



Centro de Investigación Científica de Yucatán, A. C.

Posgrado en Ciencias Biológicas

**METAGENÓMICA FUNCIONAL EN EL ACUÍFERO DE  
YUCATÁN PARA LA BÚSQUEDA DE PROTEASAS  
NOVEDOSAS**

Tesis que presenta

**MAX MIZRAÍM APOLINAR HERNÁNDEZ**

En opción al título de

**DOCTOR EN CIENCIAS**

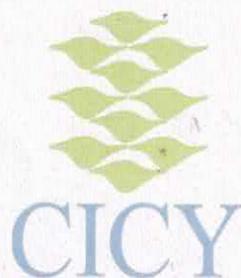
(Ciencias Biológicas: Opción Biotecnología)

Mérida, Yucatán. México

2016



*CENTRO DE INVESTIGACIÓN CIENTÍFICA DE YUCATÁN, A. C.  
POSGRADO EN CIENCIAS BIOLÓGICAS*



**RECONOCIMIENTO**

Por medio de la presente, hago constar que el trabajo de tesis de **Max Mizraím Apolinar Hernández** titulado "Metagenómica funcional en el acuífero de Yucatán para la búsqueda de proteasas novedosas" fue realizado en la Unidad de Biotecnología, línea de investigación Agrobiotecnología, laboratorio de metagenómica del Centro de Investigación Científica de Yucatán, A.C. bajo la dirección del(a) Dr.(a) Aileen O'Connor Sanchez y Yuri Jorge Jesús Peña Ramírez, dentro de la opción de Biotecnología perteneciente al Programa de Posgrado en Ciencias Biológicas de este Centro

Atentamente

---

Dr. Manuel Martínez Estevez  
Director de Docencia

Mérida, Yucatán, México, a 18 de Noviembre de 2016



## DECLARACIÓN DE PROPIEDAD

Declaro que la información contenida en la sección de Materiales y Métodos Experimentales, los Resultados y Discusión de este documento proviene de las actividades de experimentación realizadas durante el período que se me asignó para desarrollar mi trabajo de tesis, en las Unidades y Laboratorios del Centro de Investigación Científica de Yucatán, A.C., y que a razón de lo anterior y en contraprestación de los servicios educativos o de apoyo que me fueron brindados, dicha información, en términos de la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, le pertenece patrimonialmente a dicho Centro de Investigación. Por otra parte, en virtud de lo ya manifestado, reconozco que de igual manera los productos intelectuales o desarrollos tecnológicos que deriven o pudieran derivar de lo correspondiente a dicha información, le pertenecen patrimonialmente al Centro de Investigación Científica, A.C., y en el mismo tenor, reconozco que si derivaren de este trabajo productos intelectuales o desarrollos tecnológicos, en lo especial, estos se registrarán en todo caso por lo dispuesto por la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, en el tenor de lo expuesto en la presente Declaración.

Firma: \_\_\_\_\_



Nombre: M. en C. Max Mizraim Apolinar Hernández



Este trabajo se llevó a cabo en la Unidad de Biotecnología del Centro de Investigación Científica de Yucatán, y forma parte del proyecto titulado Aprovechamiento Sustentable Del Ácuifero De Yucatán Mediante Tecnología Metagenómica Para La Búsqueda De Productos Biotecnológicos Con Potencial De Alto Impacto Económico bajo la Co-dirección de la Dra. Aileen O'Connor Sabchez (Centro de Investigación científica de Yucatán) y el Dr. Yuri Jorge Jesús Peña Ramírez (El Colegio de la Frontera Sur - Unidad Campeche).

Este proyecto fue financiado por FOMIX-Yucatán proyecto N° 165026 Infraestructura-CONACYT 269833, y Fiscal-CICY 1039200015.



## **AGRADECIMIENTOS**

Al Centro de Investigación Científica de Yucatán y a la Unidad de Biotecnología por las instalaciones prestadas para el desarrollo de la parte experimental de este trabajo.

Al Consejo Nacional de Ciencia y Tecnología por la beca otorgada N° 228239.

Al proyecto FOMIX-Yucatán 165026 Titulado: Aprovechamiento Sustentable Del Ácuifero De Yucatán Mediante Tecnología Metagenómica Para La Búsqueda De Productos Biotecnológicos Con Potencial De Alto Impacto Económico, por el financiamiento del proyecto realizado en este trabajo.

A mis asesores la Dra. Aileen O'Connor Sanchez (Centro de Investigación Científica de Yucatán) y al Dr. Yuri J. Peña Ramírez (El Colegio de la Frontera Sur - Unidad Campeche) por su amistad y todo su apoyo incondicional tanto en forma académica como personal, durante la realización de este proyecto.

A mi comité tutorial y revisor de tesis; Dra. Aileen O'Connor, Dra. Blondi B. Canto Canche, Dr. Felipe Vazquez Flota, Dr. Yuri J. Peña Ramirez, Dra. Luisa A. López Ochoa, Dra. Alejandra Prieto Davo y Dr. Luis A. Saenz Carbonel, por todas las críticas constructivas que me hicieron crecer como persona y profesionalmente y por todos sus comentarios y observaciones realizadas para mejorar mi escrito de tesis.

Un agradecimiento especial al Dr. Cesar de los santos Briones por toda su ayuda tecnica en la realización de este trabajo.

Agradesco al Dr. Jorge Humberto Ramírez Prado y Dr. Ernesto Perez Rueda, por su ayuda en el desarrollo de la parte bioinformática de este trabajo.

A la IB. Abril Denisse Morales Jiménez (Instituto Tecnológico de Mazatlan), Judith H. Caamal Chan (Estudiante de la Universidad Autonoma de Yucatán), por toda su ayuda en el escrutinio de la metagenoteca.

A todos mis amigos y compañeros de los diferentes laboratorios del CICY, por su amistad y sus comentarios para mejorar este trabajo.

A Dios por darme la vida y permitirme estar aquí así como a mi madre Sara Hernández Alemán por ser una mujer de carácter y que me forjó una idea de triunfo y éxito aun contra la adversidad.

#### Congresos:

2014; Metagenomica Funcional en el acuífero de Yucatan para la búsqueda de proteasas novedosas, VII Congreso de biotecnología y bioingeniería del sureste, Yucatán México.

2013; Funtional metagenomics in the aquifer of Yucatan: search for novel proteases, 12<sup>th</sup> international symposium on the genetics of industrial microorganisms, XV national congress of biotechnology and bioengineering, Cancun México.

2012; Metagenomica Funcional en el acuífero de Yucatan para la búsqueda de proteasas novedosas, VI Congreso de biotecnología y bioingeniería del sureste, Yucatán México.

Dedicada a toda mi familia...



## El gran pez

Del mar, donde el cielo asoma,

Viene el gran pez a comerte

Porque es ley, verdad y axioma,

Que el débil lo coma al fuerte.

Y el leviatán de la muerte

Acude al puntual llamado

Donde tú eres el que pesca,

Pero tú eres el pescado.

José Quintero.



**ÍNDICE**

<b>ÍNDICE .....</b>	<b>i</b>
<b>ÍNDICE DE FIGURAS.....</b>	<b>v</b>
<b>ÍNDICE DE TABLAS .....</b>	<b>vii</b>
<b>ABREVIATURAS .....</b>	<b>ix</b>
<b>RESUMEN.....</b>	<b>xi</b>
<b>ABSTRACT .....</b>	<b>xiii</b>
<b>INTRODUCCIÓN.....</b>	<b>1</b>
<b>CAPÍTULO I .....</b>	<b>3</b>
<b>ANTECEDENTES .....</b>	<b>3</b>
<b>1.1 El Acuífero de Yucatán.....</b>	<b>3</b>
<b>1.2 Metagenómica.....</b>	<b>5</b>
<b>1.2.1 Construcción de bibliotecas metagenómicas.....</b>	<b>10</b>
<b>1.2.2 Escrutinio de metagenotecas .....</b>	<b>6</b>
<b>1.2.2.1 Escrutinio basado en secuencia .....</b>	<b>8</b>
<b>1.2.2.2 Escrutinio basado en función .....</b>	<b>9</b>
<b>1.2.3 Metagenómica y biotecnología .....</b>	<b>10</b>
<b>1.3 Enzimas .....</b>	<b>11</b>
<b>1.3.1 Nomenclatura de las enzimas .....</b>	<b>11</b>

---

1.3.2 Proteasas.....	12
1.3.3 Fuente de proteasas .....	14
1.3.4 Proteasas microbianas.....	15
1.3.5. Serina proteasas .....	17
1.4 Aislamiento de proteasas a partir de metagenotecas .....	18
1.5 OBJETIVOS .....	21
1.5.1 Objetivo general.....	21
1.5.2 Objetivos específicos .....	21
1.6 ESQUEMA METODOLÓGICO.....	23
CAPÍTULO II .....	25
MATERIALES Y MÉTODOS.....	25
2.1 Material biológico .....	25
2.2 Vectores y cepas bacterianas .....	25
2.3 Extracción de ADN metagenómico.....	26
2.4 Construcción de la metagenoteca.....	27
2.4.1 Obtención de los fragmentos de ADN metagenómico entre un tamaño de 30 a 40 Kpb.....	27
2.4.2 Reparación del ADN metagenómico.....	27
2.4.3 Reparación de los insertos utilizando el fragmento Klenow .....	28

2.4.4 Reacción de ligación .....	29
2.4.5 Empaquetamiento de la ligación en fagos .....	29
2.5 Escrutinio de la metagenoteca por funcionalidad.....	31
2.6 Extracción del ADN fosmídico .....	31
2.7 Secuenciación.....	31
2.8 Análisis <i>in silico</i> .....	32
2.8.1. Clasificación de las secuencias de proteasas .....	33
2.8.2. Modelado de la estructura terciaria de <i>PrAY5</i> y <i>PrAY6</i> .....	33
2.9. Clonación de los ORFs en el vector de expresión pLATE52 .....	33
2.9.1. Pruebas de funcionalidad de las construcciones pLATE52-PrAY5 y pLATE52-PrAY6 .....	36
CAPÍTULO III .....	37
RESULTADOS .....	37
Construcción de una metagenoteca a partir de agua subterránea del Acuífero de Yucatán .....	37
3.1 Obtención de la biomasa procariota .....	37
3.2 Extracción de ADN metagenómico.....	38
3.3 Obtención de los insertos de ADN metagenómico .....	39
3.4 Reparación de los insertos .....	40
3.5 Construcción de la metagenoteca.....	41
3.5.1 Clonación de fragmentos pegajosos.....	42

---

3.5.2 Variación en el método de extracción de ADN metagenómico.....	44
3.6 Escrutinio de la metagenoteca.....	47
3.6.1 Prueba de controles .....	47
3.6.2 Escrutinio sobre agar leche .....	48
3.6.3 Extracción del ADN fosmídico de las clonas putativas .....	50
3.7 Análisis <i>In silico</i> de los resultados de secuenciación.....	53
3.7.1 Secuenciación se los fósmidos .....	53
3.7.2 Análisis del inserto de la clona MAJ10.....	54
3.7.3 Caracterización <i>in silico</i> de las secuencias <i>PrAY5</i> y <i>PrAY6</i> .....	61
3.7.4 Relación filogenética de las PrAYs con otras proteasas. ....	65
3.7.5 Modelado 3D de PrAY5 y PrAY6 .....	67
3.8 Subclonación de los ORFs <i>PrAY5</i> y <i>PrAY6</i> en el vector de expresión pLATE52.....	74
CAPÍTULO IV.....	79
DISCUSIÓN GENERAL.....	79
CAPÍTULO V.....	85
CONCLUSIONES Y PERSPECTIVAS.....	85
BIBLIOGRAFÍA.....	89

**LISTADO DE FIGURAS**

**Figura 1.1** Anillo de cenotes de Chicxulub.. ..... 3

**Figura 1.2** Esquema de la construcción de una metagenoteca y su escrutinio..... 7

**Figura 1.3** Reacción catalizada por una hidrolasa..... 13

**Figura 3.1** Filtros después de su uso para la obtención de biomasa procariota..... 37

**Figura 3.2** ADN metagenómico extraído de muestras de agua del Acuífero de Yucatán..38

**Figura 3.3** Tamaño de los fragmentos de ADN metagenómico utilizados para la construcción de la metagenoteca..... 39

**Figura 3.4** Representación esquemática de los daños que podrían tener los fragmentos de ADN metagenómico recién aislado.. ..... 40

**Figura 3.5** ADN metagenómico después de la reparación de sus extremos..... 41

**Figura 3.6** Digestión parcial del ADN metagenómico con la enzima *Bam*HI..... 42

**Figura 3.7** ADN metagenómico utilizado para la digestión con *Bam*HI..... 43

**Figura 3.8** ADN metagenómico extraído con el paquete de extracción de agua de Epicentre<sup>MR</sup>..... 44

**Figura 3.9** ADN metagenómico utilizado para la construcción de la metagenoteca..... 45

**Figura 3.10** Fósmidos de las clonas putativas de la metagenoteca digeridos con la enzima *Bam*HI..... 46

**Figura 3.11** Fósmidos de la metagenoteca digeridos con la enzima *Bam*HI..... 47

<b>Figura 3.12</b> Control positivo y control negativo sembrados en el medio de detección de clonas secretoras de proteasas. ....	48
<b>Figura 3.13</b> Escrutinio funcional de clonas de la metagenoteca. ....	49
<b>Figura 3.14</b> Clonas sobreproductoras de proteasas extracelulares. ....	49
<b>Figura 3.15</b> ADN de las clonas que produjeron halo hidrolítico en agar leche. ....	51
<b>Figura 3.16</b> Análisis de restricción de las clonas positivas utilizando <i>Bam</i> HI. ....	52
<b>Figura 3.17</b> Análisis <i>in silico</i> del inserto del fósmino de la clona MAJ10. ....	60
<b>Figura 3.18</b> Alineamiento de las secuencias de peptidasa S8. ....	62
<b>Figura 3.19</b> Relación filogenética de PrAY5 y PrAY6 con respecto a otras secuencias de la familia subtilasa S8 y otras secuencias de muestras ambientales. ....	66
<b>Figura 3.20</b> Modelo 3D de los zimógenos de <i>PrAY5</i> y <i>PrAY6</i> . ....	68
<b>Figura 3.21</b> Plegamiento teórico, enumerando las hélices- $\alpha$ and hojas- $\beta$ en PrAY5 y PrAY6. ....	70
<b>Figura 3.22</b> Alineamiento de las secuencias modeladas y el templado 3AFG. ....	71
<b>Figura 3.23</b> Sobrelapamiento de los modelos 3D de PrAY5, PrAY6 y S8 de <i>C. koreensis</i> tomando como molde el templado 3AFG. ....	73
<b>Figura 3.24</b> Productos de PCR de los ORFs <i>PrAY5</i> y <i>PrAY6</i> . ....	74
<b>Figura 3.25</b> Amplificación por PCR de <i>PrAY5</i> y <i>PrAY6</i> en pLATE52. ....	75
<b>Figura 3.26</b> Prueba funcional de las construcciones pLATE52-PrAY5 y pLATE52-PrAY6. ....	76

## LISTADO DE TABLAS

<b>Tabla 1.1</b> Clasificación internacional de enzimas .....	12
<b>Tabla 1.2</b> Clasificación de proteasas .....	13
<b>Tabla 1.3</b> Proteasas alcalinas (bacterianas comerciales), fuentes, aplicaciones y sus proveedores.....	15
<b>Tabla 1.4</b> Proteasas obtenidas de genotecas metagenómicas.....	18
<b>Tabla 2.1</b> Reacción de reparación de los insertos .....	28
<b>Tabla 2.2</b> Reacción de reparación de los insertos con Klenow .....	29
<b>Tabla 2.3</b> Secuencia de los oligonucleótidos utilizados para subclonar <i>PrAY5</i> y <i>PrAY6</i> .....	35
<b>Tabla 3.1</b> Lista de las proteínas homólogas más cercanas a las codificadas en el inserto del fósido de la clona MAJ10. Predichas con GLIMMER3.....	55
<b>Tabla 3.2</b> Lista de ORFs codificados en el inserto del fósido de la clona MAJ10 predichos con FgenesB. ....	58
<b>Tabla 3.3</b> Análisis <i>in silico</i> de las secuencias <i>PrAY5</i> y <i>PrAY6</i> .....	64



## ABREVIATURAS

.pdb	Extensión de archivos de la base de datos <i>Protein Data Bank</i>
ADN	Acido desoxirribonucleico
Amp	Ampicilina
Asn, N	Aminoácido Asparagina
Asp, D	Aminoácido Ácido aspártico
BLAST	<i>Basic Local Alignment Search Tool</i>
CDD	<i>Conserved Domain Database</i>
Contig	Segmentos de ADN superpuestos
<i>C .koreensis</i>	<i>Chitinimonas koreensis</i>
Cn	Cloranfenicol
Glu, E	Aminoácido acido Glutámico
His, H	Aminoácido Histidina
ORF	<i>Open Reading Frame</i> (marco de lectura abierto)
PCR	<i>Polymerase chain reaction</i> (Reacción en cadena de la polimerasa)
<i>PDB</i>	<i>Phage Dilution Buffer</i> (Amortiguador de dilución de fago)
PPC	Dominio Prepeptidasa carboxi terminal
<i>PrAY5</i>	Gen <u>P</u> roteasa del <u>A</u> cuífero de <u>Y</u> ucatán <u>5</u>
<i>PrAY6</i>	Gen <u>P</u> roteasa del <u>A</u> cuífero de <u>Y</u> ucatán <u>6</u>
PrAY5	Proteína <u>P</u> roteasa del <u>A</u> cuífero de <u>Y</u> ucatán <u>5</u>
PrAY6	Proteína <u>P</u> roteasa del <u>A</u> cuífero de <u>Y</u> ucatán <u>6</u>
Ser, S	Aminoácido Serina



## RESUMEN

Las proteasas (sinónimo de peptidasas) (EC 3.4) son enzimas hidrolíticas que catalizan la ruptura de enlaces peptídicos de otras proteínas y en algunos casos tienen incluso la capacidad de auto hidrolizarse. Se utilizan para una amplia gama de aplicaciones biotecnológicas, especialmente en las industrias de detergentes, alimentos y productos farmacéuticos. En este estudio se construyó una biblioteca metagenómica de fósmidos a partir de biomasa procariota de agua subterránea del Acuífero de Yucatán. Se obtuvieron aproximadamente 250,000 clonas con insertos de ~35,000 pb. Se llevó a cabo un escrutinio funcional de 21,000 de estas clonas, para identificar aquellas que fueran sobre productoras de proteasas, resultando 23 positivas. Se analizó *in silico* la secuencia de los insertos de tres de estas clonas positivas con GLIMMER3 y FgenesB, para anotar los genes contenidos en el ADN exógeno, y solamente en una de ella se encontraron dos genes con un dominio característico de proteasas. Estos genes, fueron nombrados, *PrAY5* y *PrAY6*. Su tamaño fue de 1845 pb y 1824 pb respectivamente y tuvieron una identidad entre sí de 66.5 %. Cuando sus ORFs fueron analizados mediante BLASTp en la base de datos no redundante del NCBI, mostraron una identidad de 60% y 68% respectivamente con una peptidasa S8 de *Chitinimonas koreensis*. Un análisis filogenético utilizando la secuencia de sus respectivas proteínas maduras, mostró que *PrAY5* y *PrAY6* agrupan dentro de la familia S8\_13; la cual, de acuerdo con la base de datos de dominios conservados CDD, es una familia de proteasas no caracterizada. Un análisis estructural 3D mostró que *PrAY5* y *PrAY6* codifican zimógenos complejos, que contienen un péptido señal, un propéptido N-terminal, una zona del sitio activo y un propéptido C-terminal con un dominio de prepeptidasa carboxi terminal PPC. Su sitio activo contiene una triada catalítica compuesta por los aminoácidos D, H y S, la cual es característica de las peptidasas S8, así como un oxianión, el cual también participa en la catálisis. Cuando cada uno de los ORFs fue subclonado individualmente en el vector de expresión pLATE52, las clonas resultantes mostraron un fenotipo proteolítico, lo cual sugiere que ambos pueden ser expresados independientemente y sus productos pueden ser procesados y secretados sin necesitar uno del otro. Todos estos resultados demuestran que, con la estrategia empleada, es factible obtener genes de proteasas novedosos con posible potencial biotecnológico.

---



**ABSTRACT**

Proteases (synonym of peptidases) (EC 3.4) are hydrolytic enzymes which catalyze the cleavage of peptide bonds of other proteins, and in some cases they even have the capacity to cleave themselves. They are used for a wide range of applications, especially in detergent, food and pharmaceutical industries. In this study, a fosmid metagenomic library of Yucatán underground water was constructed. About 250,000 clones harboring ~35,000 bp environmental DNA (eDNA) inserts were recuperated, and when 21,000 of these clones were screened for protease-overproduction, 23 of them resulted positive. When the exogenous eDNA sequences contained in three of these positive clones were analyzed by GLIMMER3 and fgenesB, two genes comprising protease domains were detected in clone MAJ10. These genes, called *PrAY5* and *PrAY6*, were 1,845 bp and 1,824 bp long respectively, and shared 66.5% identity. When their translated ORFs were analyzed by BLASTp in the NCBI nr database, they showed 60% and 68% identity respectively with a peptidase S8 of *Chitinimonas koreensis*. A phylogenetic analysis of their active sites showed that PrAY5 and PrAY6 grouped within the family S8\_13, which according to the CDD database, is an uncharacterized protease family. A 3D structural analysis showed that both PrAY5 and PrAY6 encode complex zymogens, which have a signal peptide, an N-pro-peptide, a mature core enzyme, and a C-pro-peptide containing a PPC domain. Their active site contains a catalytic triade composed by the amino acids D, H, and S, which is characteristic of peptidases S8, as well as an oxyanion hole, which also participate in the catalysis. When both ORFs were individually sub-cloned in the expression vector pLATE52, the resulting sub-clones showed a proteolytic phenotype when grown on the testing medium, suggesting that they can be independently expressed, and their products can be independently secreted. All these results together point out that by the followed strategy, two novel protease genes with biotechnological potential was discovered.



## INTRODUCCIÓN

Durante las últimas décadas el mal manejo de los recursos naturales ha llevado al deterioro del medio ambiente. Si esto continúa, pronto va a ser imposible incrementar el nivel de vida de la población humana existente, e incluso se corre el riesgo de acabar destruyendo el equilibrio ecológico indispensable para que sobrevivan las personas y otros muchos seres vivos (Donohoe, 2003). Por esto, es necesario realizar un aprovechamiento sustentable de ellos. Actualmente existe una unidad política global para promover la biotecnología blanca (industrial) como un elemento central de la gestión sostenible del futuro económico y el reparto de beneficios de países industrializados (Casas, 2011). Esto requiere el desarrollo de nuevas enzimas, procesos, productos y aplicaciones que sean amigables con el medio ambiente. Una de las maneras de lograrlo es generando nuevos productos que tengan alto impacto económico en los mercados internacionales mediante tecnologías que no dañen el medio ambiente (Tang y Zhao, 2009).

México es uno de los países más diversos del mundo. En él podemos encontrar un amplio número de paisajes y ecosistemas. En su región suroeste se encuentra la península de Yucatán, la cual se caracteriza por ser una de las regiones geológicas más jóvenes del territorio mexicano. Una de las características particulares de Yucatán es su Acuífero, mismo que conforma el sistema de ríos subterráneos más extenso del mundo. Los ecosistemas de la península dependen de él, y por ende, la economía de la región (Bauer-Gottwein *et al.*, 2011). Por esta razón es necesario darle un aprovechamiento sustentable mediante la obtención de nuevos productos con potencial biotecnológico de alto impacto, y utilizando tecnología que no dañe su integridad. Para lograr esto, es necesario utilizar las nuevas herramientas biotecnológicas con las que se cuenta hoy en día. Una de ellas es la metagenómica, con la cual se tiene acceso al ADN de la mayor parte de la biodiversidad microbiana, misma que hasta hoy se desconoce. Ahora se sabe que el mundo de los microorganismos procariontes es mucho más diverso, complejo e importante de lo que antes se pensaba. Esto significa que se cuenta con una fuente más amplia de diversidad metabólica y catalítica, y por lo tanto que existe un enorme potencial de encontrar nuevas enzimas, más adecuadas para ciertas aplicaciones industriales

importantes, y que pueden permitir procesos más limpios y amigables con el ambiente (Lorenz y Eck, 2005).

Actualmente, en diversos grupos a nivel mundial, se están explorando metagenómicamente comunidades de microorganismos que habitan nichos ecológicos diferentes de los hasta ahora estudiados. Sobre todo dirigiendo la búsqueda hacia hábitats con características consideradas extremas, tales como alta salinidad (Mirete *et al.*, 2015), valores extremos de pH (Guazzaroni *et al.*, 2013), altas o bajas temperaturas y presiones (Berlemont *et al.*, 2011; Moser *et al.*, 2012), o alta contaminación (Terrón-González *et al.*, 2016); así como hacia aquellos que forman parte del microbioma de macro organismos, por ejemplo esponjas marinas (Mori *et al.*, 2016; O'Connor-Sanchez *et al.*, 2014), raíces de manglares (Simões *et al.*, 2015), de insectos (Meirelles *et al.*, 2016), intestino de peces (Hennersdorf *et al.*, 2016), aguas marinas o dulces (Behzad *et al.*, 2016; Meziti *et al.*, 2016) o simplemente aquellos que se desarrollan en ambientes con características distintas de las de los ambientes ya explorados.

El presente trabajo, se enfocó en la búsqueda de nuevas biomoléculas con potencial biotecnológico utilizando tecnología metagenómica -la cual es totalmente inocua para el ambiente-, partiendo de un ambiente acuático icónico, hasta ahora inexplorado metagenómicamente, y de enorme importancia para la península de Yucatán: su Acuífero subterráneo.

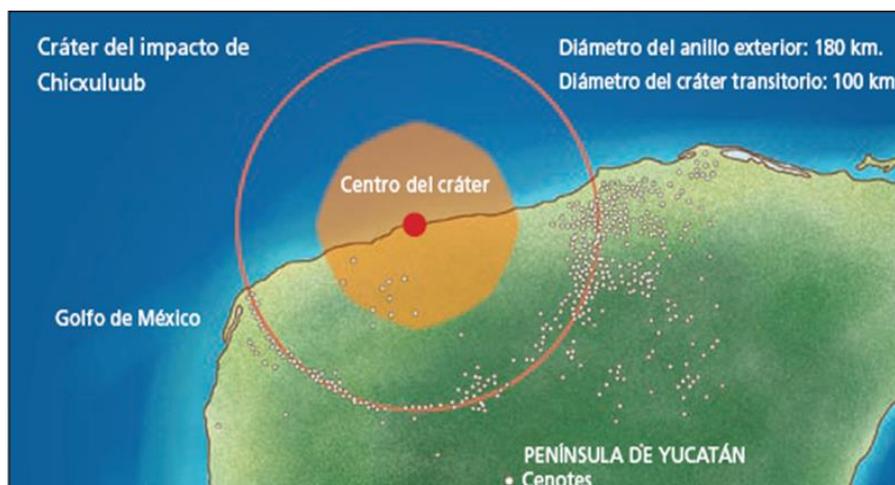
## CAPÍTULO I

### ANTECEDENTES

#### 1.1 El Acuífero de Yucatán

La península de Yucatán (PY) se localiza hacia el sureste de la República Mexicana, entre  $19^{\circ} 40'$  y  $21^{\circ} 37'$  N, y  $87^{\circ} 30'$  y  $90^{\circ} 26'$  O, rodeada por el golfo de México y el Mar Caribe, cubre un área de  $39\,340\text{ km}^2$  lo cual representa el 2% del territorio nacional (Schmitter-Soto, 2002)

La PY es una de las zonas geológicas más jóvenes de México. Se caracteriza por ser una plataforma calcárea. Se considera que, durante los periodos glaciares e interglaciares del Pleistoceno emergió del mar y durante este proceso y por acción pluvial, se formaron una serie de cavernas que fueron cambiando a lo largo de cientos de miles de años. También se sabe que, hace sesenta y cinco millones de años ocurrió el impacto de un meteorito que causo uno de los eventos de extinción masiva más importante en la historia de la vida en la Tierra. Este hecho provocó en la PY un sistema de fallas; las cuales, en conjunto con el sistema de cavernas, dio origen a uno de los sistemas hidrogeológicos más raros en el mundo, al cual se denomina anillo de cenotes de Chicxulub (Figura 1.1) (Perry *et al.*, 1995, Bauer-Gottwein *et al.*, 2011).



