62, 434-464. doi:10.1128/mmbr.62.2.434-464.1998.
- Demain, A. L., y Blander, R. P. (1999). The β -lactam antibiotics: Past, present, and future. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* 75, 5-19. doi:10.1023/A:1001738823146.
- Demple, B., y Harrison, L. (1994). Repair of Oxidative Damage to DNA: Enzymology and Biology. *Annu. Rev. Biochem.* 63, 915-948. doi:10.1146/annurev.biochem.63.1.915.
- Dib, J., Motok, J., Zenoff, V. F., Ordoñez, O., y Farías, M. E. (2008). Occurrence of resistance to antibiotics, UV-B, and arsenic in bacteria isolated from extreme environments in high-altitude (above 4400 m) Andean wetlands. *Curr. Microbiol.* 56, 510-517. doi:10.1007/s00284-008-9103-2.
- Dib, J. R., Angelov, A., Liebl, W., Döbber, J., Voget, S., ... Daniel, R. (2015a). Complete genome sequence of the linear plasmid pJD12 hosted by *Micrococcus* sp. D12, isolated from a high-altitude volcanic lake in Argentina. *Genome Announc.* 3, e00627-15. doi:10.1128/genomea.00627-15.
- Dib, J. R., Liebl, W., Wagenknecht, M., Farías, M. E., y Meinhardt, F. (2013a). Extrachromosomal genetic elements in *Micrococcus*. *Appl. Microbiol. Biotechnol.* 97, 63-75. doi:10.1007/s00253-012-4539-5.
- Dib, J. R., Perez, M. F., Schuldes, J., Poehlein, A., Wagenknecht, M., ... Daniel, R. (2018). Complete genome sequences of pLMA1 and pLMA7, two large linear plasmids of *Micrococcus* strains isolated from a high-altitude lake in Argentina. *Genome Announc.* 6, e00010-18. doi:10.1128/genomea.00010-18.
- Dib, J. R., Schuldes, J., Thurmer, A., Farias, M. E., Daniel, R., y Meinhardt, F. (2013b). Complete genome sequence of pAP13, a large linear plasmid of a *Brevibacterium* strain isolated from a saline lake at 4,200 meters above sea level in Argentina. *Genome Announc.* 1, e00878-13. doi:10.1128/genomea.00878-13.
- Dib, J. R., Schuldes, J., Thurmer, A., Farias, M. E., Daniel, R., y Meinhardt, F. (2013c). First complete sequence of a giant linear plasmid from a *Micrococcus* strain isolated from an extremely high-altitude lake. *Genome Announc.* 1, e00885-13. doi:10.1128/genomea.00885-13.
- Dib, J. R., Wagenknecht, M., Farías, M. E., y Meinhardt, F. (2015b). Strategies and approaches in

- plasmidome studies-uncovering plasmid diversity disregarding of linear elements? *Front. Microbiol.* 6, 463. doi:10.3389/fmicb.2015.00463.
- Dib, J. R., Wagenknecht, M., Hill, R. T., Farías, M. E., y Meinhardt, F. (2010a). First report of linear megaplasms in the genus *Micrococcus*. *Plasmid* 63, 40-45. doi:10.1016/j.plasmid.2009.10.001.
- Dib, J. R., Wagenknecht, M., Hill, R. T., Farías, M. E., y Meinhardt, F. (2010b). Novel linear megaplasms from *Brevibacterium* sp. isolated from extreme environment. *J. Basic Microbiol.* 50, 280-284. doi:10.1002/jobm.200900332.
- Dib, J., Weiss, A., Neumann, A., Ordonez, O., Estevez, M., y Farias, M. (2009). Isolation of Bacteria from Remote High Altitude Andean Lakes Able to Grow in the Presence of Antibiotics. *Recent Pat. Antiinfect. Drug Discov.* 4, 66-76. doi:10.2174/157489109787236300.
- Ding, J. Y., Chiang, P. W., Hong, M. J., Dyll-Smith, M., y Tang, S. L. (2014). Complete genome sequence of the extremely halophilic archaeon *Haloarcula hispanica* strain N601. *Genome Announc.* 2, 178-192. doi:10.1128/genomeA.00178-14.
- Doddamani, H. P., y Ninnekar, H. Z. (2001). Biodegradation of carbaryl by a *Micrococcus* species. *Curr. Microbiol.* 43, 69-73. doi:10.1007/s002840010262.
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460-2461. doi:10.1093/bioinformatics/btq461.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., ... Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427-D432. doi:10.1093/nar/gky995.
- Fan, Y., Dai, Y., Cheng, Q., Zhang, G., Zhang, D., ... Qin, Z. (2012). A self-ligation method for PCR-sequencing the telomeres of *Streptomyces* and *Mycobacterium* linear replicons. *J. Microbiol. Methods* 90, 105-107. doi:10.1016/J.MIMET.2012.04.012.
- Farías, M. (2020). *Microbial Ecosystems in Central Andes Extreme Environments: Biofilms, Microbial Mats, Microbialites and Endoevaporites*. Springer Nature.
- Farias, M. E. (2018). «Ecosistemas microbianos de la Puna: El inmenso valor de lo diminuto», en *Serie conservación de la naturaleza 24: La Puna argentina. Naturaleza y cultura* (Fundación e Instituto Miguel Lillo), 246-268.
- Farías, M. E., Contreras, M., Rasuk, M. C., Kurth, D., Flores, M. R., ... Visscher, P. T. (2014). Characterization of bacterial diversity associated with microbial mats, gypsum evaporites and carbonate microbialites in thalassic wetlands: Tebenquiche and La Brava, Salar de Atacama, Chile. *Extremophiles* 18, 311-329. doi:10.1007/s00792-013-0617-6.
- Farías, M. E., Fernández-Zenoff, V., Flores, R., Ordóñez, O., y Estévez, C. (2009). Impact of solar radiation on bacterioplankton in Laguna Vilama, a hypersaline Andean lake (4650 m). *J. Geophys. Res. Biogeosciences* 114. doi:10.1029/2008JG000784.
- Farías, M. E., Poiré, D. G., Arrouy, M. J., y Albarracin, V. H. (2011). «Modern Stromatolite

- Ecosystems at Alkaline and Hypersaline High-Altitude Lakes in the Argentinean Puna», en *Stromatolites: interaction of microbes with sediments*, eds. V. Tewari y J. Seckbach (Springer, Dordrecht), 427-441. doi:10.1007/978-94-007-0397-1_19.
- Farías, M. E., Rascovan, N., Toneatti, D. M., Albarracín, V. H., Flores, M. R., ... Polerecky, L. (2013). The Discovery of Stromatolites Developing at 3570 m above Sea Level in a High-Altitude Volcanic Lake Socompa, Argentinean Andes. *PLoS One* 8, e53497. doi:10.1371/journal.pone.0053497.
- Farías, M., Villafañe, P., y Central, A. L. (2020). «Integral Prospection of Andean Microbial Ecosystem Project», en *Microbial Ecosystems in Central Andes Extreme Environments*, ed. M. E. Farías (Springer International Publishing), 245-260. doi:10.1007/978-3-030-36192-1_17.
- Fernández Zenoff, V., Siñeriz, F., y Farías, M. E. (2006). Diverse responses to UV-B radiation and repair mechanisms of bacteria isolated from high-altitude aquatic environments. *Appl. Environ. Microbiol.* 72, 7857-7863. doi:10.1128/AEM.01333-06.
- Filée, J., Siguier, P., y Chandler, M. (2007). Insertion Sequence Diversity in Archaea. *Microbiol. Mol. Biol. Rev.* 71, 121-157. doi:10.1128/MMBR.00031-06.
- Flores, M. R., Ordoñez, O. F., Maldonado, M. J., y Farías, M. E. (2009). Isolation of UV-B resistant bacteria from two high altitude Andean lakes (4,400 m) with saline and non saline conditions. *J. Gen. Appl. Microbiol.* 55, 447-458. doi:10.2323/jgam.55.447.
- Foster, T. J. (1983). Plasmid-determined resistance to antimicrobial drugs and toxic metal ions in bacteria. *Microbiol. Rev.* 47, 361-409. doi:10.1128/mmbr.47.3.361-409.1983.
- Frost, L. S., Leplae, R., Summers, A. O., y Toussaint, A. (2005). Mobile genetic elements: The agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722-732. doi:10.1038/nrmicro1235.
- Gantner, S., Andersson, A. F., Alonso-Sáez, L., y Bertilsson, S. (2011). Novel primers for 16S rRNA-based archaeal community analyses in environmental samples. *J. Microbiol. Methods* 84, 12-18. doi:10.1016/j.mimet.2010.10.001.
- Garcillán-Barcia, M. P., y de la Cruz, F. (2013). Ordering the bestiary of genetic elements transmissible by conjugation. *Mob. Genet. Elements* 3, e24263. doi:10.4161/mge.24263.
- Garcillán-Barcia, M. P., Francia, M. V., y De La Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* 33, 657-687. doi:10.1111/j.1574-6976.2009.00168.x.
- Garcillán-Barcia, M. P., Redondo-Salvo, S., Vielva, L., y de la Cruz, F. (2020). MOBscan: Automated Annotation of MOB Relaxases. *Methods Mol. Biol.* 2075, 295-308. doi:10.1007/978-1-4939-9877-7_21.
- Ghosal, D., Ghosh, S., Dutta, T. K., y Ahn, Y. (2016). Current state of knowledge in microbial degradation of polycyclic aromatic hydrocarbons (PAHs): A review. *Front. Microbiol.* 7, 1369.

- doi:10.3389/fmicb.2016.01369.
- Gil, R., Sabater-Muñoz, B., Perez-Brocal, V., Silva, F. J., y Latorre, A. (2006). Plasmids in the aphid endosymbiont *Buchnera aphidicola* with the smallest genomes. A puzzling evolutionary story. *Gene* 370, 17-25. doi:10.1016/j.gene.2005.10.043.
- Gomez, F. J., Boidi, F. J., Mlewski, E. C., y Gérard, E. (2020). «The Carbonate System in Hypersaline Lakes: The Case of Laguna Negra (in the Puna Region of Catamarca, Argentina)», en *Microbial Ecosystems in Central Andes Extreme Environments*, ed. M. E. Farias (Springer International Publishing), 231-242. doi:10.1007/978-3-030-36192-1_16.
- Gomez, F. J., Kah, L. C., Bartley, J. K., y Astini, R. A. (2014). Microbialites in a high-altitude Andean lake: Multiple controls on carbonate precipitation and lamina accretion. *Palaios* 29, 233-249. doi:10.2110/PALO.2013.049.
- Guy, L., Roat Kultima, J., y Andersson, S. G. E. (2010). genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 26, 2334-2335. doi:10.1093/bioinformatics/btq413.
- Halary, S., Leigh, J. W., Cheaib, B., Lopez, P., y Baptiste, E. (2010). Network analyses structure genetic diversity in independent genetic worlds. *Proc. Natl. Acad. Sci.* 107, 127-132. doi:10.1073/pnas.0908978107.
- Hale, T. L. (1991). Genetic basis of virulence in *Shigella* species. *Microbiol. Rev.* 55, 206-224. doi:10.1128/membr.55.2.206-224.1991.
- Hansen, J. B., y Olsen, R. H. (1978). Isolation of large bacterial plasmids and characterization of the P2 incompatibility group plasmids pMG1 and pMG5. *J. Bacteriol.* 135, 227-238. doi:10.1128/jb.135.1.227-238.1978.
- Harriott, O. T., Huber, R., Stetter, K. O., Betts, P. W., y Noll', K. M. (1994). A Cryptic Miniplasmid from the Hyperthermophilic Bacterium *Thermotoga* sp. Strain RQ7. *J. Bacteriol.* 176, 2759-2762. doi:10.1128/jb.176.9.2759-2762.1994.
- Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., ... Eisen, J. A. (2010). The complete genome sequence of *Haloferax volcanii* DS2, a model Archaeon. *PLoS One* 5, e9605. doi:10.1371/journal.pone.0009605.
- Havis, S., Bodunrin, A., Rangel, J., Zimmerer, R., Murphy, J., ... Bark, S. J. (2019). A universal stress protein that controls bacterial stress survival in *Micrococcus luteus*. *J. Bacteriol.* 201, e00497-19. doi:10.1128/JB.00497-19.
- Hayakawa, T., Tanaka, T., Sakaguchi, K., Otake, N., y Yonehara, H. (1979). A linear plasmid-like DNA in *Streptomyces* sp. producing lankacidin group antibiotics. *J. Gen. Appl. Microbiol.* 25, 255-260. doi:10.2323/jgam.25.255.
- Hemme, C. L., Deng, Y., Gentry, T. J., Fields, M. W., Wu, L., ... Zhou, J. (2010). Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *ISME*

- J.* 4, 660-672. doi:10.1038/ismej.2009.154.
- Hernández-Arriaga, A. M., Chan, W. T., Espinosa, M., y Díaz-Orejas, R. (2015). «Conditional Activation of Toxin-Antitoxin Systems: Postsegregational Killing and Beyond», en *Plasmids: Biology and Impact in Biotechnology and Discovery*, eds. M. E. Tolmasky y J. C. Alonso (ASM Press), 175-192. doi:10.1128/microbiolspec.PLAS-0009-2013.
- Heuer, H., Abdo, Z., y Smalla, K. (2008). Patchy distribution of flexible genetic elements in bacterial populations mediates robustness to environmental uncertainty. *FEMS Microbiol. Ecol.* 65, 361-371. doi:10.1111/J.1574-6941.2008.00539.X.
- Heuer, H., Kopmann, C., Binh, C. T. T., Top, E. M., y Smalla, K. (2009). Spreading antibiotic resistance through spread manure: Characteristics of a novel plasmid type with low %G+C content. *Environ. Microbiol.* 11, 937-949. doi:10.1111/J.1462-2920.2008.01819.X.
- Hill, K. E., Weightman, A. J., y Fry, J. C. (1992). Isolation and screening of plasmids from the epilithon which mobilize recombinant plasmid pD10. *Appl. Environ. Microbiol.* 58, 1292-1300. doi:10.1128/AEM.58.4.1292-1300.1992.
- Hinnebusch, J., y Barbour, A. G. (1992). Linear- and circular-plasmid copy numbers in *Borrelia burgdorferi*. *J. Bacteriol.* 174, 5251-5257. doi:10.1128/jb.174.16.5251-5257.1992.
- Hishinuma, F., y Hirai, K. (1991). Genome organization of the linear plasmid, pSKL, isolated from *Saccharomyces kluyveri*. *MGG Mol. Gen. Genet.* 226, 97-106. doi:10.1007/BF00273592.
- Huang, C.-H., Chen, C.-Y., Tsai, H.-H., Chen, C., Lin, Y.-S., y Chen, C. W. (2003). Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. *Mol. Microbiol.* 47, 1563-1576. doi:10.1046/J.1365-2958.2003.03403.X.
- Hülter, N., Ilhan, J., Wein, T., Kadibalban, A. S., Hammerschmidt, K., y Dagan, T. (2017). An evolutionary perspective on plasmid lifestyle modes. *Curr. Opin. Microbiol.* 38, 74-80. doi:10.1016/j.mib.2017.05.001.
- Jacob, F., y Wollman, E. (1958). Les épisomes, éléments génétiques ajoutés. *Comptes Rendus Hebd. Des Seances L Acad. Des Sci.* 247, 154-156.
- Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., ... McArthur, A. G. (2017). CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573. doi:10.1093/nar/gkw1004.
- Johnson, T. J., y Nolan, L. K. (2009). Pathogenomics of the Virulence Plasmids of *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* 73, 750-774. doi:10.1128/mmbr.00015-09.
- Jones, B. V., y Marchesi, J. R. (2007). Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Nat. Methods* 4, 55-61. doi:10.1038/nmeth964.
- Jørgensen, T. S., Xu, Z., Hansen, M. A., Sørensen, S. J., y Hansen, L. H. (2014). Hundreds of circular novel plasmids and DNA elements identified in a rat cecum metatranscriptome. *PLoS One* 9, 1-9.

- doi:10.1371/journal.pone.0087924.
- Kado, C. I., y Liu, S. T. (1981). Rapid procedure for detection and isolation of large and small plasmids. *J. Bacteriol.* 145, 1365-1373. doi:10.1128/jb.145.3.1365-1373.1981.
- Kasting, J. F. (1991). Box models for the evolution of atmospheric oxygen: an update. *Glob. Planet. Change* 5, 125-131. doi:10.1016/0921-8181(91)90133-H.
- Kav, A. B., Benhar, I., y 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. doi:10.1016/J.MIMET.2013.09.008.
- Kav, A. B., Sasson, G., Jami, E., Doron-Faigenboim, A., Benhar, I., y Mizrahi, I. (2012). Insights into the bovine rumen plasmidome. *Proc. Natl. Acad. Sci.* 109, 5452-5457. doi:10.1073/pnas.1116410109.
- Khelaifia, S., y Drancourt, M. (2012). Susceptibility of archaea to antimicrobial agents: Applications to clinical microbiology. *Clin. Microbiol. Infect.* 18, 841-848. doi:10.1111/j.1469-0691.2012.03913.x.
- Kikuchi, Y., Hirai, K., Gunge, N., y Hishinuma, F. (1985). Hairpin plasmid--a novel linear DNA of perfect hairpin structure. *EMBO J.* 4, 1881-1886. doi:10.1002/j.1460-2075.1985.tb03864.x.
- Kim, S. R., y Komano, T. (1997). The plasmid R64 thin pilus identified as a type IV pilus. *J. Bacteriol.* 179, 3594-3603. doi:10.1128/jb.179.11.3594-3603.1997.
- Kirstahler, P., Teudt, F., Otani, S., Aarestrup, F. M., y Pamp, S. J. (2021). A Peek into the Plasmidome of Global Sewage. *mSystems* 6, e00283-21. doi:10.1128/MSYSTEMS.00283-21.
- Klassen, R., y Meinhardt, F. (2007). «Linear Protein-Primed Replicating Plasmids in Eukaryotic Microbes», en *Microbial Linear Plasmids*, eds. R. Meinhardt y F. Klassen (Springer Berlin Heidelberg), 187-226. doi:10.1007/7171_2007_095.
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., ... Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41, e1. doi:10.1093/nar/gks808.
- Kondorosi, A., Kondorosi, E., Pankhurst, C. E., Broughton, W. J., y Banfalvi, Z. (1982). Mobilization of a *Rhizobium meliloti* megaplasmid carrying nodulation and nitrogen fixation genes into other rhizobia and *Agrobacterium*. *MGG Mol. Gen. Genet.* 188, 433-439. doi:10.1007/BF00330045.
- Konieczny, I., Bury, K., Wawrzycka, A., y Wegrzyn, K. (2014). Iteron Plasmids. *Microbiol. Spectr.* 2. doi:10.1128/microbiolspec.PLAS-0026-2014.
- Koonin, E. V., y Wolf, Y. I. (2008). Genomics of bacteria and archaea: The emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36, 6688-6719. doi:10.1093/nar/gkn668.
- Kothari, A., Wu, Y. W., Chandonia, J. M., Charrier, M., Rajeev, L., ... Mukhopadhyay, A. (2019). Large circular plasmids from groundwater plasmidomes span multiple incompatibility groups

- and are enriched in multimetal resistance genes. *MBio* 10, e02899-18. doi:10.1128/mBio.02899-18.
- Kurth, D., Amadio, A., Ordoñez, O. F., Albarracín, V. H., Gärtner, W., y Farías, M. E. (2017). Arsenic metabolism in high altitude modern stromatolites revealed by metagenomic analysis. *Sci. Rep.* 7, 1-16. doi:10.1038/s41598-017-00896-0.
- Kurth, D., Belfiore, C., Gorriti, M. F., Cortez, N., Farias, M. E., y Albarracín, V. H. (2015). Genomic and proteomic evidences unravel the UV-resistome of the poly-extremophile *Acinetobacter* sp. Ver3. *Front. Microbiol.* 6, 328. doi:10.3389/FMICB.2015.00328.
- Kvint, K., Nachin, L., Diez, A., y Nyström, T. (2003). The bacterial universal stress protein: Function and regulation. *Curr. Opin. Microbiol.* 6, 140-145. doi:10.1016/S1369-5274(03)00025-0.
- Lai, Q., Cao, J., Yuan, J., Li, F., y Shao, Z. (2014). *Celeribacter indicus* sp. nov., a polycyclic aromatic hydrocarbon-degrading bacterium from deep-sea sediment and reclassification of *Huaishuia halophila* as *Celeribacter halophilus* comb. nov. *Int. J. Syst. Evol. Microbiol.* 64, 4160-4167. doi:10.1099/ij.s.0.069039-0.
- Langmead, B., y Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357-359. doi:10.1038/nmeth.1923.
- Lanka, E., y Wilkins, B. M. (1995). DNA Processing Reactions in Bacterial Conjugation. *Annu. Rev. Biochem.* 64, 141-169. doi:10.1146/annurev.bi.64.070195.001041.
- Lapaque, N., Moriyon, I., Moreno, E., y Gorvel, J. P. (2005). *Brucella lipopolysaccharide* acts as a virulence factor. *Curr. Opin. Microbiol.* 8, 60-66. doi:10.1016/J.MIB.2004.12.003.
- Le Dantec, C., Winter, N., Gicquel, B., Vincent, V., y Picardeau, M. (2001). Genomic sequence and transcriptional analysis of a 23-kilobase mycobacterial linear plasmid: Evidence for horizontal transfer and identification of plasmid maintenance systems. *J. Bacteriol.* 183, 2157-2164. doi:10.1128/JB.183.7.2157-2164.2001.
- Lederberg, J. (1952). Cell genetics and hereditary symbiosis. *Physiol. Rev.* 32, 403-430. doi:10.1152/physrev.1952.32.4.403.
- Leplae, R., Lima-Mendez, G., y Toussaint, A. (2010). ACLAME: A CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res.* 38, D57-D61. doi:10.1093/nar/gkp938.
- Lepp, P. W., Brinig, M. M., Ouverney, C. C., Palm, K., Armitage, G. C., y Relman, D. A. (2004). Methanogenic Archaea and human periodontal disease. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6176-6181. doi:10.1073/pnas.0308766101.
- Li, H., Li, P., Hua, T., Zhang, Y., Xiong, X., y Gong, Z. (2005). Bioremediation of contaminated surface water by immobilized *Micrococcus roseus*. *Environ. Technol.* 26, 931-940. doi:10.1080/09593332608618504.
- Li, L. L., Norman, A., Hansen, L. H., y Sørensen, S. J. (2012). Metamobilomics - expanding our

- knowledge on the pool of plasmid encoded traits in natural environments using high-throughput sequencing. *Clin. Microbiol. Infect.* 18, 5-7. doi:10.1111/j.1469-0691.2012.03862.x.
- Li, W., Jaroszewski, L., y Godzik, A. (2002). Sequence clustering strategies improve remote homology recognitions while reducing search times. *Protein Eng.* 15, 643-649. doi:10.1093/protein/15.8.643.
- Lilley, A., Bailey, M., Day, M., y Fry, J. (1996). Diversity of mercury resistance plasmids obtained by exogenous isolation from the bacteria of sugar beet in three successive years. *FEMS Microbiol. Ecol.* 20, 211-227. doi:10.1111/j.1574-6941.1996.tb00320.x.
- Lilly, J., y Camps, M. (2015). «Mechanisms of Theta Plasmid Replication», en *Plasmids: Biology and Impact in Biotechnology and Discovery*, eds. M. E. Tolmasky y J. C. Alonso (ASM Press), 33-44. doi:10.1128/9781555818982.ch3.
- Luo, H., y Gao, F. (2019). DoriC 10.0: An updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acids Res.* 47, D74-D77. doi:10.1093/nar/gky1014.
- Ma, Y., Paulsen, I. T., y Palenik, B. (2012). Analysis of two marine metagenomes reveals the diversity of plasmids in oceanic environments. *Environ. Microbiol.* 14, 453-466. doi:10.1111/j.1462-2920.2011.02633.x.
- Makarova, K. S., Wolf, Y. I., y Koonin, E. V. (2019). Towards functional characterization of archaeal genomic dark matter. *Biochem. Soc. Trans.* 47, 389-398. doi:10.1042/BST20180560.
- Malfatti, S., Tindall, B. J., Schneider, S., Fährnich, R., Lapidus, A., ... Chain, P. (2009). Complete genome sequence of *Halogeometricum borinquense* type strain (PR3T). *Stand. Genomic Sci.* 1, 150-158. doi:10.4056/sigs.23264.
- Mana-Capelli, S., Mandal, A. K., y Argüello, J. M. (2003). *Archaeoglobus fulgidus* CopB is a thermophilic Cu²⁺-ATPase: functional role of its histidine-rich-N-terminal metal binding domain. *J. Biol. Chem.* 278, 40534-40541. doi:10.1074/jbc.M306907200.
- Manikandan, M., Kannan, V., y Pašić, L. (2011). Extraction, purification and characterization of a protease from *Micrococcus* sp. VKMM 037. *Environ. Technol.* 32, 1487-1495. doi:10.1080/09593330.2010.540718.
- Manivasagan, P., Kang, K. H., Sivakumar, K., Li-Chan, E. C. Y., Oh, H. M., y Kim, S. K. (2014). Marine actinobacteria: An important source of bioactive natural products. *Environ. Toxicol. Pharmacol.* 38, 172-188. doi:10.1016/j.etap.2014.05.014.
- Marshall, K. C., Characklis, W. G., Filip, Z., Fletcher, M., Hirsch, P., ... Calleja, G. B. (2012). *Microbial Adhesion and Aggregation: Report of the Dahlem Workshop on Microbial Adhesion and Aggregation Berlin 1984, January 15-20*. Springer Science & Business Media.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.

- EMBnet.journal* 17, 10-12. doi:10.14806/ej.17.1.200.
- Maruyama, K., Akita, K., Naitou, C., Yoshida, M., y Kitamura, T. (2005). Purification and characterization of an esterase hydrolyzing monoalkyl phthalates from *Micrococcus* sp. YGJ1. *J. Biochem.* 137, 27-32. doi:10.1093/jb/mvi004.
- McHardy, A. C., y Rigoutsos, I. (2007). What's in the mix: phylogenetic classification of metagenome sequence samples. *Curr. Opin. Microbiol.* 10, 499-503. doi:10.1016/J.MIB.2007.08.004.
- Medema, M. H., Trefzer, A., Kovalchuk, A., Van Den Berg, M., Mü Ller, U., ... Takano, E. (2010). The Sequence of a 1.8-Mb Bacterial Linear Plasmid Reveals a Rich Evolutionary Reservoir of Secondary Metabolic Pathways. *Genome Biol. Evol.* 2, 212-224. doi:10.1093/gbe/evq013.
- Meinhardt, F., Kempken, F., Kämper, J., y Esser, K. (1990). Linear plasmids among eukaryotes: fundamentals and application. *Curr. Genet.* 17, 89-95. doi:10.1007/BF00312851.
- Meinhardt, F., Schaffrath, R., y Larsen, M. (1997). Microbial linear plasmids. *Appl. Microbiol. Biotechnol.* 47, 329-336. doi:10.1007/S002530050936.
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., ... Edwards, R. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9, 386. doi:10.1186/1471-2105-9-386.
- Miller, S. C., Porcella, S. F., Raffel, S. J., Schwan, T. G., y Barboura, A. G. (2013). Large linear plasmids of *Borrelia* species that cause relapsing fever. *J. Bacteriol.* 195, 3629-3639. doi:10.1128/JB.00347-13.
- Miranda, A. T., González, M. V., González, G., Vargas, E., Campos-García, J., y Cervantes, C. (2005). Involvement of DNA helicases in chromate resistance by *Pseudomonas aeruginosa* PAO1. *Mutat. Res. Mol. Mech. Mutagen.* 578, 202-209. doi:10.1016/j.mrfmmm.2005.05.018.
- Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., y Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41, e121. doi:10.1093/nar/gkt263.
- Moura, A., Soares, M., Pereira, C., Leitão, N., Henriques, I., y Correia, A. (2009). INTEGRALL: A database and search engine for integrons, integrases and gene cassettes. *Bioinformatics* 25, 1096-1098. doi:10.1093/bioinformatics/btp105.
- Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., ... DasSarma, S. (2000). Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12176-12181. doi:10.1073/pnas.190337797.
- Nisanian, M., Holladay, S. D., Karpuzoglu, E., Kerr, R. P., Williams, S. M., ... Gogal, R. M. (2014). Exposure of juvenile Leghorn chickens to lead acetate enhances antibiotic resistance in enteric bacterial flora. *Poult. Sci.* 93, 891-897. doi:10.3382/ps.2013-03600.
- Nishida, H. (2012). Comparative Analyses of Base Compositions, DNA Sizes, and Dinucleotide

- Frequency Profiles in Archaeal and Bacterial Chromosomes and Plasmids. *Int. J. Evol. Biol.* 2012, 342482. doi:10.1155/2012/342482.
- Norman, A., Hansen, L. H., y Sørensen, S. J. (2009). Conjugative plasmids: vessels of the communal gene pool. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 2275-2289. doi:10.1098/rstb.2009.0037.
- Norman, A., Riber, L., Luo, W., Li, L. L., Hansen, L. H., y Sørensen, S. J. (2014). An improved method for including upper size range plasmids in metamobilomes. *PLoS One* 9, e104405. doi:10.1371/journal.pone.0104405.
- Novick, R. P. (1987). Plasmid Incompatibility. *Microbiol. Rev.* 51, 381-395. doi:10.1128/mr.51.4.381-395.1987.
- Novick, R. P. (2003). Mobile genetic elements and bacterial toxins: The superantigen-encoding pathogenicity islands of *Staphylococcus aureus*. *Plasmid* 49, 93-105. doi:10.1016/S0147-619X(02)00157-9.
- Nurk, S., Meleshko, D., Korobeynikov, A., y Pevzner, P. A. (2017). MetaSPAdes: A new versatile metagenomic assembler. *Genome Res.* 27, 824-834. doi:10.1101/gr.213959.116.
- Okazaki, S., Noisangiam, R., Okubo, T., Kaneko, T., Oshima, K., ... Teatumroong, N. (2015). Genome analysis of a novel *Bradyrhizobium* sp. DOA9 carrying a symbiotic plasmid. *PLoS One* 10, e0117392. doi:10.1371/journal.pone.0117392.
- Ordoñez, O. F., Flores, M. R., Dib, J. R., Paz, A., y Farías, M. E. (2009). Extremophile culture collection from Andean lakes: extreme pristine environments that host a wide diversity of microorganisms with tolerance to UV radiation. *Microb. Ecol.* 58, 461-473. doi:10.1007/s00248-009-9527-7.
- Ordoñez, O. F., Rasuk, M. C., Soria, M. N., Contreras, M., y Farías, M. E. (2018). Haloarchaea from the Andean Puna: biological role in the energy metabolism of arsenic. *Microb. Ecol.* 76, 695-705. doi:10.1007/s00248-018-1159-3.
- Ou, J. T., y Anderson, T. F. (1970). Role of Pili in Bacterial Conjugation. *J. Bacteriol.* 102, 648-654. doi:10.1128/jb.102.3.648-654.1970.
- Page, R., y Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.* 12, 208-214. doi:10.1038/nchembio.2044.
- Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., y Larsson, D. G. J. (2014). BacMet: Antibacterial biocide and metal resistance genes database. *Nucleic Acids Res.* 42, D737-D743. doi:10.1093/nar/gkt1252.
- Parks, D. H., y Beiko, R. G. (2010). Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* 26, 715-721. doi:10.1093/bioinformatics/btq041.
- Parschat, K., Overhage, J., Strittmatter, A. W., Henne, A., Gottschalk, G., y Fetzner, S. (2007). Complete Nucleotide Sequence of the 113-Kilobase Linear Catabolic Plasmid pAL1 of

- Arthrobacter nitroguajacolicus* R61a and Transcriptional Analysis of Genes Involved in Quinaldine Degradation. *J. Bacteriol.* 189, 3855–3867. doi:10.1128/JB.00089-07.
- Peimbert, M., Alcaraz, L. D., Bonilla-Rosso, G., Olmedo-Alvarez, G., García-Oliva, F., ... Souza, V. (2012). Comparative metagenomics of two microbial mats at Cuatro Ciénegas Basin I: Ancient lessons on how to cope with an environment under severe nutrient stress. *Astrobiology* 12, 648-658. doi:10.1089/ast.2011.0694.
- Pfeiffer, F., Schuster, S. C., Broicher, A., Falb, M., Palm, P., ... Oesterhelt, D. (2008). Evolution in the laboratory: The genome of *Halobacterium salinarum* strain R1 compared to that of strain NRC-1. *Genomics* 91, 335-346. doi:10.1016/j.ygeno.2008.01.001.
- Poirel, L., Kieffer, N., Liassine, N., Thanh, D., y Nordmann, P. (2016). Plasmid-mediated carbapenem and colistin resistance in a clinical isolate of *Escherichia coli*. *Lancet Infect. Dis.* 16, 281. doi:10.1016/S1473-3099(16)00006-2.
- Prieto-Barajas, C. M., Valencia-Cantero, E., y Santoyo, G. (2018). Microbial mat ecosystems: Structure types, functional diversity, and biotechnological application. *Electron. J. Biotechnol.* 31, 48-56. doi:10.1016/J.EJBT.2017.11.001.
- Pring, D. R., Levings, C. S., Hut, W. W. L., y Timothy, D. H. (1977). Unique DNA associated with mitochondria in the «S»-type cytoplasm of male-sterile maize. *Proc. Natl. Acad. Sci.* 74, 2904-2908. doi:10.1073/pnas.74.7.2904.
- Rangseekaew, P., y Pathom-Aree, W. (2019). Cave actinobacteria as producers of bioactive metabolites. *Front. Microbiol.* 10, 387. doi:10.3389/fmicb.2019.00387.
- Rascovan, N., Maldonado, J., Vazquez, M. P., y Farías, M. E. (2016). Metagenomic study of red biofilms from Diamante Lake reveals ancient arsenic bioenergetics in haloarchaea. *ISME J.* 10, 299-309. doi:10.1038/ismej.2015.109.
- Rasuk, M. C., Fernández, A. B., Kurth, D., Contreras, M., Novoa, F., ... Farías, M. E. (2016). Bacterial Diversity in Microbial Mats and Sediments from the Atacama Desert. *Microb. Ecol.* 71, 44-56. doi:10.1007/s00248-015-0649-9.
- Rasuk, M. C., Kurth, D., Flores, M. R., Contreras, M., Novoa, F., ... Farias, M. E. (2014). Microbial Characterization of Microbial Ecosystems Associated to Evaporites Domes of Gypsum in Salar de Llamara in Atacama Desert. *Microb. Ecol.* 68, 483-494. doi:10.1007/s00248-014-0431-4.
- Ravel, J., Wellington, E. M. H., y Hill, R. T. (2000). Interspecific transfer of *Streptomyces* giant linear plasmids in sterile amended soil microcosms. *Appl. Environ. Microbiol.* 66, 529-534. doi:10.1128/AEM.66.2.529-534.2000.
- Redenbach, M., Bibb, M., Gust, B., Seitz, B., y Spychaj, A. (1999). The linear plasmid SCP1 of *Streptomyces coelicolor* A3(2) possesses a centrally located replication origin and shows significant homology to the transposon Tn4811. *Plasmid* 42, 174-185.