**Figura 1.1 Anillo de cenotes de Chicxulub.** Mapa esquemático del norte de la península Yucatán mostrando la distribución de los cenotes (en el círculo) y el punto rojo al centro que representa la zona de impacto del meteorito  $21.4^{\circ}\text{N}/89.52^{\circ}\text{O}$ .

Los cenotes son el resultado del hundimiento del techo de cavernas subterráneas que colapsaron. Por lo tanto, el agua que contienen es parte de los ríos subterráneos, los cuales a su vez conforman el Acuífero Yucateco. El acuífero es un gran sistema cavernoso que se intercomunica por medio de canales, que distribuyen agua a lo largo de la PY; la cual, por ser una plataforma permeable, presenta escasos ríos superficiales. El Acuífero Yucateco es el sistema cárstico más extenso del mundo; abarca aproximadamente 165,000 km<sup>2</sup> incluyendo México, Belice y Guatemala (Bauer-Gottwein *et al.*, 2011), lo que lo hace de vital importancia para los ecosistemas y las poblaciones humanas que se han asentado en esta zona y que sostuvieron a una de las civilizaciones más prósperas de la historia: La civilización Maya.

Una característica del Acuífero Yucateco es la convergencia de dos grandes masas de agua, una de origen pluvial (dulce) y otra salada que entra del mar. El agua dulce se encuentra por encima del agua salada debido a su menor densidad, formándose una zona de transición entre las dos, llamada haloclina (Perry *et al.*, 1995). Entre otras de las características particulares del acuífero destacan su pH ligeramente alcalino, su riqueza en carbonatos, su calidad oligotrófica y su casi total falta de luz (Bauer-Gottwein *et al.*, 2011; Gondwe *et al.*, 2010).

En Yucatán, el Acuífero se ha estudiado principalmente desde el punto de vista geofísico. El conocimiento desde un panorama ecológico y el conocimiento de su microbiota son muy pobres. Los estudios de su microbiota incluyen solamente al fitoplancton de algunos cenotes, donde las poblaciones dominantes son algas verdes como *Coelastrum sp.*, *Microspora sp.*, y *Scenedemus quadricuadra* (Herrera-Silveira, 1988), que varían sus poblaciones en relación con la estación del año.

El escaso conocimiento sobre las poblaciones microbianas del Acuífero y la constante búsqueda de nuevas biomoléculas en ambientes no explorados, que sean potencialmente utilizables en la industria, hace de esta microbiota una fuente novedosa para la realización de proyectos científicos en busca de productos naturales con potencial biotecnológico. En este sentido, uno de trabajos pioneros desde este enfoque, fue el realizado por De la Rosa-García y colaboradores (2007), quienes por medio de métodos tradicionales aislaron 256 cepas bacterianas de cenotes de Yucatán, de las cuales, 46 presentaron actividad antimicrobiana, concluyendo que las bacterias nativas del Acuífero de Yucatán muestran

una interesante actividad antimicrobiana, diverso modo de acción y moderada halotolerancia, con respecto a las de otros ambientes.

Aunado a los resultados mostrados por De la Rosa-García, existe evidencia de cenotes y porciones del Acuífero que aún pueden ser relativamente prístinas, en el sentido de no haber tenido contacto significativo con actividades humanas, desechos o productos. Además, la composición y conformación del suelo, y la rica diversidad biológica de la región en general, hacen de este un ambiente muy singular y por lo tanto atractivo para el análisis metagenómico, ya que puede contener una riqueza genética igualmente particular (Rojas-Herrera *et al.*, 2011).

Buscando en la gran diversidad microbiana existe la posibilidad de encontrar nuevos genes que produzcan enzimas con mejores propiedades y adecuadas para su uso comercial. La multitud de hábitats diferentes fisicoquímicamente ha hecho que los microorganismos desarrollen una gran conjunto de adaptaciones, por lo cual la diversidad de microorganismos es un recurso importante para la biotecnología de productos y procesos (Lorenz y Eck, 2005).

## 1.2 Metagenómica

Durante los últimos 5000 años las bacterias cultivables han servido como fuente de muchos productos útiles en biotecnología. Sin embargo, las bacterias cultivables solo representan el 1% del total existente (Rappé y Giovannoni, 2003; Amann *et al.*, 1995; Torsvik *et al.*, 1994), dejando el 99 % de la diversidad microbiana restante como una poderosa fuente de nuevos productos naturales esperando a ser descubiertos. Debido a que la mayoría de las bacterias no se han podido cultivar, hay una limitante para caracterizar estos microorganismos. Es aquí donde interviene la metagenómica, la cual ofrece el acceso a esta diversidad aun inexplorada. El término metagenómica fue definido por Handelsman (Handelsman *et al.*, 1998) como el acceso a los genomas de toda una comunidad bacteriana sin necesidad de la fase de cultivo *in vitro*. La metagenómica es una herramienta muy poderosa para el aislamiento de genes novedosos con potencial biotecnológico (Handelsman *et al.*, 1998).

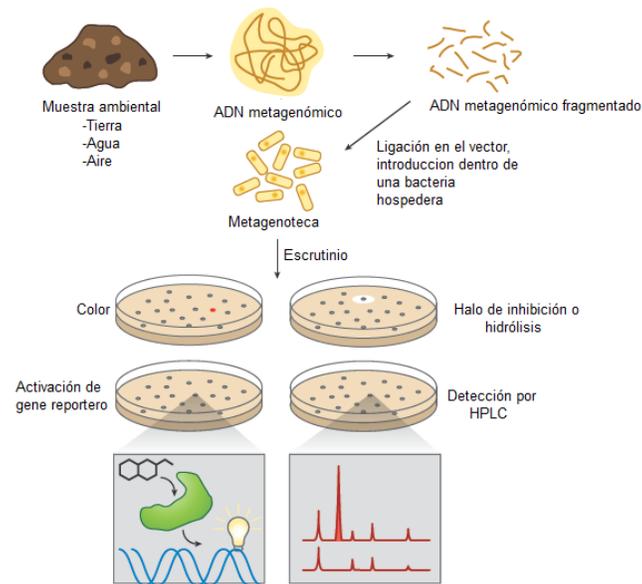
Las técnicas metagenómicas parten de la extracción en bloque de los genomas de todos los organismos que se encuentran en una muestra ambiental (e. g. suelo, flora intestinal, biopelículas, aire, agua, etc.), esta colección de todos los genomas de la comunidad se conoce como metagenoma y en él se encuentran todos los genes de todos los microorganismos de dicha comunidad. Posteriormente, este conjunto de genes se puede analizar utilizando diversas estrategias. Una de ellas es clonando el metagenoma en genotecas (que para referencia en este trabajo se llamarán metagenotecas) y realizando escrutinios funcionales para una determinada actividad, otra estrategia consiste en secuenciar directamente el ADN metagenómico para realizar análisis de diversidad mediante el ARN ribosomal 16S (Rondon *et al.*, 2000), y otras se basan en secuenciar al azar todos los genes presentes y hacer análisis tanto de diversidad taxonómica, como de diversidad de grupos funcionales (Martín-Cuadrado, 2007).

Desde un enfoque metagenómico, en los últimos diez años se ha empezado a analizar una gran diversidad de ambientes en todo el mundo, por citar algunos se pueden mencionar los siguientes: El Mar de los Sargazos (Venter *et al.*, 2004), las marismas de California (Córdoba-Kreylos *et al.*, 2006), el mar y costas del Mediterráneo (Caumette *et al.*, 1994; Martín-Cuadrado *et al.*, 2007), sedimentos marinos en Perú (Biddle *et al.*, 2008), lagunas hipersaladas en Chile (Da Silva *et al.*, 2008), estromatolitos de *Shark bay* en Australia (Papineau *et al.*, 2005), ventilas hidrotermales en el Tíbet (Lau *et al.*, 2009), tapetes microbianos en Tailandia (Portillo *et al.*, 2009) y la cuenca de Cuatrociénegas Coahuila en México (Peimbert *et al.*, 2012; Bonilla-Rosso *et al.*, 2012; Breitbar *et al.*, 2009). La mayoría de estos estudios se han realizado para determinar la diversidad microbiana y genética de las muestras, y aportar una nueva perspectiva para el análisis de la ecología y los ciclos biogeoquímicos de los distintos lugares.

### **1.2.1 Construcción de bibliotecas Metagenómicas**

Para realizar un estudio metagenómico funcional, el primer paso es el aislamiento del ADN de la muestra, posteriormente se realiza la construcción de la biblioteca metagenómica (Figura 1.2). El microorganismo más utilizado para tal fin es *Escherichia coli* (Brady *et al.*, 2009), aunque esta bacteria es un hospedero idóneo para realizar

expresión heteróloga de ADN, por su versatilidad genética y otros factores, también es una limitante para la expresión de algunos tipos de proteínas que le pueden resultar tóxicos (Wang *et al.*, 2000), por lo cual, se han construido bibliotecas metagenómicas en otros hospederos, tales como *Streptomyces sp.* y *Ralstonia metallidurans* (Wang *et al.*, 2000; Craig *et al.*, 2009; Craig *et al.*, 2010).



**Figura 1.2 Esquema de la construcción de una metagenoteca y su escrutinio.**

Para realizar la construcción se parte de cualquier muestra ambiental, se obtiene su metagenoma y se clona en algún hospedero para posteriormente por medio de ensayos de funcionalidad encontrar aquellas clonas que tengan algún inserto interesante (figura tomada de Piel, 2011)

Otro aspecto muy importante en la construcción de una metagenoteca para análisis funcionales es la selección del vector a utilizar. Se debe considerar que, entre más pequeños sean los fragmentos clonados, mayor será el esfuerzo del escrutinio que se necesite hacer. Esto es debido a que se necesitarán un mayor número de clonas para una mejor representación del metagenoma.

Un tipo de vector son los plásmidos los cuales son muy versátiles, sin embargo, solo pueden contener insertos pequeños, de entre 2 a 10 Kpb. Esto es una limitante si se desean aislar operones o rutas biosintéticas completas. Para ello, se utilizan vectores que

soporten tamaños de insertos más grandes, dentro de estos se encuentran los BAC (por sus siglas en inglés: *Bacterial Artificial Chromosome*) en los cuales se pueden insertar fragmentos de hasta 750 Kpb. Sin embargo, la problemática que enfrentan estos es la dificultad de obtener insertos tan grandes ya que solo se obtienen en el orden de los 100 a 300 Kpb después de la digestión del ADN (Handelsman, 2004), y otro problema es la estabilidad de estos vectores en el organismo hospedero. Pese a esto, los vectores BAC han sido utilizados con éxito en algunos trabajos (MacNeil *et al.*, 2001; Long *et al.*, 2005). Sin embargo, hasta ahora los vectores que mejor han funcionado en la construcción de metagenotecas son los cósmidos (Collins *et al.*, 1978) y los fósmidos (King *et al.*, 2009), los cuales pueden recibir insertos de entre 35 y 45 Kpb, pueden ser introducidos en *E. coli* obteniéndose varios millones de clonas en genotecas muy estables, gracias a que se puede controlar el número de copias. La ventaja de estos vectores con respecto a los BACs es que los fragmentos de ADN que se clonan son más cortos, por lo que son fáciles de obtener y manipular; y además los fragmentos que se clonan son más largos que los de los plásmidos pudiendo soportar hasta 45 Kpb.

Para fines prácticos estos últimos vectores tienen un tamaño idóneo para hacer estudios de metagenómica funcional. La diferencia entre ellos, es que los fósmidos contienen un origen de replicación del plásmido F con lo cual el fragmento clonado es mantenido en una copia por célula, esto mantiene la estabilidad de la clona (King *et al.*, 2009) y evita los problemas de inestabilidad en la metagenoteca lo cual sí se ha observado cuando se utilizan cósmidos con alto número de copias (Sudek *et al.*, 2007).

### **1.2.2 Escrutinio de metagenotecas**

Después de la construcción de las metagenotecas el siguiente paso es el escrutinio de estas, lo cual se realiza para identificar clones que tengan genes de interés. El escrutinio se puede abordar de dos maneras (i) análisis por secuencia y (ii) análisis por funcionalidad. Ambas se describirán brevemente en los siguientes apartados.

### 1.2.2.1 Escrutinio basado en secuencia

Una alternativa para abordar el escrutinio de una metagenoteca es basándose en la detección de secuencias específicas de ADN. Esta estrategia tiene una ventaja significativa, ya que se pueden detectar genes que no se expresen en la metagenoteca. Se parte de una familia de biomoléculas de interés y se buscan sus homólogos en bases de datos. Con las secuencias obtenidas se realizan alineamiento y los consensos resultantes se utilizan para diseñar oligonucleótidos degenerados los cuales sirven para amplificar diferentes regiones las cuales posteriormente son secuenciadas y con base en estas, se diseñan oligonucleótidos específicos o sondas que sirven para realizar el escrutinio fino de la biblioteca. Sin embargo, debido a que el diseño de los primeros oligonucleótidos se hace sobre secuencias conservadas y conocidas, este tipo de escrutinio solo identifica secuencias ya conocidas, por lo que, si lo que se busca es descubrir nuevos genes, es una metodología no muy exitosa (Simon y Daniel, 2009).

### 1.2.2.2 Escrutinio basado en función

Una de las maneras más utilizadas para lograr la detección de clonas de interés es observando los fenotipos resultantes, cuando se siembran en un medio de selectivo adecuado. Para ello el escrutinio puede basarse por ejemplo en cambios de coloración (producción de pigmentos), en la formación de halos de inhibición o halos de degradación alrededor de las clonas positivas (Rondón *et al.*, 2000; Wang *et al.*, 2000), o en ensayos relacionados con genes reporteros (Williamson *et al.*, 2005). Para ello, es necesario que los genes insertados en el vector se expresen y sean activos.

Como ejemplos podemos mencionar el escrutinio basado en la observación de halos de degradación donde el medio de cultivo tiene un sustrato que solo puede ser transformado en productos si la clona contiene un gen, expresado heterológamente, cuyo producto tiene la capacidad para transformarlo (Waschkowitz *et al.*, 2009). Mientras que para el escrutinio utilizando genes reporteros se menciona el ejemplo de los ensayos METREX (por sus siglas en inglés: *METabolite-Regulated EXpression*) (Williamson *et al.*, 2005) y

SIGEX (por sus siglas en inglés: *Substrate-Induced Gene EXpression*) (Uchiyama y Watanabe, 2008), ambos ensayos consisten en que el producto del gen buscado, con la función buscada, induzca un promotor que dispara la síntesis de un gen reportero, como por ejemplo la proteína verde fluorescente (GFP).

Utilizando el escrutinio por funcionalidad se han obtenido nuevas enzimas por ejemplo lipasas, estererasas, quitinasas y proteasas (Hjort *et al.*, 2014; Privé *et al.*, 2015; Vester *et al.*, 2015).

### 1.2.3 Metagenómica y biotecnología

En paralelo a las importantes contribuciones que la metagenómica está haciendo en el entendimiento básico del funcionamiento de los ecosistemas, esta disciplina ha emergido también como una poderosa herramienta para el descubrimiento de genes novedosos que puedan ser utilizados en diferentes procesos industriales (Lorenz y Eck, 2005).

Este tipo de biotecnología se cataloga como paralela de la biotecnología roja y/o verde (médica y agrícola respectivamente). A la fecha diferentes sectores de la industria están interesados en explotar la fuente de microorganismos no cultivables debido a que hasta ahora se han descubierto genes que codifican enzimas de interés biotecnológico, como amilasas, lipasas, proteasas, entre otras (Steele *et al.*, 2009), con nuevas características respecto de las ya existentes. Cuando una nueva enzima es descubierta por metagenómica es catalogada mediante una calificación de uno a seis, usando como criterios de selección actividad *in vitro*, eficiencia, especificidad y estabilidad. Esto determina las fortalezas y debilidades de la enzima, de tal forma que las más promisorias pueden ser seleccionadas para desarrollo de nuevos procesos utilizando ingeniería de proteínas o métodos de evolución dirigida (Lorenz y Eck, 2005).

Las enzimas son ampliamente utilizadas en la industria (Kirk *et al.*, 2002), tan solo en 2003 el mercado de enzimas fue estimado en 2.3 billones de dólares (Global Industry Analysts. Industrial Enzymes, 2004).

### 1.3 Enzimas

Por definición textual, según la real academia de la lengua española, una enzima es una proteína (aunque también pueden haber enzimas de ARN como las ribozimas), que cataliza específicamente cada una de las reacciones bioquímicas del metabolismo, ampliando el concepto, son aquellas moléculas (catalizadores biológicos) que aceleran reacciones químicas. El término que se le da a este tipo de reacciones es reacciones enzimáticas o bioquímicas, en las cuales se parte de una molécula sustrato, la cual después de ser catalizada por la enzima, forma productos (Grisham *et al.* 1999).

Todas las enzimas son producidas por células vivas. A pesar de que todas se producen inicialmente dentro de la célula (endoenzimas) también existen aquellas que actúan extracelularmente a las que se les llama exoenzimas. Se han descubierto una gran cantidad de ellas en los organismos de los tres dominios de la vida actuando en solución acuosa a temperaturas y pH variables.

#### 1.3.1 Nomenclatura de las Enzimas

Las enzimas muchas veces se nombran de acuerdo con el sustrato o reacción que catalizan adicionando el sufijo “asa”. Sin embargo, de acuerdo con la Unión internacional de bioquímica y biología molecular (por sus siglas en inglés: IUBMB), se ha creado un sistema de nomenclatura y clasificación de enzimas (Tabla 1.1). Este sistema agrupa a las diferentes enzimas en seis clases, cada una de ellas con diferentes subclases, de acuerdo con la reacción que catalizan (Lehninger, 1993).

**Tabla 1.1** Clasificación internacional de enzimas

Nº	Clase	Tipo de reacción catalizada
1	Oxidoreductasa	Transferencia de electrones (iones hidruro o átomos de H)
2	Transferasas	Reacciones de transferencia de grupos
3	Hidrolasas	Reacciones de hidrólisis (transferencia de grupos funcionales al agua)
4	Liasas	Adición de grupos a dobles enlaces, o formación de dobles enlaces por eliminación de grupos
5	Isomerasas	Transferencia de grupos dentro de moléculas dando formas isoméricas
6	Ligasas	Formación de enlaces- C-C, C-S, C-O y C-N mediante reacciones de condensación acopladas a la ruptura del ATP

Modificado de Lehninger, 1993.

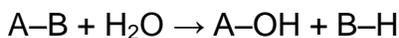
A cada enzima se le asigna un nombre sistemático que identifica la reacción catalizada y un número clasificador de cuatro dígitos. Por ejemplo, el nombre sistemático de una enzima es ATP: Glucosa fosfotransferasa lo que indica que cataliza la transferencia de un grupo fosforilo desde el ATP a la glucosa. El número de clasificación de esta enzima (numero E.C.) es 2.7.1.1. El primer número (2) denota el nombre de la clase (transferasa); el segundo (7), la subclase fosfotransferasa; el tercer (1), fosfotransferasas con un grupo hidroxilo como aceptor y el cuarto digito (1), D-glucosa como aceptor del grupo fosforilo (Lehninger, 1993).

### 1.3.2 Proteasas

Las proteasas son ubicuas en los tres dominios de la vida (Sharma *et al.*, 2004) y no solo son de enorme importancia para los seres vivos, sino también para la industria. Se utilizan en detergentes, preparación de cueros, procesamiento de alimentos, productos farmacéuticos, tratamiento de algunos desechos industriales, entre otras aplicaciones. Las proteasas constituyen el 45 % de todas las enzimas utilizadas en la industria y se espera que este número siga creciendo (Gupta *et al.*, 2002). Se estima que para el año 2019, el

mercado mundial de proteasas alcanzará los \$2,767 millones de dólares anuales (Markets and Market, 2016).

Las proteasas pertenecen al grupo de las hidrolasas, también se les conoce como peptidasas, ya que catalizan la ruptura de enlaces peptídicos de las proteínas utilizando para ello una molécula de agua (Figura 2.3).

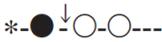
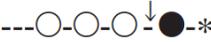
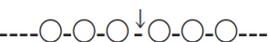


**Figura 1.3 Reacción catalizada por una hidrolasa.** A y B pueden ser cualquier residuo de aminoácido.

Las proteasas se clasifican con base en tres criterios (1) la reacción que catalizan, (2) naturaleza del sitio catalítico y (3) la relación evolutiva de su estructura (Barett, 1994). En general, primero son divididas en dos grandes grupos las endopeptidasas y las exopeptidasas (Tabla 1.2) dependiendo del sitio de acción en la cadena polipeptídica. Las endopeptidasas actúan dentro de la cadena polipeptídica mientras que la exopeptidasas en los extremos de la cadena, ya sea en el amino terminal o en el carboxilo terminal. Con base en el grupo funcional presente en el sitio activo las endopeptidasas se clasifican en cuatro grupos: serina proteasas, ácido aspártico proteasas, cisteína proteasas y metaloproteasas (Hartley, 1960).

**Tabla 1.2** Clasificación de proteasas

Proteasas	Modo de acción <sup>a</sup>	Número EC.
<b>Exopeptidasas</b>		
Aminopeptidasas	●-↓○-○-○-○---	3.4.11
Dipeptidil-peptidasas	●-●-↓○-○-○---	3.4.14
Tripeptidil-peptidasas	●-●-●-↓○-○---	3.4.14
Carboxipeptidasas	---○-○-○-○-○↓●	3.4.16-3.4.18
Carboxipeptidasas tipo serina		3.4.16
Metalocarboxipeptidasas		3.4.17
Carboxipeptidasas tipo cisteína		3.4.18
Peptidil-dipeptidasas	---○-○-○-○↓●-●	3.4.15

Dipeptidasas		3.4.13
Peptidasas Omega		3.4.19
		3.4.19
Endopeptidasas		3.4.21-3.4.34
Serina endopeptidasas		3.4.21
Cisteína endopeptidasas		3.4.22
Ácido aspártico endopeptidasas		3.4.23
Metaloendopeptidasas		3.4.24
Treonina endopeptidasas		3.4.25
Endopeptidasas con mecanismo catalítico desconocido		3.4.99

---

<sup>a</sup>Los círculos representan los residuos de aminoácidos en la cadena polipeptídica. Los círculos rellenos indican los aminoácidos terminales y los asteriscos representan el bloqueo de la cadena. Las flechas indican el sitio de corte de la enzima. (Modificado de Rao *et al.*, 1998).

### 1.3.3 Fuente de proteasas

Las proteasas son necesarias para la sobrevivencia de cualquier organismo vivo (Rao *et al.*, 1998), por esta razón las podemos obtener de diferentes fuentes como plantas, animales y microorganismos. Algunas de las proteasas que se obtienen de plantas son: la papaína, la bromelina, y la queratinasa; algunas de las proteasas que se obtienen de animales son proteasas pancreáticas, la tripsina, la quimiotripsina, la pepsina y la renina. La obtención de proteasas de plantas implica un proceso de obtención muy lento, que depende de la disponibilidad del área para el cultivo de las plantas y de las condiciones climáticas, mientras que, para la obtención de proteasas de origen animal el proceso es muy caro ya que se debe de contar con animales que sean sacrificados (Rao *et al.*, 1998). Debido a las limitantes que se tienen para la obtención de proteasas de plantas y animales; y a la incapacidad de poder satisfacer la demanda mundial de las proteasas, actualmente se tiene un mucho mayor interés en las proteasas de microorganismos.

### 1.3.4 Proteasas microbianas

Los microorganismos representan una excelente fuente de enzimas debido a su amplia diversidad bioquímica, a su fácil manejo, rápido crecimiento y relativamente fácil manipulación genética. Las proteasas microbianas (PM) representan aproximadamente el 40% del total del mercado mundial de enzimas (Godfrey y West, 1996)

La mayoría de las proteasas comerciales, neutras y alcalinas, son producidas por bacterias, principalmente del genero *Bacillus* sp. Las proteasas bacterianas neutras son activas solo en un intervalo estrecho de pH (5 a 8) y tienen baja termotolerancia. Estas enzimas son importantes sobre todo porque generan menos amargura en los hidrolizados de alimentos a diferencia de las de origen animal, por lo cual son más valiosas en la industria alimenticia. La neutrasa, una proteasa neutra, es insensible a los inhibidores de proteinasas de plantas, por lo cual es muy útil en la elaboración de cerveza. Su baja termotolerancia es ventajosa para controlar su reactividad durante la producción de hidrolizados proteicos en alimentos con un bajo grado de hidrólisis. Algunas de las proteasas neutras pertenecen al tipo metaloproteasa y requieren iones metálicos divalentes para su actividad, mientras que otras son serina proteasas, que no se ven afectadas por agentes quelantes (Rao M. *et al.*, 1998).

Las proteasas alcalinas provenientes de bacterias, se caracterizan por su alta actividad a pH alcalinos y su amplia gama de sustratos, su temperatura óptima está aproximadamente en 60 °C. Estas propiedades las hacen idóneas en la industria de los detergentes. En la Tabla 1.3 se muestra un resumen de diferentes proteasas alcalinas bacterianas utilizadas comúnmente en la industria.

**Tabla 1.3** Proteasas alcalinas (bacterianas comerciales), fuentes, aplicaciones y sus proveedores.

Proveedor	Nombre comercial del producto	Fuente microbiana	Aplicación
Novo Nordisk, Dinamarca	Alcalasa	<i>B. licheniformis</i>	Detergentes, desgomado de seda
	Savinasa	<i>Bacillus</i> sp.	Detergentes, textiles
	Esperasa	<i>B. lentus</i>	Detergentes, alimentos, desgomado de seda
	Biofeed pro	<i>B. licheniformis</i>	Alimento para animales
	Durazym	<i>Bacillus</i> sp.	Detergentes

## CAPÍTULO I

---

	Novozyme 471MP	s.e.	Hidrólisis de gelatina fotográfica
	Novosyme 243 B	<i>B. licheniformis</i>	Limpiadores de dentaduras postizas
	Nue	<i>Bacillus</i> sp.	Curtido de cuero
Genencor International, EE.UU.	Purafact	<i>B. lentus</i>	Detergentes
	Primatán	s.e.	Limpiadores de dentaduras postizas
Gist-Brocades, Países Bajos	Subtilisina	<i>B. alcalofilus</i>	Detergentes
	Maxacal	<i>Bacillus</i> sp.	Detergentes
	Maxatasa	<i>Bacillus</i> sp.	Detergentes
Solvay Enzymes, Alemania	Opticlean	<i>B. alcalofilus</i>	Detergentes
	Optimasa	<i>B. licheniformis</i>	Detergentes
	Maxapem	Proteína modificada de <i>Bacillus</i> sp.	Detergentes
	HT-proteolytic	<i>B. subtilis</i>	Alcohol, pasteles, cerveza, comida, alimentos, curtido de cuero, residuos fotográficos
	Proteasa	<i>B. licheniformis</i>	Alimentos, residuos
Pharmaceuticals Amano, Japón	Proleather	<i>Bacillus</i> sp.	Alimentos
	Collagenasa	<i>Clostridium</i> sp.	De carácter técnico
	Amano proteasa S	<i>Bacillus</i> sp.	Alimentos
Enzyme Development, EE.UU.	Enzeco alcalino proteasa	<i>B. licheniformis</i>	Industrial
	Enzeco alcalino proteasa-L FG	<i>B. licheniformis</i>	Alimentos
	Enzeco high alcalino proteasa	<i>Bacillus</i> sp.	Industrial
Nagase Biochemicals, Japón	Bioprasa concentrate	<i>B. subtilis</i>	Cosméticos, biofarmaceutica
	P.s proteasa	<i>P. aeuriginosa</i>	En investigación
	P.s elastasa	<i>P. aeuriginosa</i>	En investigación
	Cryst. proteasa	<i>B. subtilis</i> (K2)	En investigación
	Cryst. proteasa	<i>B. subtilis</i> ( <i>bioteus</i> )	En investigación
	Bioprasa	<i>B. subtilis</i>	Detergente, limpieza
	Bioprasa SP-10	<i>B. subtilis</i>	Alimentos
Godo Shusei, Japón	Godo-Bap	<i>B. licheniformis</i>	Detergentes, alimentos
Rohm, Alemania	Corolasa 7089	<i>B. subtilis</i>	Alimentos
Wuxi Synder Bioproducts, China	Wuxi	<i>Bacillus</i> sp.	Detergentes
Advance Biochemicals, India	Protosol	<i>Bacillus</i> sp.	Detergentes

s.e. Sin especificar (Modificado de Gupta *et al.*, 2002)

Durante las últimas décadas se han realizado muchos esfuerzos para identificar nuevas proteasas, así como para optimizar las ya existentes, para satisfacer la gran gama de

requerimientos específicos de los procesos que se requieren en la industria (Herbots, 2007).