- doi:10.1006/PLAS.1999.1419.
- Reniero, D., Galli, E., y Barbieri, P. (1995). Cloning and comparison of mercury- and organomercurial-resistance determinants from a *Pseudomonas stutzeri* plasmid. *Gene* 166, 77-82. doi:10.1016/0378-1119(95)00546-4.
- Reuther, J., Gekeler, C., Tiffert, Y., Wohlleben, W., y Muth, G. (2006). Unique conjugation mechanism in mycelial streptomycetes: a DNA-binding ATPase translocates unprocessed plasmid DNA at the hyphal tip. *Mol. Microbiol.* 61, 436-446. doi:10.1111/J.1365-2958.2006.05258.X.
- Riadi, G., Medina-Moenne, C., y Holmes, D. S. (2012). TnpPred: A web service for the robust prediction of prokaryotic transposases. *Comp. Funct. Genomics* 2012, 678761. doi:10.1155/2012/678761.
- Roh, S. W., Nam, Y. Do, Nam, S. H., Choi, S. H., Park, H. S., y Bae, J. W. (2010). Complete genome sequence of *Halalkalicoccus jeotgali* B3T, an extremely halophilic archaeon. *J. Bacteriol.* 192, 4528-4529. doi:10.1128/JB.00663-10.
- Rothschild, L. J., Giver, L. J., White, M. R., y Mancinelli, R. L. (1994). Metabolic activity of microorganisms in evaporites. *J. Phycol.* 30, 431-438. doi:10.1111/J.0022-3646.1994.00431.X.
- Rozov, R., Brown Kav, A., Bogumil, D., Shterzer, N., Halperin, E., ... Shamir, R. (2016). Recycler: an algorithm for detecting plasmids from de novo assembly graphs. *Bioinformatics* 33, 475-482. doi:10.1093/bioinformatics/btw651.
- Ruiz-Masó, J. A., Machón, C., Bordanaba-Ruiseco, L., Espinosa, M., Coll, M., y Del Solar, G. (2015). «Plasmid Rolling-Circle Replication», en *Plasmids: Biology and Impact in Biotechnology and Discovery*, eds. M. E. Tolmasky y J. C. Alonso (ASM Press), 45-69. doi:10.1128/9781555818982.ch4.
- Ruvindy, R. (2014). Metagenomic Analyses of Modern Stromatolitic Mats from Hamelin Pool, Shark Bay, Western Australia.
- Ruvindy, R., White, R. A., Neilan, B. A., y Burns, B. P. (2016). Unravelling core microbial metabolisms in the hypersaline microbial mats of Shark Bay using high-throughput metagenomics. *ISME J.* 10, 183-196. doi:10.1038/ismej.2015.87.
- Sabaté, M., y Prats, G. (2002). Structure and function of integrons. *Enferm. Infecc. Microbiol. Clin.* 20, 341-345. doi:10.1016/S0213-005X(02)72813-9.
- Sahl, J. W., Pace, N. R., y Spear, J. R. (2008). Comparative molecular analysis of endoevaporitic microbial communities. *Appl. Environ. Microbiol.* 74, 6444-6446. doi:10.1128/AEM.00879-08.
- Sakaguchi, K. (1990). Invertrons, a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of adeno-type viruses. *Microbiol. Rev.* 54, 66-74. doi:10.1128/mr.54.1.66-74.1990.

- Salas, M. (1991). Protein-priming of DNA replication. *Annu. Rev. Biochem.* 60, 39-71. doi:10.1146/annurev.bi.60.070191.000351.
- Sancho-Tomás, M., Somogyi, A., Medjoubi, K., Bergamaschi, A., Visscher, P. T., ... Philippot, P. (2020). Geochemical evidence for arsenic cycling in living microbialites of a high altitude Andean lake (Laguna Diamante, Argentina). *Chem. Geol.* 549, 119681. doi:10.1016/j.chemgeo.2020.119681.
- Sancho-Tomás, M., Somogyi, A., Medjoubi, K., Bergamaschi, A., Visscher, P. T., ... Philippot, P. (2018). Distribution, redox state and (bio)geochemical implications of arsenic in present day microbialites of Laguna Brava, Salar de Atacama. *Chem. Geol.* 490, 13-21. doi:10.1016/j.chemgeo.2018.04.029.
- Saona Acuña, L. A., Soria, M. N., Villafañe, P. G., Stepanenko, T., y Farías, M. E. (2020). «Arsenic and its biological role: from early Earth to current Andean microbial ecosystems», en *Microbial Ecosystems in Central Andes Extreme Environments*, ed. M. Farías (Springer International Publishing), 275-284.
- Saona, L. A., Valenzuela-Díaz, S., Kurth, D., Contreras, M., Meneses, C., ... Farías, M. E. (2019). Analysis of co-regulated abundance of genes associated with arsenic and phosphate metabolism in Andean Microbial Ecosystems. *bioRxiv*, 870428. doi:10.1101/870428.
- Saunders, E., Tindall, B. J., Fährnich, R., Lapidus, A., Copeland, A., ... Kyrpides, N. C. (2010). Complete genome sequence of *Haloterrigena turkmenica* type strain (4k T). *Stand. Genomic Sci.* 2, 107-116. doi:10.4056/sigs.681272.
- Schloss, P. D., y Handelsman, J. (2005). Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biol.* 6, 1-4. doi:10.1186/GB-2005-6-8-229.
- Schlüter, A., Krause, L., Szczepanowski, R., Goesmann, A., y Pühler, A. (2008). Genetic diversity and composition of a plasmid metagenome from a wastewater treatment plant. *J. Biotechnol.* 136, 65-76. doi:10.1016/j.jbiotec.2008.03.017.
- Schmieder, R., y Edwards, R. (2012). Insights into antibiotic resistance through metagenomic approaches. *Future Microbiol.* 7, 73-89. doi:10.2217/fmb.11.135.
- Schopf, J. W. (2006). Fossil evidence of Archaean life. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 869-885. doi:10.1098/RSTB.2006.1834.
- Schröder, G., y Lanka, E. (2005). The mating pair formation system of conjugative plasmids - A versatile secretion machinery for transfer of proteins and DNA. *Plasmid* 54, 1-25. doi:10.1016/j.plasmid.2005.02.001.
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068-2069. doi:10.1093/bioinformatics/btu153.
- Sekine, M., Tanikawa, S., Omata, S., Saito, M., Fujisawa, T., ... Harayama, S. (2006). Sequence

- analysis of three plasmids harboured in *Rhodococcus erythropolis* strain PR4. *Environ. Microbiol.* 8, 334-346. doi:10.1111/j.1462-2920.2005.00899.x.
- Sengupta, M., y Austin, S. (2011). Prevalence and significance of plasmid maintenance functions in the virulence plasmids of pathogenic bacteria. *Infect. Immun.* 79, 2502-2509. doi:10.1128/IAI.00127-11.
- Sentchilo, V., Mayer, A. P., Guy, L., Miyazaki, R., Tringe, S. G., ... Van Der Meer, J. R. (2013). Community-wide plasmid gene mobilization and selection. *ISME J.* 7, 1173-1186. doi:10.1038/ismej.2013.13.
- Sforza, M. C., Philippot, P., Somogyi, A., Van Zuilen, M. A., Medjoubi, K., ... Visscher, P. T. (2014). Evidence for arsenic metabolism and cycling by microorganisms 2.7 billion years ago. *Nat. Geosci.* 7, 811-815. doi:10.1038/ngeo2276.
- Shi, Y., Zhang, H., Tian, Z., Yang, M., y Zhang, Y. (2018). Characteristics of ARG-carrying plasmidome in the cultivable microbial community from wastewater treatment system under high oxytetracycline concentration. *Appl. Microbiol. Biotechnol.* 102, 1847-1858. doi:10.1007/s00253-018-8738-6.
- Shimizu, S., Kobayashi, H., Masai, E., y Fukuda, M. (2001). Characterization of the 450-kb Linear Plasmid in a Polychlorinated Biphenyl Degradar, *Rhodococcus* sp. Strain RHA1. *Appl. Environ. Microbiol.* 67, 2021-2028. doi:10.1128/AEM.67.5.2021-2028.2001.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., ... Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi:10.1038/msb.2011.75.
- Silver, S., y Phung, L. T. (1996). Bacterial heavy metal resistance: New Surprises. *Annu. Rev. Microbiol.* 50, 753-789. doi:10.1146/annurev.micro.50.1.753.
- Smalla, K., Haines, A. S., Jones, K., Krögerrecklenfort, E., Heuer, H., ... Thomas, C. M. (2006). Increased abundance of IncP-1 β plasmids and mercury resistance genes in mercury-polluted river sediments: First discovery of IncP-1 β plasmids with a complex mer transposon as the sole accessory element. *Appl. Environ. Microbiol.* 72, 7253-7259. doi:10.1128/AEM.00922-06.
- Smalla, K., Heuer, H., Gotz, A., Niemeyer, D., Krogerrecklenfort, E., y Tietze, E. (2000a). 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. doi:10.1128/AEM.66.11.4854-4862.2000.
- Smalla, K., Osborn, M., y Wellington, E. M. H. (2000b). «Isolation and Characterisation of Plasmids from Bacteria», en *Horizontal Gene Pool*, ed. C. M. Thomas (CRC Press), 234-273. doi:10.1201/9781482283549-12.
- Smalla, K., y Sobecky, P. A. (2002). The prevalence and diversity of mobile genetic elements in

- bacterial communities of different environmental habitats: insights gained from different methodological approaches. *FEMS Microbiol. Ecol.* 42, 165-175. doi: 10.1111/J.1574-6941.2002.TB01006.X.
- Søborg, D. A., Hendriksen, N. B., Kilian, M., Christensen, J. H., y Kroer, N. (2016). Bacterial Human Virulence Genes across Diverse Habitats As Assessed by In silico Analysis of Environmental Metagenomes. *Front. Microbiol.* 7, 1712. doi:10.3389/fmicb.2016.01712.
- Stecker, C., Johann, A., Herzberg, C., Aaverhoff, B., y Gottschalk, G. (2003). Complete nucleotide sequence and genetic organization of the 210-kilobase linear plasmid of *Rhodococcus erythropolis* BD2. *J. Bacteriol.* 185, 5269-5274. doi:10.1128/JB.185.17.5269-5274.2003.
- Stepanenko, T., Soria, M. N., Saona Acuña, L. A., Lencina, A. I., y Farías, M. E. (2020). «A Unique Natural Laboratory to Study Polyextremophile Microorganisms: Diamante Lake as a Window to the Origin of Life», en *Microbial Ecosystems in Central Andes Extreme Environments*, ed. M. E. Farias (Springer International Publishing), 113-120. doi:10.1007/978-3-030-36192-1_8.
- Stivaletta, N., y Barbieri, R. (2009). Endolithic microorganisms from spring mound evaporite deposits (southern Tunisia). *J. Arid Environ.* 73, 33-39. doi:10.1016/J.JARIDENV.2008.09.024.
- Szczepanowski, R., Bekel, T., Goesmann, A., Krause, L., Krömeke, H., ... Schlüter, A. (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. doi:10.1016/j.jbiotec.2008.03.020.
- Szmulka, A., Lestár, B., Pászti, J., Fekete, P. Z., y Nagy, B. (2015). Conjugative IncF and IncI1 plasmids with tet(A) and class 1 integron conferring multidrug resistance in F18+ porcine Enterotoxigenic *E. coli*. *Acta Vet. Hung.* 63, 425-443. doi:10.1556/004.2015.040.
- Szostková, M., y Horáková, D. (1998). The effect of plasmid DNA sizes and other factors on electrotransformation of *Escherichia coli* JM109. *Bioelectrochemistry Bioenerg.* 47, 319-323. doi:10.1016/S0302-4598(98)00203-7.
- Szuplewska, M., Czarnecki, J., y Bartosik, D. (2015). Autonomous and non-autonomous Tn 3 -family transposons and their role in the evolution of mobile genetic elements. *Mob. Genet. Elements* 4, 1-4. doi:10.1080/2159256x.2014.998537.
- Thoma, L., y Muth, G. (2012). Conjugative DNA transfer in *Streptomyces* by TraB: is one protein enough? *FEMS Microbiol. Lett.* 337, 81-88. doi:10.1111/1574-6968.12031.
- Toneatti, D. M., Albarracín, V. H., Flores, M. R., Polerecky, L., y Farías, M. E. (2017). Stratified bacterial diversity along physico-chemical gradients in high-altitude modern stromatolites. *Front. Microbiol.* 8, 646. doi:10.3389/FMICB.2017.00646/FULL.
- Tripathi, M., Singh, D. N., Vikram, S., Singh, V. S., y Kumar, S. (2018). Metagenomic Approach towards Bioprospection of Novel Biomolecule(s) and Environmental Bioremediation. *Annu. Res.*

- Rev. Biol.*, 1-12. doi:10.9734/ARRB/2018/38385.
- Tully, B. J., Graham, E. D., y Heidelberg, J. F. (2018). The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci. data* 5, 170203. doi:10.1038/sdata.2017.203.
- UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506-D515. doi:10.1093/NAR/GKY1049.
- Valero-Garcés, B. L., Arenas, C., y Delgado-Huertas, A. (2001). Depositional environments of Quaternary lacustrine travertines and stromatolites from high-altitude Andean lakes, northwestern Argentina. *Can. J. Earth Sci.* 38, 1263-1283. doi:10.1139/E01-014.
- Van Goethem, M. W., Pierneef, R., Bezuidt, O. K. I., Van De Peer, Y., Cowan, D. A., y Makhalanyane, T. P. (2018). A reservoir of «historical» antibiotic resistance genes in remote pristine Antarctic soils. *Microbiome* 6, 40. doi:10.1186/s40168-018-0424-5.
- Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., ... van Sinderen, D. (2007). Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum. *Microbiol. Mol. Biol. Rev.* 71, 495-548. doi:10.1128/mmbr.00005-07.
- Vignale, F. A., Lencina, A. I., Stepanenko, T. M., Soria, M. N., Saona, L. A., ... Farías, M. E. (2021). Lithifying and non-lithifying microbial ecosystems in the wetlands and salt flats of the Central Andes. *Microb. Ecol.*, 1-17. doi:10.1007/s00248-021-01725-8.
- Villafañe, P. G., Cónsole-Gonella, C., Cury, L. F., y Farías, M. E. (2021b). Short-term microbialite resurgence as indicador of ecological resilience against crises (Catamarca, Argentine Puna). *Environ. Microbiol. Rep.* 13, 659-667. doi: 10.1111/1758-2229.12977.
- Villafañe, P., Lencina, A., Soria, M., Saona, L., Gómez, F., ... Farías, M. (2021a). Las Quínoas oncoids: a new deposit of microbialites in the Salar de Antofalla (Catamarca, Argentina). *Andean Geol.* 48, 281-302. doi:10.5027/andgeoV48n2-3292.
- Wagenknecht, M., Dib, J. R., Thürmer, A., Daniel, R., Farías, M. E., y Meinhardt, F. (2010). Structural peculiarities of linear megaplasmid, pLMA1, from *Micrococcus luteus* interfere with pyrosequencing reads assembly. *Biotechnol. Lett.* 32, 1853-1862. doi:10.1007/s10529-010-0357-y.
- Wagenknecht, M., Pérez, M. F., y Dib, J. R. (2020). «Linear megaplasmsids spreading in the Andean resistome», en *Microbial Ecosystems in Central Andes Extreme Environments*, ed. M. E. Farias (Springer International Publishing), 261-273. doi: 10.1007/978-3-030-36192-1_18.
- Wagner-Döbler, I., Ballhausen, B., Berger, M., Brinkhoff, T., Buchholz, I., ... Simon, M. (2010). The complete genome sequence of the algal symbiont *Dinoroseobacter shibae*: A hitchhiker's guide to life in the sea. *ISME J.* 4, 61-77. doi:10.1038/ismej.2009.94.
- Wang, G., Kennedy, S. P., Fasiludeen, S., Rensing, C., y DasSarma, S. (2004). Arsenic resistance in *Halobacterium* sp. strain NRC-1 examined by using an improved gene knockout system. *J.*

- Bacteriol.* 186, 3187-3194. doi:10.1128/JB.186.10.3187-3194.2004.
- Warburton, P. J., Allan, E., Hunter, S., Ward, J., Booth, V., ... Mullany, P. (2011). Isolation of bacterial extrachromosomal DNA from human dental plaque associated with periodontal disease, using transposon-aided capture (TRACA). *FEMS Microbiol. Ecol.* 523, 349-354. doi:10.1111/J.1574-6941.2011.01166.X.
- Warren, R., Hsiao, W. W. L., Kudo, H., Myhre, M., Dosanjh, M., ... Eltis, L. D. (2004). Functional characterization of a catabolic plasmid from polychlorinated-biphenyl-degrading *Rhodococcus* sp. strain RHA1. *J. Bacteriol.* 186, 7783-7795. doi:10.1128/JB.186.22.7783-7795.2004.
- Werner, J., Ferrer, M., Michel, G., Mann, A. J., Huang, S., ... Teeling, H. (2014). *Halorhabdus tiamatea*: Proteogenomics and glycosidase activity measurements identify the first cultivated euryarchaeon from a deep-sea anoxic brine lake as potential polysaccharide degrader. *Environ. Microbiol.* 16, 2525-2537. doi:10.1111/1462-2920.12393.
- White, R. A., Chan, A. M., Gavelis, G. S., Leander, B. S., Brady, A. L., ... Suttle, C. A. (2016). Metagenomic Analysis Suggests Modern Freshwater Microbialites Harbor a Distinct Core Microbial Community. *Front. Microbiol.* 6, 1531. doi:10.3389/fmicb.2015.01531.
- White, R. A., Gavelis, G., Soles, S. A., Gosselin, E., Slater, G. F., ... Suttle, C. A. (2018). The complete genome and physiological analysis of the microbialite-dwelling *Agrococcus pavilionensis* sp. Nov; reveals genetic promiscuity and predicted adaptations to environmental stress. *Front. Microbiol.* 9, 2180. doi:10.3389/fmicb.2018.02180.
- White, R. A., Soles, S. A., Gavelis, G., Gosselin, E., Slater, G. F., ... Suttle, C. A. (2019). The complete genome and physiological analysis of the eurythermal firmicute *Exiguobacterium chiriqhucha* strain RW2 isolated from a freshwater microbialite, widely adaptable to broad thermal, pH, and salinity ranges. *Front. Microbiol.* 10, 3189. doi:10.3389/fmicb.2018.03189.
- Wijetunge, D. S. S., Karunathilake, K. H. E. M., Chaudhari, A., Katani, R., Dudley, E. G., ... Kariyawasam, S. (2014). Complete nucleotide sequence of pRS218, a large virulence plasmid, that augments pathogenic potential of meningitis-associated *Escherichia coli* strain RS218. *BMC Microbiol.* 14, 203. doi:10.1186/s12866-014-0203-9.
- Williams, P. A., y Murray, K. (1974). Metabolism of benzoate and the methylbenzoates by *Pseudomonas putida* (arvilla) mt-2: evidence for the existence of a TOL plasmid. *J. Bacteriol.* 120, 416-423. doi:10.1128/jb.120.1.416-423.1974.
- Wilson, C., Caton, T. M., Buchheim, J. A., Buchheim, M. A., Schneegurt, M. A., y Miller, R. V. (2004). DNA-repair potential of *Halomonas* spp. from the salt plains microbial observatory of Oklahoma. *Microb. Ecol.* 48, 541-549. doi:10.1007/s00248-004-0243-z.
- Wu, X., y Roy, K. L. (1993). Complete nucleotide sequence of a linear plasmid from *Streptomyces clavuligerus* and characterization of its RNA transcripts. *J. Bacteriol.* 175, 37-52.

- doi:10.1128/JB.175.1.37-52.1993.
- Xie, Y., Wei, Y., Shen, Y., Li, X., Zhou, H., ... Ou, H. Y. (2018). TADB 2.0: An updated database of bacterial type II toxin-antitoxin loci. *Nucleic Acids Res.* 46, D749-D753. doi:10.1093/nar/gkx1033.
- Xie, Z., y Tang, H. (2017). ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33, 3340-3347. doi:10.1093/bioinformatics/btx433.
- Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., ... Glöckner, F. O. (2014). The SILVA and «all-species Living Tree Project (LTP)» taxonomic frameworks. *Nucleic Acids Res.* 42, D643-D648. doi:10.1093/nar/gkt1209.
- Yim, G., Wang, H. H., y Davies, J. (2007). Antibiotics as signalling molecules. *Philos. Trans. R. Soc. B Biol. Sci.* 362, 1195-1200. doi:10.1098/rstb.2007.2044.
- Zenoff, V. F., Heredia, J., Ferrero, M., Siñeriz, F., y Farías, M. E. (2006). Diverse UV-B resistance of culturable bacterial community from high-altitude wetland water. *Curr. Microbiol.* 52, 359-362. doi:10.1007/s00284-005-0241-5.
- Zhang, J., Kobert, K., Flouri, T., y Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614-620. doi:10.1093/bioinformatics/btt593.
- Zhang, R., Xia, H., Guo, P., y Qin, Z. (2009). Variation in the replication loci of *Streptomyces* linear plasmids. *FEMS Microbiol. Lett.* 290, 209-216. doi:10.1111/J.1574-6968.2008.01432.X.
- Zhang, T., Zhang, X. X., y Ye, L. (2011). Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. *PLoS One* 6, 1-7. doi:10.1371/journal.pone.0026041.

ANEXO

Tabla A1. Parámetros fisicoquímicos medidos en el suelo de Puquio de Campo Naranja, Catamarca, Argentina.

Parámetro	Invierno	Verano
pH	7,3	8,4
Conductividad Eléctrica 1:5 (uS cm ⁻¹)	5.190	17.820
Humedad (% p/p)	59,5	20,8
Cloruro soluble en medio acuoso (mg Kg ⁻¹)	19.700	30.545
Fosfato soluble en medio acuoso (mg Kg ⁻¹)	< 50,0	< 50,0
Nitrato soluble en medio acuoso (mg Kg ⁻¹)	90	< 50,0
Nitrito soluble en medio acuoso (mg Kg ⁻¹)	< 1,0	< 1,0
Nitrógeno Amoniacal (mg Kg ⁻¹)	400	30
Sulfato soluble en medio acuoso (mg Kg ⁻¹)	670	470
Arsénico (mg Kg ⁻¹)	109	82,2
Boro (mg Kg ⁻¹)	7.385	369
Calcio (mg Kg ⁻¹)	18.000	24.700
Litio (mg Kg ⁻¹)	259	100
Fósforo total (mg Kg ⁻¹)	370	115
Magnesio (mg Kg ⁻¹)	12.100	8.000
Carbono Orgánico total (mg Kg ⁻¹)	9,5	2,1
Potasio (mg Kg ⁻¹)	12.700	1.660
Sodio (mg Kg ⁻¹)	77.200	29.200
Bromo (mg Kg ⁻¹)	< 50,0	< 50,0
Manganeso (mg Kg ⁻¹)	4.430	44,9
Hierro (mg Kg ⁻¹)	10.200	684
Cobre total (mg Kg ⁻¹)	28,3	< 5,0
Sílice soluble en medio ácido (mg Kg ⁻¹)	2.810	6.090
Níquel (mg Kg ⁻¹)	3,7	< 0,6
Estaño (mg Kg ⁻¹)	< 1,25	< 1,25
Mercurio (mg Kg ⁻¹)	< 0,1	< 0,1
Plomo (mg Kg ⁻¹)	4,7	< 0,6
Cromo (mg Kg ⁻¹)	51,3	2,7
Cadmio (mg Kg ⁻¹)	16,3	1,5
Aluminio (mg Kg ⁻¹)	1.240	617
Cianuro total (mg Kg ⁻¹)	< 0,25	< 0,25

Tabla A2. Genes plasmídicos de la base de datos ACLAME presentes en el plasmidoma de Puquio de Campo Naranja.

ID Prokka*	ID ACLAME**	Identidad (%)	Longitud (pb)	E-value	Bitscore	Cobertura (%)	Anotación de Prokka	Función de Aclame	Plásmido	Huésped
yourlocust ag_04860	gene:plasmid: 139559	84,250	2927	0.0	2760.0	100,000	23S ribosomal RNA	rRNA-23S	pGMI1000MP	<i>Ralstonia solanacearum</i> GMI1000
yourlocust ag_11730	gene:plasmid: 140775	77,902	1249	0.0	761.0	94,193	Glutamate synthase [NADPH] large chain precursor	NA	p42e	<i>Rhizobium etli</i> CFN 42
yourlocust ag_14410	gene:plasmid: 154792	75,284	1056	3.9e-128	459.0	95,913	Alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnI	phosphonate metabolic process	1	<i>Paracoccus denitrificans</i> PD1222
yourlocust ag_22940	gene:plasmid: 146204	70,851	1434	2.69e-66	254.0	96,177	Glycerol kinase	transferase activity, transferring phosphorus-containing groups	1	<i>Deinococcus geothermalis</i> DSM 11300
yourlocust ag_36690	gene:plasmid: 21818	74,290	1972	0.0	765.0	96,101	Elongation factor G	transferase activity, transferring phosphorus-containing groups	pGMI1000MP	<i>R. solanacearum</i> GMI1000
yourlocust ag_39290	gene:plasmid: 83411	77,837	1304	0.0	791.0	92,482	Ribulose biphosphate carboxylase large chain 2	NA (sin proteína asociada)	pRSPA02	<i>Rhodobacter sphaeroides</i> ATCC 17025
yourlocust ag_48520	gene:plasmid: 134908	93,506	77	5.01e-26	115.0	100,000	tRNA-Pro	NA (sin proteína asociada)	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_50360	gene:plasmid: 116643	79,058	573	1.22e-105	383.0	100,000	hypothetical protein	NA	p42b	<i>R. etli</i> CFN 42
yourlocust ag_65410	gene:plasmid: 85653	75,933	1313	4.53e-178	625.0	96,191	(Dimethylallyl)adenosine tRNA methyltransferase MiaB	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025

yourlocust ag_72830	gene:plasmid: 134974	100,000	87	7.49e -40	161.0	100,000	tRNA-Leu	NA (sin proteína asociada)	TM1040	Silicibacter sp. TM1040
yourlocust ag_75370	gene:plasmid: 134777	74,708	941	2.73e -104	379.0	97,716	L-Ala-D/L-Glu epimerase	isomerase activity	TM1040	Silicibacter sp. TM1040
yourlocust ag_75380	gene:plasmid: 134776	76,899	961	7.41e -150	531.0	95,908	hypothetical protein	NA	TM1040	Silicibacter sp. TM1040
yourlocust ag_90760	gene:plasmid: 130665	81,563	1356	0.0	1099.0	92,813	Ribulose biphosphate carboxylase large chain	NA	1	Nitrobacter hamburgensis X14
yourlocust ag_91160	gene:plasmid: 85668	83,165	891	0.0	806.0	96,117	Tryptophan synthase beta chain	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_91960	gene:plasmid: 154805	81,108	704	6.56e -159	560.0	100,000	High-affinity branched-chain amino acid transport ATP-binding protein LivF	ATP-binding cassette (ABC) transporter activity	1	P. denitrificans PD1222
yourlocust ag_91970	gene:plasmid: 154806	79,497	756	5.78e -150	531.0	96,183	Lipopolysaccharide export system ATP- binding protein LptB	ATP-binding cassette (ABC) transporter activity	1	P. denitrificans PD1222
yourlocust ag_91980	gene:plasmid: 154807	79,872	934	0.0	671.0	97,597	leucine/isoleucine/val ine transporter permease subunit	ATP-binding cassette (ABC) transporter activity	1	P. denitrificans PD1222
yourlocust ag_92000	gene:plasmid: 154809	78,098	799	7.85e -139	494.0	98,278	hypothetical protein	ATP-binding cassette (ABC) transporter complex, substrate- binding subunit- containing	1	P. denitrificans PD1222
yourlocust ag_92080	gene:plasmid: 134892	74,800	1377	1.68e -162	573.0	100,000	Electron transfer flavoprotein- ubiquinone oxidoreductase	NA	TM1040	Silicibacter sp. TM1040

yourlocust ag_93890	gene:plasmid: 85997	72,422	1610	7.68e -122	438.0	97,576	<i>Alcohol dehydrogenase [acceptor]</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides ATCC 17025</i>
yourlocust ag_96040	gene:plasmid: 85694	76,464	1691	0.0	850.0	99,588	<i>2',3'-cyclic-nucleotide 2'- phosphodiesterase/3'- nucleotidase precursor</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides ATCC 17025</i>
yourlocust ag_97300	gene:plasmid: 83371	75,660	1364	0.0	634.0	97,989	<i>Ribulose bisphosphate carboxylase</i>	NA (sin proteína asociada)	pRSPA02	<i>R. sphaeroides ATCC 17025</i>
yourlocust ag_97420	gene:plasmid: 138740	84,282	1107	0.0	1074.0	99,461	<i>S- (hydroxymethyl)gluta thione dehydrogenase</i>	<i>oxidoreductase activity</i>	<i>megaplasmid</i>	<i>Silicibacter pomeroi DSS-3</i>
yourlocust ag_10502 0	gene:plasmid: 134633	79,749	1195	0.0	857.0	100,000	<i>Succinyl-CoA ligase [ADP-forming] subunit beta</i>	NA	TM1040	<i>Silicibacter sp. TM1040</i>
yourlocust ag_12008 0	gene:plasmid: 26502	76,809	940	4.14e -142	505.0	100,000	<i>hypothetical protein</i>	<i>oxidoreductase activity; D-nopaline dehydrogenase activity</i>	pSymA	<i>Sinorhizobiu m meliloti 1021</i>
yourlocust ag_12146 0	gene:plasmid: 86344	78,076	447	1.01e -75	283.0	98,676	<i>Dipeptide transport system permease protein DppC</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides ATCC 17025</i>
yourlocust ag_12405 0	gene:plasmid: 135153	73,958	1271	1.31e -128	460.0	96,070	<i>Beta-glucosidase A</i>	NA	TM1040	<i>Silicibacter sp. TM1040</i>
yourlocust ag_12487 0	gene:plasmid: 22713	75,046	541	1.26e -60	233.0	98,543	<i>hypothetical protein</i>	<i>nitrogenase activity</i>	p42d	<i>R. etli</i> CFN 42
yourlocust ag_12556 0	gene:plasmid: 83409	71,290	930	3.0e- 49	196.0	94,801	<i>hypothetical protein</i>	NA (sin proteína asociada)	pRSPA02	<i>R. sphaeroides ATCC 17025</i>
yourlocust ag_12786 0	gene:plasmid: 154816	86,096	187	1.14e -51	202.0	97,396	<i>Bifunctional protein PaaZ</i>	<i>oxidoreductase activity</i>	1	<i>P. denitrificans PD1222</i>

yourlocust ag_12789 0	gene:plasmid: 154813	88,419	1278	0.0	1539.0	97,483	Phenylacetate- coenzyme A ligase	NA	1	<i>P. denitrificans</i> PD1222
yourlocust ag_12790 0	gene:plasmid: 154812	84,906	424	1.55e -118	425.0	95,496	Acyl-coenzyme A thioesterase PaaI	NA	1	<i>P. denitrificans</i> PD1222
yourlocust ag_12791 0	gene:plasmid: 154811	80,952	756	2.65e -168	592.0	95,094	1,2- epoxyphenylacetyl- CoA isomerase	catalytic activity	1	<i>P. denitrificans</i> PD1222
yourlocust ag_14112 0	gene:plasmid: 134649	86,122	771	0.0	830.0	98,846	Succinate dehydrogenase iron- sulfur subunit	NA	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_14272 0	gene:plasmid: 29622	74,791	718	2.33e -74	279.0	90,657	hypothetical protein	NA	pSLA2-L	<i>Streptomyces rochei</i>
yourlocust ag_14441 0	gene:plasmid: 134630	80,841	2615	0.0	2034.0	99,962	2-oxoglutarate dehydrogenase E1 component	NA	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_15419 0	gene:plasmid: 86016	78,745	1689	0.0	1109.0	100,000	Catalase-peroxidase	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_15525 0	gene:plasmid: 154821	87,762	286	6.01e -91	333.0	100,000	1,2-phenylacetyl-CoA epoxidase, subunit B	NA	1	<i>P. denitrificans</i> PD1222
yourlocust ag_15526 0	gene:plasmid: 154822	85,791	929	0.0	979.0	92,438	1,2-phenylacetyl-CoA epoxidase, subunit A	NA	1	<i>P. denitrificans</i> PD1222
yourlocust ag_15553 0	gene:plasmid: 154737	75,187	1334	3.71e -164	579.0	94,409	NAD/NADP- dependent betaine aldehyde dehydrogenase 1	oxidoreductase activity	1	<i>P. denitrificans</i> PD1222
yourlocust ag_15831 0	gene:plasmid: 85663	85,655	1213	0.0	1269.0	98,859	30S ribosomal protein S1	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_16134	gene:plasmid: 81467	77,641	729	6.1e- 120	431.0	91,698	Glutamine transport ATP-binding protein	NA (sin proteína asociada)	pSMED03	<i>Sinorhizobiu m medicae</i>

0	<i>GlnQ</i>									WSM419
yourlocust ag_16272 0	gene:plasmid: 83537	81,172	802	0.0	636.0	91,240	<i>Chromosomal replication initiator protein DnaA</i>	NA (sin proteína asociada)	pRSPA02	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_16331 0	gene:plasmid: 84372	78,413	1399	0.0	881.0	95,756	<i>Mercuric reductase</i>	NA (sin proteína asociada)	pMATVIM-7	<i>Pseudomonas aeruginosa</i>
yourlocust ag_16741 0	gene:plasmid: 91662	81,061	1489	0.0	1175.0	90,242	<i>60 kDa chaperonin</i>	NA (sin proteína asociada)	pSN254	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport str. SL254
yourlocust ag_17069 0	gene:plasmid: 93519	82,209	697	1.84e -169	595.0	97,211	<i>IS2 transposase TnpB</i>	NA (sin proteína asociada)	pRSPA05	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_17778 0	gene:plasmid: 154730	76,774	620	5.43e -90	331.0	91,043	<i>Alpha-D-ribose 1- methylphosphonate 5-triphosphate synthase subunit PhnL</i>	<i>ATP-binding cassette (ABC) transporter activity</i>	1	<i>P. denitrificans</i> PD1222
yourlocust ag_17889 0	gene:plasmid: 93228	73,068	427	5.19e -29	128.0	90,658	<i>hypothetical protein</i>	NA (sin proteína asociada)	pBBta01	<i>Bradyrhizobi um</i> sp. BTAi1
yourlocust ag_18710 0	gene:plasmid: 125962	74,091	1953	0.0	728.0	100,000	<i>Catalase-peroxidase</i>	NA	p42f	<i>R. etli</i> CFN 42
yourlocust ag_18918 0	gene:plasmid: 139013	80,067	1194	0.0	865.0	100,000	<i>S-adenosylmethionine synthase</i>	NA	megaplasmid	<i>S. pomeroyi</i> DSS-3
yourlocust ag_19543 0	gene:plasmid: 134882	78,845	1177	0.0	785.0	100,000	<i>Acetyl-CoA acetyltransferase</i>	<i>acetyl-CoA C- acyltransferase activity</i>	TM1040	<i>Silicibacter</i> sp. TM1040

yourlocust ag_19724 0	gene:plasmid: 93523	80,967	1592	0.0	1214.0	98,821	Transposase IS66 family protein	NA (sin proteína asociada)	pRSPA05	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_19771 0	gene:plasmid: 154774	84,000	1650	0.0	1568.0	100,000	60 kDa chaperonin 1	NA	1	<i>P. denitrificans</i> PD1222
yourlocust ag_19884 0	gene:plasmid: 139557	87,356	87	5.96e -21	99.0	100,000	tRNA-Ser	NA (sin proteína asociada)	pGMI1000MP	<i>R. solanacearum</i> GMI1000
yourlocust ag_20214 0	gene:plasmid: 89088	76,112	1147	6.83e -166	584.0	94,403	Acyl-CoA dehydrogenase	NA (sin proteína asociada)	pSMED01	<i>S. medicae</i> WSM419
yourlocust ag_20775 0	gene:plasmid: 27447	74,781	1487	5.25e -178	625.0	94,233	Heme-binding protein A precursor	ATP-binding cassette (ABC) transporter activity	pSymB	<i>S. meliloti</i> 1021
yourlocust ag_20896 0	gene:plasmid: 140776	78,100	621	1.93e -104	379.0	90,393	GXGXX motif protein	NA	p42e	<i>R. etli</i> CFN 42
yourlocust ag_21406 0	gene:plasmid: 131374	86,228	1641	0.0	1749.0	92,036	23S ribosomal RNA (partial)	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_21468 0	gene:plasmid: 154741	77,720	1535	0.0	881.0	99,353	hypothetical protein	NA	1	<i>P. denitrificans</i> PD1222
yourlocust ag_21531 0	gene:plasmid: 86221	80,107	749	2.8e- 153	542.0	91,119	Phosphate-import ATP-binding protein PhnC	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_21532 0	gene:plasmid: 86220	80,621	805	4.48e -171	601.0	99,383	Phosphate-import permease protein PhnE	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_22078 0	gene:plasmid: 134501	94,366	71	2.29e -24	110.0	93,421	tRNA-Thr	NA (sin proteína asociada)	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_22377 0	gene:plasmid: 83360	73,913	989	2.3e- 105	383.0	94,191	Transposase IS116/IS110/IS902 family protein	NA (sin proteína asociada)	pRSPA02	<i>R. sphaeroides</i> ATCC 17025