Las proteasas que se utilizan en la industria hasta el momento, han sido obtenidas de colecciones de microorganismos o utilizando técnicas microbiológicas tradicionales, que involucran necesariamente tener cultivado el microorganismo a estudiar. Este enfoque limita a aquellos microorganismos que son capaces de crecer en condiciones de laboratorio, y como ya se mencionó, se ha estimado que solo el 1% de los microorganismos pueden ser cultivados por medio de técnicas estándares de laboratorio (Amann *et al.*, 1995). Debido a esta situación, es necesario utilizar otras técnicas que nos permitan tener acceso a los microorganismos no cultivables, por lo cual, el escrutinio de metagenotecas se ha convertido en una alternativa atractiva para la recuperación de nuevas enzimas provenientes de microorganismos no cultivables. Durante los últimos años se han obtenido varias nuevas enzimas utilizando esta metodología (Eck *et al.*, 2009).

### 1.3.5. Serina proteasas

Con base en su secuencia de aminoácidos, las proteasas están agrupadas en más de 50 familias, que a su vez se agrupan en 14 clanes, los cuales están definidos de acuerdo con la similitud de sus estructuras terciarias, el orden de sus residuos catalíticos y su secuencia (Rawlings y Barrett., 1993). Una de las familias más extensas e importantes desde el punto de vista comercial son las serina proteasas que se encuentran dentro del grupo de las hidrolasas y pertenece al clan SB (Siezen y Leunissen., 1997).

De acuerdo con MEROPS – *The Peptidase Database*, el clan SB incluye a las familias S8 y S53. La familia S8 se caracteriza por tener una triada catalítica en el orden D, H, S. Estos aminoácidos se encuentran embebidos en los motivos **D(T/S)G**, **HGTH** y **GTSXaaXaaXbbP**. En este último motivo, Xaa representa un aminoácido alifático y Xbb a un aminoácido pequeño. En la familia S53 el mecanismo catalítico es muy diferente respecto a la S8 y hay mayor complejidad en los residuos catalíticos, teniendo una tétrada de G, D, D, S (Rawling y Barrett., 2004)

**1.4 Aislamiento de proteasas a partir de metagenotecas**

Hasta la fecha se ha utilizado la metagenómica funcional para identificar varias nuevas enzimas como: amilasas, esterases, lipasas y quitinasas. Sin embargo, hasta ahora solo existen 10 reportes de enzimas proteasas obtenidas por esta metodología (Tabla 1.4).

**Tabla 1.4** Proteasas obtenidas por escrutinio funcional de genotecas metagenómicas.

Ambiente	Vector	T.I Kpb	Hospedero	Clonas/ Escrutinio	Produjeron Halo / Caracterizadas	Nombre	Referencia
Arena del Valle de la Muerte	pBSKII+	6	<i>E.coli</i> DH10B	30,000	1/1	DV1	
Arena del desierto de Gobi	pCC1FOS	32	<i>E.coli</i> EPI300	17,000	16/1	M30C4	(Neveu <i>et al.</i> , 2011)
Suelo de la mina Shaft	pBSKII+	3.9		10,000	0		
	pCR2.1-Topo	4.6		37,000/10,000	1/1	pTW4	
Suelo de composta	pBSKII+	2.4	<i>E.coli</i> TOP 10	32,000/10,000	0		
	pCR2.1-Topo	3.5		60,000/10,000	0		
Muestra mixta A	pBSKII+	4.2		52,000/10,000	0		
	pCR2.1-Topo	5.3		78,000/10,000	1/1	pTW1	
Muestra mixta B	pBSKII+	3.9		55,000/10,000	2/2	pTW2 pTW3	
	pCR2.1-Topo	4.0		65,000/10,000	0		(Waschkowitz <i>et al.</i> , 2009)
Superficie de la piel de cabra	pUC19	3.2	<i>E.coli</i> DH10B	70,000	1/1	AS-protease	(Pushpam <i>et al.</i> , 2011)
Sedimento marino	pWEB-TNC	35	<i>E.coli</i> EPI100	7000/5000	6/1	ACPRO001	(Zhang <i>et al.</i> , 2011)
Suelo del Artico	pCC1BAC-EcoRI	5.2	<i>E.coli</i> EPI300 DH10B	113,742	3		(Berlemont <i>et al.</i> , 2011)
Suelo del Bosque	pHT01	4-9	<i>E.coli</i> DH10B	35,000	1	SBcas3.3	(Biver <i>et al.</i> , 2013)
Rio Tinto	pBSKII+	1-8	<i>E.coli</i> DH10B				(Guazzaroni <i>et al.</i> , 2013)
Suelo	pET21a+	0.5–2.8	<i>E. coli</i> Top 10		1	HM219181	(Purohit y Singh 2013)
Barro	pCC1FOS	30–40	<i>E.coli</i> EPI300	30,000	1	pES63H9	(Lee <i>et al.</i> , 2007)
Lodo activado	pCC1FOS	38	<i>E.coli</i> EPI300	28032	1	M1-1	(Morris y Marchesi 2015)

Abreviaturas y símbolos: T.I (Tamaño de inserto); Kpb (Kilopares de bases); Pos. (Positivas).

El método de identificación de clonas sobreproductoras de proteasas de metagenotecas mediante escrutinio funcional ha sido utilizando agar LB (Sambrook *et al.*, 1989) suplementado con 1 % (p/v) de leche descremada, el cual ha tenido éxito (Rondón *et al.*, 2000). Sin embargo, se ha reportado que este método a veces produce falsos positivos (Jones y Marchesi, 2007). Recientemente se ha publicado un método más robusto, muy simple y fácil para el escrutinio de enzimas proteasas, que consiste en la preparación del medio de cultivo agregando como sustrato leche en polvo libre de grasas y lactosa [1% (p/v)]. La ausencia de lactosa elimina la posibilidad de enmascarar falsos positivos con lipasas y otras enzimas glucosidasas, que pueden producir ácido acético derivado de la fermentación de la lactosa, causando un descenso en el pH, y producir el halo de hidrólisis muy similar al característico de las enzimas proteasas (Morris *et al.*, 2012).

Trabajos recientes, utilizando metagenómica funcional, reportan el aislamiento de metaloproteasas. En uno de estos trabajos se utilizó composta, como fuente, para construir una metagenoteca de 80,000 clonas en el plásmido pBluescrip SKII+, las cuales tenían insertos aproximadamente de 2 Kpb, el hospedero fue *E. coli*. En esta metagenoteca se hallaron cuatro clonas positivas productoras de proteasas de las cuales dos eran completamente nuevas (Waschkowitz *et al.*, 2009). Otro reporte cita el aislamiento de dos serina proteasas novedosas de metagenotecas provenientes de muestras de suelo del Valle de la Muerte y del desierto de Gobi, uno ubicado entre el norte de China y sur de Mongolia y el otro en el este de California. Las genotecas fueron construidas utilizando el plásmido pBluescrip SKII+ y el fósido pCC1FOS, y como hospedero se usó *E. coli*, las clonas contenían insertos de 6 Kpb y 30 a 40 Kpb respectivamente. De la genoteca de plasmidos se obtuvo una clona positiva de 30 000 y de la de fósidos se obtuvieron 16 clonas positivas de 17 000 analizadas. De estas últimas 16, una fue caracterizada dando como resultado una nueva proteasa (Neveu *et al.*, 2011).

Hasta el 2007 solo había un reporte de una proteasa aislada de un metagenoma marino, el cual involucraba el aislamiento y caracterización de una metaloproteasa proveniente de una metagenoteca de sedimentos de aguas profundas. La muestra fue tomada al sur de la isla de Lihir, en Papúa, Nueva Guinea. Con esta muestra se construyó una

---

metagenoteca de 30,000 clonas utilizando el fósido pCC1FOS y como hospedero a *E. coli*. Después del escrutinio en el medio LB adicionado con 1 % (p/v) de leche descremada, se logró aislar una enzima con un porcentaje de identidad de 46% a una metalloproteasa de *Dechloromonas aromatica*, el bajo porcentaje de identidad, con lo ya reportado, hace que esta enzima sea novedosa (Lee *et al.*, 2007).

Como se observa en la tabla 1.4, la mayoría de proteasas encontradas mediante un abordaje de metagenómica funcional se han encontrado utilizando muestras de suelo. Sin embargo, la necesidad de obtener nuevas enzimas, que presenten características novedosas, hace necesaria la exploración de nuevos ambientes que puedan contener otras comunidades de microorganismos que aporten nuevas biomoléculas.

El Acuífero de Yucatán es uno de los sistemas hídricos más importantes de la península, el escaso estudio de su microbiota y el antecedente de que puede contener una diversidad microbiana particular (De la Rosa-García *et al.*, 2007; Rojas-Herrera *et al.*, 2011), lo hacen de gran atractivo para realizar un análisis metagenómico que aporte nueva información acerca su composición así como para la búsqueda de nuevas enzimas. Por todo esto, en el presente trabajo se plantea el análisis metagenómico desde una perspectiva funcional para la búsqueda de proteasas usando como material de partida agua del Acuífero de Yucatán.

## 1.5 OBJETIVOS

### 1.5.1 Objetivo General

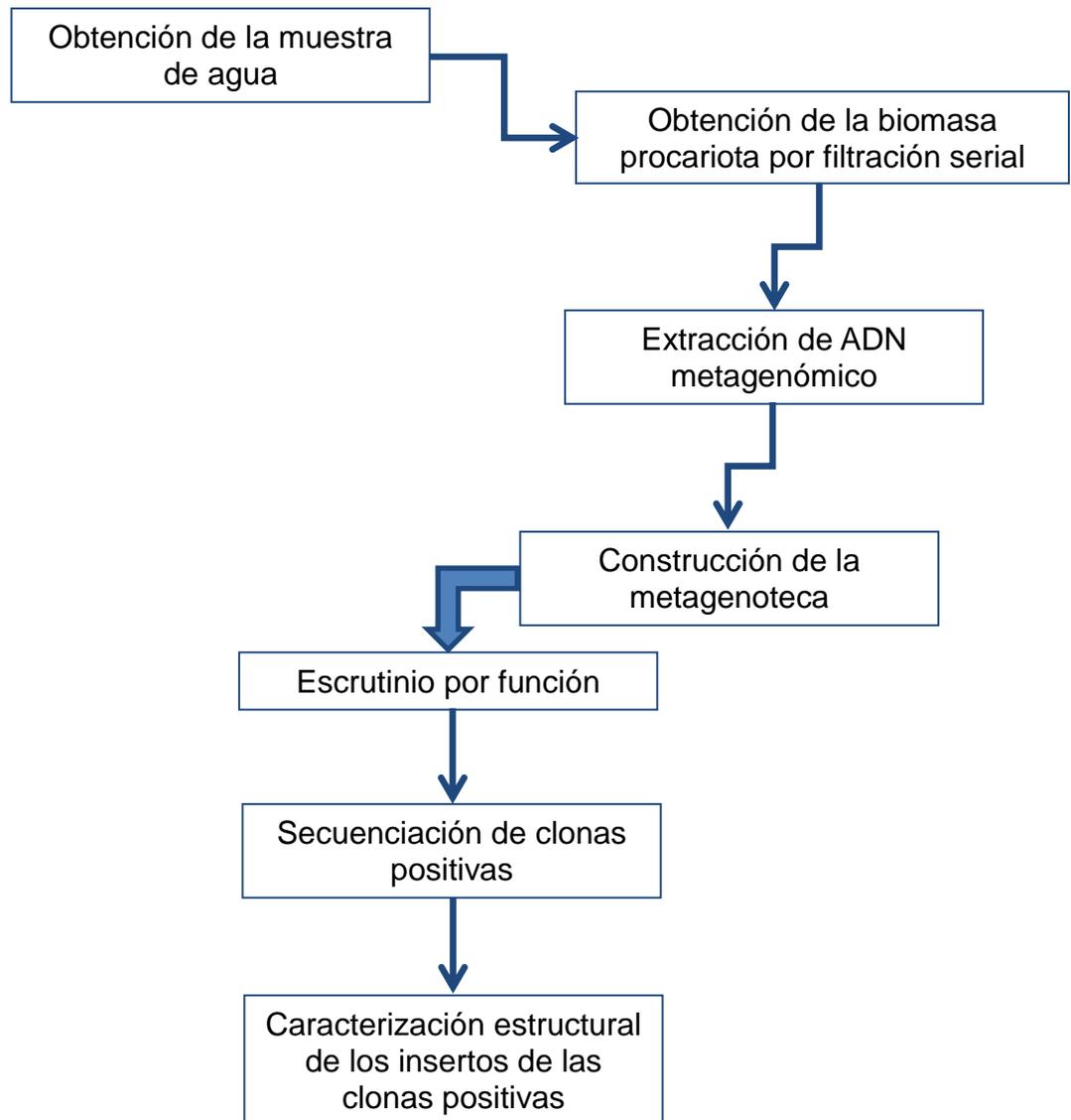
- Identificar funcionalmente y caracterizar genes de proteasas en una biblioteca metagenómica de fósidos del Acuífero de Yucatán.

### 1.5.2 Objetivos Específicos

- Construir una metagenoteca de fósidos a partir de una muestra de agua del Acuífero de Yucatán.
- Realizar un escrutinio funcional de la metagenoteca, para identificar clonas sobreproductoras de proteasas extracelulares.
- Hacer una caracterización *in silico* de los insertos de las clonas sobreproductoras de proteasas extracelulares.
- Caracterizar estructuralmente la secuencia de los ORFs traducidos responsables de la actividad.



## 1.6 ESQUEMA METODOLÓGICO





## CAPÍTULO II

### MATERIALES Y MÉTODOS

#### 2.1 Material biológico

Para obtener la biomasa procariota se tomó una muestra de agua del Acuífero de Yucatán (21° 01' 44'' Norte y 89° 38' 19'' Oeste), mediante bombeo aproximadamente a 30 metros de profundidad.

El agua obtenida se filtró secuencialmente, por bombeo peristáltico, primero a través de un filtro Millipore de 5 µm de diámetro de poro (Merck Millipore, Billerica, MA, USA) para eliminar los sedimentos, partículas mayores, agregados celulares y células microeucariotas; posteriormente a través de un filtro Sterivex– GV<sup>MR</sup> de 0.22 µm de diámetro de poro, para retener las células procariotas (Merck Millipore, Billerica, MA, USA). Para conservar la biomasa obtenida se adicionó RNAlater<sup>MR</sup> (Applied Biosystems, Carlsbad, CA, USA) y se mantuvo en refrigeración hasta el día que se extrajo el ADN.

#### 2.2 Vectores y cepas bacterianas

La metagenoteca se construyó utilizando el sistema de fósmidos pCC2FOS<sup>MR</sup> (Epicentre Madison, WI, USA). Una característica de este vector es que se puede controlar el número de copias; es decir, en un medio basal el vector puede estar en bajo número de copias por célula hospedera, y con la adición de 0.01% de arabinosa, se induce un alto número de copias (10 a 200). Como hospedero se utilizó *Escherichia coli* EPI300<sup>MR</sup>. Para el ensayo funcional de proteasas se utilizó como control positivo a *Bacillus subtilis* (ATCC 465). La subclonación y expresión de los ORFs se realizó en el vector de expresión pLATE52 (Thermo Scientific, Waltham, MA, USA) y las construcciones fueron transformadas en *E. coli* Top10 y BL21 (DE3) (New England Biolabs, Ipswich, MA, USA).

### 2.3 Extracción de ADN metagenómico

Para realizar la extracción del ADN metagenómico se utilizaron dos métodos:

- 1) Método reportado por Martín-Cuadrado (Martín-Cuadrado *et al.*, 2007).

Este método se utilizó con ligeras modificaciones y consistió en: (i) Añadir al filtro Sterivex<sup>MR</sup> con la biomasa bacteriana, 1.8 mL de amortiguador de lisis (40 mM EDTA, 50 mM Tris pH 8.3, 0.75 M sacarosa) y 50  $\mu$ L de lisozima (2 mg·mL<sup>-1</sup>) e incubar a 37°C durante 1h con agitación suave, ii) agregar 50  $\mu$ L de proteinasa K (0.4 mg·mL<sup>-1</sup>) y 200  $\mu$ L de SDS (10%) e incubar a 55°C durante 1h con movimiento suave. (iii) Con una jeringa recuperar el lisado del Sterivex<sup>MR</sup> y pasarlo a un tubo Falcon de 50 mL. (iv) Extraer dos veces con igual volumen de fenol: cloroformo: alcohol isoamílico (25:24:1 pH 8), centrifugando 10 min a 3000 g y recuperando la fase acuosa. (v) Extraer una vez con un volumen igual cloroformo: alcohol isoamílico (24:1). (vi) Concentrar la fase acuosa mediante un microconcentrador (Centricón 100, Amicón), hasta un volumen aproximado de 200  $\mu$ L; (vii) Añadir 1 mL de TE y volver a concentrar a un volumen de 100  $\mu$ L, (viii) invertir el Centricón y coleccionar 100 a 200  $\mu$ L.

- 2) El segundo método utilizado fue siguiendo las especificaciones del paquete de extracción de agua *Epicentre Metagenomic DNA Isolation Kit For Water* (Cat. No. MGD08420). Este método agiliza los pasos para la construcción de la metagenoteca eliminando la necesidad de purificar de gel para obtener el tamaño de los insertos adecuados y se recomienda que se utilice en conjunto con el paquete *Epicentre CopyControlMR HTP Fosmid Library Production* (Cat. No. CCFOS059).

En ambos métodos la concentración del ADN metagenómico se determinó mediante lectura en un espectrofotómetro NanoDrop (absorbancia a 230 y 260 nm UV) y su integridad fue analizada en geles de agarosa al 1%. Para la electroforesis se utilizó el amortiguador TBE 1X (Tris base 90 mM, ácido bórico 90 mM, EDTA 2.5 mM, pH 8.0). Los geles fueron teñidos con bromuro de etidio ( $0.5 \text{ mg}\cdot\text{mL}^{-1}$ ) y visualizados en un transiluminador de UV.

## **2.4 Construcción de la metagenoteca**

El ADN metagenómico obtenido, como se describió anteriormente, se utilizó para la construcción de la metagenoteca siguiendo las especificaciones del fabricante del paquete *Epicentre CopyControlMR HTP Fosmid Library Production* (Cat. No. CCFOS059).

### **2.4.1 Obtención de los fragmentos de ADN metagenómico entre un tamaño de 30 a 40 Kpb**

Debido a que se utilizaron dos métodos de extracción diferentes para la obtención del ADN metagenómico, fue necesario realizar dos estrategias diferentes para la preparación de los fragmentos. Un paso en común entre ambos métodos fue la ruptura mecánica del ADN por pipeteos subsecuentes.

Para preparar los insertos a partir del método de Martín-Cuadrado, fue necesaria la separación en gel de agarosa y posteriormente la purificación de los fragmentos que estuviesen en el tamaño de 30 a 40 Kpb. Para el método en el que se empleó el paquete de Epicentre<sup>MR</sup>, el ADN fue precipitado con etanol y se utilizó directamente en el siguiente paso de la construcción de la genoteca.

### **2.4.1 Reparación del ADN metagenómico**

Este paso es necesario ya que, debido a la ruptura mecánica, los extremos de los insertos pueden estar dañados, es decir, que haya extremos protuyentes, de manera que es

---

necesario rellenarlos para generar extremos romos y fosforilar los extremos 5'. Esto se realiza utilizando una mezcla enzimática que contiene T4 ADN polimerasa y T4 polinucleótido cinasa.

La reacción de reparación se llevó a cabo siguiendo las especificaciones del fabricante, los componentes de la reacción se muestran en la Tabla 2.1.

**Tabla 2.1** Reacción de reparación de los insertos

Reactivos	Cantidades
Agua estéril	2 $\mu$ L
Amortiguador " <i>End-repair</i> " 10 X	8 $\mu$ L
dNTPs 2.5 mM (c/u)	8 $\mu$ L
ATP 10 mM	8 $\mu$ L
ADN 5 $\mu$ g	50 $\mu$ L
Mezcla de las enzimas	4 $\mu$ L
Volumen de reacción	80 $\mu$ L

La reacción se incubó 45 min a temperatura ambiente. Al término de este lapso fue necesario inactivar la mezcla enzimática, para lo cual se incubó 10 min a 70 °C.

### 2.4.2 Reparación de los Insertos Utilizando el Fragmento Klenow

Un método alternativo para la reparación de los fragmentos de ADN fue utilizando el fragmento grande de la ADN polimerasa I de *E. coli*; llamado fragmento Klenow, el cual tiene actividad polimerasa 5'  $\rightarrow$  3' (esta ayuda a rellenar) y exonucleasa 3'  $\rightarrow$  5' la cual recorta cadena sencilla. Ambas actividades son necesarias para crear los extremos romos. Esta reacción se llevó a cabo bajo los parámetros que se muestran en la Tabla 2.2.

**Tabla 2.2** Reacción de reparación de los insertos con Klenow

Reactivos	Cantidades
ADN 3 µg	30 µL
Amortiguador (NEB2 10X)	5 µL
dNTPs (0.1 mM c/u)	5 µL
Agua estéril	9 µL
Enzima Klenow	1 µL
Volumen de reacción	50 µL

Tiempo de incubación 3 h. a 37 °C

Al final del tiempo de incubación para desactivar a la enzima se adicionó EDTA a una concentración final de 10 mM y se calentó a 75 °C por 20 min. Posteriormente se realizó la precipitación del ADN utilizando 1/10 de volumen de acetato de sodio 3 M pH 7 y 2.5 volúmenes de etanol absoluto helado, al final se resuspendió en 20 µL de TE. El ADN resultante se utilizó para seguir el protocolo del paquete *Epicentre CopyControl<sup>MR</sup> HTP Fosmid Library Production* según las especificaciones del fabricante.

### 2.4.3 Reacción de Ligación

La reacción de ligación se llevó a cabo según las especificaciones del proveedor. La relación molar óptima para realizar la ligación es 10:1 (vector-inserto), las cantidades utilizadas fueron 500 ng del vector (aproximadamente 0.09 pmol) y 250 ng de ADN metagenómico de aproximadamente 40 Kpb (0.009 pmol), la reacción se llevó a cabo en un volumen de 10 µL donde 3.5 µL eran agua estéril, 1 µL amortiguador 10X, 1 µL ATP 10mM, 1 µL ADN ligasa (Thermo Scientific, Waltham, MA, USA), 1 µL vector (0.5 µg µL<sup>-1</sup>) y 2.5 µL inserto (100 ng µL<sup>-1</sup>). Esta reacción se incubó toda la noche a 16 °C.

### 2.4.4 Empaquetamiento de la ligación en fagos

El resultado de la ligación se empacó en fagos. Para ello, células de *E. coli* EPI300<sup>MR</sup> (Epicentre, Madison, WI, USA) se estriaron en una caja de Petri en medio LB sin antibiótico. Un día antes del empaquetamiento, de dicha caja se seleccionó una sola

## CAPÍTULO II

---

colonia con la cual se inocularon 50 mL de medio LB adicionado con 10 mM MgSO<sub>4</sub> y 0.2 % de maltosa y se cultivó toda la noche a 37 °C en agitación constante (200 rpm).

El día del empaquetamiento se tomaron 0.5 mL del cultivo crecido toda la noche y con ello se inocularon 50 mL medio LB adicionado con 10 mM MgSO<sub>4</sub> y 0.2 % de maltosa. El cultivo se incubó a 37 °C y 200 rpm por un lapso de 2 h, esto se realizó para llevar a las células a un crecimiento exponencial (absorbancia de 0.8 a 1.0, a 600 nm).

Posteriormente se descongeló en hielo una alícuota (50 µL) del *MaxPlax<sup>MR</sup> Lambda Packaging Extract* (este es necesario para encapsidar los fragmentos de ADN que posteriormente son transfectados a la célula hospedera), y una vez descongelada se adicionaron 25 µL a los 10 µL de la reacción de ligación y se incubó durante 2 h a 30 °C (los 25 µL restantes se mantuvieron a -70 °C), posteriormente se adicionaron los 25 µL restantes y se incubó nuevamente 2 h a 30 °C. Al final de esta última incubación se adicionaron 25 µL de cloroformo y 0.5 mL del amortiguador de dilución de fago (amortiguador *PDB* [10 mM Tris-Cl, 100 mM NaCl y 10 mM MgCl<sub>2</sub>]).

El siguiente paso fue determinar el título viral, el cual es necesario para determinar la cantidad de Fago a utilizar para transfectar las células hospederas y tener una cantidad de células que se puedan separar y no se saturen las cajas de Petri al ser cultivadas.

El título viral se realizó utilizando las siguientes diluciones:

- A. Dilución 1:10<sup>1</sup>.- se diluyeron 10 µL del fago empaquetado en 90 µL de *PDB*
- B. Dilución 1:10<sup>2</sup>.- se diluyeron 100 µL de la dilución 1:10<sup>1</sup> en 900 µL de *PDB*
- C. Dilución 1:10<sup>3</sup>.- se diluyeron 100 µL de 1:10<sup>2</sup> en 900 µL de *PDB*

De cada una de estas diluciones, se adicionaron 10 µL a 100 µL de las células de *E. coli* crecidas exponencialmente y se incubaron a 37 °C durante 1 h. Adicionalmente también se incubaron 100 µL de células con 10 µL de fagos sin diluir y como control se utilizaron 100 µL de células sin fagos.

Al término del tiempo de incubación los 100 µL de células transfectadas se plaquearon en medio LB adicionado con 12.5 µg mL<sup>-1</sup> de Cloranfenicol (Cn). Esto se realizó para

---

seleccionar aquellas células que tuviesen el fósido. Estas cajas fueron incubadas a 37 °C toda la noche.

## 2.5 Escrutinio de la metagenoteca por funcionalidad

El escrutinio de la metagenoteca construida fue realizado utilizando el medio agar leche (medio de escrutinio para actividad proteolítica). El cual consiste en utilizar el medio Luria-Bertani (LB) suplementado con 12.5  $\mu\text{g}\cdot\text{mL}^{-1}$  de cloranfenicol como agente de selección, 0.001% de arabinosa (p/v) (inductor de multicopia) y 2% (p/v) de leche descremada en polvo (Svelty figura 0%, Nestlé, Vevey, Suiza). Después de 24 a 72 horas de incubación a 37 °C, la actividad de proteasa fue detectada por la presencia de halos de hidrólisis alrededor de las colonias.

## 2.6 Extracción del ADN fosmídico

Para verificación la presencia de los insertos en los fósidos se utilizó el método de lisis alcalina. La extracción de ADN fosmídico para la secuenciación se realizó utilizando el paquete de extracción Epicentre *FosmidMAX™ DNA purification* (Cat. No. FMAX046).

## 2.7 Secuenciación

Los insertos de cada fósido fueron secuenciados y ensamblados por el *USU Center for Integrated Biosystems* de la Universidad Estatal de Utah, EE. UU., utilizando la plataforma Ion Torrent chip 314 (Life Technologies).

Todas las muestras fueron ajustadas a una concentración de 100 ng  $\mu\text{L}^{-1}$ . Se usaron 5 etiquetas, bajo la siguiente nomenclatura: Etiqueta Proteasas 1 (PR-1) (clona MAJ1), PR-2 (clona MAJ10), PR-3 (clona MAJ19), PR-4 (clonas MAJ: 7, 8, 14, 20, 21) y PR-5 (clonas MAJ: 2, 3, 4, 5, 6, 9, 11, 12, 13, 15, 16, 17, 18, 22 y 23).

### 2.8 Análisis *in silico*

Los posibles genes presentes en los insertos fueron predichos *in silico* utilizando el algoritmo GLIMMER3 (Salzberg *et al.*, 1998) (alojado en el software suite GENEIOUS) y los resultados fueron verificados con fgenesB Sofberry (disponible en <http://www.softberry.com/berry.phtml?topic=fgenesb>), utilizando los parámetros predeterminados (Solovyev y Salamov, 2011).

La búsqueda de proteínas putativas se realizó utilizando la base de datos no redundante del Centro Nacional de Información Biotecnológica de los EE. UU. (NCBI por sus siglas en inglés) con la ayuda de la herramienta BLASTx (Gish y States, 2013) (utilizando parámetros predeterminados) traduciendo a aminoácidos con el código genético de bacterias y arqueas.

Las siguientes bases de datos y herramientas bioinformáticas fueron utilizadas para obtener información adicional acerca de la naturaleza de las secuencias obtenidas:

Base de datos de dominios conservados (por sus siglas en inglés CDD: *Conserved domain database*): es una fuente de anotación de proteínas la cual cuenta con una colección de modelos de alineamientos múltiples donde se identifican dominios proteicos (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) (Marchler-Bauer *et al.*, 2011).