yourlocust ag_22443 0	gene:plasmid: 23615	83,494	830	0.0	769.0	94,104	Nitrogenase iron protein	arsenite transmembrane- transporting ATPase activity	pNGR234a	Rhizobium sp. NGR234
yourlocust ag_23116 0	gene:plasmid: 21942	71,429	1022	6.75e -56	219.0	99,031	hypothetical protein	efflux transmembrane transporter activity	pGMI1000MP	R. solanacearum GMI1000
yourlocust ag_23162 0	gene:plasmid: 90302	72,567	1757	8.18e -137	488.0	98,102	Arsenical pump- driving ATPase	NA (sin proteína asociada)	pMTH1	Paracoccus methylytens
yourlocust ag_23448 0	gene:plasmid: 118513	81,081	1147	0.0	902.0	98,540	3-hydroxy-D- aspartate aldolase	NA	1	Mesorhizobiu m sp. BNC1
yourlocust ag_23733 0	gene:plasmid: 154791	80,338	829	6.13e -175	614.0	96,283	Alpha-D-ribose 1- methylphosphonate 5-phosphate C-P lyase	phosphonate metabolic process	1	P. denitrificans PD1222
yourlocust ag_24313 0	gene:plasmid: 125170	78,520	1243	0.0	804.0	95,249	UDP-glucose 6- dehydrogenase TuaD	oxidoreductase activity	unnamed	Silicibacter sp. TM1040
yourlocust ag_24924 0	gene:plasmid: 28052	72,917	672	1.71e -50	200.0	92,562	7-cyano-7- deazaguanine synthase	NA	pSymB	S. meliloti 1021
yourlocust ag_25129 0	gene:plasmid: 154817	80,532	714	1.54e -150	532.0	94,444	1,2-phenylacetyl-CoA epoxidase, subunit A	NA	1	P. denitrificans PD1222
yourlocust ag_25510 0	gene:plasmid: 121712	82,932	996	0.0	893.0	95,954	hypothetical protein	transferase activity, transferring phosphorus- containing groups	pTB2	Roseobacter denitrificans OCh 114
yourlocust ag_26065 0	gene:plasmid: 135602	94,118	85	2.05e -30	130.0	100,000	tRNA-Tyr	NA (sin proteína asociada)	pSbal01	Shewanella baltica OS155
yourlocust ag_26409 0	gene:plasmid: 131330	98,667	75	1.33e -31	134.0	100,000	tRNA-Gly	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025

yourlocust ag_27539 0	gene:plasmid: 135146	96,000	75	2.99e -28	122.0	97,403	tRNA-Arg	NA (sin proteína asociada)	TM1040	Silicibacter sp. TM1040
yourlocust ag_28164 0	gene:plasmid: 22010	77,419	341	1.3e- 48	193.0	99,708	hypothetical protein	oxidoreductase activity	pGMI1000MP	R. solanacearum GMI1000
yourlocust ag_28217 0	gene:plasmid: 85639	75,436	1205	7.54e -151	534.0	91,915	Cytochrome bd ubiquinol oxidase subunit 1	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_28761 0	gene:plasmid: 85794	86,223	421	1.86e -127	455.0	100,000	hypothetical protein	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_29003 0	gene:plasmid: 131858	100,000	75	2.86e -33	139.0	100,000	tRNA-Asn	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_29455 0	gene:plasmid: 85613	78,603	1117	0.0	719.0	99,025	Integrase core domain protein	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_29648 0	gene:plasmid: 134940	76,316	1102	2.4e- 160	566.0	93,232	hypothetical protein	oxidoreductase activity	TM1040	Silicibacter sp. TM1040
yourlocust ag_29668 0	gene:plasmid: 130851	83,461	1046	0.0	961.0	99,053	Ribulose biphosphate carboxylase large chain 1	NA	1	N. hamburgensis X14
yourlocust ag_30039 0	gene:plasmid: 86330	77,551	735	1.68e -115	416.0	94,231	hypothetical protein	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_31668 0	gene:plasmid: 138839	90,868	219	2.12e -79	294.0	100,000	Translation initiation factor IF-1	NA	megaplasmid	S. pomeroyi DSS-3
yourlocust ag_32123 0	gene:plasmid: 134997	73,675	1113	1.54e -107	390.0	95,128	Lysine/ornithine decarboxylase	NA	TM1040	Silicibacter sp. TM1040
yourlocust ag_32458	gene:plasmid: 83368	74,312	436	1.04e -40	167.0	97,539	hypothetical protein	NA (sin proteína asociada)	pRSPA02	R. sphaeroides

0	ATCC 17025									
yourlocust ag_34102 0	gene:plasmid: 138883	100,000	75	2.86e -33	139.0	100,000	tRNA-Phe	NA (sin proteína asociada)	megaplasmid	S. pomeroyi DSS-3
yourlocust ag_34112 0	gene:plasmid: 81525	75,243	1236	2.72e -155	549.0	93,213	Integrase core domain protein	NA (sin proteína asociada)	pSMED03	S. medicae WSM419
yourlocust ag_34553 0	gene:plasmid: 92570	73,540	582	1.73e -49	196.0	100,000	hypothetical protein	NA (sin proteína asociada)	pSWIT02	Sphingomona s wittichii RW1
yourlocust ag_35402 0	gene:plasmid: 154920	83,826	1082	0.0	1018.0	96,435	Sarcosine oxidase subunit beta	oxidoreductase activity; D-nopaline dehydrogenase activity	1	P. denitrificans PD1222
yourlocust ag_35503 0	gene:plasmid: 28525	75,216	924	7.55e -105	381.0	96,250	Methylthioribose-1- phosphate isomerase	NA	pSymB	S. meliloti 1021
yourlocust ag_35727 0	gene:plasmid: 83247	74,869	1333	2.32e -156	553.0	91,239	hypothetical protein	NA (sin proteína asociada)	pSWIT01	S. wittichii RW1
yourlocust ag_37074 0	gene:plasmid: 134985	77,749	391	9.77e -61	233.0	90,509	Sulfite exporter TauE/SafE	plasmid function unknown	TM1040	Silicibacter sp. TM1040
yourlocust ag_37461 0	gene:plasmid: 90534	75,969	645	6.84e -84	311.0	97,727	hypothetical protein	NA (sin proteína asociada)	pBVIE02	Burkholderia vietnamiensis G4
yourlocust ag_38169 0	gene:plasmid: 131155	91,111	90	3.7e- 28	122.0	100,000	tRNA-Ser	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_38302 0	gene:plasmid: 22462	79,156	403	3.17e -70	265.0	100,000	hypothetical protein	NA	pGMI1000MP	R. solanacearum GMI1000
yourlocust ag_38633 0	gene:plasmid: 84281	82,763	818	0.0	725.0	99,878	hypothetical protein	NA (sin proteína asociada)	pOANT01	Ochrobactru m anthropi ATCC 49188

yourlocust ag_38637 0	gene:plasmid: 90278	79,305	633	6.79e -119	427.0	92,140	hypothetical protein	NA (sin proteína asociada)	pSMED01	<i>S. medicae</i> WSM419
yourlocust ag_38800 0	gene:plasmid: 85828	82,008	1006	0.0	848.0	90,876	Transposase DDE domain protein	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_41225 0	gene:plasmid: 154742	74,103	1143	3.16e -114	412.0	100,000	Putative nucleotidyltransferas e substrate binding domain protein	oxidoreductase activity	1	<i>P. denitrificans</i> PD1222
yourlocust ag_42799 0	gene:plasmid: 85937	76,622	817	3.09e -118	425.0	94,560	Long-chain-fatty- acid--CoA ligase	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_43250 0	gene:plasmid: 134993	83,221	1037	0.0	948.0	96,286	hypothetical protein	NA	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_43675 0	gene:plasmid: 123601	77,059	510	2.6e- 77	289.0	94,444	hypothetical protein	NA	1	<i>Polaromonas</i> sp. JS666
yourlocust ag_44613 0	gene:plasmid: 13101	94,737	76	3.8e- 27	119.0	100,000	tRNA-Ala	NA	pBM400	<i>Bacillus</i> <i>megaterium</i>
yourlocust ag_45864 0	gene:plasmid: 86324	81,776	642	1.83e -149	529.0	92,241	Galactose/methyl galactoside import ATP-binding protein MglA	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_45955 0	gene:plasmid: 82581	80,159	882	0.0	649.0	99,324	Glutamate synthase [NADPH] large chain precursor	NA (sin proteína asociada)	pSMED02	<i>S. medicae</i> WSM419
yourlocust ag_46982 0	gene:plasmid: 26456	83,854	384	8.82e -101	366.0	92,754	Electron transfer flavoprotein subunit alpha	electron carrier activity; nitrogen fixation	pSymA	<i>S. meliloti</i> 1021
yourlocust ag_47481 0	gene:plasmid: 16555	72,614	964	6.21e -71	268.0	99,177	Catalase-peroxidase	NA	pNRC200	<i>Halobacteriu</i> <i>m</i> sp. NRC-1

yourlocust ag_47804 0	gene:plasmid: 87643	87,882	1015	0.0	1194.0	99,803	ISXO2-like transposase domain protein	NA (sin proteína asociada)	pACRY03	<i>Acidiphilium cryptum</i> JF-5
yourlocust ag_50592 0	gene:plasmid: 154915	77,919	865	6.58e -145	514.0	97,740	<i>Formyltetrahydrofolate deformylase</i>	NA	1	<i>P. denitrificans</i> PD1222
yourlocust ag_54791 0	gene:plasmid: 134515	76,733	851	4.05e -127	455.0	96,158	<i>putative metallo- hydrolase</i>	<i>hydrolase activity</i>	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_55086 0	gene:plasmid: 81524	77,117	555	2.15e -83	309.0	95,855	<i>hypothetical protein</i>	NA (sin proteína asociada)	pSMED03	<i>S. medicae</i> WSM419
yourlocust ag_55375 0	gene:plasmid: 28503	74,784	694	1.33e -71	270.0	92,533	<i>Lipopolysaccharide export system ATP- binding protein LptB</i>	<i>ATP-binding cassette (ABC) transporter activity</i>	pSymB	<i>S. meliloti</i> 1021
yourlocust ag_55905 0	gene:plasmid: 85860	82,079	731	1.85e -174	612.0	100,000	<i>hypothetical protein</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_56296 0	gene:plasmid: 138806	80,342	468	1.04e -95	350.0	97,500	<i>hypothetical protein</i>	NA	megaplasmid	<i>S. pomeroyi</i> DSS-3
yourlocust ag_57839 0	gene:plasmid: 151922	74,532	534	2.81e -57	222.0	94,180	<i>hypothetical protein</i>	<i>two-component sensor activity; two-component signal transduction system (phosphorelay); two-component response regulator activity</i>	megaplasmid	<i>Desulfovibrio vulgaris</i> subsp. <i>vulgaris</i> str. Hildenboroug h
yourlocust ag_58401 0	gene:plasmid: 138912	77,151	372	8.62e -51	200.0	98,413	<i>hypothetical protein</i>	<i>ion transmembrane transporter activity</i>	megaplasmid	<i>S. pomeroyi</i> DSS-3
yourlocust ag_58914 0	gene:plasmid: 155082	77,627	885	8.6e- 144	510.0	98,993	<i>3-keto-5- aminohexanoate cleavage enzyme</i>	<i>plasmid function unknown</i>	1	<i>P. denitrificans</i> PD1222

yourlocust ag_60623 0	gene:plasmid: 83528	92,793	333	6.68e -136	483.0	100,000	<i>HicB family protein</i>	NA (sin proteína asociada)	pRSPA02	<i>R. sphaeroides ATCC 17025</i>
yourlocust ag_61497 0	gene:plasmid: 156517	75,464	754	1.08e -92	340.0	93,086	<i>hypothetical protein</i>	NA	pPNAP02	<i>Polaromonas naphthaleniv orans CJ2</i>
yourlocust ag_65893 0	gene:plasmid: 92550	82,282	745	0.0	636.0	98,938	<i>hypothetical protein</i>	NA (sin proteína asociada)	pSWIT02	<i>S. wittichii RW1</i>
yourlocust ag_66247 0	gene:plasmid: 154980	79,755	326	2.62e -60	231.0	99,694	<i>hypothetical protein</i>	<i>oxidoreductase activity</i>	1	<i>P. denitrificans PD1222</i>
yourlocust ag_68954 0	gene:plasmid: 21136	74,819	830	5.17e -96	351.0	98,810	<i>Methylisocitrate lyase</i>	NA	pGMI1000MP	<i>R. solanacearum GMI1000</i>
yourlocust ag_69049 0	gene:plasmid: 86133	90,036	281	2.35e -100	364.0	90,064	<i>Transposase IS66 family protein</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides ATCC 17025</i>
yourlocust ag_70082 0	gene:plasmid: 115548	76,808	733	7.85e -104	377.0	94,703	<i>hypothetical protein</i>	NA	RPME01	<i>Methylibium petroleiphilu m PM1</i>
yourlocust ag_71825 0	gene:plasmid: 151861	84,859	568	7.53e -163	573.0	90,159	<i>Nitrogenase iron protein 1</i>	<i>arsenite transmembrane- transporting ATPase activity</i>	<i>megaplasmid</i>	<i>Desulfovibrio vulgaris subsp. vulgaris str. Hildenboroug h</i>
yourlocust ag_72114 0	gene:plasmid: 86123	91,095	539	0.0	730.0	97,117	<i>DNA replication protein DnaC</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides ATCC 17025</i>
yourlocust ag_73589 0	gene:plasmid: 85593	77,483	604	1.08e -96	353.0	94,523	<i>putative tRNA/rRNA methyltransferase</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides ATCC 17025</i>
yourlocust ag_73772 0	gene:plasmid: 158676	87,415	294	1.9e- 91	335.0	91,589	<i>hypothetical protein</i>	<i>DDE-based recombinase activity;</i>	pSD25	<i>Ruegeria sp. PR1b</i>

								<i>transpositional DNA recombination; IS5 family</i>		
yourlocust ag_742490	gene:plasmid: 83413	77,010	622	1.1e-96	353.0	95,988	<i>Denitrification regulatory protein NirQ</i>	NA (sin proteína asociada)	pRSPA02	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_749600	gene:plasmid: 131373	97,368	76	1.76e-30	130.0	100,000	<i>tRNA-Ala</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_763350	gene:plasmid: 27836	86,102	295	6.26e-86	316.0	100,000	<i>hypothetical protein</i>	<i>oxidoreductase activity</i>	pSymB	<i>S. meliloti</i> 1021
yourlocust ag_766540	gene:plasmid: 134653	81,609	261	5.88e-56	217.0	97,753	<i>Transposase</i>	<i>DDE-based recombinase activity; IS3 family; transpositional DNA recombination</i>	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_799710	gene:plasmid: 22073	73,051	449	3.23e-31	135.0	91,820	<i>"D-beta-hydroxybutyrate dehydrogenase</i>	<i>oxidoreductase activity</i>	pGMI1000MP	<i>R. solanacearum</i> GMI1000
yourlocust ag_799770	gene:plasmid: 130722	74,093	386	3.22e-35	148.0	100,000	<i>hypothetical protein</i>	<i>toxin; interaction with host via protein secreted by type III secretion system</i>	1	<i>N. hamburgensis</i> X14
yourlocust ag_837700	gene:plasmid: 19088	74,046	393	4.24e-34	145.0	100,000	<i>hypothetical protein</i>	<i>transferase activity, transferring phosphorus-containing groups</i>	pNL1	<i>Novosphingobium aromaticivorans</i>
yourlocust ag_858340	gene:plasmid: 109387	78,030	528	3.51e-86	318.0	92,632	<i>putative transporter</i>	<i>transporter activity</i>	D	<i>R. sphaeroides</i> 2.4.1
yourlocust ag_860560	gene:plasmid: 158678	90,634	331	4.07e-123	440.0	99,399	<i>Toxin YhaV</i>	NA	pSD25	<i>Ruegeria</i> sp. PR1b

yourlocust ag_86094 0	gene:plasmid: 133565	82,521	595	9.8e- 147	520.0	95,813	transposase/IS protein	IS21 family; positive regulation of DNA transposition	pCAR3	<i>Sphingomona</i> s sp. KA1
yourlocust ag_89477 0	gene:plasmid: 86249	85,381	472	7.44e -137	486.0	100,000	Transposase IS116/IS110/IS902 family protein	NA (sin proteína asociada)	pRSPA01	<i>R.</i> <i>sphaeroides</i> ATCC 17025
yourlocust ag_94042 0	gene:plasmid: 23216	77,070	314	1.66e -42	172.0	93,452	hypothetical protein	transaminase activity	pRi1724	<i>Agrobacteriu</i> <i>m rhizogenes</i>
yourlocust ag_94231 0	gene:plasmid: 26457	79,406	471	8.2e- 87	320.0	97,516	hypothetical protein	electron carrier activity; nitrogen fixation	pSymA	<i>S. meliloti</i> 1021
yourlocust ag_94330 0	gene:plasmid: 82580	85,417	144	3.09e -36	150.0	97,959	hypothetical protein	NA (sin proteína asociada)	pSMED02	<i>S. medicae</i> WSM419
yourlocust ag_94482 0	gene:plasmid: 26163	74,492	541	2.63e -52	206.0	100,000	Sulfate adenyltransferase subunit 2	sulfur metabolic process	pSymA	<i>S. meliloti</i> 1021
yourlocust ag_95635 0	gene:plasmid: 85459	78,495	279	3.08e -44	178.0	94,898	Copper-transporting P-type ATPase	NA (sin proteína asociada)	pXAUT01	<i>Xanthobacter</i> <i>autotrophicus</i> Py2
yourlocust ag_96445 0	gene:plasmid: 86055	83,133	166	3.84e -36	150.0	93,785	hypothetical protein	NA (sin proteína asociada)	pRSPA01	<i>R.</i> <i>sphaeroides</i> ATCC 17025
yourlocust ag_98083 0	gene:plasmid: 134943	100,000	76	8.11e -34	141.0	100,000	tRNA-Ala	NA (sin proteína asociada)	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_99879 0	gene:plasmid: 139052	84,502	542	1.21e -150	532.0	90,787	Acryloyl-CoA reductase (NADH)	oxidoreductase activity	megaplasmid	<i>S. pomeroyi</i> DSS-3
yourlocust ag_10481 10	gene:plasmid: 156246	75,263	380	7.11e -42	171.0	95,960	Transposase, Mutator family	DDE-based recombinase activity; transpositional DNA	p42a	<i>R. etli</i> CFN 42

								<i>recombination; IS256 family</i>		
yourlocust ag_10686 70	gene:plasmid: 158746	92,966	327	3.05e -134	477.0	100,000	<i>Antitoxin PrIF</i>	NA	pSD25	<i>Ruegeria</i> sp. PR1b
yourlocust ag_10745 90	gene:plasmid: 83527	92,248	129	2.68e -46	183.0	97,727	<i>hypothetical protein</i>	NA (sin proteína asociada)	pRSPA02	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_10920 40	gene:plasmid: 112656	93,190	279	2.64e -114	411.0	100,000	<i>hypothetical protein</i>	<i>ATPase activity; plasmid partitioning</i>	pNG600	<i>Haloarcula marismortui</i> ATCC 43049
yourlocust ag_10928 50	gene:plasmid: 149985	85,714	70	8.19e -14	75.0	93,333	<i>tRNA-Arg</i>	NA (sin proteína asociada)	pSymB	<i>S. meliloti</i> 1021
yourlocust ag_11170 60	gene:plasmid: 138953	79,735	528	1.12e -100	366.0	100,000	<i>hypothetical protein</i>	<i>transferase activity</i>	<i>megaplasmid</i>	<i>S. pomeroyi</i> DSS-3
yourlocust ag_11243 80	gene:plasmid: 86313	80,952	378	1.11e -79	296.0	96,923	<i>Enoyl-[acyl-carrier- protein] reductase [NADH] FabI</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_11272 20	gene:plasmid: 134629	84,301	465	2.68e -126	451.0	100,000	<i>Dihydrolipoyllysine- residue succinyltransferase component of 2- oxoglutarate dehydrogenase complex</i>	<i>oxidoreductase activity</i>	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_11857 60	gene:plasmid: 158662	84,774	243	8.86e -64	243.0	98,781	<i>hypothetical protein</i>	<i>relaxase activity</i>	pSD25	<i>Ruegeria</i> sp. PR1b
yourlocust ag_11886 40	gene:plasmid: 131375	92,727	110	3.64e -39	159.0	99,099	<i>5S ribosomal RNA</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_11886 50	gene:plasmid: 134946	96,104	77	2.31e -29	126.0	100,000	<i>tRNA-Met</i>	NA (sin proteína asociada)	TM1040	<i>Silicibacter</i> sp. TM1040

yourlocust ag_11898 60	gene:plasmid: 86219	84,198	405	1.83e -107	388.0	100,000	Phosphate-import permease protein PhnE	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_11910 80	gene:plasmid: 145257	74,157	267	3.09e -19	95.3	93,684	NAD(P) transhydrogenase subunit alpha	NA	megaplasmid	Ralstonia eutropha JMP134
yourlocust ag_11965 60	gene:plasmid: 138232	79,567	323	1.25e -58	226.0	96,131	Sulfoacetaldehyde acetyltransferase	transferase activity, transferring aldehyde or ketonic groups	B	R. sphaeroides 2.4.1
yourlocust ag_12057 90	gene:plasmid: 125495	85,012	407	3.94e -114	411.0	100,000	Integrase core domain protein	NA (sin proteína asociada)	C	R. sphaeroides 2.4.1
yourlocust ag_12068 40	gene:plasmid: 138778	82,564	195	3.27e -42	171.0	100,000	hypothetical protein	oxidoreductase activity	megaplasmid	S. pomeroyi DSS-3
yourlocust ag_12082 00	gene:plasmid: 86226	78,049	369	3.06e -60	231.0	97,619	hypothetical protein	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_12315 40	gene:plasmid: 92582	81,062	433	9.22e -91	333.0	100,000	transposase/IS protein	NA (sin proteína asociada)	pSWIT02	S. wittichii RW1
yourlocust ag_12330 80	gene:plasmid: 158680	93,907	279	1.22e -117	422.0	100,000	hypothetical protein	NA	pSD25	Ruegeria sp. PR1b
yourlocust ag_12546 10	gene:plasmid: 134925	98,667	75	1.33e -31	134.0	100,000	tRNA-Thr	NA (sin proteína asociada)	TM1040	Silicibacter sp. TM1040
yourlocust ag_12578 90	gene:plasmid: 93518	80,000	215	1.01e -37	156.0	100,000	IS2 repressor TnpA	NA (sin proteína asociada)	pRSPA05	R. sphaeroides ATCC 17025
yourlocust ag_12656 40	gene:plasmid: 85523	89,920	377	5.91e -137	486.0	99,735	Transposase DDE domain protein	NA (sin proteína asociada)	pXAUT01	X. autotrophicus Py2

yourlocust ag_13055 30	gene:plasmid: 85826	90,730	356	1.27e -133	475.0	94,933	hypothetical protein	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_13341 30	gene:plasmid: 158744	93,122	189	1.81e -74	278.0	100,000	hypothetical protein	NA	pSD25	<i>Ruegeria</i> sp. PR1b
yourlocust ag_13359 40	gene:plasmid: 134439	100,000	75	2.97e -33	139.0	97,403	tRNA-Arg	NA (sin proteína asociada)	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_13432 70	gene:plasmid: 134636	79,741	232	1.92e -40	165.0	96,667	Malate dehydrogenase	NA	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_13856 30	gene:plasmid: 27394	76,433	157	1.23e -15	82.4	100,000	hypothetical protein	ATP-binding cassette (ABC) transporter activity	pSymB	<i>S. meliloti</i> 1021
yourlocust ag_13862 80	gene:plasmid: 97700	84,416	231	9.29e -59	226.0	90,588	Alginate biosynthesis protein AlgA	NA (sin proteína asociada)	pSD20	<i>Ruegeria</i> sp. PR1b
yourlocust ag_13940 90	gene:plasmid: 85664	81,752	274	8.02e -60	230.0	97,163	Integration host factor subunit beta	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_14324 70	gene:plasmid: 16869	89,355	310	1.38e -107	388.0	100,000	hypothetical protein	IS4 family; transpositional DNA recombination	pNRC100	<i>Halobacteriu</i> <i>m</i> sp. NRC-1
yourlocust ag_14528 40	gene:plasmid: 154071	99,259	270	1.14e -137	488.0	100,000	hypothetical protein	two-component sensor activity; two-component signal transduction system (phosphorelay); two-component response regulator activity	R478	<i>Serratia</i> <i>marcescens</i>
yourlocust ag_14538 80	gene:plasmid: 154647	82,153	353	2.87e -80	298.0	96,448	hypothetical protein	NA	1	<i>P. denitrificans</i> PD1222

yourlocust ag_14860 80	gene:plasmid: 134978	88,073	109	3.05e -30	130.0	93,162	<i>Fe/S biogenesis protein NfuA</i>	<i>oxidoreductase activity</i>	TM1040	<i>Silicibacter sp. TM1040</i>
yourlocust ag_14861 00	gene:plasmid: 118440	78,571	210	1.81e -30	132.0	93,333	<i>hypothetical protein</i>	<i>ATP-binding cassette (ABC) transporter activity</i>	1	<i>Mesorhizobiu m sp. BNC1</i>
yourlocust ag_15160 70	gene:plasmid: 85749	78,431	255	2.66e -39	161.0	100,000	<i>hypothetical protein</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides ATCC 17025</i>
yourlocust ag_15497 80	gene:plasmid: 156128	97,273	330	2.97e -159	560.0	100,000	<i>hypothetical protein</i>	<i>ATPase activity; plasmid partitioning</i>	pLVPK	<i>Klebsiella pneumoniae</i>
yourlocust ag_15622 80	gene:plasmid: 91183	88,298	282	4.7e- 92	337.0	97,917	<i>putative HTH-type transcriptional regulator YgaV</i>	NA (sin proteína asociada)	pYR1	<i>Yersinia ruckeri</i>
yourlocust ag_15801 50	gene:plasmid: 93245	81,034	290	5.11e -57	220.0	97,643	<i>Glutamine transport ATP-binding protein GlnQ</i>	NA (sin proteína asociada)	pBBta01	<i>Bradyrhizobi um sp. BTAi1</i>
yourlocust ag_15811 60	gene:plasmid: 130898	78,832	137	1.9e- 18	91.6	93,197	<i>hypothetical protein</i>	<i>oxidoreductase activity</i>	1	<i>N. hamburgensis X14</i>
yourlocust ag_16167 40	gene:plasmid: 158652	95,169	207	1.96e -89	327.0	100,000	<i>hypothetical protein</i>	<i>intercellular transfer by conjugation; Type IV secretion complex VirB4 family; type IV protein secretion/mating pair formation system complex</i>	pSD25	<i>Ruegeria sp. PR1b</i>
yourlocust ag_16923 00	gene:plasmid: 28434	77,515	338	6.32e -47	187.0	94,678	<i>Acryloyl-CoA reductase (NADH)</i>	<i>oxidoreductase activity</i>	pSymB	<i>S. meliloti 1021</i>
yourlocust ag_16996	gene:plasmid: 17914	86,164	159	7.23e -43	172.0	100,000	<i>TrbC/VIRB2 family protein</i>	<i>intercellular transfer by</i>	pMLb	<i>Mesorhizobiu m loti</i>

20								<i>conjugation; pilin</i>		MAFF30309 9
yourlocust ag_17048 80	gene:plasmid: 138785	83,654	208	2.11e -49	195.0	99,048	Outer membrane protein precursor	integral to cell outer membrane; porin activity	megaplasmid	<i>S. pomeroyi</i> DSS-3
yourlocust ag_17331 10	gene:plasmid: 121718	82,007	289	9.07e -65	246.0	90,881	Alginate biosynthesis protein AlgA	NA	pTB2	<i>R.</i> <i>denitrificans</i> OCh 114
yourlocust ag_17334 40	gene:plasmid: 28312	79,336	271	1.39e -47	189.0	94,097	ornithine cyclodeaminase	ornithine cyclodeaminase activity	pSymB	<i>S. meliloti</i> 1021
yourlocust ag_17415 90	gene:plasmid: 83372	81,180	356	2.89e -75	281.0	97,268	hypothetical protein	NA (sin proteína asociada)	pRSPA02	<i>R.</i> <i>sphaeroides</i> ATCC 17025
yourlocust ag_17429 80	gene:plasmid: 85726	79,325	237	2.41e -39	161.0	100,000	hypothetical protein	NA (sin proteína asociada)	pRSPA01	<i>R.</i> <i>sphaeroides</i> ATCC 17025
yourlocust ag_17578 40	gene:plasmid: 85759	77,273	242	9.12e -34	143.0	99,589	hypothetical protein	NA (sin proteína asociada)	pRSPA01	<i>R.</i> <i>sphaeroides</i> ATCC 17025
yourlocust ag_17724 60	gene:plasmid: 139080	72,848	302	2.72e -20	99.0	94,081	FKBP-type peptidyl- prolyl cis-trans isomerase SlyD	NA	megaplasmid	<i>S. pomeroyi</i> DSS-3
yourlocust ag_17884 00	gene:plasmid: 156439	89,610	77	5.04e -21	99.0	100,000	tRNA-Met	NA (sin proteína asociada)	pPNAP02	<i>P.</i> <i>naphthaleniv</i> <i>orans</i> CJ2
yourlocust ag_18075 90	gene:plasmid: 134955	94,444	90	3.67e -33	139.0	100,000	tRNA-Ser	NA (sin proteína asociada)	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_18141 60	gene:plasmid: 134872	82,310	277	1.35e -62	239.0	97,193	1,2-phenylacetyl-CoA epoxidase, subunit B	NA	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_18437 40	gene:plasmid: 154451	93,909	197	1.46e -80	298.0	99,495	hypothetical protein	transpositional DNA recombination; DDE-based	pFBAOT6	<i>Aeromonas</i> <i>punctata</i>

								<i>recombinase activity; Tn3 family</i>		
yourlocust ag_18591 80	gene:plasmid: 155206	90,909	77	1.4e- 21	100.0	100,000	<i>tRNA-Met</i>	NA (sin proteína asociada)	pSymA	<i>S. meliloti</i> 1021
yourlocust ag_18912 90	gene:plasmid: 138750	89,071	183	2.56e -58	224.0	91,045	<i>Ureidoglycolate lyase</i>	<i>transferase activity</i>	<i>megaplasmid</i>	<i>S. pomeroyi</i> DSS-3
yourlocust ag_19003 80	gene:plasmid: 138876	84,436	257	5.94e -66	250.0	94,139	<i>Geraniol dehydrogenase</i>	<i>oxidoreductase activity</i>	<i>megaplasmid</i>	<i>S. pomeroyi</i> DSS-3
yourlocust ag_19102 50	gene:plasmid: 158679	90,984	244	6.88e -90	329.0	95,686	<i>hypothetical protein</i>	NA	pSD25	<i>Ruegeria</i> sp. PR1b
yourlocust ag_19198 80	gene:plasmid: 134942	100,000	77	2.3e- 34	143.0	100,000	<i>tRNA-Ile</i>	NA (sin proteína asociada)	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_19222 60	gene:plasmid: 131376	97,403	77	4.97e -31	132.0	100,000	<i>tRNA-Met</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_04850	gene:plasmid: 13101	90,909	77	3.83e -22	102.0	100,000	<i>tRNA-Ala</i>	NA	pBM400	<i>B. megaterium</i>
yourlocust ag_05160	gene:plasmid: 22712	76,686	845	6.67e -130	464.0	96,132	<i>Nitrogenase iron protein</i>	<i>arsenite transmembrane- transporting ATPase activity</i>	p42d	<i>R. etli</i> CFN 42
yourlocust ag_39950	gene:plasmid: 131155	86,957	92	4.96e -22	102.0	100,000	<i>tRNA-Ser</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_80160	gene:plasmid: 91662	79,238	525	5.88e -99	361.0	90,206	<i>hypothetical protein</i>	NA (sin proteína asociada)	pSN254	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport str. SL254

yourlocust ag_88970	gene:plasmid: 134955	86,813	91	1.76e -21	100.0	100,000	tRNA-Ser	NA (sin proteína asociada)	TM1040	Silicibacter sp. TM1040
yourlocust ag_11426 0	gene:plasmid: 85668	76,734	1182	1.45e -177	623.0	97,044	Tryptophan synthase beta chain	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_16848 0	gene:plasmid: 134908	92,208	77	2.33e -24	110.0	100,000	tRNA-Pro	NA (sin proteína asociada)	TM1040	Silicibacter sp. TM1040
yourlocust ag_17233 0	gene:plasmid: 138740	77,011	957	4.29e -147	521.0	99,068	hypothetical protein	oxidoreductase activity	megaplasmid	S. pomeroyi DSS-3
yourlocust ag_21404 0	gene:plasmid: 134943	100,000	76	8.11e -34	141.0	100,000	tRNA-Ala	NA (sin proteína asociada)	TM1040	Silicibacter sp. TM1040
yourlocust ag_21533 0	gene:plasmid: 86219	81,349	252	5.48e -51	200.0	100,000	hypothetical protein	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_25748 0	gene:plasmid: 146149	88,546	227	2.97e -73	274.0	97,009	Cell wall-associated hydrolase	NA (sin proteína asociada)	1	Deinococcus geothermalis DSM 11300
yourlocust ag_27419 0	gene:plasmid: 127297	73,968	1648	1.22e -174	614.0	97,055	hypothetical protein	oxidoreductase activity	pRL8	Rhizobium leguminosaru m bv. viciae 3841
yourlocust ag_27740 0	gene:plasmid: 131376	89,474	76	1.81e -20	97.1	98,701	tRNA-Met	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_29572 0	gene:plasmid: 86016	78,686	1689	0.0	1083.0	100,000	Catalase-peroxidase	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_32357 0	gene:plasmid: 139557	86,207	87	2.77e -19	93.5	100,000	tRNA-Ser	NA (sin proteína asociada)	pGMI1000MP	R. solanacearum GMI1000
yourlocust ag_32795 0	gene:plasmid: 23616	84,199	462	4.69e -124	444.0	95,062	Nitrogenase molybdenum-iron protein alpha chain	nitrogenase activity	pNGR234a	Rhizobium sp. NGR234

yourlocust ag_33427 0	gene:plasmid: 151861	77,559	811	8.03e -134	477.0	98,303	Nitrogenase iron protein 1	arsenite transmembrane- transporting ATPase activity	megaplasmid	Desulfovibrio vulgaris subsp. vulgaris str. Hildenboroug h
yourlocust ag_34279 0	gene:plasmid: 137860	79,549	709	3.41e -142	505.0	92,680	transposase/IS protein	IS21 family; positive regulation of DNA transposition	pRL7	R. leguminosaru m bv. viciae 3841
yourlocust ag_38482 0	gene:plasmid: 83411	80,835	407	9.7e- 86	316.0	91,667	hypothetical protein	NA (sin proteína asociada)	pRSPA02	R. sphaeroides ATCC 17025
yourlocust ag_42000 0	gene:plasmid: 28436	83,019	212	8.07e -49	193.0	95,496	hypothetical protein	NA	pSymB	S. meliloti 1021
yourlocust ag_42969 0	gene:plasmid: 86220	82,946	516	2.51e -127	455.0	92,473	Phosphate-import permease protein PhnE	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_42970 0	gene:plasmid: 86219	81,000	300	3.05e -59	228.0	100,000	hypothetical protein	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_44027 0	gene:plasmid: 154739	78,294	926	8.95e -164	577.0	96,761	Acetyl-coenzyme A synthetase	oxidoreductase activity	1	P. denitrificans PD1222
yourlocust ag_44513 0	gene:plasmid: 141404	90,311	289	7.79e -105	379.0	99,313	hypothetical protein	NA (sin proteína asociada)	pBM400	B. megaterium
yourlocust ag_45142 0	gene:plasmid: 28525	77,138	573	3.9e- 86	318.0	90,952	Methylthioribose-1- phosphate isomerase	NA	pSymB	S. meliloti 1021
yourlocust ag_45221 0	gene:plasmid: 86123	78,400	625	1.72e -109	396.0	100,000	hypothetical protein	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_46822	gene:plasmid: 109387	84,674	261	3.55e -68	257.0	95,604	hypothetical protein	transporter activity	D	R. sphaeroides