ExPasy ProtParam: permite obtener varios parámetros físicos y químicos para una determinada proteína almacenada en la base de datos de Swiss-Prot o TrEMBL o para una secuencia introducida por el usuario (<http://web.expasy.org/protparam/>) (Gasteiger *et al.*, 2005).

HMMER: Busca en las bases de datos secuencias homólogas de proteínas y hace alineamiento de proteínas (<http://www.ebi.ac.uk/Tools/hmmer/>) (Finn *et al.*, 2011).

TMHMM: predice hélices transmembranales en proteínas (<http://www.cbs.dtu.dk/services/TMHMM/>) (Krogh *et al.*, 2001).

SignalP 4.4 Server: esta herramienta se utilizó para la predicción de posibles péptidos señales en los ORFs (<http://www.cbs.dtu.dk/services/SignalP/>) (Petersen *et al.*, 2011).

### 2.8.1. Clasificación de las secuencias de proteasas

La clasificación de las secuencias de proteasas encontradas en este estudio fue realizada con ayuda de la base de datos de peptidasas MEROPS – *The Peptidase Database* (<http://merops.sanger.ac.uk/>). La relación filogenética se realizó mediante un alineamiento entre las secuencias obtenidas en este trabajo, homólogos depositados en la base de datos no redundante y muestras ambientales del NCBI, así como con otras secuencias de diferentes subconjuntos de la base de datos de dominios conservados (CDD) pertenecientes a la familia de las proteasa S8. El alineamiento se llevó a cabo utilizando la herramienta MUSCLE, incluida en el software MEGA6 (Tamura *et al.*, 2013). Posteriormente se construyó un árbol filogenético utilizando el método de unión de vecinos (Neighbor-Joining) (Saitou y Nei, 1987) con un bootstrap (Felsenstein, 1985) de 1000, en el cual la distancia evolutiva fue computada utilizando la corrección de Poisson (Zuckerkandl y Pauling, 1965).

### 2.8.2. Modelado de la Estructura Terciaria de *PrAY5* y *PrAY6*

Para realizar el modelado tridimensional (3D) de las secuencias se utilizó el servidor I-TASSER v4.4 (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>), (Yang *et al.*, 2015) los resultados obtenidos en formato .pdb fueron visualizados en el sistema de gráficos moleculares PyMOL (<https://www.pymol.org/>) (DeLano, 2002). Para realizar el sobrelapamiento de las estructuras tridimensionales se utilizó Mustang-MR Structural Sieving Server (<http://pxgrid.med.monash.edu.au:8080/mustangserver/>) (Konagurthu *et al.*, 2010).

### 2.9. Clonación de los ORFs en el vector de expresión pLATE52

Para realizar la expresión de los ORFs de *PrAY5* y *PrAY6* se utilizó como vector el plásmido pLATE52 (Thermo Scientific, Waltham, MA, USA), el cual tiene un sistema de clonación direccional LIC. Este sistema utiliza una T4 ADN polimerasa para crear cadenas

---

sencillas de 14 a 21 nucleótidos que se sobrelapan en el vector. Después del reconocimiento entre el inserto y el vector el producto es transformado en células competentes de *E. coli* dentro de las cuales ocurre la ligación. Las secuencias clonadas en este sistema se encuentran bajo la regulación del promotor T7 el cual es un promotor inducible que permite la expresión controlada de productos que puedan ser tóxicos para *E. coli*. Para facilitar la purificación de las proteínas expresadas en este sistema, se realizó una fusión traduccional con una etiqueta de histidinas, la cual posteriormente fue requerida para la purificación en columna por afinidad.

Para llevar a cabo las construcciones pLATE52-PrAY5 y pLATE52-PrAY6, se realizó el diseño de los oligonucleótidos siguiendo las especificaciones de sistema LIC. PrAY5 y PrAY6 fueron amplificados por PCR, tomando como templado el fósido de la clona MAJ10 y utilizando el par de oligonucleótidos Fw-PR5 y Rv-PR5, y Fw-PR6 y Rv-PR6 respectivamente (Tabla 2.3). Las reacciones de PCR fueron realizadas utilizando la enzima de alta fidelidad *Phusion Green Hot Start II High-Fidelity* (Thermo Scientific). Posteriormente se realizó la purificación de los productos de PCR y se llevó a cabo el tratamiento con la enzima T4 ADN polimerasa y el reconocimiento entre el vector y los insertos.

**Tabla 2.3** Secuencia de los oligonucleótidos utilizados para subclonar PrAY5 y PrAY6 en el vector pLATE52 y verifica las construcciones.

Nombre	Secuencia	Tamaño (Nucleótidos)
Fw-PR5	5' <i>ggttgggaattgcaa</i> ATGAAGCAATTCAATACAAATAAGCCTGT 3'	44
Rv-PR5	5' <i>ggagatgggaagtca</i> TTACCAGTTGAACTTCACCGTGAC 3'	40
Fw-PR6	5' <i>ggttgggaattgcaa</i> ATGAAGCAAACCAATAAGCCTGCG 3'	39
Rv-PR6	5' <i>ggagatgggaagtca</i> TTACCAGCTGAGTTTCAGCGAGTA 3'	39
Fw-LIC	5' TAATACGACTCACTATAGGG 3'	20
Rv-LIC	5' GAGCGGATAACAATTTTCACACAGG 3'	24

Abreviaturas: Fw, sentido; Rv, antisentido. Las letras en itálicas y minúsculas representan a los nucleótidos utilizados para llevar acabo el reconocimiento de los insertos con el vector pLATE52.

Las construcciones fueron introducidas en células competentes de *E. coli* Top10 y BL21(DE3) genotipo *phuA2* [lon] *ompT gal* ( $\lambda$  DE3) [dcm]  $\Delta$ *hsdS*  $\lambda$  DE3 =  $\lambda$  *sBamHlo*  $\Delta$ *EcoRI-B* int::(*lac*::*PlacUV5*::T7 gene1) *i21*  $\Delta$ *nin5* (New England Biolabs). La transformación de las células se realizó utilizando el método de choque térmico y las clonas transformadas fueron seleccionadas en cajas de Petri en medio LB semisólido suplementado con 100  $\mu$ g mL<sup>-1</sup> de Ampicilina (Amp).

Las clonas que presentaron resistencia a Amp fueron incubadas toda la noche en 1.5 mL de medio líquido LB con 100  $\mu$ g. mL<sup>-1</sup> Amp y al día siguiente se cosecharon y se extrajo el ADN plasmídico, utilizando el paquete *GeneJET Plasmid Miniprep* (Thermo Scientific). Este ADN se utilizó como templado para verificar que las clonas tuviesen el inserto, para ello se realizó una reacción de PCR utilizando los oligonucleótidos Fw-LIC y Rv-LIC (Tabla 2.3), cuyos sitios blancos están dentro del vector pLATE52. Las clonas seleccionadas se enviaron a secuenciar por *primer walking* a Macrogen Co. (Seoul, South Korea).

### **2.9.1. Pruebas de funcionalidad de las construcciones pLATE52-PrAY5 y pLATE52-PrAY6**

Las clonas resultantes de la transformación con las construcciones de PrAY5 y PrAY6 en el vector de expresión pLATE52. Fueron crecidas en medio LB + leche bajo presión de selección con  $100 \text{ mg mL}^{-1}$  de Amp. y se incubaron a  $37 \text{ }^{\circ}\text{C}$  durante 3 días.

## CAPÍTULO III

### RESULTADOS

#### Construcción de una metagenoteca a partir de agua subterránea del Acuífero de Yucatán

##### 3.1 Obtención de la biomasa procariota

Para realizar la construcción de la metagenoteca el primer paso fue la obtención de la biomasa procariota. Para esto, se tomó una muestra de agua de un pozo ubicado dentro de las instalaciones del Centro de Investigación Científica de Yucatán (CICY). Al realizar el filtrado del agua, cada filtro Sterivex<sup>MR</sup> se saturó con el paso de 40 a 60 L, sin embargo, esto sucedió durante los días soleados mientras que los días lluviosos bastaba con filtrar solo 5 a 10 L. En la figura 3.1 se muestran los filtros utilizados en la obtención de la biomasa después de haber filtrado agua.



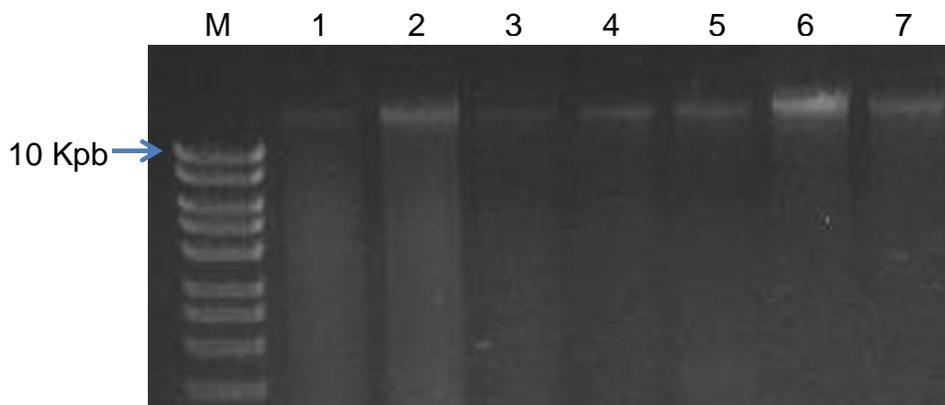
**Figura 3.1 Filtros después de su uso para la obtención de biomasa procariota.**  
A) Millipore<sup>MR</sup> de 5 µm y B) Sterivex–GV<sup>MR</sup> de 0.22 µm.

Como se observa en la figura 3.1, los dos filtros presentan una coloración marrón oscura, la cual fue característica del agua filtrada en días en los cuales hubo lluvias, mientras que en los filtros utilizados en días soleados solo el filtro de 5 µm presentaba esta coloración

(en menor intensidad), y el filtro de 0.22  $\mu\text{m}$  no presentó coloración alguna. Esto sugiere que durante las lluvias el río subterráneo se agita teniendo como consecuencia una mayor concentración de partículas y microorganismos en suspensión, lo cual podemos relacionar con que los filtros se tapan con un menor volumen de agua. Debido a este fenómeno en los días soleados necesitamos filtrar más agua que en los días lluviosos para la obtención de la biomasa.

### 3.2 Extracción de ADN metagenómico

Los filtros Sterivex<sup>MR</sup> que contenían la biomasa procariota fueron utilizados como contenedores para realizar la extracción de ADN metagenómico utilizando el método de Martín-Cuadrado (Martín-Cuadrado *et al.*, 2007) descrito en la sección de Materiales y Métodos. Recurriendo a este procedimiento se logró obtener ADN con buena integridad (Figura 3.2) a una concentración de 50 ng  $\mu\text{L}^{-1}$ . Posteriormente al ser pasado por un microconcentrador la concentración final fue 100 ng  $\mu\text{L}^{-1}$ .

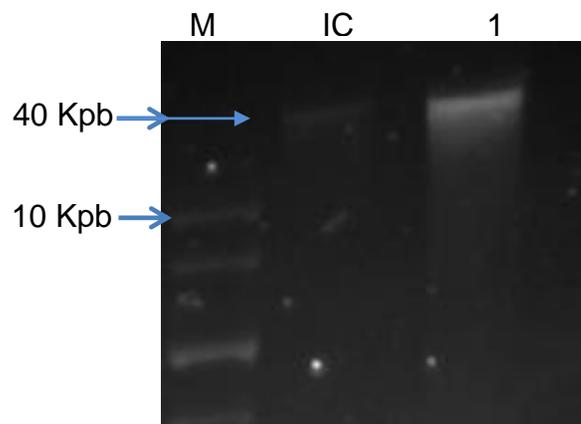


**Figura 3.2 ADN metagenómico extraído de muestras de agua del Acuífero de Yucatán.** Gel de agarosa al 1 %. Carril 1 a 7 muestras de ADN metagenómico. M: marcador de peso molecular de 1 Kpb.

El ADN metagenómico obtenido se utilizó para seguir el procedimiento de construcción de la metagenoteca según las instrucciones del fabricante (Epicentre<sup>MR</sup>).

### 3.3 Obtención de los Insertos de ADN metagenómico

Los fragmentos de ADN metagenómico se generaron por ruptura mecánica. Con frecuencia el ADN metagenómico después de ser extraído se encuentra roto en fragmentos de 30 a 40 Kpb, por lo que no necesita un tratamiento de digestión adicional. Para verificar lo anterior, se realizó un gel de agarosa al 1%, el cual se dejó correr toda la noche a 35 Volts (Figura 3.3)

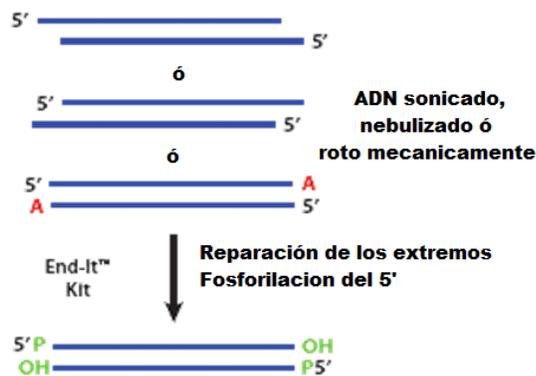


**Figura 3.3 Tamaño de los fragmentos de ADN metagenómico utilizados para la construcción de la metagenoteca.** Gel de agarosa al 1%. Carril 1: ADN metagenómico obtenido del acuífero; IC: ADN control proporcionado por el paquete con peso molecular de 40 Kpb y M: marcador de peso molecular de 1 Kpb.

El paquete de Epicentre<sup>MR</sup> proporciona un ADN control que corresponde a un fragmento del cromosoma X humano, el cual tiene un peso molecular de 40 Kpb. Este tiene dos utilidades (i) servir como marcador de tamaño para la selección de los fragmentos a clonar y (ii) como control de ligación en el protocolo. Como se puede observar en la figura 3.3, el ADN metagenómico de la muestra del Acuífero migró cerca de la altura a la que lo hizo el ADN control. Lo cual sugiere que los fragmentos obtenidos tienen un tamaño de aproximadamente 40 Kpb.

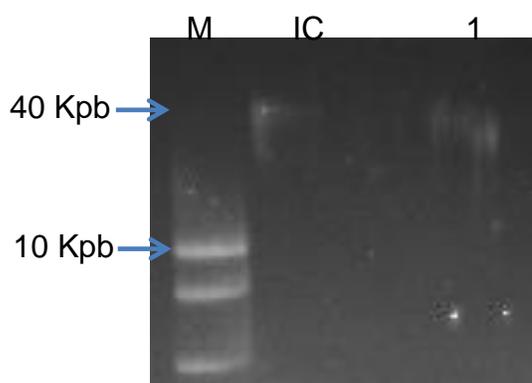
### 3.4 Reparación de los Insertos

El paso posterior a la obtención de los insertos es su reparación. Este paso es muy importante ya que debido al proceso de ruptura del ADN de forma mecánica (pipeteos subsiguientes) los extremos de los insertos quedan dañados en uno o ambos extremos (Figura 3.4).



**Figura 3.4. Representación esquemática de los daños que podrían tener los fragmentos de ADN metagenómico recién aislado.** El proceso de reparación consiste en rellenar mediante una polimerasa los extremos protuyentes del ADN y adicionar el grupo hidroxilo y fosfato para que se lleve a cabo la ligación del inserto con el vector.

El proceso de reparación, además de rellenar los extremos, también logra que el extremo 5' de cada fragmento quede con su grupo fosfato y el extremo 3' con el grupo hidroxilo (OH) correspondiente. Ambos grupos funcionales son necesarios para que se lleve a cabo la ligación con el vector. El paquete proporciona una mezcla de dos enzimas para lograr la reparación, en una ADN polimerasa del fago T4 (la cual rellena los extremos protuyentes) y una T4 polinucleótido cinasa que fosforila el extremo 5'. Una vez reparado el ADN metagenómico, se volvió a analizar por electroforesis para verificar su integridad (Figura 3.5).



**Figura 3.5 ADN metagenómico después de la reparación de sus extremos.** Gel de agarosa al 1%. Carril 1: ADN metagenómico obtenido del acuífero después del proceso de reparación; IC: ADN control proporcionado por el paquete 40 Kpb y M: marcador de peso molecular de 1 Kpb (Bioline, Taunton, MA, USA).

El ADN metagenómico reparado se utilizó para llevar a cabo la ligación en el vector pCC2FOS<sup>MR</sup>, lo cual es el siguiente paso del protocolo del paquete.

### 3.5 Construcción de la metagenoteca

Posteriormente a la reacción de ligación se llevó a cabo el empaquetamiento en los fagos y los fósmidos fueron transfectados en las células EPI300<sup>MR</sup> (según se describe en Materiales y Métodos). En un primer ensayo, utilizando el método de extracción reportado por Martín-Cuadrado (Martín-Cuadrado *et al.*, 2007), no se obtuvo crecimiento de ninguna colonia de ninguna dilución. Al utilizar el inserto control se obtuvo una eficiencia de  $2 \times 10^7$  unidades formadoras de colonia (UFC) mL<sup>-1</sup>, lo cual era de esperarse según las especificaciones del proveedor. Se repitió una vez más el protocolo descrito logrando obtener una clona en la caja Petri donde se plaquearon los 100  $\mu$ L de cultivo al que se le adicionaron 10  $\mu$ L del fago sin diluir.

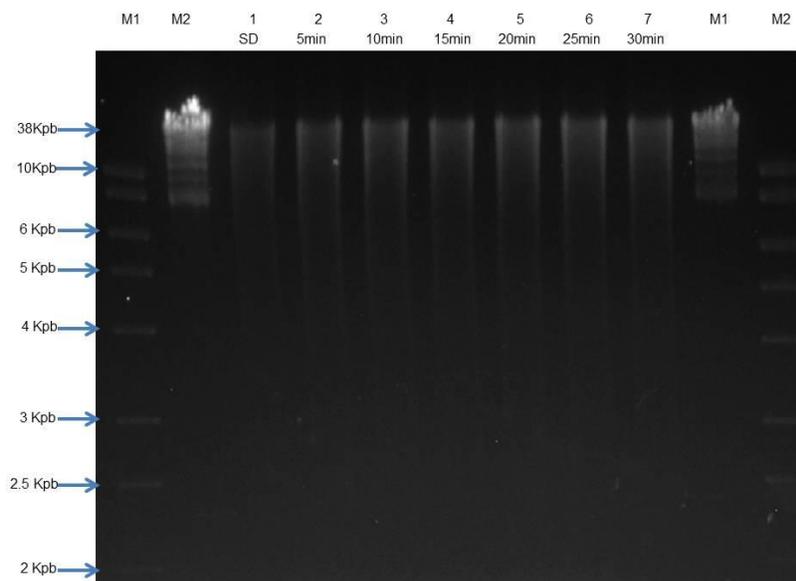
Debido a que en ADN control, el cual ya viene reparado, funcionó adecuadamente, se supuso que había una reparación inadecuada de los fragmentos metagenómicos, y por consiguiente su ligación era deficiente. Ante esto, se utilizó un método alternativo de reparación (sección 2.4.2). Al final de este experimento, una vez más se obtuvo un resultado negativo, ya que solo se logró recuperar una clona. Teniendo comunicación con

el proveedor y asistencia técnica se decidió seguir otra estrategia la cual se describe en el apartado siguiente.

### 3.5.1 Clonación de fragmentos pegajosos

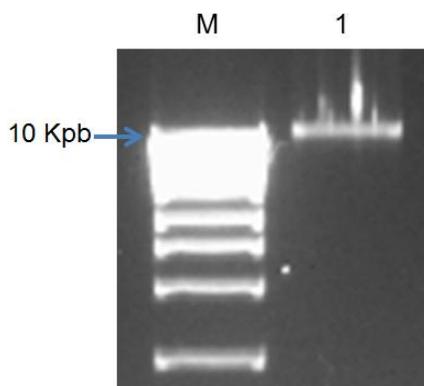
Para lograr una nueva ligación en el vector se crearon extremos pegajosos utilizando una enzima de restricción. Para lo cual, de acuerdo al mapa del pCC2FOS<sup>MR</sup>, se utilizó la enzima *Bam*HI.

El primer paso para desarrollar esta estrategia fue realizar una curva de digestión parcial del ADN metagenómico, para saber el tiempo adecuado de digestión y obtener fragmentos de entre 30 y 40 Kpb. Para ello se incubaron 800 ng de ADN metagenómico con 1U de *Bam*HI, a 37 °C en un volumen final de 30 µL. Se tomaron alícuotas de 5 µL cada 5 min durante un lapso de 30 min. El resultado de la digestión se visualizó en un gel de agarosa al 1% (Figura 3.6).



**Figura 3.6 Digestión parcial del ADN metagenómico con la enzima *Bam*HI.** M1 marcador de peso molecular de 1 Kpb; M2 marcador de peso molecular lambda Mix Marker, de 19 hasta 48 Kpb; carril 1 ADN metagenómico sin digerir (SD); carriles 2 a 7 ADN digerido en incrementos de 5 minutos (min).

Este resultado mostró que aun digiriendo por 30 min, el ADN metagenómico tenía un alto peso molecular (entre 10 y 40 Kpb). Así que, partiendo de 5  $\mu\text{g}$  de ADN metagenómico (Figura 3.7) se realizó una digestión utilizando la enzima *Bam*HI incubando a 37 °C durante 30 min. Posteriormente se realizó un gel preparativo y se purificó el ADN de la banda comprendida entre 30 y 40 Kpb, el cual fue utilizado para generar insertos en los fósmidos.



**Figura 3.7 ADN metagenómico utilizado para la digestión con *Bam*HI.** M marcador 1 Kpb. Carril 1 ADN metagenómico.

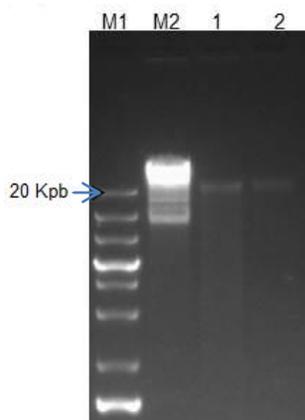
El vector pCC2FOS<sup>MR</sup> también fue digerido con la enzima *Bam*HI. La reacción se llevó a cabo partiendo de 1  $\mu\text{g}$  de ADN. Posteriormente a la digestión el vector fue desfosforilado, utilizando fosfatasa alcalina, para evitar la recircularización del vector.

Se realizaron los ajustes de las concentraciones de ADN para tener una relación 10:1 vector: inserto y se realizó la ligación según el protocolo del paquete hasta llegar a la titulación del fago. Desafortunadamente, empleando esta estrategia, tampoco se logró obtener clona alguna.

Una alternativa más era utilizar un método de extracción de ADN metagenómico diferente, ya que cabía la posibilidad de que el que hasta ese momento se había utilizado, estuviera acarreado algún contaminante que inhibiera ciertas reacciones.

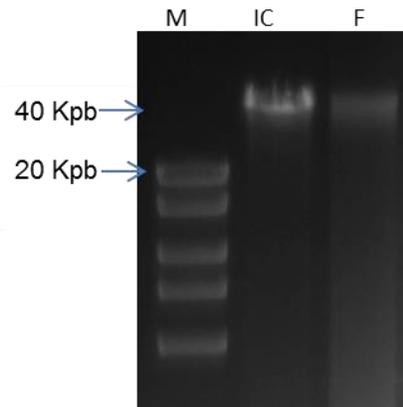
### 3.5.2 Variación en el método de extracción de ADN metagenómico

Dado que aun utilizando un método diferente de ligación no se logró obtener colonias, se utilizó otro método de extracción de ADN metagenómico. Para ello se eligió un método (el cual llamamos método dos) basado en el paquete de extracción de agua Epicentre<sup>MR</sup>, *Metagenomic DNA Isolation Kit For Water* (Cat. No. MGD08420) (descrito en la sección de Materiales y Métodos). Los filtros Sterivex<sup>MR</sup>– GV 0.22  $\mu$ M que contenían la biomasa procariota fueron lavados según las especificaciones del fabricante para realizar la extracción de ADN metagenómico. Recurriendo a este procedimiento, por medio de ruptura mecánica, se logró obtener ADN con buena integridad de un tamaño aproximado a los 30 a 40 Kpb (Figura 3.8).



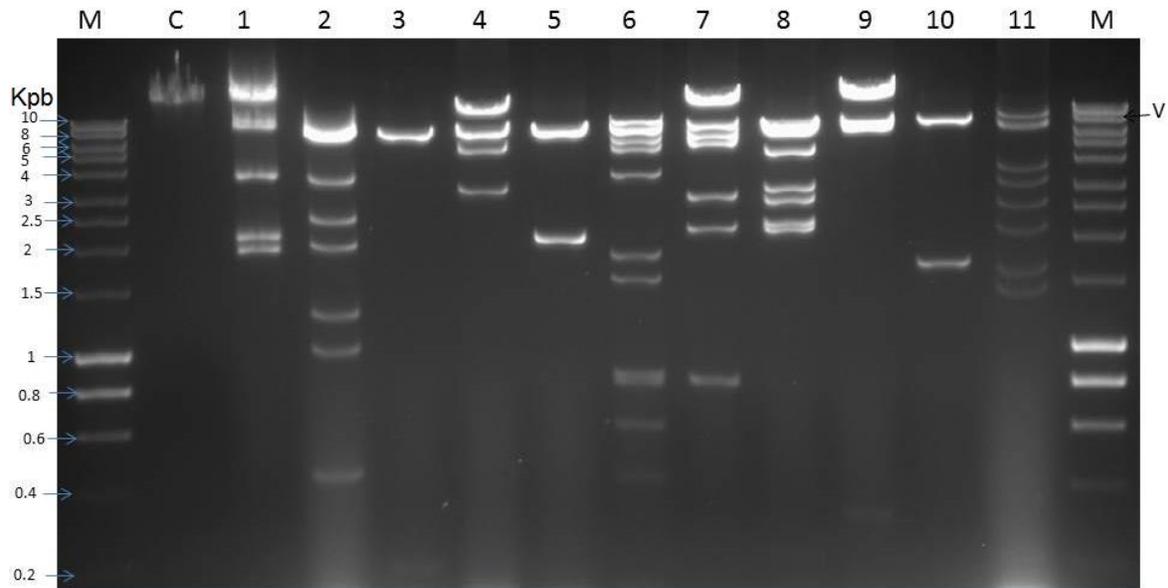
**Figura 3.8 ADN metagenómico extraído con el paquete de extracción de agua Epicentre<sup>MR</sup>.** Gel de agarosa al 1 %. Carriles 1 y 2 muestras de ADN metagenómico. M1: marcador de peso molecular de 1 Kpb; M2: marcador de peso molecular de 40 Kpb.

El ADN metagenómico obtenido se utilizó para seguir el procedimiento de construcción de la metagenoteca según el fabricante *Epicentre CopyControl<sup>MR</sup> HTP Fosmid Library Production* Cat. No. CCFOS059. Para llevar a cabo el protocolo del paquete, el ADN obtenido fue fraccionado en un gel de agarosa al 1%, para verificar que el tamaño de los fragmentos fuera de entre 30 y 40 Kpb (Figura 3.9).



**Figura 3.9** ADN metagenómico utilizado para la construcción de la metagenoteca. Gel de agarosa al 1%. M, marcador de peso molecular 1 Kpb; IC, control de inserto 40 Kpb; F, ADN obtenido del filtro Sterivex<sup>MR</sup>– GV 0.22  $\mu$ M.

Como se muestra en la Figura 3.9, comparando el ADN obtenido con el inserto control (IC, suministrado por el fabricante) se observa que los fragmentos del ADN extraído tienen el peso molecular deseado (entre los 30 a 40 Kpb) aunque también se nota un barrido en la parte inferior del gel el cual puede ser debido a la fragmentación mecánica. Estos fragmentos posteriormente de ser reparados fueron ligados en el vector pCC2FOS<sup>MR</sup>. El fósforo ya con los insertos fue introducido en la cepa de *E. coli* EPI300<sup>MR</sup>, de este experimento se obtuvieron 1000 clonas, las cuales fueron seleccionadas con cloranfenicol (Cm), posteriormente once clonas fueron seleccionadas al azar y se les extrajo el ADN fosmídico para que; mediante una digestión con *Bam*HI (Figura 3.10), se corroborara el tamaño de inserto de cada clona, se verificara el patrón de restricción y para evaluar que las clonas obtenidas portaran insertos con distinta secuencia.