0	2.4.1									
yourlocust ag_47529 0	gene:plasmid: 134942	88,235	68	4.89e -16	82.4	90,667	<i>tRNA-Ile</i>	NA (sin proteína asociada)	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_49209 0	gene:plasmid: 141404	89,686	223	1.41e -76	285.0	92,917	<i>hypothetical protein</i>	NA (sin proteína asociada)	pBM400	<i>B.</i> <i>megaterium</i>
yourlocust ag_50369 0	gene:plasmid: 28496	81,436	404	1.87e -87	322.0	100,000	<i>hypothetical protein</i>	<i>chromosome segregation; DNA translocase activity</i>	pSymB	<i>S. meliloti</i> 1021
yourlocust ag_51073 0	gene:plasmid: 135075	79,798	198	7.37e -34	143.0	98,508	<i>hypothetical protein</i>	NA	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_52241 0	gene:plasmid: 85663	76,600	906	3.36e -133	475.0	94,671	<i>hypothetical protein</i>	NA (sin proteína asociada)	pRSPA01	<i>R.</i> <i>sphaeroides</i> ATCC 17025
yourlocust ag_55499 0	gene:plasmid: 140775	84,219	640	9.73e -177	619.0	99,689	<i>hypothetical protein</i>	NA	p42e	<i>R. etli</i> CFN 42
yourlocust ag_57708 0	gene:plasmid: 125495	84,545	220	1.7e- 55	215.0	100,000	<i>hypothetical protein</i>	NA (sin proteína asociada)	C	<i>R.</i> <i>sphaeroides</i> 2.4.1
yourlocust ag_60776 0	gene:plasmid: 85828	81,358	869	0.0	701.0	100,000	<i>Transposase DDE domain protein</i>	NA (sin proteína asociada)	pRSPA01	<i>R.</i> <i>sphaeroides</i> ATCC 17025
yourlocust ag_61883 0	gene:plasmid: 141404	87,534	369	3.54e -119	427.0	100,000	<i>Cell wall-associated hydrolase</i>	NA (sin proteína asociada)	pBM400	<i>B.</i> <i>megaterium</i>
yourlocust ag_66543 0	gene:plasmid: 141405	84,786	723	0.0	706.0	90,488	<i>16S ribosomal RNA (partial)</i>	NA (sin proteína asociada)	pBM400	<i>B.</i> <i>megaterium</i>
yourlocust ag_69026 0	gene:plasmid: 154774	82,885	707	3.84e -176	617.0	100,000	<i>60 kDa chaperonin</i>	NA	1	<i>P.</i> <i>denitrificans</i> PD1222

yourlocust ag_69510 0	gene:plasmid: 85663	82,143	280	1.39e -62	239.0	95,238	hypothetical protein	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_70711 0	gene:plasmid: 154774	88,153	287	1.07e -93	342.0	94,720	hypothetical protein	NA	1	<i>P. denitrificans</i> PD1222
yourlocust ag_72864 0	gene:plasmid: 89236	81,315	289	3.77e -58	224.0	100,000	hypothetical protein	NA (sin proteína asociada)	pSMED01	<i>S. medicae</i> WSM419
yourlocust ag_78216 0	gene:plasmid: 141405	83,534	249	6.05e -61	233.0	90,217	hypothetical protein	NA (sin proteína asociada)	pBM400	<i>B. megaterium</i>
yourlocust ag_79885 0	gene:plasmid: 121712	83,279	610	2.04e -158	558.0	100,000	hypothetical protein	transferase activity, transferring phosphorus- containing groups	pTB2	<i>R. denitrificans</i> OCh 114
yourlocust ag_80347 0	gene:plasmid: 141405	81,696	224	5.45e -46	183.0	91,057	hypothetical protein	NA (sin proteína asociada)	pBM400	<i>B. megaterium</i>
yourlocust ag_80391 0	gene:plasmid: 86220	79,174	581	1.07e -106	387.0	90,498	Phosphate-import permease protein PhnE	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_80673 0	gene:plasmid: 154739	83,595	573	3.47e -151	534.0	93,171	Acetyl-coenzyme A synthetase	oxidoreductase activity	1	<i>P. denitrificans</i> PD1222
yourlocust ag_80864 0	gene:plasmid: 154647	88,435	147	1.42e -44	178.0	100,000	hypothetical protein	NA	1	<i>P. denitrificans</i> PD1222
yourlocust ag_81985 0	gene:plasmid: 82529	84,672	685	0.0	671.0	93,579	hypothetical protein	NA (sin proteína asociada)	pSMED02	<i>S. medicae</i> WSM419
yourlocust ag_87657 0	gene:plasmid: 134649	77,536	414	5.84e -63	241.0	95,833	Succinate dehydrogenase iron- sulfur subunit	NA	TM1040	<i>Silicibacter</i> sp. TM1040
yourlocust ag_88483	gene:plasmid: 13101	93,421	76	1.77e -25	113.0	100,000	tRNA-Ala	NA	pBM400	<i>B. megaterium</i>

0										
yourlocust ag_91415 0	gene:plasmid: 131155	86,957	92	4.96e -22	102.0	100,000	tRNA-Ser	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_93589 0	gene:plasmid: 141405	86,752	234	2.55e -69	261.0	91,765	hypothetical protein	NA (sin proteína asociada)	pBM400	B. megaterium
yourlocust ag_95116 0	gene:plasmid: 85828	86,446	332	2.57e -100	364.0	97,935	hypothetical protein	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_95251 0	gene:plasmid: 138740	83,607	122	1.14e -24	111.0	100,000	S- (hydroxymethyl)gluta thione dehydrogenase	oxidoreductase activity	megaplasmid	S. pomeroyi DSS-3
yourlocust ag_95928 0	gene:plasmid: 158676	88,971	136	7.91e -42	169.0	98,551	hypothetical protein	DDE-based recombinase activity; transpositional DNA recombination; IS5 family	pSD25	Ruegeria sp. PR1b
yourlocust ag_96079 0	gene:plasmid: 134955	87,500	88	1.63e -21	100.0	100,000	tRNA-Ser	NA (sin proteína asociada)	TM1040	Silicibacter sp. TM1040
yourlocust ag_96377 0	gene:plasmid: 131155	86,667	90	1.73e -21	100.0	100,000	tRNA-Ser	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_96681 0	gene:plasmid: 28436	85,156	128	1.07e -30	132.0	90,780	hypothetical protein	NA	pSymB	S. meliloti 1021
yourlocust ag_97479 0	gene:plasmid: 86138	86,328	256	2.53e -74	278.0	100,000	hypothetical protein	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_98440 0	gene:plasmid: 87643	86,283	226	6.29e -65	246.0	99,123	ISXO2-like transposase domain protein	NA (sin proteína asociada)	pACRY03	A. cryptum JF-5

yourlocust ag_10438 20	gene:plasmid: 137860	82,203	236	4.29e -52	204.0	93,651	transposase/IS protein	IS21 family; positive regulation of DNA transposition	pRL7	R. leguminosaru m bv. viciae 3841
yourlocust ag_10536 40	gene:plasmid: 141403	83,036	112	2.93e -20	97.1	100,000	5S ribosomal RNA	NA (sin proteína asociada)	pBM400	B. megaterium
yourlocust ag_10543 20	gene:plasmid: 131375	87,273	110	1.3e- 28	124.0	100,000	5S ribosomal RNA	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_10543 30	gene:plasmid: 134946	96,104	77	2.31e -29	126.0	100,000	tRNA-Met	NA (sin proteína asociada)	TM1040	Silicibacter sp. TM1040
yourlocust ag_10592 10	gene:plasmid: 154774	85,825	388	3.75e -114	411.0	100,000	hypothetical protein	NA	1	P. denitrificans PD1222
yourlocust ag_10921 20	gene:plasmid: 81467	81,139	281	1.1e- 58	226.0	94,613	Glutamine transport ATP-binding protein GlnQ	NA (sin proteína asociada)	pSMED03	S. medicae WSM419
yourlocust ag_11002 90	gene:plasmid: 134944	83,007	153	9.28e -32	135.0	98,077	hypothetical protein	NA (sin proteína asociada)	TM1040	Silicibacter sp. TM1040
yourlocust ag_11071 90	gene:plasmid: 91662	84,022	363	2.91e -95	348.0	96,032	hypothetical protein	NA (sin proteína asociada)	pSN254	Salmonella enterica subsp. enterica serovar Newport str. SL254
yourlocust ag_11078 70	gene:plasmid: 125933	89,916	119	1.87e -37	154.0	99,167	hypothetical protein	electron transporter, transferring electrons within cytochrome c oxidase complex activity	p42f	R. etli CFN 42

yourlocust ag_11254 90	gene:plasmid: 134974	96,552	87	2.63e -34	143.0	100,000	tRNA-Leu	NA (sin proteína asociada)	TM1040	Silicibacter sp. TM1040
yourlocust ag_11568 50	gene:plasmid: 85523	83,590	390	3.87e -99	361.0	100,000	hypothetical protein	NA (sin proteína asociada)	pXAUT01	X. autotrophicus Py2
yourlocust ag_11611 20	gene:plasmid: 139562	81,743	241	2.69e -49	195.0	92,337	hypothetical protein	NA (sin proteína asociada)	pGMI1000MP	R. solanacearum GMI1000
yourlocust ag_11885 50	gene:plasmid: 141405	89,362	235	2.41e -79	294.0	95,529	hypothetical protein	NA (sin proteína asociada)	pBM400	B. megaterium
yourlocust ag_12533 10	gene:plasmid: 83247	83,459	133	5.92e -28	122.0	100,000	hypothetical protein	NA (sin proteína asociada)	pSWIT01	S. wittichii RW1
yourlocust ag_12613 10	gene:plasmid: 138839	89,498	219	7.67e -74	276.0	100,000	Translation initiation factor IF-1	NA	megaplasmid	S. pomeroyi DSS-3
yourlocust ag_12656 30	gene:plasmid: 13101	92,208	77	8.22e -24	108.0	100,000	tRNA-Ala	NA	pBM400	B. megaterium
yourlocust ag_12682 50	gene:plasmid: 138851	80,211	379	1.45e -73	276.0	96,438	3-oxoacyl-[acyl- carrier-protein] synthase 1	oxidoreductase activity	megaplasmid	S. pomeroyi DSS-3
yourlocust ag_12749 70	gene:plasmid: 141404	81,860	215	6.09e -45	180.0	99,537	hypothetical protein	NA (sin proteína asociada)	pBM400	B. megaterium
yourlocust ag_12814 70	gene:plasmid: 134946	93,506	77	1.8e- 25	113.0	100,000	tRNA-Met	NA (sin proteína asociada)	TM1040	Silicibacter sp. TM1040
yourlocust ag_13370 50	gene:plasmid: 125495	80,741	405	3.18e -85	315.0	100,000	hypothetical protein	NA (sin proteína asociada)	C	R. sphaeroides 2.4.1
yourlocust ag_13457 70	gene:plasmid: 154647	80,347	346	1.27e -68	259.0	100,000	hypothetical protein	NA	1	P. denitrificans PD1222

yourlocust ag_14599 20	gene:plasmid: 131858	91,892	74	3.83e -22	102.0	97,368	tRNA-Asn	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_14665 20	gene:plasmid: 154916	82,787	244	2.08e -50	198.0	93,487	Bifunctional protein Fold protein	NA	1	P. denitrificans PD1222
yourlocust ag_14916 40	gene:plasmid: 138981	82,703	370	1.39e -88	326.0	96,354	3-hydroxy-D- aspartate aldolase	NA	megaplasmid	S. pomeroyi DSS-3
yourlocust ag_15105 40	gene:plasmid: 158651	88,529	340	4.33e -113	407.0	97,701	Soluble lytic murein transglycosylase precursor	lytic transglycosylase activity	pSD25	Ruegeria sp. PR1b
yourlocust ag_15110 60	gene:plasmid: 27836	80,905	398	4.11e -84	311.0	98,272	Acyl-CoA dehydrogenase, short-chain specific	oxidoreductase activity	pSymB	S. meliloti 1021
yourlocust ag_15739 00	gene:plasmid: 93523	82,781	302	5.16e -72	270.0	98,693	hypothetical protein	NA (sin proteína asociada)	pRSPA05	R. sphaeroides ATCC 17025
yourlocust ag_15853 10	gene:plasmid: 129361	80,083	241	1.24e -42	172.0	93,411	hypothetical protein	NA	p42c	R. etli CFN 42
yourlocust ag_15883 60	gene:plasmid: 135602	91,765	85	4.43e -27	119.0	100,000	tRNA-Tyr	NA (sin proteína asociada)	pSbal01	S. baltica OS155
yourlocust ag_16044 40	gene:plasmid: 86016	81,892	370	3.75e -84	311.0	99,462	Catalase-peroxidase	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_16098 90	gene:plasmid: 134649	87,984	258	1.22e -82	305.0	96,629	Succinate dehydrogenase iron- sulfur subunit	NA	TM1040	Silicibacter sp. TM1040
yourlocust ag_16242 70	gene:plasmid: 131376	89,610	77	1.81e -20	97.1	100,000	tRNA-Met	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025
yourlocust ag_16254 60	gene:plasmid: 131375	91,743	109	6.03e -37	152.0	99,091	5S ribosomal RNA	NA (sin proteína asociada)	pRSPA01	R. sphaeroides ATCC 17025

yourlocust ag_16461 40	gene:plasmid: 92721	82,353	255	4.43e -57	220.0	97,701	hypothetical protein	NA (sin proteína asociada)	pNL2	<i>N. aromaticivor ans</i> DSM 12444
yourlocust ag_16608 60	gene:plasmid: 125270	87,037	216	2.13e -64	244.0	100,000	hypothetical protein	intercellular transfer by conjugation; Type IV secretion complex VirB4 family; type IV protein secretion/mating pair formation system complex	unnamed	<i>Silicibacter</i> sp. TM1040
yourlocust ag_17141 50	gene:plasmid: 26090	84,828	145	1.54e -34	145.0	92,949	hypothetical protein	NA	pSymA	<i>S. meliloti</i> 1021
yourlocust ag_17928 70	gene:plasmid: 131376	89,474	76	1.81e -20	97.1	98,701	tRNA-Met	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_17958 30	gene:plasmid: 158676	85,507	138	1.33e -34	145.0	100,000	hypothetical protein	DDE-based recombinase activity; transpositional DNA recombination; IS5 family	pSD25	<i>Ruegeria</i> sp. PR1b
yourlocust ag_18321 60	gene:plasmid: 137860	81,858	226	1.15e -47	189.0	93,004	transposase/IS protein	IS21 family; positive regulation of DNA transposition	pRL7	<i>R. leguminosaru m</i> bv. <i>viciae</i> 3841
yourlocust ag_18336 70	gene:plasmid: 125933	92,188	128	9.37e -46	182.0	99,225	hypothetical protein	electron transporter, transferring electrons within cytochrome c	p42f	<i>R. etli</i> CFN 42

								<i>oxidase complex activity</i>		
yourlocust ag_18449 30	gene:plasmid: 131155	85,556	90	2.9e- 19	93.5	100,000	<i>tRNA-Ser</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_18901 70	gene:plasmid: 91662	88,435	294	1.75e- 96	351.0	98,000	<i>hypothetical protein</i>	NA (sin proteína asociada)	pSN254	<i>Salmonella enterica</i> subsp. <i>enterica</i> <i>serovar</i> Newport str. SL254
yourlocust ag_18910 20	gene:plasmid: 89088	82,270	141	6.74e- 28	122.0	95,918	<i>hypothetical protein</i>	NA (sin proteína asociada)	pSMED01	<i>S. medicae</i> WSM419
yourlocust ag_19211 80	gene:plasmid: 131155	84,783	92	1.07e- 18	91.6	100,000	<i>tRNA-Ser</i>	NA (sin proteína asociada)	pRSPA01	<i>R. sphaeroides</i> ATCC 17025
yourlocust ag_19215 50	gene:plasmid: 154739	85,795	176	3.19e- 47	187.0	91,667	<i>hypothetical protein</i>	<i>oxidoreductase activity</i>	1	<i>P. denitrificans</i> PD1222
yourlocust ag_19222 50	gene:plasmid: 138740	86,066	122	1.13e- 29	128.0	100,000	<i>S-(hydroxymethyl)gluta thione dehydrogenase</i>	<i>oxidoreductase activity</i>	<i>megaplasmid</i>	<i>S. pomeroyi</i> DSS-3

ID: identificación

NA: no asignada

*Prokka: software de anotación rápida de genomas procariotas

**ACLAME (*A CLAssification of Mobile genetic Elements*): base de datos de elementos genéticos móviles

Tabla A3. Genes plasmídicos de la base de datos ACLAME presentes en el plasmidoma de Laguna Diamante.