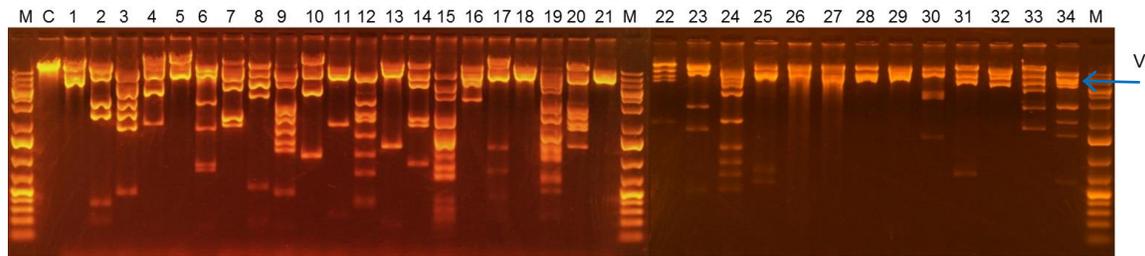


**Figura 3.10 Fósmidos de las clonas putativas de la metagenoteca digeridos con la enzima *Bam*HI.** Gel de agarosa al 1%. Digestión de clonas putativas con la enzima *Bam*HI. M, marcador de peso molecular de 1 Kpb; C, ADN control 40 Kpb; Carriles 1 a 11; clonas seleccionadas al azar; V, vector pCC2FOS<sup>MR</sup> 8.1 Kpb.

Como se observa en la Figura 3.10, las once clonas dieron un patrón de restricción diferente, lo cual sugiere que la metagenoteca resultante posee una diversidad de insertos.

En un primer ejercicio de clonación se obtuvieron aproximadamente 1000 colonias. Sin embargo, en una nueva prueba se aumentó la cantidad de ADN en la ligación cinco veces más y la eficiencia aumentó. De este segundo experimento se obtuvo una biblioteca metagenómica de  $2.5 \times 10^5$  clonas.

Para verificar la calidad de esta segunda metagenoteca, 34 nuevas clonas fueron sometidas a un análisis de restricción con la enzima *Bam*HI (Figura 3.11).



**Figura 3.11 Fósmidos de la metagenoteca digeridos con la enzima *Bam*HI.** Gel de agarosa al 1%. M, marcador de peso molecular de 1 Kpb; C, ADN control 40 Kpb; Carriles 1 a 34; productos de restricción de 34 clonas seleccionadas al azar; V, vector pCC2FOS<sup>MR</sup> 8.1 Kpb.

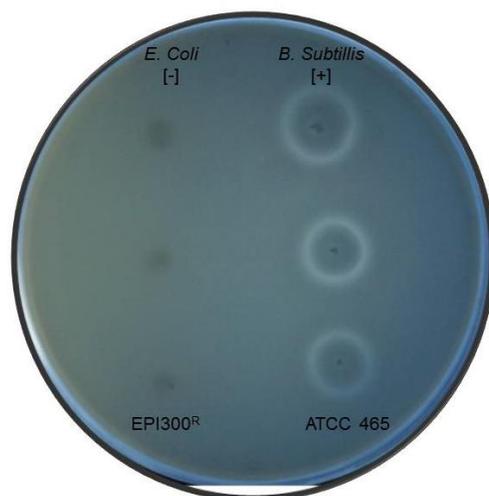
Con base en el patrón de restricción de *Bam*HI, el tamaño de inserto en la metagenoteca fue estimado entre 30 y 40 Kpb.

En relación con otras metagenotecas reportadas, en las que también se usó el fósido pCCFOS y como hospedera la cepa de *E. coli* EPI300<sup>MR</sup>, la construida en este trabajo presentó un número de clonas mayor por ejemplo, Lee *et al.* (2007) obtuvieron 30,000; Neveu *et al.* (2011) 17,000; Morris y Marchesi, (2015) 28,032. Sin embargo al igual que en estas tres metagenotecas, el tamaño de insertos fue de entre 30 y 38 Kpb lo cual es explicable, por que es el intervalo de tamaño de ADN que pueden empaquetar los fósidos.

### 3.6 Escrutinio de la metagenoteca

#### 3.6.1 Prueba de controles

Para descartar que por sí misma la cepa bacteriana EPI300<sup>MR</sup>, receptora de los fósidos, secretara proteasas que fuesen detectadas en el bioensayo con medio agar leche, se realizó un experimento en el cual se sembró la cepa hospedera en el medio de detección junto con *B. subtilis* como control positivo, ya que se sabe que esta especie secreta proteasas al medio (Figura 3.12).

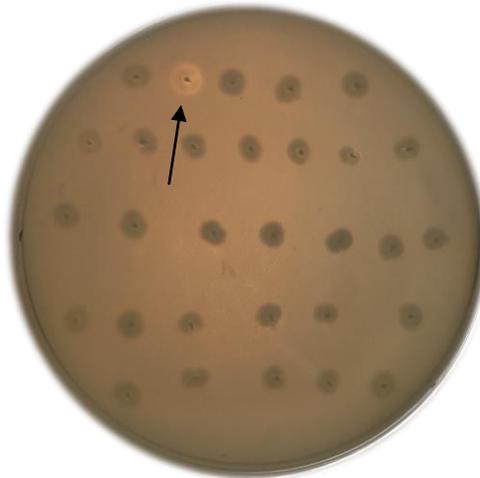


**Figura 3.12 Control positivo y control negativo sembrados en el medio de detección de clonas secretoras de proteasas.** Fotografía de una caja de Petri con agar leche al 2 %. Izquierda: EPI300<sup>MR</sup>; Derecha: *B. subtilis* (ATCC 465). Nótese los halos de hidrólisis rodeando las colonias de *B. subtilis* y su ausencia en las células hospederas EPI300<sup>®</sup>.

Como se puede observar en la Figura 3.12, no se produjo halo de hidrólisis en la cepa de *E. coli* EPI300<sup>MR</sup>, hospedera de la metagenoteca, pero sí en la de *B. subtilis*. Con esto se verificó que la cepa EPI300<sup>MR</sup> por sí sola no secretara proteasas al medio, detectables en este bioensayo, y por lo tanto fuera adecuada para realizar el escrutinio de las clonas.

### 3.6.2 Escrutinio sobre agar leche

El escrutinio de la metagenoteca se realizó en agar leche como se describe en Materiales y Métodos. Después de realizar el escrutinio de 21,000 clonas se seleccionaron aquellas que produjeron halo de hidrólisis alrededor de la colonia (Figura 3.13).



**Figura 3.13 Escrutinio funcional de clonas de la metagenoteca.** Fotografía de una caja de Petri con agar leche en la que se muestra una clona (flecha) que claramente se distingue del resto por formar el halo de hidrólisis característico de las colonias secretoras de proteasas crecidas en medio agar leche.

Siguiendo este criterio de selección, se aislaron 23 clonas candidatas secretoras de proteasas, las cuales fueron nombradas MAJ1 a MAJ23. En la Figura 3.14 se muestran nueve de las clonas de las seleccionadas por presentar el fenotipo hidrolítico.



**Figura 3.14 Clonas sobreproductoras de proteasas extracelulares.** Fotografía de una caja Petri con agar leche en la que se muestran nueve clonas seleccionadas

como posibles productoras de proteasas,. Como control positivo (+) se utilizó *B. subtilis* (ATCC 465) y como control negativo (-) *E. coli* (EPI300<sup>MR</sup>).

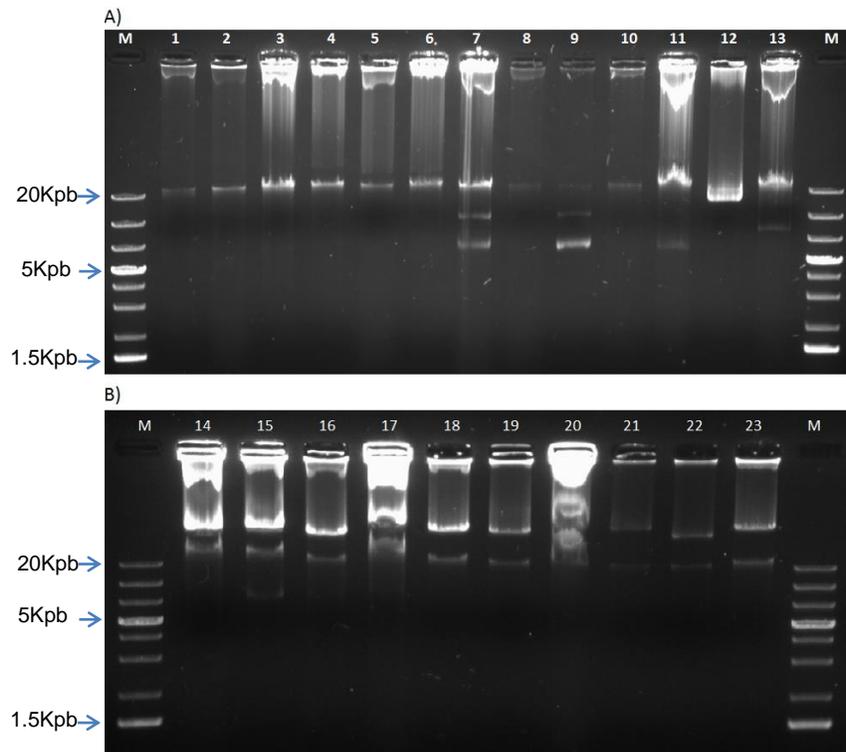
Las clonas seleccionadas, a diferencia del control negativo, presentan un halo de hidrólisis alrededor de la colonia, el cual es muy similar al observado en el control positivo (*B. subtilis*), por lo que fueron catalogadas como clonas putativas sobreproductoras de proteasas extracelulares (Figura 3.14).

El número de clonas positivas obtenidas del escrutinio de este trabajo fue muy similar al obtenido en una metagenoteca de arena del desierto de Gobi (Neveu *et al.*, 2011), la cual fue construida en fósmidos y también se empleó a *E. coli* EPI300<sup>MR</sup> como hospedero, los autores reportan que de 17,000 clonas se obtuvieron 16 que presentaron halos hidrolíticos. En ambos casos, la frecuencia de clonas positivas fue aproximadamente de una por cada mil sometidas a escrutinio. Tal vez esta sea la eficiencia aproximada que se obtendría en cualquier escrutinio similar, independientemente del origen del ADN metagenómico –debido a la frecuencia de genes de proteasas extracelulares capaces de expresarse en *E.coli*, entre el número de genes totales-. Sin embargo, para llegar a esta conclusión, se requerirían más trabajos equivalentes, y de momento son todavía muy escasos.

### **3.6.3 Extracción del ADN fosmídico de las clonas positivas**

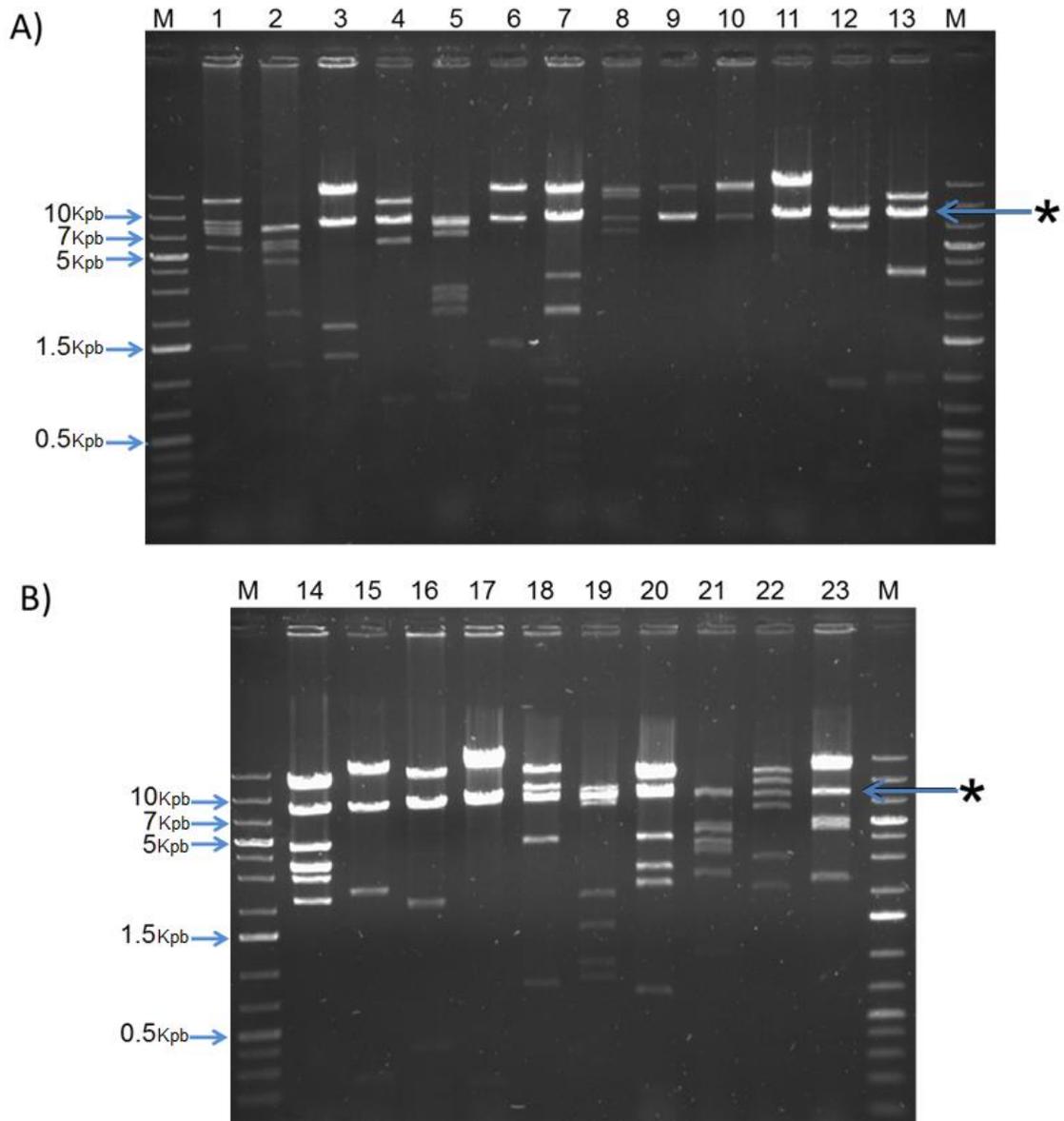
Utilizando el paquete de extracción *Epicentre FosmidMAX<sup>TM</sup>*, y partiendo de cultivos celulares de 1.5 mL, se obtuvo ADN fosmídico de las 23 clonas positivas de la prueba funcional de secreción de proteasas (seleccionadas durante el escrutinio de 21,000 clonas de la biblioteca metagenómica).

En la Figura 3.15 se muestra el análisis electroforético del ADN fosmídico resultado de la extracción, en donde puede apreciarse que se logró obtener ADN de buena calidad, íntegro, de alto peso molecular y a buena concentración.



**Figura 3.15 ADN de las clonas que produjeron halo hidrolítico en agar leche.** Fotografía del gel de electroforesis en el que se analizó el ADN fosmídico de las clonas positivas. A) Carriles 1 a 13: clonas MAJ1 a MAJ13 y B) Carriles 14 al 23: clonas MAJ14 a MAJ23. M: marcador de peso molecular de 1 Kpb.

Para determinar si cada clona era distinta de las demás y para conocer el tamaño aproximado de cada inserto, se realizó una digestión utilizando la enzima de restricción *Bam*HI (Figura 3.16). Los sitios de corte de esta enzima, dentro del fosmido pCC2FOS<sup>MR</sup>, se encuentran exclusivamente flanqueando el inserto, por lo cual al cortar con esta enzima, se liberan el inserto, y si en éste se encuentran sitios de restricción para esta endonucleasa, el patrón de restricción nos indica si cada inserto es igual o diferente de los demás.



**Figura 3.16 Análisis de restricción de las clonas positivas utilizando *Bam*HI.** A) Carriles 1 a 13 clonas MAJ1 a MAJ13; B) carriles 14 a 23 clonas MAJ14 a MAJ23; M: marcador de peso molecular 1 Kpb; Flecha del lado derecho con \*: indica la banda correspondiente al peso molecular del vector pCC2FOS<sup>MR</sup> (8.1 Kpb).

Como se observa en la Figura 3.16, la digestión de las clonas MAJ1 a MAJ23 produjo un patrón de restricción diferente para cada clona analizada, por lo cual podemos inferir que cada una de ellas es diferente de las otras. La suma de las bandas observadas indica que el tamaño de los insertos liberados del vector es de aproximadamente 30 a 40 Kpb.

### 3.7 Análisis *in silico* de los resultados de secuenciación

#### 3.7.1 Secuenciación de los fósmidos

Los insertos de los fósmidos de las 23 clonas obtenidas del escrutinio de la biblioteca metagenómica fueron secuenciados y ensamblados como servicio externo por la Universidad Estatal de Utah (Como se describe en Materiales y Métodos).

Las clonas MAJ7, MAJ8, MAJ14, MAJ20 y MAJ21 (mezcla de cinco clonas) se secuenciaron utilizando una etiqueta conjunta; al igual que las clonas MAJ2, MAJ3, MAJ4, MAJ5, MAJ6, MAJ9, MAJ11, MAJ12, MAJ13, MAJ15, MAJ16, MAJ17, MAJ18, MAJ22 y MAJ23 (mezcla de 15 clonas). Como resultado de la secuenciación de estas clonas, se obtuvieron 21 y 24 **contigs** respectivamente. Dado que la secuenciación fue en lotes, hasta este momento ha sido difícil discriminar cuál **contig** corresponde a cada clona.

Tres clonas fueron secuenciadas individualmente MAJ1, MAJ10 y MAJ19. Como resultado se obtuvo el inserto completo de la clona MAJ1 en un solo **contig** de 35,616 pb. Mientras que de las clonas MAJ10 y MAJ19 se lograron ensamblar cinco **contigs** con tamaños de 14,843; 5,490; 7,749; 12,189 y 1,343 pb y tres **contigs** con tamaños de 23,073; 539; 6,926 pb respectivamente. El tamaño promedio de los insertos de estas tres clonas fue de ~35 Kpb, lo cual concuerda con el resultado estimado por los patrones de restricción realizados con *Bam*HI (Figura 3.16).

El análisis *in silico* de las secuencias obtenidas, para la anotación de los marcos de lectura abiertos (ORFs), se realizó utilizando los algoritmos para la predicción de genes GLIMMER3 (Salzberg *et al.*, 1998) y fgenesB (Solovyev and Salamov, 2011). Posteriormente se realizó la traducción de los ORFs y con BlastP (Altschul *et al.*, 1990) se realizó la búsqueda de las proteínas homólogas, más cercanas reportadas, en la base de datos del NCBI.

Para el inserto de las clonas MAJ1 y MAJ19 se predijeron 38 y 36 genes respectivamente. Sin embargo, ninguno de ellos presentó homología con genes relacionados con proteasas. Los genes homólogos más cercanos encontrados para los 38 ORFs de la

clona MAJ1 correspondieron a los géneros bacterianos *Sphingomonas* sp., *Sphingopyxis* sp. y *Sphingobium* sp., estos pertenecen al filo proteobacteria, clase alfa proteobacteria, la cual es encontrada en diferentes ambientes, tanto terrestres como acuáticos. Estos géneros se han utilizado en aplicaciones biotecnológicas tales como biorremediación ambiental y en la industria alimenticia (por su producción de exopolisacáridos). Por otro lado, los 36 genes homólogos predichos para la clona MAJ19 también están relacionados con el género *Sphingomonas* sp. y con el género *Novosphingobium* sp., este último estrechamente relacionado con microorganismos degradadores de compuestos aromáticos.

Dado que por homología en las clonas MAJ1 y MAJ19 no se encontraron proteínas putativas relacionadas con proteasas, será necesario implementar otras estrategias para la identificación de las secuencias causantes del fenotipo proteolítico. Una posibilidad consiste en realizar la sub clonación del fósido en un plásmido y realizar el escrutinio funcional sobre las clonas resultantes. Otra es seguir una estrategia similar a la reportada por Culligan *et al.* (2013), quienes utilizando el sistema de mutagénesis con transposones, fueron capaces de aislar un nuevo gen de tolerancia a salinidad (*stIA*).

### **3.7.2 Análisis del inserto de la clona MAJ10**

Los posibles genes de la clona MAJ10 se anotaron utilizando el predictor de genes GLIMMER3, con el cual se predijeron 50 marcos de lectura abiertos (ORFs) y sus respectivas proteínas putativas codificadas utilizando la herramienta bioinformática BlastX del NCBI (Altschul *et al.*, 1990), los resultados se muestran en la Tabla 3.1.

**Tabla 3.1** Lista de las proteínas homólogas más cercanas a las codificadas en el inserto del fósido de la clona MAJ10. Predichas con GLIMMER3

ORF	Número de aa	Proteínas putativas	Organismo	% C	Valor E	% ID	Accesión
1	166	P.H	<i>Algicola sagamiensis</i>	98%	1e-17	36%	WP_018692576.1
2	232	Metiltransferasa	<i>Mariprofundus ferrooxydans</i>	97%	1e-36	33%	WP_009850620.1
3	333	P.H	<i>Shewanella loihica</i> PV-4	94%	5e-74	45%	YP_001093436.1
4	41	P.H	<i>Alteromonas macleodii</i> ATCC 27126	92%	2e-04	55%	YP_006747845.1
<b>5*</b>	<b>614</b>	<b>Peptidasa S8</b>	<b><i>Chitinimonas koreensis</i></b>	<b>92%</b>	<b>0.0</b>	<b>61%</b>	<b>WP_028445624.1</b>
6	65	GA15457	<i>Drosophila pseudoobscura</i>	56%	0.81	46%	XP_001358411.2
7	52	Acetolactato sintasa II subunidad grande	<i>Candidatus Puniceispirillum marinum</i> IMCC1322	55%	1.9	48%	YP_003551104.1
<b>8*</b>	<b>607</b>	<b>Peptidasa S8</b>	<b><i>Chitinimonas koreensis</i></b>	<b>99%</b>	<b>0.0</b>	<b>68%</b>	<b>WP_028445624.1</b>
9	162	N-acetiltransferasa GCN5	<i>Granulicella tundricola</i> MP5ACTX9	89%	1e-40	48%	YP_004219044.1
10	219	P.H	<i>Melioribacter roseus</i> P3M-2	96%	5e-29	36%	YP_006528722.1
11	281	Fosfatidilinositol Fosfolipasa C	<i>Photobacterium</i> sp. AK15	76%	4e-77	58%	WP_007463772.1
12	44	Fosfatidilinositol Fosfolipasa C región X	<i>Clostridium cellulovorans</i> 743B	93%	1e-15	76%	YP_003844187.1
13	52	P.H	<i>Bacillus</i> sp. 123MFChir2	69%	0.002	47%	WP_020060820.1
<b>14</b>	<b>45</b>	<b>S.S</b>	<b>n/a</b>	<b>n/a</b>	<b>n/a</b>	<b>n/a</b>	<b>n/a</b>
15	167	P.H	<i>Acidovorax avenae</i>	95%	8e-55	60%	WP_019703938.1
16	165	P.H	<i>Variovorax paradoxus</i>	93%	4e-33	42%	WP_019655369.1
<b>17</b>	<b>51</b>	<b>S.S</b>	<b>n/a</b>	<b>n/a</b>	<b>n/a</b>	<b>n/a</b>	<b>n/a</b>
18	270	P.H	<i>Labrenzia alexandrii</i>	86%	2e-48	38%	WP_008190717.1
19	314	Factor de elongación G	<i>Pseudogulbenkiania ferrooxidans</i>	98%	2e-172	79%	WP_008952344.1
20	1052	P.H	<i>Sulfuritalea hydrogenivorans</i> sk43H	44%	7e-27	30%	BAO29611.1
21	153	P.H	<i>Thalassiosira oceanica</i>	26%	1e-04	68%	EJK55459.1

## CAPÍTULO III

22	120	Proteína de unión a calcio	Paracoccus sp. TRP	99%	2e-14	54%	WP_010400424.1
23	52	Proteína de ramificación de glucógeno	<i>Lactobacillus mucosae</i>	69%	1-00	42%	WP_0064499197.1
24	86	Regulador de respuesta de quimiotaxis que contiene un dominio receptor CheY y un dominio metilesterasa	<i>Desulfomonile tiedjei</i> DSM 6799	28%	5.6	48%	YP_006445055.1
25	59	Proteína de unión a calcio de tipo hemolisina	<i>Serratia marcescens</i> FGI94	64%	2.6	45%	YP_0073444647.1
26	41	P.H	<i>Cardiobacterium valvarum</i>	90%	3.1	41%	WP_006984185.1
27	38	Proteína putativa de unión a calcio	<i>Xenococcus</i> sp. PCC 7305	84%	0.002	66%	WP_006507483.1
28	40	Hemolisina	<i>Xylella fastidiosa</i>	97%	4.2	44%	WP_023907444.1
29	258	P.H	<i>Leeia oryzae</i>	58%	1e-06	28%	WP_018152984.1
<b>30</b>	<b>51</b>	<b>S.S</b>	<b>n/a</b>	<b>n/a</b>	<b>n/a</b>	<b>n/a</b>	<b>n/a</b>
31	314	P.H	<i>Candidatus Nitrospira defluvii</i>	61%	8e-23	34%	YP_003799066.1
32	52	P.H	<i>Escherichia coli</i>	69%	5.7	42%	ADZ45142.1
33	45	Enzima de restricción EcoRV	<i>Candidatus Accumolibacter</i> sp. BA-91	44%	7.7	55%	EXI96166.1
34	237	P.H	<i>Methylomonas methanica</i> MC09	66%	8e-05	27%	YP_004512476.1
35	313	P.H	<i>Kingella oralis</i>	97%	3e-95	50%	WP_003796942.1
<b>36</b>	<b>43</b>	<b>S.S</b>	<b>n/a</b>	<b>n/a</b>	<b>n/a</b>	<b>n/a</b>	<b>n/a</b>
37	77	Dominio proteico hélice-giro-hélice	<i>Methylomonas methanica</i> MC09	88%	1e-29	78%	YP_004510965.1
38	47	P.H	<i>Caenorhabditis briggsae</i>	61%	6.0	45%	XP_002646845.1
39	399	Proteína de la familia 3 Aciltransferasa	<i>Janthinobacterium</i> sp. HH01	92%	2e-134	57%	WP_008447439.1
40	53	Fosfo pantotenoil cisteina decarboxilasa	Marine actinobacterium PHSC20C1	83%	4.2	36%	WP_009772248.1
41	58	P.H	<i>Chitiniphilus shinanonensis</i>	96%	5e-26	79%	WP_018748194.1
42	63	P.H	<i>Chitiniphilus shinanonensis</i>	92%	4e-26	83%	WP_018748194.1
43	470	Proteína putativa de unión a calcio hemolisina	<i>Bradyrhizobium oligotrophicum</i> S58	72%	1e-29	39%	YP_007515571.1

44	60	Factor de iniciación 2	<i>Candida tenuis</i> ATCC 10573	50%	7.5	47%	XP_006684972.1
45	2013	Proteína hemolisina HlyJ	<i>Pseudomonas cichorii</i> JBC1	93%	3e-157	36%	AHF66653.1
46	228	Transposasa	<i>Burkholderia</i> sp. Ch1-1	96%	1e-108	82%	WP_007176182.1
47	59	Proteína IS Transposasa	<i>Ralstonia solanacearum</i> GMI1000	84%	3e-18	80%	NP_518369.1
48	364	Proteína de elemento móvil	<i>Burkholderia vietnamiensis</i>	96%	2e-169	67%	WP_021163056.1
49	84	Receptor de la membrana exterior de ferri entero queelina y colicina	<i>Pseudoalteromonas luteoviolacea</i>	90%	0.11	35%	WP_023399318.1
50	477	Factor de elongación G	<i>Pseudogulbenkiania</i> sp. NH8B	99%	0.0	78%	YP_004848819.1

Abreviaturas y símbolos: ORFs (marcos de lectura abierto); aa (aminoácidos); %C (porcentaje de cobertura; %ID (porcentaje de identidad); P.H (proteína hipotética); S.S (sin similitud); n/a (no aplica); Los ORFs marcados con asterisco (\*) indican genes putativos de proteasas. El texto en negritas significa que no se encontraron proteínas putativas utilizando BlastX en el NCBI.

De los 50 genes putativos analizados (Tabla 3.1), 19 están codificados en la cadena positiva (sentido) y 31 en la cadena complementaria (antisentido). Cincuenta por ciento de los genes codifican proteínas, con diversos grados de identidad con proteínas homólogas de varios géneros microbianos (el intervalo de identidad de nucleótidos va de 33% a 82%). Para el 8% de las proteínas codificantes (4 de 50) no se encontró homología con secuencias previamente reportadas en las bases de datos del NCBI, lo cual podría deberse a que corresponden a genes erróneamente predichos o nuevos genes. Finalmente, el 42% restante de las proteínas son hipotéticas, es decir, que a pesar de tener similitud con alguna de las ya reportadas en bases de datos, se desconoce su función. Estos resultados, aunados a que los porcentajes de identidad a nivel de proteína están en un intervalo del 27% al 82% y a que la mayoría de las proteínas predichas son hipotéticas (predichas en otros trabajos, pero sin demostración experimental de su existencia) o no tienen homología con secuencias previamente reportadas, sugieren que el fragmento clonado en MAJ10 proviene de un microorganismo hasta ahora no secuenciado, y por lo tanto sus genes se encuentran aun sin reportar en las bases de datos.

Los resultados obtenidos con GLIMMER3 fueron verificados utilizando fgenesB, del programa en red Softberrey (Solovyev y Salamov, 2011). A diferencia de GLIMMER3, fgenesB (Solovyev y Salamov, 2011) utiliza un set de secuencias relacionadas con un grupo de organismos en específico, lo que lo hace más preciso al realizar predicciones (Kunin *et al.*, 2008). En este caso se utilizó el grupo de bacterias y arqueas. Al realizar el análisis se encontraron 35 proteínas putativas (Tabla 3.2). Es normal encontrar diferencias cuando se usan distintos programas informáticos para la predicción de genes, debido a que manejan diferentes parámetros en sus algoritmos, por ejemplo, algunos toman diferentes codones de inicio (ATG) para el inicio lo cual causa que haya variación en el tamaño del ORF, teniendo como consecuencia diferentes valores de E, el cual es el valor estadístico que representa que el alineamiento no se debió al azar, por lo tanto valores bajos del valor E representan un mejor alineamiento entre las secuencias.

**Tabla 3.2** Lista de proteínas homologas codificados en el inserto del fósido de la clona MAJ 10 predichos con fgenesB.

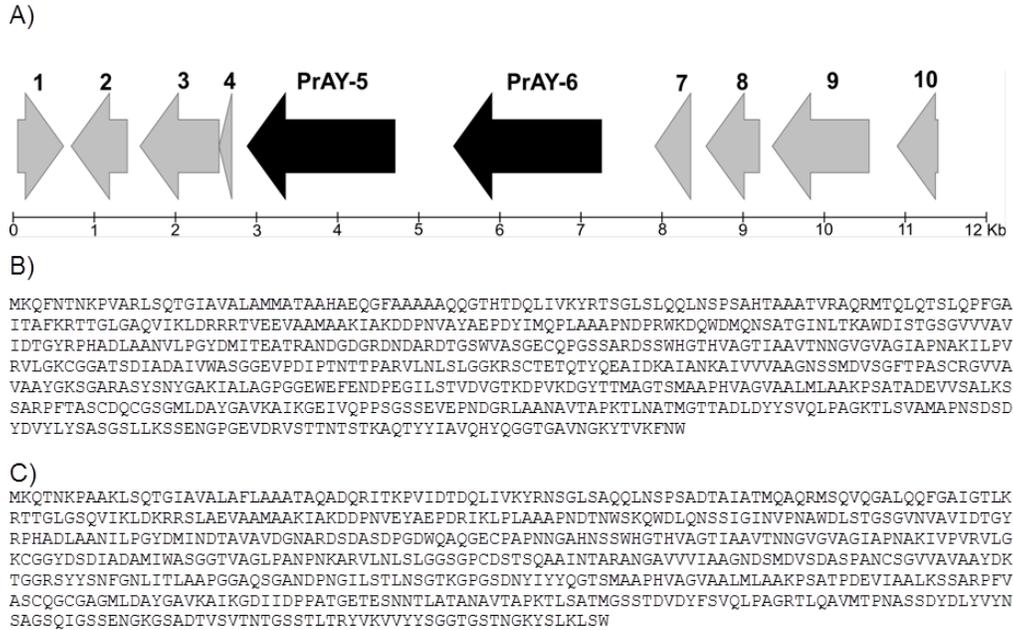
ORF	Número de aa	Proteínas putativas	Organismo	% C	Valor E	% ID	Accesión
1	195	P.H	<i>Burkholderia ubonensis</i>	98%	3e-46	46%	KWZ58790.1
2	231	P.H	<i>Chitinimonas koreensis</i>	96%	5e-44	45%	WP_051319001.1
3	332	Proteína caracterizada	no <i>Achromobacter 2789STDY5608618</i> sp.	92%	1e-116	56%	CUJ34213.1
4	61	Proteína caracterizada	no <i>Achromobacter 2789STDY5608618</i> sp.	98%	3e-22	82%	CUJ34213.1
<b>5*</b>	<b>614</b>	<b>Peptidasa s8</b>	<b><i>Chitinimonas koreensis</i></b>	<b>94%</b>	<b>0.0</b>	<b>60%</b>	<b>WP_028445624.1</b>
<b>6*</b>	<b>607</b>	<b>Peptidasa s8</b>	<b><i>Chitinimonas koreensis</i></b>	<b>100%</b>	<b>0.0</b>	<b>68%</b>	<b>WP_028445624.1</b>
7	150	acetiltransferasa	<i>Acidobacterium PMMR2</i> sp.	100%	6e-42	46%	WP_026441689.1
8	218	P.H	<i>Chitinimonas koreensis</i>	91%	1e-82	60%	WP_051319279.1
9	397	Fosfatidilinositol Fosfolipasa C	<i>Photobacterium marinum</i>	84%	4e-117	54%	WP_007463772.1
10	166	P.H	<i>Acidovorax</i>	96%	5e-59	59%	WP_056410286.1
11	185	P.H	<i>Acidovorax</i> sp. <i>Leaf78</i>	91%	7e-33	38%	WP_056163330.1
12	269	Transportador ABC	<i>Magnetospirillum gryphiswaldense MSR-1</i>	96%	5e-38	34%	CAM74722.1
13	72	S.S	S.S	n/a	n/a	n/a	n/a

14	313	Factor de elongación g	<i>Pseudogulbenkiania ferrooxidans</i>	98%	6e-172	79%	WP_008952344.1
15	1078	P.H	<i>Candidatus Nitrospira nitrosa</i>	87%	2e-45	25%	CUS38404.1
16	787	Proteína de unión a calcio	<i>Leptolyngbya sp. Heron Island J</i>	99%	4e-49	36%	WP_023070311.1
17	222	P.H	<i>Ottowia sp. oral taxon 894</i>	68%	4e-09	34%	WP_050714282.1
18	50	S.S	S.S	n/a	n/a	n/a	n/a
19	313	P.H	<i>Nitrospira defluvii</i>	61%	1e-22	34%	WP_013249961.1
20	47	S.S	S.S	n/a	n/a	n/a	n/a
21	236	S.S	S.S	n/a	n/a	n/a	n/a
22	312	P.H	<i>Methylomonas sp. LW13</i>	99%	2e-122	56%	WP_033159067.1
23	76	Regulador transcripcional	<i>Methylomonas methanica</i>	89%	2e-29	78%	WP_013816738.1
24	387	Aciltransferasa	<i>Chromobacterium</i>	98%	1e-139	56%	WP_043572702.1
25	66	P.H	<i>Chitiniphilus shinanonensis</i>	84%	7e-26	79%	WP_018748194.1
26	76	P.H	<i>Chitiniphilus shinanonensis</i>	76%	1e-25	83%	WP_018748194.1
27	499	P.H	<i>Methylobacter tundripaludum</i>	90%	1e-36	30%	WP_031435808.1
28	2055	Hemolisina	<i>Pseudomonas aeruginosa</i>	94%	0.0	37%	WP_011270159.1
29	218	Transposasa	<i>Burkholderia phenoliruptrix</i>	97%	9e-121	82%	WP_015004346.1
30	58	Transposasa	<i>Burkholderia sp. PAMC 26561</i>	86%	1e-18	78%	AME26330.1
31	367	Transposasa	<i>Ralstonia solanacearum</i>	97%	1e-165	66%	WP_011004245.1
32	116	Asparagina sintetasa b	<i>Terrisporobacter othiniensis</i>	27%	1.3	41%	WP_039677874.1
33	476	Factor de elongación g	<i>Pseudogulbenkiania sp.</i>	100%	0.0	78%	WP_024304203.1
34	194	P.H	<i>Nitrospira defluvii</i>	84%	2e-28	39%	WP_013249960.1
35	184	P.H	<i>Nitrospira defluvii</i>	82%	7e-13	33%	WP_013249961.1

Abreviaturas y símbolos: ORF (marco de lectura abierto); aa (aminoácidos); %C (porcentaje de cobertura; %ID (porcentaje de identidad a nivel aminoácido); P.H (proteína hipotética); S.S (sin similitud); n/a (no aplica); Los ORFs marcados con asterisco (\*) indican genes putativos de proteasas. El texto en negritas significa que no se encontraron proteínas homologas utilizando BlastP en el NCBI.

Algo importante que resultó consistente en ambas predicciones fue que con los dos se identificaron dos ORFs que codifican proteínas con dominios de proteasas (proteína 5 y 6.

Tabla 9). Estos fueron nombrados Proteasa Acuífero de Yucatán 5 (*PrAY5*) y 6 (*PrAY6*) respectivamente (Figura 3.17).



**Figura 3.17 Análisis *in silico* del inserto del fósido de la clona MAJ10.** A) Se muestran los primeros 12,000 pb del *contig* y sobre él se esquematizan con flechas 10 de las proteínas codificadas (predichas con *fgenesB*); B) se muestran las secuencias de las proteínas traducidas de *PrAY5*; y C) de *PrAY6*.

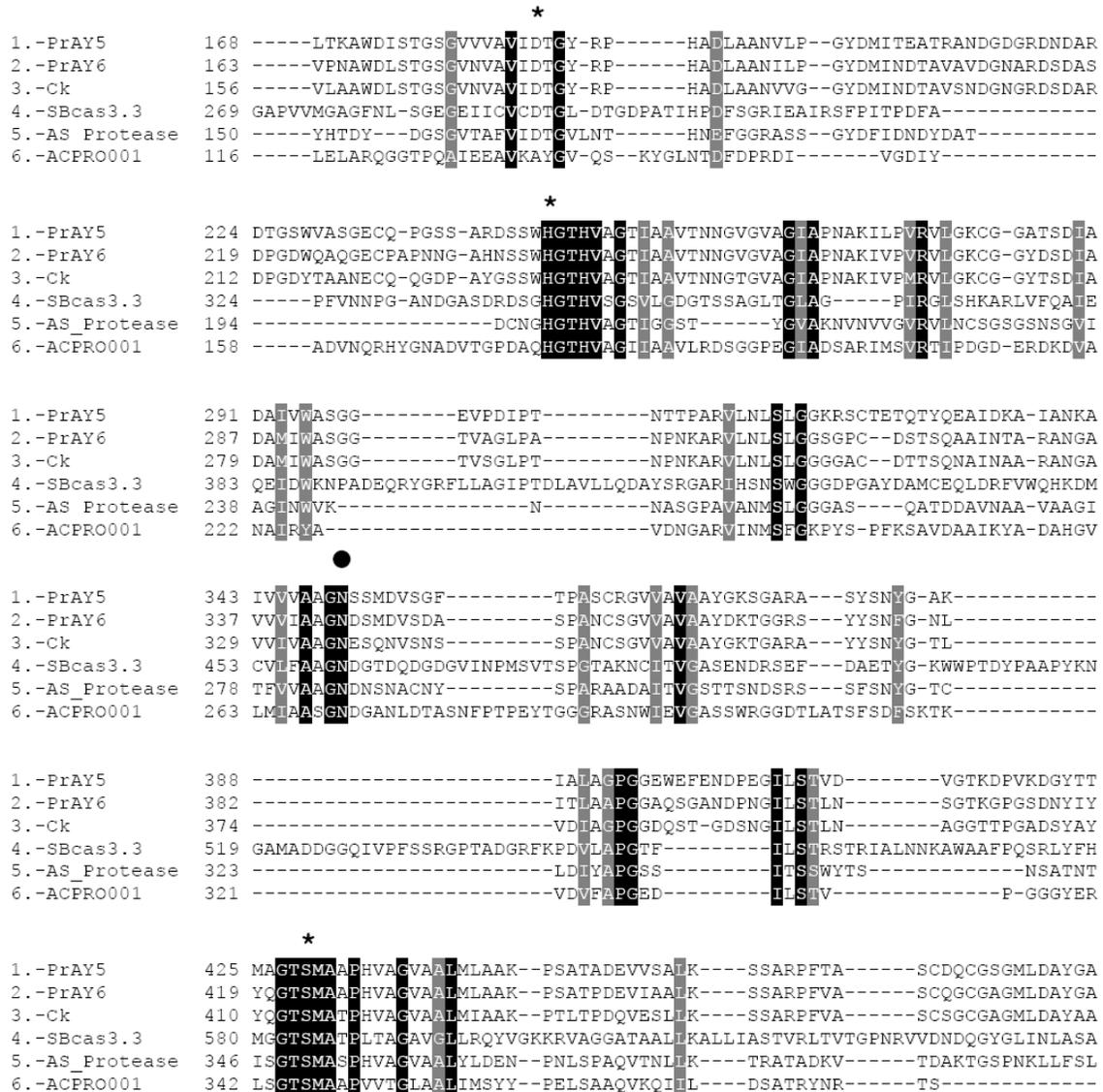
Como se muestra en las Tablas 3.1 y 3.2, las proteínas codificadas por las secuencias *PrAY5* y *PrAY6* tienen como homólogo más cercano una peptidasa S8 de *Chitinimonas koreensis* (con identidades de 60% y 68% respectivamente). Esta bacteria pertenece al filo Proteobacteria, clase Betaproteobacteria, fue descrita por Kim y colaboradores (2006), es aerobia, Gram-negativa, de vida libre y la mayoría de sus genes se reportaron como proteínas hipotéticas. Dado el bajo porcentaje de identidad a nivel de aminoácidos, y a que el valor E es cero las secuencias de *PrAY5* y *PrAY6* posiblemente no pertenecen a esta especie. El alineamiento entre *PrAY5* y *PrAY6* mostró que entre ambas hay 66.5%

de identidad a nivel de aminoácidos, por lo que se consideran dos proteínas diferentes (Figura 3.17, Tabla 3.2).

### 3.7.3 Caracterización *in silico* de las secuencias *PrAY5* y *PrAY6*

Los ORFs correspondientes a *PrAY5* y *PrAY6* se encuentran localizados en el **contig** de 14,843 pb de la clona MAJ10, *PrAY5* está muy cercano a *PrAY6* (Figura 3.17), separado únicamente por 718 pb. *PrAY5* consta de un marco de lectura abierto de 1845 pb y *PrAY6* de 1824 pb, ambos tienen el codón de inicio ATG y el codón de paro TAA, por lo que las proteínas que codifican deben ser de aproximadamente 614 una y de 607 aminoácidos la otra (Figura 3.17).

Como ya se había mencionado, las proteínas homólogas más cercanas a *PrAY5* y *PrAY6* son las codificadas por *C. koreensis*. Para realizar la búsqueda de dominios proteolíticos se utilizó la base de datos de dominios conservados CDD del NCBI. Tanto *PrAY5* como *PrAY6* comparten el dominio proteasa de la familia S8, subfamilia 13 sin caracterizar (Peptidasa\_S8\_13). La familia S8 es un miembro del clan SB, al que se le denomina de las peptidasas S8 o serina subtilasas, las cuales pueden ser exo o endo peptidasas. Estas peptidasas se caracterizan por tener una triada catalítica constituida por los aminoácidos D / H / S (Figura 3.18) similar a la que se encuentra en la tripsina. Sin embargo, no tienen la misma estructura tridimensional ni presentan homología con tripsina. La estabilidad de este tipo de peptidasas suele incrementarse adicionando calcio, y algunos miembros de este clan presentan puentes disulfuro. Las enzimas con este dominio pueden ser intra o extra celulares y algunas pueden funcionar a temperaturas y valores de pH extremos (Marchler-Bauer *et al.*, 2011).



**Figura 3.18 Alineamiento de secuencias de peptidasa S8.** Los asteriscos muestran los aminoácidos característicos de la triada catalítica de las peptidasas S8. El círculo muestra el oxianión Asn (N).

*PrAY5* y *PrAY6* también comparten un dominio C-terminal pre-peptidasa (PCC), que normalmente se encuentra en peptidasas bacterianas secretadas inmaduras, pero no se encuentra en las peptidasas activas (Figura 3.20).

La presencia de este dominio PCC aunada al fenotipo proteolítico de las clonas, hace suponer que *PrAY5* y *PrAY6* son proteasas que se transportan al exterior celular. Para saber si había indicios de que *PrAY5* y *PrAY6* fueran proteasas transmembranales se utilizó la herramienta TMHMM (Sonnhammer *et al.*, 1998) la cual predice la presencia de posibles hélices transmembranales. Como resultado se obtuvo que estas proteínas no las tienen, y, aunque este análisis es producto de una predicción teórica apunta a que las proteínas son extracelulares. Otra observación que fortalece esta hipótesis, es la presencia, en ambas secuencias, de un péptido señal el cual fue predicho utilizando SignalP.

Otra herramienta utilizada fue ExPASy ProtParam (Gasteiger *et al.*, 2005) con esta se logró estimar la masa molecular y el punto isoeléctrico (pI) teórico, que para *PrAY5* fue de 63,515 Da y pI 5.83; mientras que para *PrAY6* es de 61,803 Da y su pI es de 5.62 (Tabla 3.3). En la tabla 3.3 se muestra un resumen de las herramientas bioinformáticas utilizadas para los análisis *in silico* que se llevaron a cabo.

**Tabla 3.3** Análisis *in silico* de las secuencias PrAY5 y PrAY6

BD/Herramienta Utilizada	Descripción	Características Identificadas		Ref.
		PrAY5	PrAY6	
Base de datos de dominios conservados (CDD)	Identificación de dominios proteicos	Dominio Peptidasa_S8_13	Dominio PCC	Marchler-Bauer <i>et.al.</i> , 2011
Expasy protparam	Permite obtener varios parámetros físicos y químicos para una determinada proteína almacenada en la base de datos de Swiss-Prot o TrEMBL o para una secuencia introducida por el usuario	Peso molecular 63,515 Da Punto isoeléctrico 5.83	Peso molecular 61,803 Da Punto isoeléctrico 5.62	Wilkins <i>et. al.</i> , 1999
HMMER	Busca en las bases de datos secuencias homólogas de proteínas y hace alineamiento de proteínas	Proteína sin caracterizar de una especie sin clasificar	Proteína sin caracterizar de una especie sin clasificar	Finn <i>et.al.</i> , 2011
TMHMM	Predice hélices transmembranales en proteínas	Proteína secretable	Proteína secretable	Sonnhammer <i>et.al.</i> , 1998
BProm	Predice promotores bacterianos	n/e	n/e	www.softberry.com

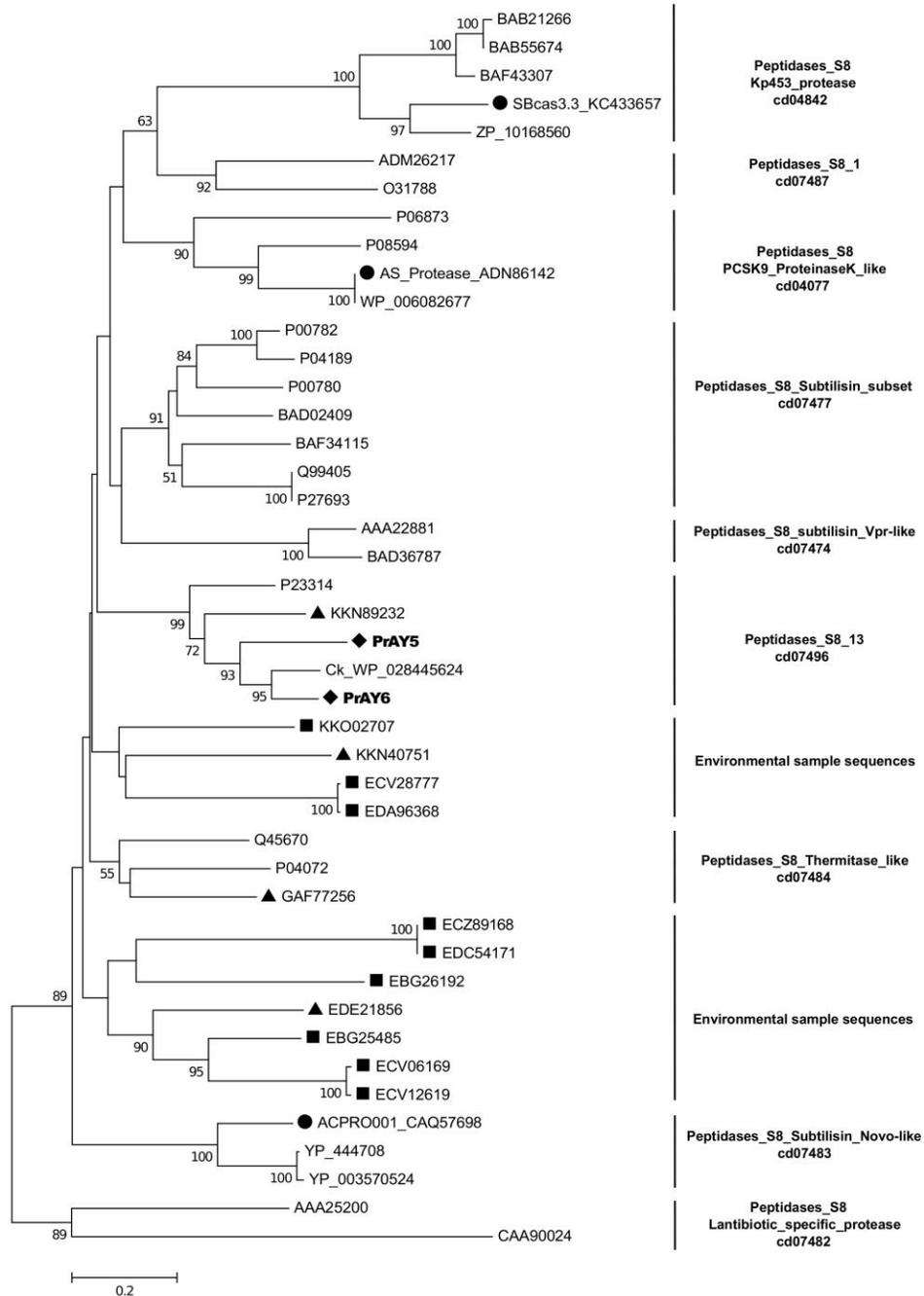
Abreviaturas: BD (base de datos); Ref. (Referencia); CDD (*Conserved domain database*); n/e (no encontrados).

### 3.7.4 Relación filogenética de las PrAYs con otras proteasas

Para evaluar la relación filogenética de PrAY5 y PrAY6 con otras proteasas se realizó un alineamiento en el cual se incluyeron: el homólogo más cercano a ambas secuencias (una peptidasa S8 de *Chitinimonas koreensis*), tres secuencias de proteasas que han sido encontradas y caracterizadas por metagenómica funcional SBcas3.3 (Biver *et al.*, 2013), ACPRO001 (Zhang *et al.*, 2011) y proteasa-AS (Pushpam *et al.*, 2011).

El alineamiento de estas seis secuencias mostró que en todas ellas se encuentra conservada la triada catalítica D, H, S y el oxianión (N); aminoácidos que están involucrados en el mecanismo catalítico de las peptidasas S8 (Figura 3.18).

La relación de PrAYs con otros miembros de la familia S8 se obtuvo comparándola con secuencias de la base de datos de dominios conservados del NCBI sub set cd00306. A este análisis se anexaron secuencias homólogas encontradas en la base de datos ambientales del NCBI. Para seleccionar estas últimas se tomaron en cuenta tres criterios i) que tuvieran una cobertura de más del 70%, ii) que presentaran un valor E menor o igual a  $10^{-6}$  y iii) la presencia de la triada catalítica característica de las peptidasas S8. Cuando se utilizó PrAY5 como referencia se encontraron cuatro secuencias y cuando se usó PrAY6, se encontraron nueve. En total 44 secuencias se alinearon y posteriormente se creó un dendograma utilizando el método de unión de vecinos (Neighbor-Joining) (Figura 3.19).



**Figura 3.19** Relación filogenética de PrAY5 y PrAY6 con respecto a otras secuencias de la familia subtilasa S8 y otras secuencias de muestras ambientales. PrAY5 y PrAY6 (◆); Homólogos ambientales de PrAY5 (▲); Homólogos ambientales de PrAY6 (■). Además para este análisis se tomaron secuencias del subconjunto cd00306 de la familia de dominios proteicos conservados que pertenece a las peptidasas\_S8 y proteasas obtenidas mediante metagenómica funcional (●). Para este análisis se aplicaron 1000 ciclos de ligamiento. La distancia evolutiva fue computada mediante el método de corrección de Poisson. El análisis se llevó a cabo en MEGA6 (Tamura *et al.*, 2013).

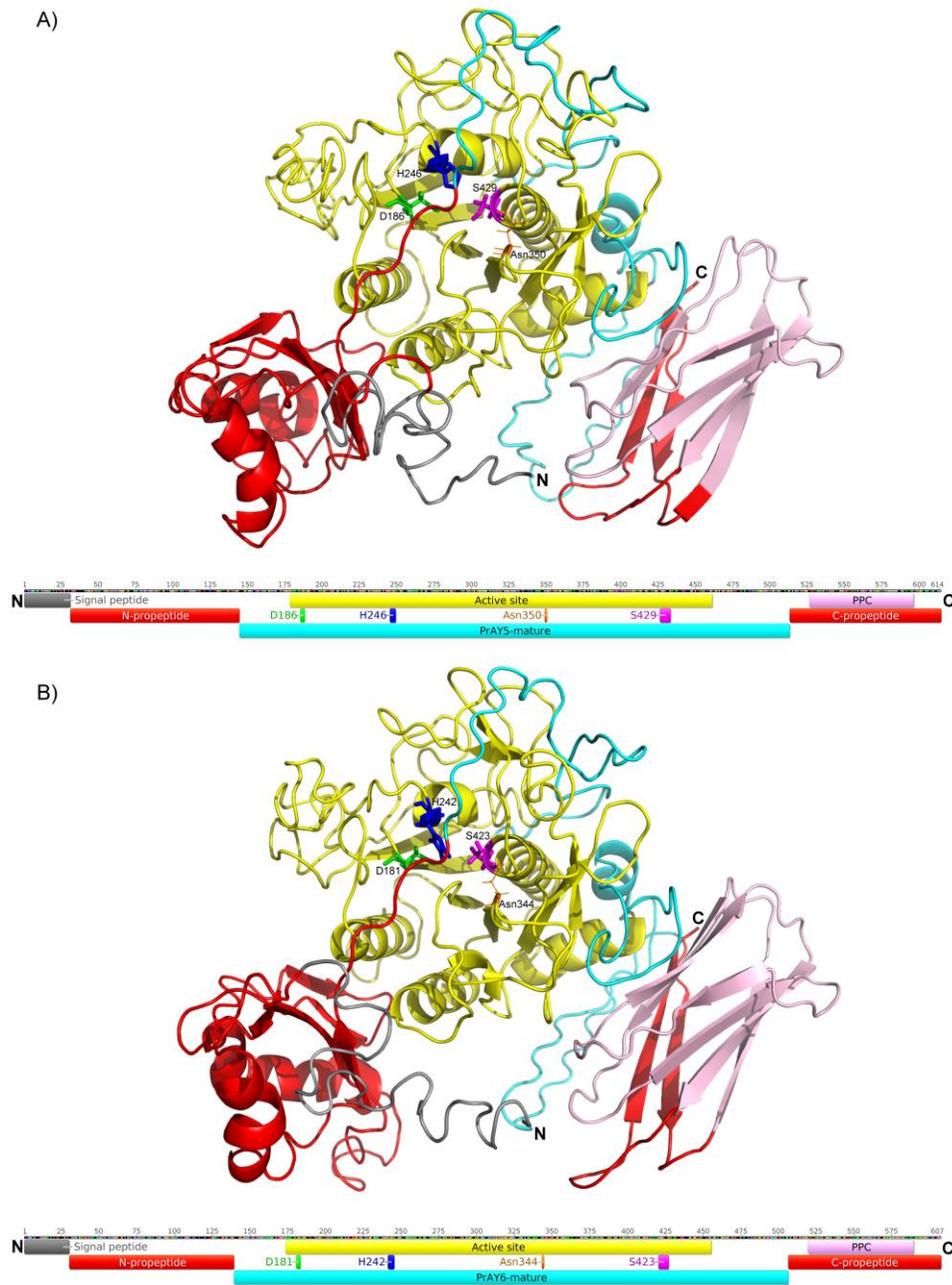
Otra observación interesante es que PrAY5 y PrAY6 agrupan en clados diferentes a sus homólogos encontrados en otros ambientes, lo cual posiblemente sea un indicio de que las bacterias que habitan en el Acuífero de Yucatán puedan tener biomoléculas específicas, diferentes a las que se encuentran en otros ambientes.