ID Prokka*	ID ACLAME**	Identidad (%)	Longitud (pb)	E-value	Bitscore	Cobertura (%)	Anotación de Prokka	Función de Aclame	Plásmido	Huésped
JDIPpm3_00700	gene:plasmid:16634	86,585	410	9.37e-126	449.0	90,508	hypothetical protein	NA	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_02610	gene:plasmid:16870	90,663	996	0.0	1319.0	100,000	hypothetical protein	NA	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_02800	gene:plasmid:112672	89,333	675	0.0	848.0	100,000	hypothetical protein	NA	pNG600	<i>Haloarcula marismortui</i> ATCC 43049
JDIPpm3_02820	gene:plasmid:112656	92,308	819	0.0	1164.0	100,000	hypothetical protein	ATPase activity; plasmid partitioning	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_02830	gene:plasmid:112640	90,441	408	1.74e-152	538.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_05690	gene:plasmid:16796	85,793	2182	0.0	2300.0	100,000	RNA polymerase-associated protein <i>RapA</i>	DNA helicase activity	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_05700	gene:plasmid:16797	80,510	549	4.16e-115	414.0	100,000	TATA-box-binding protein 2	transcription factor activity	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_06330	gene:plasmid:140315	79,755	815	6.4e-165	580.0	92,090	hypothetical protein	NA	pNG400	<i>H. marismortui</i> ATCC 43049
JDIPpm3_10560	gene:plasmid:16820	79,935	1849	0.0	1334.0	96,002	Putative arsenical pump-driving ATPase	arsenite transmembrane-transporting ATPase activity	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_13990	gene:plasmid:112576	83,993	1618	0.0	1550.0	99,142	RNA polymerase-associated protein <i>RapA</i>	DNA helicase activity	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_15070	gene:plasmid:137546	73,645	664	7.07e-64	244.0	100,000	hypothetical protein	NA	PL131	<i>Natronomonas pharaonis</i> DSM 2160
JDIPpm3_15100	gene:plasmid:137545	80,333	961	0.0	710.0	95,622	hypothetical protein	NA	PL131	<i>N. pharaonis</i> DSM 2160

JDIPpm3_17160	gene:plasmid:128518	78,076	1081	0.0	665.0	97,387	Agmatinase	amino acid catabolic process	pNG700	<i>H. marismortui</i> ATCC 43049
JDIPpm3_18620	gene:plasmid:16594	83,333	1284	0.0	1170.0	100,000	hypothetical protein	IS4 family; transpositional DNA recombination	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_18900	gene:plasmid:128415	84,533	2140	0.0	2106.0	99,767	Aminomethyltransferase	oxidoreductase activity; D-nopaline dehydrogenase activity	pNG700	<i>H. marismortui</i> ATCC 43049
JDIPpm3_18910	gene:plasmid:128570	82,848	2542	0.0	2272.0	99,569	Aminomethyltransferase	oxidoreductase activity; D-nopaline dehydrogenase activity	pNG700	<i>H. marismortui</i> ATCC 43049
JDIPpm3_20860	gene:plasmid:16479	82,132	319	1.98e-71	268.0	98,457	hypothetical protein	NA	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_20880	gene:plasmid:112585	77,949	712	2.74e-123	442.0	91,990	DNA polymerase 1	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_21290	gene:plasmid:16555	79,309	2083	0.0	1402.0	95,902	Catalase-peroxidase 2	NA	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_24050	gene:plasmid:16853	78,158	380	4.19e-59	228.0	95,238	hypothetical protein	NA	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_31770	gene:plasmid:112573	86,152	3293	0.0	3544.0	99,970	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_31780	gene:plasmid:112574	91,667	240	4.97e-91	333.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_32850	gene:plasmid:128371	90,678	118	1.79e-37	154.0	100,000	5S ribosomal RNA	NA (sin proteína asociada)	pNG700	<i>H. marismortui</i> ATCC 43049
JDIPpm3_41990	gene:plasmid:112575	77,306	2851	0.0	1657.0	97,973	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_42400	gene:plasmid:112577	83,437	483	1.02e-125	449.0	98,171	ORC1-type DNA replication protein 2	DNA replication	pNG600	<i>H. marismortui</i> ATCC 43049

	initiation									
JDIPpm3_47930	gene:plasmid:112580	92,910	1227	0.0	1784.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_49640	gene:plasmid:137472	78,593	981	2.0e-180	632.0	96,460	hypothetical protein	NA	PL131	<i>N. pharaonis</i> DSM 2160
JDIPpm3_52260	gene:plasmid:128524	81,602	1511	0.0	1242.0	100,000	Formate-- tetrahydrofolate ligase	NA	pNG700	<i>H. marismortui</i> ATCC 43049
JDIPpm3_52270	gene:plasmid:128299	82,594	879	0.0	765.0	92,429	Formate dehydrogenase subunit alpha	oxidoreductase activity	pNG700	<i>H. marismortui</i> ATCC 43049
JDIPpm3_55240	gene:plasmid:16790	79,259	1080	0.0	750.0	99,723	UvrABC system protein B	DNA helicase activity; DNA repair	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_55450	gene:plasmid:112645	84,906	530	9.1e-152	536.0	91,065	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_55460	gene:plasmid:112652	88,976	381	1.67e-132	472.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_65110	gene:plasmid:16879	84,777	1143	0.0	1140.0	90,499	hypothetical protein	NA	pZMX101	<i>Halorubrum</i> <i>saccharovorum</i>
JDIPpm3_66340	gene:plasmid:112616	85,894	397	1.78e-117	422.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_66730	gene:plasmid:133430	73,016	882	4.54e-77	289.0	95,146	hypothetical protein	NA	pNG200	<i>H. marismortui</i> ATCC 43049
JDIPpm3_67670	gene:plasmid:124010	79,954	1317	0.0	952.0	100,000	hypothetical protein	IS605 family; transposition, DNA-mediated; tyrosine-based recombinase activity	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_69020	gene:plasmid:128373	75,778	1028	2.37e-140	499.0	90,653	Carnitine monooxygenase oxygenase subunit	dioxygenase activity	pNG700	<i>H. marismortui</i> ATCC 43049
JDIPpm3_74730	gene:plasmid:126269	82,338	402	1.12e-94	346.0	100,000	hypothetical protein	NA	PL47	<i>Haloquadratum</i> <i>walsbyi</i> DSM 16790
JDIPpm3_76130	gene:plasmid:16699	83,481	339	7.3e-86	316.0	100,000	hypothetical protein	NA	pNRC100	<i>Halobacterium</i> sp. NRC-1

JDIPpm3_76140	gene:plasmid:16698	87,082	449	1.51e-143	508.0	99,778	hypothetical protein	NA	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_78070	gene:plasmid:16813	75,275	546	1.77e-64	246.0	91,000	hypothetical protein	NA	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_81660	gene:plasmid:112664	84,738	439	2.0e-122	438.0	98,210	hypothetical protein	site-specific DNA recombination; tyrosine-based recombinase activity	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_82270	gene:plasmid:112736	86,335	483	1.64e-148	525.0	98,773	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_82280	gene:plasmid:112668	88,406	552	0.0	665.0	100,000	hypothetical protein	plasmid function unknown	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_84560	gene:plasmid:123958	88,048	1347	0.0	1585.0	100,000	hypothetical protein	NA	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_90930	gene:plasmid:16614	78,958	1036	0.0	691.0	94,872	hypothetical protein	DNA endonuclease activity	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_91830	gene:plasmid:16605	96,244	426	0.0	699.0	99,301	hypothetical protein	IS605 family; transposition, DNA-mediated; tyrosine-based recombinase activity	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_98190	gene:plasmid:16566	80,303	528	2.51e-107	388.0	97,238	Aspartate carbamoyltransferase	NA	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_105870	gene:plasmid:112615	85,542	332	2.66e-95	348.0	95,402	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_108710	gene:plasmid:16546	84,307	873	0.0	852.0	98,980	hypothetical protein	NA	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_108910	gene:plasmid:16804	91,422	886	0.0	1214.0	100,000	hypothetical protein	NA	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_109000	gene:plasmid:137573	95,819	885	0.0	1430.0	100,000	hypothetical protein	NA	PL131	<i>N. pharaonis</i> DSM 2160

JDIPpm3_109220	gene:plasmid:123986	81,975	638	2.77e-152	538.0	100,000	hypothetical protein	NA	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_112190	gene:plasmid:123914	82,317	492	4.79e-119	427.0	100,000	hypothetical protein	transpositional DNA recombination; DDE-based recombinase activity; IS630 family	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_114420	gene:plasmid:112613	85,070	355	3.49e-99	361.0	100,000	hypothetical protein	DNA endonuclease activity	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_118790	gene:plasmid:137497	79,625	427	3.39e-80	298.0	99,534	hypothetical protein	IS605 family; transposition, DNA-mediated; tyrosine-based recombinase activity	PL131	<i>N. pharaonis</i> DSM 2160
JDIPpm3_128540	gene:plasmid:16488	78,627	641	2.97e-117	422.0	98,920	hypothetical protein	chaperone activity	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_129560	gene:plasmid:137536	92,697	178	2.33e-68	257.0	95,699	hypothetical protein	NA	PL131	<i>N. pharaonis</i> DSM 2160
JDIPpm3_129790	gene:plasmid:16844	88,440	545	0.0	658.0	96,120	hypothetical protein	plasmid function unknown	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_130460	gene:plasmid:111278	72,452	628	5.31e-45	182.0	100,000	hypothetical protein	NA	pRHL1	<i>Rhodococcus</i> sp. RHA1
JDIPpm3_134350	gene:plasmid:16825	94,697	132	5.72e-53	206.0	100,000	hypothetical protein	NA	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_135440	gene:plasmid:124006	93,827	405	3.62e-174	610.0	100,000	hypothetical protein	IS605 family; transposition, DNA-mediated; tyrosine-based recombinase activity	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_137520	gene:plasmid:112581	96,082	536	0.0	874.0	99,814	hypothetical protein	NA	pNG600	<i>H. marismortui</i>

										ATCC 43049
JDIPpm3_140060	gene:plasmid:112624	89,226	297	1.33e-102	372.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_140070	gene:plasmid:112696	92,262	168	7.5e-63	239.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_140400	gene:plasmid:112595	87,702	618	0.0	721.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_143710	gene:plasmid:128558	81,250	368	1.49e-78	292.0	90,196	hypothetical protein	oxidoreductase activity	pNG700	<i>H. marismortui</i> ATCC 43049
JDIPpm3_144510	gene:plasmid:16811	72,155	413	8.17e-22	104.0	93,651	hypothetical protein	transcription factor activity	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_146720	gene:plasmid:137567	93,210	162	7.19e-63	239.0	100,000	hypothetical protein	NA	PL131	<i>N. pharaonis</i> DSM 2160
JDIPpm3_152160	gene:plasmid:16843	84,459	444	1.21e-119	429.0	98,667	hypothetical protein	transposase activity; transpositional DNA recombination; IS66 family	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_153590	gene:plasmid:16472	87,273	495	1.27e-159	562.0	100,000	hypothetical protein	IS605 family; transposition, DNA-mediated; tyrosine-based recombinase activity	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_156100	gene:plasmid:16712	86,979	192	4.5e-56	217.0	91,429	hypothetical protein	gas vesicle	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_161120	gene:plasmid:112605	90,237	338	1.15e-123	442.0	99,705	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_167290	gene:plasmid:123946	83,641	379	1.74e-97	355.0	100,000	hypothetical protein	IS110 family; site-specific DNA recombination	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_170330	gene:plasmid:16780	87,500	184	1.99e-54	211.0	91,542	hypothetical protein	plasmid function unknown	pNRC100	<i>Halobacterium</i> sp. NRC-1

JDIPpm3_179210	gene:plasmid:16863	82,166	314	2.05e-71	268.0	93,452	hypothetical protein	hydrolase activity	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_180360	gene:plasmid:16553	82,732	388	1.39e-93	342.0	100,000	hypothetical protein	DDE-based recombinase activity; transpositional DNA recombination; IS6 family	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_184650	gene:plasmid:123899	95,062	243	4.94e-106	383.0	100,000	hypothetical protein	NA	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_186260	gene:plasmid:137505	88,060	201	9.29e-63	239.0	98,529	hypothetical protein	NA	PL131	<i>N. pharaonis</i> DSM 2160
JDIPpm3_195960	gene:plasmid:137529	90,625	288	5.95e-106	383.0	100,000	hypothetical protein	IS5 family; DDE-based recombinase activity; transpositional DNA recombination	PL131	<i>N. pharaonis</i> DSM 2160
JDIPpm3_207970	gene:plasmid:16445	80,297	269	4.63e-52	204.0	99,630	hypothetical protein	IS4 family; transpositional DNA recombination	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_00650	gene:plasmid:112605	89,644	309	3.88e-108	390.0	99,039	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_16130	gene:plasmid:16594	79,827	1274	0.0	917.0	97,625	hypothetical protein	IS4 family; transpositional DNA recombination	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_27780	gene:plasmid:124015	85,343	423	2.44e-121	435.0	100,000	hypothetical protein	NA	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_30210	gene:plasmid:112672	86,095	676	0.0	726.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_32940	gene:plasmid:16796	78,231	2917	0.0	1827.0	99,624	RNA polymerase-associated protein	DNA helicase activity	pNRC100	<i>Halobacterium</i> sp. NRC-1

							<i>RapA</i>			
JDIPpm3_34060	gene:plasmid:16870	79,055	974	0.0	647.0	98,087	hypothetical protein	NA	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_38140	gene:plasmid:16796	89,891	366	1.6e-132	472.0	100,000	hypothetical protein	DNA helicase activity	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_41240	gene:plasmid:16606	87,799	418	6.59e-137	486.0	99,524	hypothetical protein	NA	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_50890	gene:plasmid:112576	82,438	1583	0.0	1378.0	99,372	RNA polymerase-associated protein	DNA helicase activity	pNG600	<i>H. marismortui</i> ATCC 43049
							<i>RapA</i>			
JDIPpm3_57580	gene:plasmid:16870	79,055	931	4.26e-177	621.0	93,193	hypothetical protein	NA	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_63380	gene:plasmid:16796	82,044	724	1.42e-175	616.0	100,000	hypothetical protein	DNA helicase activity	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_71110	gene:plasmid:112573	83,072	1465	0.0	1323.0	99,660	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_72410	gene:plasmid:112575	77,188	1280	0.0	732.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_84950	gene:plasmid:140349	81,519	1277	0.0	1046.0	100,000	hypothetical protein	IS4 family; transpositional DNA recombination	pNG400	<i>H. marismortui</i> ATCC 43049
JDIPpm3_98550	gene:plasmid:16796	86,498	474	2.85e-146	518.0	93,491	hypothetical protein	DNA helicase activity	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_100660	gene:plasmid:16847	90,210	429	1.41e-158	558.0	100,000	hypothetical protein	IS4 family; transpositional DNA recombination	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_102150	gene:plasmid:128299	84,709	412	3.99e-114	411.0	100,000	hypothetical protein	oxidoreductase activity	pNG700	<i>H. marismortui</i> ATCC 43049
JDIPpm3_102610	gene:plasmid:112573	89,163	609	0.0	760.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_102620	gene:plasmid:112573	87,234	329	1.18e-103	375.0	97,917	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_113110	gene:plasmid:112576	92,208	231	2.22e-89	327.0	100,000	hypothetical protein	DNA helicase activity	pNG600	<i>H. marismortui</i> ATCC 43049

JDIPpm3_118400	gene:plasmid:16591	85,370	581	1.65e-169	595.0	90,078	hypothetical protein	NA	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_124880	gene:plasmid:16804	90,877	285	5.88e-106	383.0	100,000	hypothetical protein	NA	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_124970	gene:plasmid:16870	84,514	381	4.97e-103	374.0	97,692	hypothetical protein	NA	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_135470	gene:plasmid:124015	86,353	425	4.02e-129	460.0	100,000	hypothetical protein	NA	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_142640	gene:plasmid:124002	88,367	447	2.48e-151	534.0	100,000	hypothetical protein	NA	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_148460	gene:plasmid:16591	87,196	453	3.27e-145	514.0	100,000	hypothetical protein	NA	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_148820	gene:plasmid:16729	93,007	429	3.81e-179	627.0	100,000	hypothetical protein	IS4 family; transpositional DNA recombination	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_149110	gene:plasmid:16653	85,099	557	5.11e-159	560.0	100,000	hypothetical protein	NA	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_155520	gene:plasmid:124002	90,000	270	5.62e-96	350.0	100,000	hypothetical protein	NA	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_156780	gene:plasmid:112616	85,642	397	8.27e-116	416.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_157820	gene:plasmid:112576	87,589	282	2.76e-89	327.0	100,000	hypothetical protein	DNA helicase activity	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_161160	gene:plasmid:124007	93,197	294	5.96e-121	433.0	100,000	hypothetical protein	IS605 family; transposition, DNA-mediated; tyrosine-based recombinase activity	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_164680	gene:plasmid:112577	90,196	306	6.98e-111	399.0	91,071	hypothetical protein	DNA replication initiation	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_167260	gene:plasmid:112575	95,669	254	8.59e-114	409.0	99,608	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_167970	gene:plasmid:112573	89,640	222	4.66e-	283.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i>

				76						ATCC 43049
JDIPpm3_168910	gene:plasmid:112616	83,627	397	1.82e-102	372.0	100,000	hypothetical protein	NA	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_174820	gene:plasmid:112613	88,947	190	1.14e-61	235.0	97,436	hypothetical protein	DNA endonuclease activity	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_174970	gene:plasmid:123942	86,408	206	7.36e-59	226.0	99,517	hypothetical protein	NA	pNG500	<i>H. marismortui</i> ATCC 43049
JDIPpm3_176770	gene:plasmid:112594	87,671	292	3.96e-93	340.0	93,590	hypothetical protein	DNA replication initiation; DNA primase activity	pNG600	<i>H. marismortui</i> ATCC 43049
JDIPpm3_177290	gene:plasmid:16591	87,783	442	8.85e-146	516.0	100,000	hypothetical protein	NA	pNRC200	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_193270	gene:plasmid:16796	91,720	157	9.16e-57	219.0	98,742	hypothetical protein	DNA helicase activity	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_204260	gene:plasmid:16831	89,815	216	2.2e-74	278.0	96,000	hypothetical protein	NA	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_207190	gene:plasmid:16847	91,935	248	1.85e-95	348.0	99,598	hypothetical protein	IS4 family; transpositional DNA recombination	pNRC100	<i>Halobacterium</i> sp. NRC-1
JDIPpm3_208620	gene:plasmid:16793	84,574	188	3.19e-47	187.0	97,917	hypothetical protein	plasmid function unknown	pNRC100	<i>Halobacterium</i> sp. NRC-1

ID: identificación

NA: no asignada

*Prokka: software de anotación rápida de genomas procariotas

**ACLAME (*A CLAssification of Mobile genetic Elements*): base de datos de elementos genéticos móviles