Estos resultados refuerzan la hipótesis que las proteasas identificadas en este trabajo representan nuevos miembros de la familia S8 y además sugieren que el Acuífero es una buena fuente para la búsqueda de nuevas proteasas, aunque es necesario seguir trabajando en este último punto para poder ser concluyentes.

### **3.7.5 Modelado 3D de PrAY5 y PraY6**

Para conocer la posible estructura tridimensional de los péptidos codificados por los ORFs *PraY5* y *PraY6*, se realizó su modelamiento por homología. Al tomar todo el ORF (o sea la proteína inmadura, también llamada precursor o zimógeno), el modelado dio como resultado que para ambas secuencias el mejor templado, encontrado por I-TASSER (Yang *et al.*, 2015), fue una proteasa de *Thermococcus kodakaraensis*, la cual tiene el número PDB 3AFG (Foophow *et al.*, 2010), una especie que interesantemente pertenece a una arquea y no a una eubacteria.

PrAY5, con respecto a 3AFG, tiene 78% de cobertura, 25% de identidad y un valor Z de 2.40. Este último valor indica el nivel de confianza del alineamiento entre la secuencia problema y el templado. Alineamientos con valores mayores a 1 indican un buen alineamiento. El valor C fue de -0.48, lo cual indica la confianza en el modelado 3D (los valores se encuentran en un intervalo de -2 a 5, y entre mayores sean, más confiable es el modelo). PrAY6 tuvo 79% de cobertura, 28% de identidad, un valor Z de 2.42 y un valor C de -0.66. con base en estos resultados, se puede concluir que en ambos casos el porcentaje de cobertura fue bueno y los otros valores indican que los alineamientos con sus templados fueron admisibles, y por lo tanto, el modelo es aceptable. En la Figura 3.20 se muestran los modelos con los mejores valores C en cada caso.

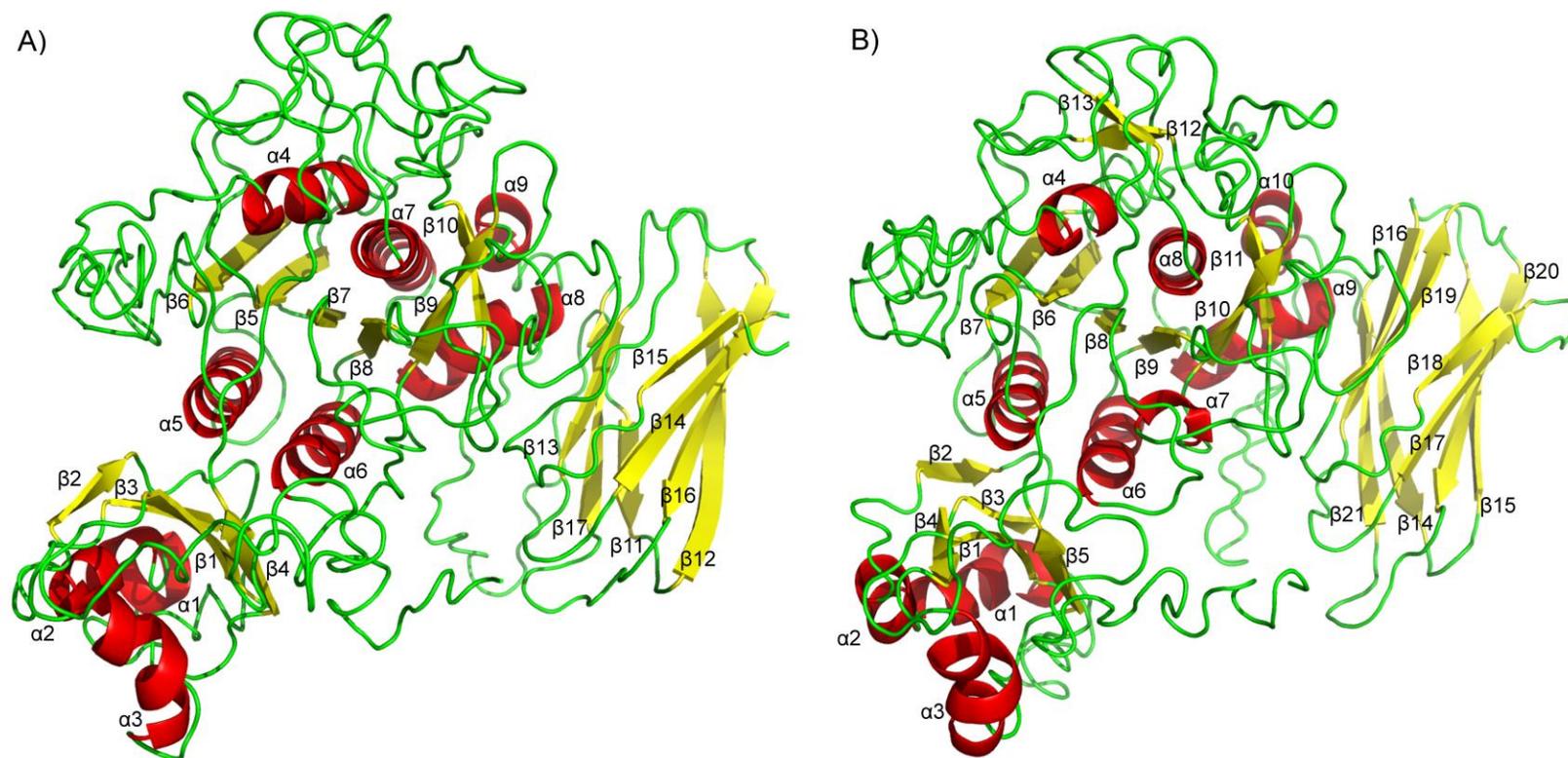


**Figura 3.20 Modelo 3D de los zimógenos de PrAY5 y PrAY6.** A) Símbolos: PrAY5 madura (aminoácidos [aa] 145 a 513), cian; péptido señal (aa 1 a 31), gris; N-propéptido (aa 32 a 144) y C-propéptido (aa 514 a 614), rojo; dominio PPC (aa 527 a 569), rosa; sitio activo (aa 179 a 461), amarillo. Los aa que forman la triada catalítica (D186, H246, S429) y el oxianión (N350), verde, azul, magenta, naranja respectivamente. B) PrAY6 tiene el mismo código de color que PrAY5. PrAY6 madura (aa 140 a 506); Péptido señal (aa 1 a 130); N-propéptido (aa 31 a 139) y C-propéptido (aa 507 a 607); dominio PPC (aa 520 a 589); sitio activo (aa 174 a 455); triada catalítica (aa D181, H242, S423); oxianión (N344).

Como referencia, la proteasa S8 de *C. koreensis* (WP\_028445624.1), también fue modelada en I-TASSER. En este caso el resultado fue muy similar a lo encontrado para nuestras secuencias, el mejor templado para *C. koreensis* también fue 3AFG con un 79% de cobertura, 28% de identidad, un valor Z de 2.43 y un valor C de -0.46. Este resultado sustenta la conclusión de que el PDB 3AFG es el mejor templado posible que existe hasta ahora en la base de datos PDB, aun siendo una secuencia de arquea y no de eubacteria. Esto sugiere que algunas estructuras tridimensionales de proteasas entre eubacterias y arqueas son muy similares, aunque en su secuencia tengan relativamente baja identidad, lo cual ya ha sido discutido por otros autores (Siezen y Leunissen, 1997).

El modelamiento mostró que el zimógeno de PrAY5 y PrAY6 tiene una estructura compleja, con una pre-región (péptido señal) y dos pro-regiones, una en el lado amino terminal y otra en el carboxilo terminal (dominio PPC), que flanquean el dominio catalítico (Figura 3.20). El propéptido en el extremo amino terminal de PrAY5 tiene un tamaño de 81 aminoácidos y se encuentra del E32 a la L113; mientras; para PrAY6 el tamaño del propéptido es de 78 aminoácidos y va de D31 a la L109. Es común que las peptidasas S8 se produzcan como zimógenos (también llamados pre-pro-enzimas), los cuales son guiados por el péptido señal al exterior celular, donde son activados por el corte del péptido señal y de la pro-región (propéptido). El propéptido cumple dos funciones, por un lado, asiste al correcto plegamiento de la enzima y por otro funciona como un fuerte inhibidor de su actividad (Eder *et al.*, 1993, Subbian *et al.*, 2005). El dominio PPC (pre-peptidasa dominio terminal C) es encontrado usualmente en el estado inmaduro de las proteasas secretadas (Yeats *et al.*, 2003). En PrAY5 tiene un tamaño de 42 aminoácidos y en PrAY6 de 69 aminoácidos.

El modelo mostró además que PrAY5 tiene 9 hélices- $\alpha$  y 17 hojas- $\beta$ , de las cuales 6 hélices- $\alpha$  y 6 hojas- $\beta$  se encuentran en el dominio catalítico (Figura 3.21A). PrAY6 tiene en total 10 hélices- $\alpha$  y 21 hojas- $\beta$  y su dominio catalítico tiene 7 hélices- $\alpha$  y 8 hojas- $\beta$  (Figura 3.21B). De acuerdo con la clasificación de SCOPe (Fox *et al.*, 2014), las proteasas S8 tienen una estructura característica de tres capas  $\alpha / \beta / \alpha$  (Laskar *et al.*, 2011) la cual es observada claramente en el modelo.

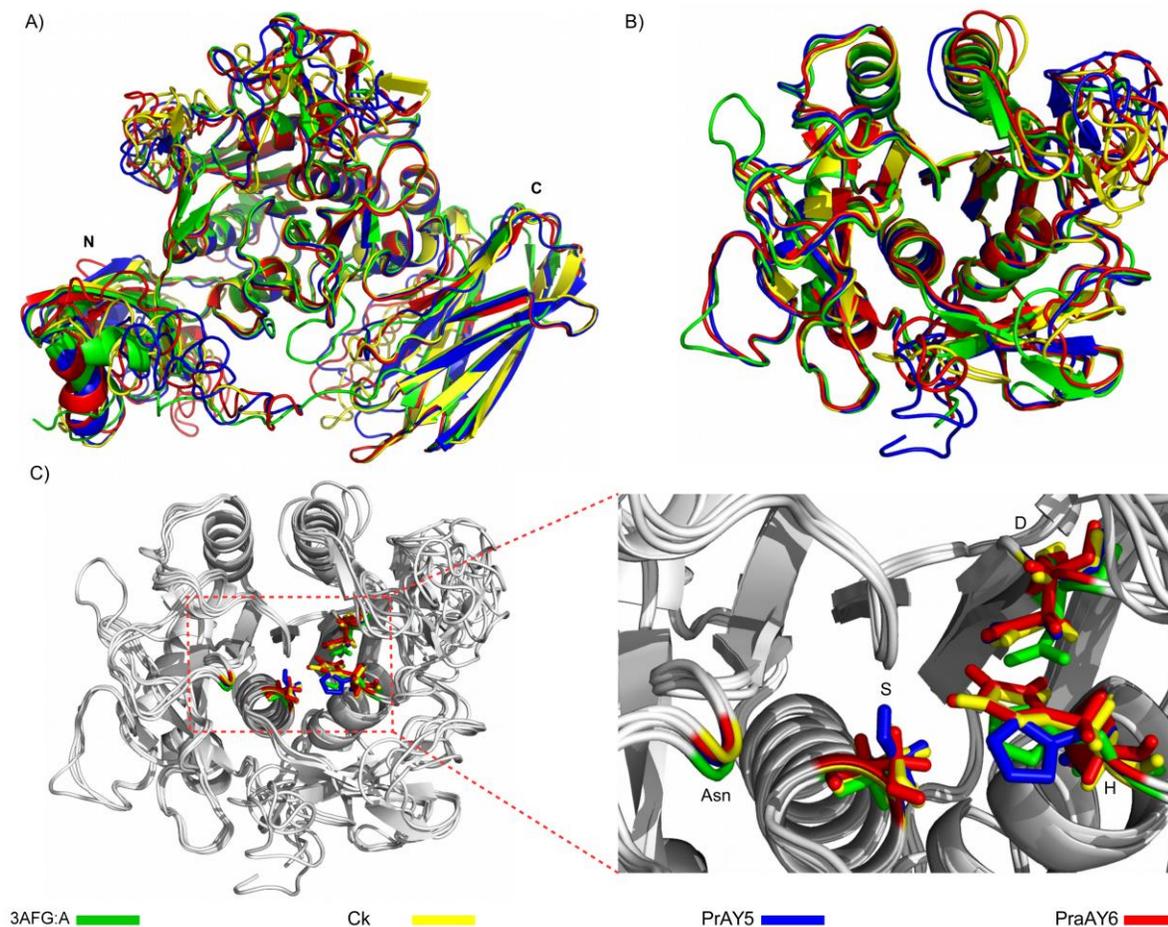


**Figura 3.21 Plegamiento teórico, enumerando las hélices- $\alpha$  and hojas- $\beta$  en PrAY5 y PrAY6.** Símbolos: hélices- $\alpha$ , se muestran en rojo; hojas- $\beta$ , se muestran en amarillo. A) PrAY5 las hojas- $\beta$  del sitio activo comienzan con  $\beta 5$ - $\beta 10$ ; B) En PrAY6 comienzan en  $\beta 6$ - $\beta 13$ . En ambas figuras se muestra la enumeración de las hélices- $\alpha$  y las hojas- $\beta$  amino y carboxilo.



Una idea más clara de la ubicación y posición de los aminoácidos que participan en la catálisis se puede observar al realizar el solapamiento de los modelos 3D de las tres secuencias con respecto al moldeado (Figura 3.23). Lo primero que se observa es que el extremo amino terminal es muy variable, mientras que el extremo C-terminal está muy conservado, y la estructura del sitio activo es prácticamente idéntica; (tiene una diferencia de solo 0.4 Å en 64 residuos). A pesar de que los aminoácidos de la triada catalítica en la secuencia se encuentran separados en la estructura primaria, al llevarse a cabo el plegamiento, estos se acercan entre sí, formando parte del sitio activo, lo cual sugiere que, a nivel estructural (tridimensional), el sitio activo está sujeto a una gran presión de selección que promueve su conservación. Y en cambio, en las regiones que flanquean el sitio activo hay cierta flexibilidad.

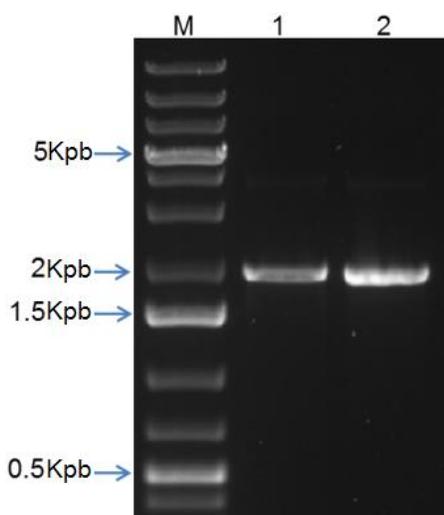
Cuando se realiza un acercamiento sobre la triada catalítica (Figura 3.23C), se puede observar que tanto esta como el oxianión se hallan altamente conservados. Una diferencia interesante entre los modelos 3D de PrAY5 y PrAY6 es que, mientras las proteasas S8 típicamente tienen una estructura de “sándwich” de 7 hojas-β entre dos capas de hélices-α, PrAY5 mostró 6 hojas-β y PrAY6 8 hojas-β ambas entre dos capas de hélices-α. Sin embargo, hay que tener en cuenta que esto podría ser un artefacto del modelamiento. Un análisis cristalográfico sería necesario para estar totalmente convencidos acerca del número de hojas-β en el sitio activo.



**Figura 3.23** Sobrelapamiento de los modelos 3D de PrAY5, PrAY6 y S8 de *C. korensis* tomando como molde el templado 3AFG. Se muestran los ORFs completos incluyendo la pre-región (péptido señal), la pro-región (pro-péptido), el dominio catalítico y el dominio PPC. Como se puede observar la estructura es diferente en las cuatro secuencias en el extremo amino terminal, mientras que está más conservada en el dominio catalítico y el dominio PPC del extremo carboxilo terminal. B) dominio catalítico en el cual se muestra C) un acercamiento a la triada catalítica DHS y el oxianión N.

### 3.8 Subclonación de los ORFs *PrAY5* y *PrAY6* en el vector de expresión pLATE52

Después de la caracterización *in silico* los resultados sugirieron fuertemente que estructuralmente las secuencias de *PrAY5* y *PrAY6* eran proteasas. Para probar que el fenotipo proteolítico mostrado en la clona MAJ10 se debía a la expresión de estos genes, se llevó a cabo la subclonación individual de cada una de las secuencias en el vector pLATE52. Para lo cual, se utilizó como templado ADN del fósido MAJ10, ambos ORFs fueron amplificados mediante PCR utilizando una  $T_m$  de 58 °C (Figura 3.24)

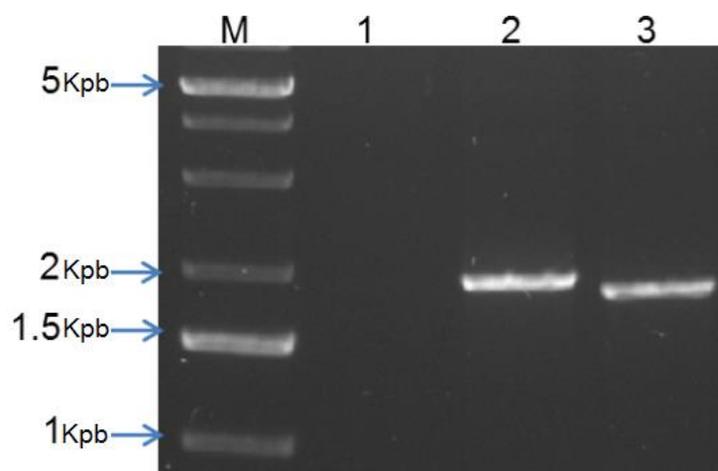


**Figura 3.24 Productos de PCR de los ORFs *PrAY5* y *PrAY6*.** Gel de agarosa al 1%. Carril 1: *PrAY5*; Carril 2: *PrAY6*; M: marcador de peso molecular de 1 Kpb.

Como se observa en la figura 3.24, el producto de amplificación fue del tamaño esperado, 1845 pb para *PrAY5* (carril 1) y 1824 pb para *PrAY6* (carril 2). Estos productos de PCR fueron purificados del gel y posteriormente clonados en el vector de expresión pLATE52.

Los productos de reconocimiento (aligning) entre el vector y el inserto fueron transformados en la cepa *E. coli* Top10. Las clonas resultantes fueron seleccionadas en medio LB adicionado con Amp. Aquellas clonas resistentes se cultivaron y se extrajo el plásmido el cual se utilizó como templado para verificar por PCR la clonación del inserto

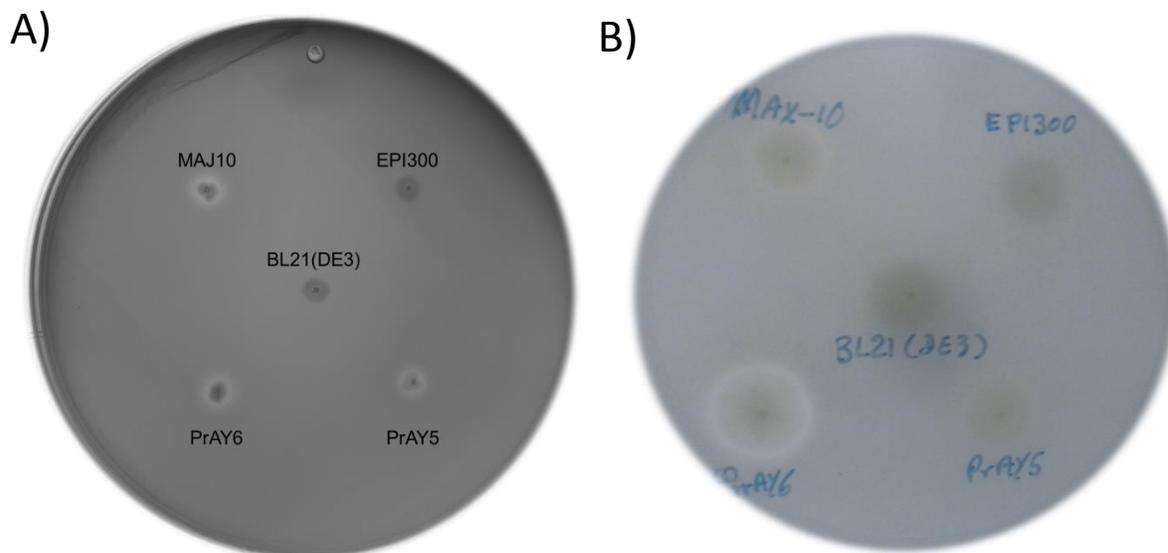
en pLATE52. Para este ensayo se utilizaron los oligonucleótidos específicos del vector de expresión (LIC-Fw y LIC-Rv, Tabla 2.3). En la Figura 3.25 se muestra que la amplificación fue exitosa.



**Figura 3.25 Amplificación por PCR de PrAY5 y PrAY6 en pLATE52.** Gel de agarosa al 0.8%. Carril 1, control negativo; Carril 2, PrAY5; Carril 3, PrAY6; M, marcador de peso molecular de 1 Kpb

En la Figura 28 se pudo corroborar, que se logró la subclonación de ambos ORFs. Se seleccionó una clona de cada construcción y se secuenció, el resultado mostró la integridad de las secuencias, las cuales dieron un porcentaje de identidad del 100% con respecto a la secuencia del fósido MAJ10.

Después de verificar las construcciones pLATE52-PrAY5 y pLATE52-PrAY6 estas fueron transformadas en la cepa *E. coli* BL21 (DE3) y analizadas sobre el medio de escrutinio agar leche. En ambos casos las clonas mostraron un fenotipo proteolítico, confirmando que ambos ORFs codifican proteasas funcionales (Figura 3.26).



**Figura 3.26 Prueba funcional de las construcciones pLATE52-PrAY5 y pLATE52-PrAY6.** Medio LB más leche sin antibiótico. A) 24 h; B) 48 h.

Como se puede apreciar en la figura 3.26, el halo de hidrólisis alrededor de la subclona con la construcción pLATE52-PrAY5 fue más pequeño y menos hialino con respecto al mostrado por la subclona con la construcción pLATE52-PrAY6. Esto podría deberse a que la proteína codificada por PrAY5 sea menos estable o menos activa bajo las condiciones de nuestro ensayo, o bien a que se secreta en menor cantidad. Sin embargo, este resultado muestra que cada uno de los ORFs codifica una proteína activa y capaz de secretarse cuando se encuentran separados uno de otro. Lo cual sugiere que no depende una proteasa de la otra para tener actividad, ni para ser secretada. Además, se pone de manifiesto que *E. coli* tiene la maquinaria necesaria para el plegamiento y secreción de estas enzimas. Sin embargo, es necesario realizar más experimentos para poder sobreexpresar y purificar estas enzimas y así tener una idea más clara de su función y fundamentar su posible aprovechamiento biotecnológico.

Los resultados obtenidos en este trabajo representan el primer análisis por metagenómica funcional de biomasa del agua subterránea del Acuífero de Yucatán, además, cabe resaltar que de las 23 clonas obtenidas en solo una de ellas se obtuvieron secuencias de proteasas con homología con motivos de proteasas ya conocidas, y esta homología resulta relativamente baja con respecto a las secuencias depositadas en las bases de

datos. Las restantes 22 clonas podrían tener otros genes de proteasas novedosos que posteriormente pueden ser analizados.



## CAPITULO IV

### Discusión General

Hasta ahora, los ambientes acuáticos, especialmente los de agua dulce, han sido pobremente estudiados, desde un enfoque metagenómico. En este trabajo reportamos la construcción y análisis funcional de una biblioteca metagenómica partiendo de una muestra de un punto del acuífero de Yucatán México.

Con base en el patrón de restricción generado con la enzima *Bam*HI y confirmado por la secuenciación, el tamaño de inserto de las clonas de la biblioteca es ~35,000 pb. Las clonas analizadas en este trabajo representan 735 Mb de información genética, lo cual es equivalente a 150 veces el genoma de *E. coli* (4.9 Mpb) y 175 veces el de *Bacillus lichiniformis* (4.2 Mpb). En esta cantidad de ADN podrían estar codificados hasta 735,000 genes si tomamos en cuenta que un gen bacteriano tiene un tamaño aproximado de 1Kpb.

Comparada con otras bibliotecas metagenómicas, construidas utilizando el fósmidos pCCFOS y utilizando como hospedero *E. coli* EPI300<sup>MR</sup>, la realizada en este trabajo es 8.3 veces más grande que la reportada por Lee *et al.*, 2007; 14.7 que la de Neveu *et al.*, 2011 y 8.9 que la de Morris y Marchesi 2015. Hasta ahora no existe un tamaño determinado para una biblioteca aunque se estima que para lograr acceder al metagenoma completo de una comunidad es necesario tener clonados 30,000 Mpb (Sjöling *et al.*, 2007), de manera que entre más grande sea una metagenoteca mejor representada se tendrá la comunidad de un ambiente.

En trabajos previos se ha reportado la dificultad técnica que implica el hacer un análisis funcional como método de escrutinio de clonas sobresecretoras de proteasas, ya que en algunas ocasiones se han detectado falsos positivos (Jones *et al.*, 2007). Sobre todo, se considera que esto ocurría debido a que al principio se utilizaba como sustrato leche con un alto contenido de grasas y lactosa, lo cual daba falsos positivos debido a la presencia de lipasas y glicósido hidrolasas. Por esta razón, en este trabajo se utilizó el método más robusto que hasta ahora se ha reportado, en el cual se utiliza como sustrato leche en

polvo baja en grasa, deslactosada y con un alto porcentaje de proteína de leche (caseína), tal como lo propuso Morris y colaboradores (Morris *et al.*, 2012). Esta prueba con leche es un método fácil, rápido y eficiente, ya que por un lado, las clonas sobreproductoras de proteasas forman halos de hidrólisis alrededor de la colonia debido a la degradación de la caseína, los cuales son muy fáciles de detectar a simple vista, y por otro, se requiere que las colonias tengan una cantidad y/o actividad extracelular de proteasa muy superior a la del control para que resulten positivas (cabe recordar que todas las células producen proteasas, pero este método es poco sensible a proteasas endógenas, por lo que no se detecta el fondo producido por las cepas hospederas). De hecho, con base en esto, se puede suponer que tanto PrAY5, como PrAY6, tienen un buen potencial biotecnológico, ya que sus halos son muy fácilmente detectables.

Las cepas hospederas EPI300<sup>MR</sup>, BL21(DE3) y TOP10, no presentaron fenotipo proteolítico al cultivarse en el medio de escrutinio, lo que sugiere que el método utilizado fue el adecuado para realizar las pruebas funcionales, ya que no se tiene el riesgo de un fenotipo positivo derivado del fondo genético de la bacteria hospedera. En contraste, nuestro control positivo; *B. subtilis*, que se sabe que es una cepa productora de proteasas que se secretan (Rao *et al.* 1998) sí mostró halos hidrolíticos. Este resultado indica que los halos obtenidos en las 23 clonas seleccionadas se deben a alguna secuencia que se encuentra en el inserto de los fósmidos, y aunque en el análisis bioinformático inicial no se encontraron dominios de proteasas en MAJ1 y MAJ19, se podrían buscar mediante otras metodologías como la realizada por Culligan y col., (2013) quienes implementaron un sistema de mutagénesis con transposones para identificar un nuevo gen de tolerancia a salinidad.

En cuanto a proporción de clonas obtenidas por número de clonas analizadas, lo obtenido en este trabajo es similar a lo reportado por Neveu y col. (2011) quienes después de realizar el escrutinio de 17,000 clonas de fósmidos, obtuvieron 16 clonas que formaban halos de hidrólisis en el medio de escrutinio. Es decir, en ambos casos la eficiencia fue similar, encontrándose aproximadamente una clona positiva por cada mil sometidas a escrutinio. Debido a que hasta ahora son muy pocas las bibliotecas metagenómicas de fósmidos en las que se han hecho escrutinios para la búsqueda de clonas sobresecretoras de proteasas, no podemos ser concluyentes acerca de si esta eficiencia

es la misma en cualquier biblioteca, sin importar el ambiente o el tipo de muestra de la que provenga, sin embargo, los datos de que se disponen hoy en día sugieren que así es. A pesar de ello, hacen falta más pruebas, para apuntalar esta hipótesis.

A pesar de la relativamente alta tasa de éxito en la búsqueda de otras enzimas, muy pocas proteasas han sido encontradas con herramientas de metagenómica funcional (Pushpam *et al.*, 2011). Posiblemente esto se deba a la citotoxicidad de las proteínas que se pretenden sobre expresar, a la falta de un hospedero correcto para su expresión y procesamiento, y/o a problemas de modificaciones postraduccionales o de secreción (Piel., 2011). Sin embargo, en este trabajo se obtuvieron 23 clonas positivas, con fragmentos exógenos que están siendo expresados en *E. coli* EPI300<sup>MR</sup> en las que se observan claramente halos de hidrólisis. Esto sugiere que se trata de proteasas con actividad extracelular y que *E. coli* tiene la maquinaria necesaria para transcribir los genes recombinantes, traducir los mensajeros, produciendo proteasas funcionales y secretar los péptidos producidos correctamente. Además del sesgo que quizá impone la expresión en *E. coli*, también hay que tener en mente que con el método que utilizamos seguramente no detectamos las proteasas sobresecretadas que tengan otros requerimientos distintos a los suministrados en el medio agar leche, por ejemplo, distintos valores de pH o temperatura, presencia de iones metálicos o concentraciones distintas de otras sales. Todo esto implica que si bien esta metodología permite detectar genes de proteasas de una amplia gama de microorganismos hasta ahora desconocidos, tal como se ha demostrado en diversos trabajos (Lee *et al.*, 2007; Waschkowitz *et al.*, 2009; Berlemont *et al.*, 2011; Neveu *et al.*, 2011; Niehaus *et al.*, 2011; Pushpam *et al.*, 2011; Zhang *et al.*, 2011; Biver *et al.*, 2013; Guazzaroni *et al.*, 2013; Purohit y Singh, 2013; Morris y Marchesi, 2015), se debe estar conscientes de que, al igual que el método de extracción de ADN metagenómico, puede ser que imponga cierto sesgo.

El análisis de la secuencia del fósido MAJ10 sugiere que el inserto posiblemente pertenezca a un microorganismo del filo Proteobacteria, ya que el 50% de las secuencias predichas en él corresponden a este filo. Los microorganismos pertenecientes al filo Proteobacterias son los más abundantes en ambientes acuáticos. En el fondo marino representan el 79 % de la biomasa bacteriana, en la superficie del mar un 64% y en agua dulce el 40% (Battistuzzi y Hedges 2008). Sin embargo, el bajo porcentaje de identidad

(en aminoácidos) de todas ellas con su respectivo homólogo más cercano, aunado a que el 50 % de los genes predichos tienen como homólogos más cercanos proteínas hipotéticas, o bien que no se encontraron secuencias similares en las bases de datos, indica que el inserto de este fósido proviene del genoma de un microorganismo desconocido, en el sentido de que su genoma hasta ahora no ha sido reportado; probablemente sea de una bacteria que hasta ahora no se ha cultivado, por lo que su identidad taxonómica es solo una suposición sustentada en los resultados de similitud de las 50 proteínas predichas en el inserto de MAJ10.

Otra forma de esbozar el posible origen de una secuencia de ADN es mediante su contenido de G+C el cual en el inserto del fósido MAJ10 es de 56.2%. Se ha reportado que en bacterias Gram positivas del filo Firmicutes es por ejemplo, de 33 % en *Staphylococcus aureus*, 38% en *Clostridium perfringens* y 25 % en *Mycoplasma capricolum*; mientras que en bacterias Gram-negativas del filo Proteobacteria es de 50 % en *E. coli*, 58 % en *Serratia marcescens* y 60 % en *Pseudomonas fluorescens* (Muto A. y Osawa S, 1987), lo cual podría sugerir que el inserto de MAJ10 pertenece a una bacteria Gram-negativa, lo cual se refuerza con que PrAY5 y PrAY6 muestran mayor identidad por una proteasa codificadas por *Chitinomonas koerencis*, la cual es una bacteria Gram negativa que tiene 65% de contenido G+C (Kim *et al*, 2006). Sin embargo, el contenido de G+C es algo que se debe tomar con cautela; solo como un indicio y no como algo concluyente.

Los análisis *in silico*, apuntan fuertemente a que los genes *PrAY5* y *PrAY6* codifican isoformas de una proteína, ya que son muy similares presentando un 66.5% de identidad entre sí y a que ambos presentan las características más distintivas de la familia peptidasa S8 subfamilia S8A. En la cual existen más de 200 miembros, que se dividen en seis subfamilias (Seizen y Leunissen 1997). Los miembros de la subfamilia S8A se denominan subtilasas y se caracterizan por tener residuos de los aminoácidos D, H y S dentro del sitio activo; formando lo que se denomina triada catalítica. Conjuntamente estos aminoácidos y el oxianión se acercan entre sí en la estructura tridimensional e intervienen coordinadamente en la catálisis que lleva a la ruptura de enlaces peptídicos (Gupta *et al.*, 2002).

La familia S8, pertenece al clan (SB) de las proteasas tipo subtilisinas de acuerdo tanto con MEROPS – *The Peptidase Database* como con el NCBI, dentro del cual se encuentra también la familia de las sedelosinas (S53). Sin embargo, las dos familias se diferencian entre sí ya que el orden de la triada catalítica en la S8 está conformada por los residuos de aminoácidos D, H, S mientras que la S53 tiene una tétrada conformada por los residuos de aminoácidos E, E, D, S (Rawling y Barrett., 2004).

La presencia de un péptido señal tanto en PrAY5 como en PrAY6, la carencia de estructuras transmembranales, y el fenotipo proteolítico que generan, hace suponer que estas dos enzimas son secretadas extracelularmente, posiblemente este sea su destino en el microorganismo al que pertenecen. Hay otras serina proteasas extracelulares que pertenecen a un grupo muy importante de enzimas industriales, como por ejemplo la subtilisina Carlsberg y subtilisina BPN' (Gupta *et al.*, 2002; Kumar y Takagi, 1999). Las serina proteasas de la familia S8 también han sido candidatos muy populares para la ingeniería de proteínas y evolución dirigida, que han dado resultados muy exitosos comercialmente (Bryan, 2000). Todo lo cual alienta la posibilidad de que PrAY5 y PrAY6 tengan un buen potencial biotecnológico.

Otra característica importante es que tanto PrAY5 como PrAY6, cuentan con un dominio PPC en el extremo C-terminal. Se ha sugerido que este dominio participa en la adhesión al sustrato (Miyoshi *et al.*, 2001) o en la secreción de la proteína en bacterias Gram negativas (Gray *et al.*, 1986). Sin embargo, se ha observado que la extensión del extremo C-terminal no está presente en la proteasa ya madura (Yeats *et al.*, 2003), y esto indica que no se requiere para su actividad.

Una forma de clasificación de proteasas es mediante su estructura tridimensional, ya que muchas veces a nivel secuencia puede haber diferencias y al plegarse adquieren una estructura similar (Lesk y Chothia, 1980). Por esta razón es importante modelar la posible estructura de una secuencia nueva. Tomando las secuencias de los ORFs de PrAY5 y PrAY6, desde el ATG hasta el TAA, se modeló por comparación la posible estructura de estas proteínas. El análisis arrojó que ambas secuencias tienen como mejor templado dentro de las existentes, a una proteasa 3AFG (Foophow *et al.*, 2010), en la cual se conserva muy estable el orden y orientación de la triada catalítica (D, H, S) característica

de las proteasas S8. Esta estructura conservada ha sido reportada previamente por otros autores (Rawlings y Barrett, 2004; Laskar *et al.*, 2011, Zhang *et al.*, 2015).

En el modelo 3D fue posible observar que, al igual que en algunas otras proteasas, el extremo N-terminal o pro-péptido se halla bloqueando el sitio activo. La mayoría de serina proteasas se caracteriza por sintetizarse como preproteínas (también llamados zimógenos o precursores), en las cuales el extremo N juega un papel muy importante como chaperona, inhibiendo la actividad de la enzima hasta su maduración (Tang *et al.*, 2003; Shinde y Inonye., 2000). Este es un mecanismo de protección, ya que si fueran funcionales desde que se sintetizan, podrían dañar otras proteínas dentro de la célula.

Además de conservar la estructura en el sitio activo, PrAY5 y PraY6 mostraron el plegamiento  $\alpha / \beta / \alpha$  característico de la subtilasas (Rawlings y Barrett, 2004; Kennan *et al.*, 2010; Laskar *et al.*, 2011). Sin embargo, a pesar de conservar esta estructura, la variación en cuanto al número de hélices- $\alpha$  y hojas- $\beta$  en el sitio activo, podría explicar la diferencia de actividad de PrAY5 con respecto a PrAY6. Es necesaria la purificación de la enzima para realizar su caracterización bioquímica hacer un análisis de cristalografía para confirmar esta hipótesis, lo cual además permitiría esbozar un posible uso biotecnológico.

## CAPITULO V

### Conclusiones y perspectivas

- 1.- El volumen de agua necesario para saturar un filtro Sterivex<sup>MR</sup>, el cual sirve para acumular la biomasa procariota, depende del día en que se tome la muestra, siendo necesarios de 5 a 10 L en épocas lluviosas y de 40 a 60 L en épocas de secas. Lo cual sugiere que durante la época de lluvias el río subterráneo se agita más y arrastra mayor contenido de partículas –de 0.22 a 5 µm- en suspensión.
- 2.- El mejor método de extracción para la obtención del metagenoma fue el método comercial de Epicentre<sup>MR</sup>, lo cual sugiere que es mejor utilizar un método planeado integralmente, desde la extracción hasta la construcción de la genoteca.
- 3.- El aumentar la relación de inserto: vector en la ligación, hasta 5:1, mejoró el título de la biblioteca.
- 4.- Se logró la construcción de una metagenoteca con una mayor representación (2.1 veces) respecto hasta las ahora reportadas.
- 6.- El escrutinio de proteasas basado en agar leche permitió aislar 23 clonas con el fenotipo buscado. Sin embargo, harían falta algunas otras pruebas para verificar que en efecto en los insertos de todas ellas se codifiquen proteasas, ya que se ha reportado que en ocasiones los halos de hidrólisis se originan debido a cambios en el pH.
7. La escasa o nula similitud entre las secuencias de los insertos de las clonas MAJ1 y MAJ19 con respecto a las secuencias de proteasas depositadas en las bases de datos, hace prometedor el descubrimiento de nuevas proteasas.
8. Los dos ORFs identificados con dominios de peptidasas S8 en la clona MAJ10 presentan 66.5 % de identidad entre sí a nivel de aminoácidos y muestran una estructura tridimensional típica de proteasas S8

9. El análisis *in silico* mostró que PrAY5 y PrAY6 tienen los aminoácidos que conforman la triada catalítica y el oxianión característico de las subtilasas, los cuales son de suma importancia en su mecanismo de acción.

10.- Al mantenerse el fenotipo esperado cuando las secuencias de *PrAY5* y *PrAY6* fueron subclonadas, se soporta fuertemente la idea de que son las responsables del fenotipo proteolítico en MAJ10.

11. Las secuencias de *PrAY5* y *PrAY6* codifican proteasas funcionales extracelulares y no necesitan estar juntas ni para transcribirse, ni para procesarse, ni para secretarse, ni para tener actividad.

### Perspectivas

En este trabajo se describe la identificación, por medio de metagenómica funcional, de dos ORFs que codifican proteasas S8 extracelulares, así como la caracterización *in silico* de sus secuencias. Esto permite tener una guía para desarrollar estrategias posteriores que permitan el aislamiento y la caracterización bioquímica de estas dos proteínas y las que se pudieran obtener de los otros 20 fósmidos, que resultaron positivos en el ensayo funcional. Una tarea que puede ser complicada si no se conoce la estructura de los ORFs y las enzimas que codifican. El resultado de la filogenia indica que *PrAY5* y *PrAY6* agruparon en clados diferentes a otras proteasas obtenidas mediante metagenómica funcional. Lo cual sugiere que el agua del Acuífero de Yucatán podría contener una comunidad microbiana diferente, y por lo tanto interesante, de otros ambientes estudiados hasta ahora.

El siguiente paso de esta investigación es purificar y caracterizar bioquímicamente ambas enzimas, para determinar sus características fisicoquímicas y sus parámetros cinéticos, y así poder estimar su posible utilización en algún proceso industrial. Para la purificación de las enzimas se sugiere emplear el método de sulfato de amonio y combinar con columnas de intercambio iónico o exclusión molecular, ya que con base en los análisis *in silico* realizados en este trabajo, se observa que ambas enzimas sufren la escisión de los

extremos amino terminal y carboxilo terminal, por lo cual no se podrían purificar por adición de etiquetas de histidina en alguno de los extremos del ORF.

Para los insertos de los fósidos MAJ1 y MAJ19, los cuales no presentaron similitud con dominios de proteasa, es necesario el abordaje desde otras perspectivas las cuales podrían ser: i) la construcción de sub-bibliotecas de plásmidos y/o ii) realizar mutagénesis etiquetada sobre estas clonas para identificar los posibles nuevos genes responsables del fenotipo.

En los 20 fósidos que fueron secuenciados en lotes existen **contigs** completos los cuales podrían ser caracterizados mediante patrones de restricción *in silico*, los cuales posteriormente serían corroborados experimentalmente.

Otra alternativa para la identificación de las secuencias de proteasas que pudieran ser las responsables del fenotipo en los fósidos restantes sería la secuenciación individual de cada fósido para hacer un escrutinio exhaustivo por secuencia.



**BIBLIOGRAFÍA**

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215:403-410
- Amann RI., Ludwig W., Schleifer KH. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59:143–169.
- Anithajothi, R., Nagarani, N., Umagowsalya, G., Duraikannu, K., Ramakritinan, C. (2014). Screening, isolation and characterization of protease producing moderately halophilic microorganism *Halomonas meridiana* associated with coral mucus. *Toxicol. Environ. Chem.* 96, 296–306.
- Barett, A.J. (1994). Proteolytic enzymes: serine and cysteine peptidases. *Method. Enzymol.* 244:1–15.
- Battistuzzi FU., Hedges SB. (2009). A Major Clade of Prokaryotes with Ancient Adaptations to Life on Land. *Mol. Biol. Evol.* 26 (2): 335-343.
- Bauer-Gottwein, P., Gondwe, B. R., Charvet, G., Marín, L. E., Rebolledo-Vieyra, M., & Merediz-Alonso, G. (2011). Review: The Yucatán Peninsula karst aquifer, Mexico. *Hydrogeol J.* 19(3), 507-524.
- Behzad, H., Ibarra, M. A., Mineta, K., & Gojobori, T. (2016). Metagenomic studies of the Red Sea. *Gene.* 576(2), 717-723.
- Berlemont R, Pipers D, Delsaute M, Angiono F, Feller G, Galleni M, Power P. (2011). Exploring the Antarctic soil metagenome as a source of novel cold-adapted enzymes and genetic mobile elements. *Rev. Argent. Microbiol.* 43(2):94-103.
- Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E., House, C. H. (2008). Metagenomic signatures of the Peru margin seafloor biosphere show a genetically distinct environment. *Proc. Natl. Acad. Sci. USA.* 105 (30):10583-10588.

## BIBLIOGRAFÍA

---

- Biver S, Portetelle D, Vandenbol M. (2013). Characterization of a new oxidant-stable serine protease isolated by functional metagenomics. *Springerplus*. 2:410.
- Bonilla-Rosso, G., Peimbert, M., Alcaraz, L. D., Hernández, I., Eguiarte, L. E., Olmedo-Alvarez, G., & Souza, V. (2012). Comparative metagenomics of two microbial mats at Cuatro Ciénegas Basin II: community structure and composition in oligotrophic environments. *Astrobiology*. 12(7), 659-673.
- Brady SF, Simmons L, Kim JH, Schmidt EW. (2009). Metagenomic approaches to natural products from free-living and symbiotic organisms. *Nat. Prod. Rep.* 26:1488–503.
- Breitbart, M., Hoare, A., Nitti, A., Siefert, J., Haynes, M., Dinsdale, E., & Hollander, D. (2009). Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Ciénegas, Mexico. *Env. Microbiol.* 11(1), 16-34.
- Bryan PN. (2000). Protein engineering of subtilisin *Biochim Biophys Acta*. 1543: 203 [PMID:11150607].
- Casas, F. (2011). Protocolo de Nagoya: por un reparto justo y equitativo de los beneficios derivados de la biodiversidad. *Ambienta: La revista del Ministerio de Medio Ambiente*, (94), 3-5.
- Caumette, P., Matheron, R., Raymond, N., Relexans, J. (1994). Microbial mats in the hypersaline ponds of Mediterranean salterns (Salins-de-Giraud, France). *FEMS Microbiol. Ecol.* 13 (4):273-286.
- Collins J, Hohn B. (1978). Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. *Proc. Natl. Acad. Sci. USA* 75:4242–46.
- Córdova-Kreylos, A. L., Cao, Y., Green, P. G., Hwang, H. M., Kuivila, K. M., LaMontagne, M. G., Van De Werfhorst, L. C., Holden, P. A., Scow, K. M. (2006). Diversity, composition, and geographical distribution of microbial communities in California salt marsh sediments. *Appl. Environ. Microb.* 72 (5):3357-3366.
- Craig JW, Chang FY, Kim JH, Obiajulu SC, Brady SF. (2010). Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid

environmental DNA libraries in diverse proteobacteria. *Appl. Environ. Microbiol.* 76:1633–41.

Craig, J.W., F.Y. Chang y S.F. Brady. (2009). Natural products from environmental DNA hosted in *Ralstonia metallidurans*. *ACS Chem. Biol.* 4: 23-28.

Culligan EP, Sleator RD, Marchesi JR, Hill C. (2013). Functional environmental screening of a metagenomic library identifies *stlA*; a unique salt tolerance locus from the human gut microbiome. *PLoS ONE.* 8(12):e82985.

Da Silva E Silva, L. H., Borghi de Almeida, L. F., Cavalcanti Iespa, A. A., Damazio Lespa, C. M. (2008). Análise de esteiras microbianas e cianobactérias da Laguna Amarga, Parque Nacional de Torres del Paine, Chile. *J. Geosci.* 4 (1):32-37.

De la Rosa-García, D., Muñoz-García, A. A., Barahona-Pérez, L. F., & Gamboa-Angulo, M. M. (2007). Antimicrobial properties of moderately halotolerant bacteria from cenotes of the Yucatan peninsula. *Lett. Appl. Microbiol.* 45(3), 289-294.

DeLano WL. (2002). The PyMOL molecular graphics system. DeLano Scientific, San Carlos, CA, USA.

Donohoe, M. (2003). Causes and health consequences of environmental degradation and social injustice. *Soc. Sci. Med.* 56(3), 573-587.

Eck, J., Gabor, E., Liebeton, K., Meurer, G., and Niehaus, F. (2009). From prospecting to product: industrial metagenomics is coming of age. In *Protein Engineering Handbook*. Lutz, S., and Bornscheuer, U.T. (eds). Weinheim.

Eder, J., Rheinnecker, M. & Fersht, A. R. (1993). Folding of subtilisin BPN': role of the pro-sequence. *J. Mol. Biol.* 233: 293-304.

Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evol.* 39:783-791.

Finn RD, Clements J, Eddy SR. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39: W29-W37. doi:10.1093/nar/gkr367. PubMed: 21593126.

- Foophow T, Tanaka S, Angkawidjaja C, Koga Y, Takano K, Kanaya S. (2010). Crystal structure of a subtilisin homologue, Tk-SP, from *Thermococcus kodakaraensis*: requirement of a C-terminal beta-jelly roll domain for hyperstability. *J Mol. Biol.* 23; 400(4): 865-77. Doi: 10.1016/j.jmb.2010.05.064.
- Fox, N. K., Brenner, S. E., & Chandonia, J. M. (2014). SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res.* 42(D1), D304-D309.
- Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press. pp. 571-607.
- Germany: Wiley-VCH, pp. 295–324. Global Industry Analysts. Industrial enzymes (2004) A global multi-client market research project. (GIA, San José, California, USA).
- Ghai, R., Rodríguez-Valera, F., McMahon, K. D., Toyama, D., Rinke, R., de Oliveira, T. C. S., ... & Henrique-Silva, F. (2011). Metagenomics of the water column in the pristine upper course of the Amazon river. *PloS one*, 6(8), e23785.
- Gish, W., & States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nat. Genet.* 3(3), 266-272.
- Godfrey, T., and S. West (1996). *Industrial enzymology*, Macmillan Publishers Inc., New York, N.Y. 2nd ed., p. 3.
- Gondwe, B. R., Lerer, S., Stisen, S., Marín, L., Rebolledo-Vieyra, M., Merediz-Alonso, G., & Bauer-Gottwein, P. (2010). Hydrogeology of the south-eastern Yucatan Peninsula: new insights from water level measurements, geochemistry, geophysics and remote sensing. *J Hydrol.* 389(1), 1-17.
- Gray, L., N. Mackman, J. M. Nicaud, and I. B. Holland (1986). The carboxyterminal region of haemolysin is required for secretion of the toxin from *Escherichia coli*. *Mol. Gen. Genet.* 205:127–133.

- Guazzaroni, M. E., Morgante, V., Mirete, S., & González-Pastor, J. E. (2013). Novel acid resistance genes from the metagenome of the Tinto River, an extremely acidic environment. *Environ. Microbiol.* 15(4), 1088-1102.
- Gupta R, Beg QK, Lorenz P. (2002). Bacterial alkaline protease: molecular approaches and industrial application. *Appl Microbial Biotechnol.* 59:15–32.
- Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5: R245–R249.
- Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68: 669-685.
- Hang-Yeon Weon, Byung-Yong Kim, Seung-Beom Hong, Young-Ah Jeon, Bon-Sung Koo, Soon-Wo Kwon and Erko Stackebrandt. (2009). *Massilia niabensis* sp. nov. and *Massilia niastensis* sp. nov., isolated from air samples. *Int. J. Syst. Evol. Microbiol.* 59, 1656–1660
- Hartley, B. S. (1960). Proteolytic enzymes. *Annu. Rev. Biochem.* 29:45–72.
- Hennersdorf, P., Mrotzek, G., Abdul-Aziz, M. A., & Saluz, H. P. (2016). Metagenomic analysis between free-living and cultured *Epinephelus fuscoguttatus* under different environmental conditions in Indonesian waters. *Marine pollution bulletin.*
- Herbots, I. (2007). Enzymes in household detergents. In *Enzymes in Industry*. Ähle, W. (ed.). Weinheim, Germany.
- Herrera-Silveira, J., (1988). Productividad primaria fitoplanctónica del estero de Celestún, Yucatán, México. Tesis de Maestría. CINVESTAV-IPN, Unidad Mérida. Wiley-VCH, pp. 154–178.
- Hjort K, Presti I, Elväng A, Marinelli F, Sjöling S. (2014). Bacterial chitinase with phytopathogen control capacity from suppressive soil revealed by functional metagenomics. *Appl Microbiol. Biotechnol.* 98: 2819-2828. Doi: 10.1007/s00253-013-5287-x.

- Jones, B.V. and Marchesi, J.R. (2007). Using skimmed milk agar to functionally screen a gut metagenomic library for proteases may lead to false positives. *Lett. Appl. Microbiol.* 45, 418–420.
- Kim BY, Weon HY, Yoo SH, Chen WM, Kwon SW, Go SJ, Stackebrandt E. (2006). *Chitinimonas koreensis* sp. nov., isolated from greenhouse soil in Korea. *Int. J. Syst. Evol. Microbiol.* 56: 1761-1764. Doi: 10.1099/ijs.0.64163-0.
- King RW, Bauer JD, Brady SF. (2009). An environmental DNA-derived type II polyketide biosynthetic pathway encodes the biosynthesis of the pentacyclic polyketide erdacin. *Angew. Chem. Int. Ed.* 48:6257–61.
- Kirk, O., Borchert, T. V. and Fuglsang C. C. (2002). Industrial enzyme applications. *Curr. Opin. Biotechnol.* 13, 345–351.
- Konagurthu AS, Reboul CF, Schmidberger JS, Irving, JA, Lesk AM, Stuckey PJ, Whisstock JC, and Buckle AM. (2010). MUSTANG-MR Structural Sieving Server: Applications in Protein Structural Analysis and Crystallography. *PLoS One.* Apr 6; 5(4): e10048.
- Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L., (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305(3), 567-580.
- Kumar, C. G., and H. Takagi. (1999). Microbial alkaline proteases: from a bioindustrial viewpoint. *Biotechnol. Adv.* 17:561–594.
- Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P. (2008). A Bioinformatician's Guide to Metagenomics. *Microbiol. Mol. Biol. Rev.* 72 (4), 557-578.
- Laskar M, James RE, Chatterjee A, Mandal C., (2011). Modeling and structural analysis of evolutionarily diverse S8 family serine proteases. *Bioinformation.* 7(5), 239-245.
- Lau, M., Aitchison, J., Pointing, S. (2009). Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. *Extremophiles.* 13 (1):139-149.

- Lee, D.G., Jeon, J.H., Jang, M.K., Kim, N.Y., Lee, J.H., Kim, S.J., Kim, G.D. and Lee, S.H. (2007). Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library. *Biotechnol. Lett.* 29, 465–472.
- Lesk AM., Chothia C. (1980), How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.* 136-225-70.
- Long PF, Dunlap WC, Battershill CN, Jaspars M. (2005). Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production. *Chem. Bio. Chem.* 6:1760–65.
- Lorenz, P., and Eck J. (2005). Metagenomics and industrial applications. *Nat. Rev. Microbiol.* 3:510–516.
- MacNeil, I. A., Tiong, C. L., Minor, C., August, P. R., Grossman, T. H., Loiacono, K. A. & Tyler, A. (2001). Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J. Mol. Microbiol. Biotechnol.* 3(2), 301-308.
- Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C. & Gwadz, M. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic. Acids. Res.* 39: D225-D229. doi:10.1093/nar/gkq1189. PubMed: 21109532.
- Markets and Markets, 2016. Protease Enzymes Market worth \$2,767 Million by 2019 [Online] Disponible en: <http://www.marketsandmarkets.com/PressReleases/protein-hydrolysis-enzymes.asp> [Acceso 3 Noviembre 2016].
- Martín-Cuadrado, A. B., López-García, P., Alba, J. C., Moreira, D., Monticelli, L., Strittmatter, A., & Rodríguez-Valera, F. (2007). Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. *PLoS One*, 2(9), e914.
- Meirelles, L. A., McFrederick, Q. S., Rodrigues, A., Mantovani, J. D., de Melo Rodovalho, C., Ferreira, H., Bacci, M. and Mueller, U. G. (2016). Bacterial microbiomes from vertically transmitted fungal inocula of the leaf-cutting ant *Atta texana*. *Environ. Microbiol. Rep.* doi:10.1111/1758-2229.12415

- Meziti, A., Tsementzi, D., Ar. Kormas, K., Karayanni, H. and Konstantinidis, K. T. (2016). Anthropogenic effects on bacterial diversity and function along a river-to-estuary gradient in Northwest Greece revealed by metagenomics. *Environ. Microbiol.* doi:10.1111/1462-2920.13303
- Mirete, S., Mora-Ruiz, M. R., Lamprecht-Grandío, M., de Figueras, C. G., Rosselló-Móra, R., & González-Pastor, J. E. (2015). Salt resistance genes revealed by functional metagenomics from brines and moderate-salinity rhizosphere within a hypersaline environment. *Front.Microbiol.* 6.
- Miyoshi, S., K. Kawata, K. Tomochika, S. Shinoda, and S. Yamamoto (2001). The C-terminal domain promotes the hemorrhagic damage caused by *Vibrio vulnificus* metalloprotease. *Toxicon.* 39:1883–1886.
- Mori, T., Iwamoto, K., Wakaoji, S., Araie, H., Kohara, Y., Okamura, Y., & Takeyama, H. (2016). Characterization of a novel gene involved in cadmium accumulation screened from sponge-associated bacterial metagenome. *Gene.* 576(2), 618-625.
- Morris L.S., Evans J., Marchesi J.R. (2012). A robust plate assay for detection of extracellular microbial protease activity in metagenomic screens and pure cultures. *J Microbiol. Meth.* 91, 144-146.
- Morris LS, Marchesi JR. (2015). Current Functional Metagenomic Approaches Only Expand the Existing Protease Sequence Space, but does not Presently Add Any Novelty to it. *Curr. Microbiol.* 70(1):19-26.
- Moser, M. J., DiFrancesco, R. A., Gowda, K., Klingele, A. J., Sugar, D. R., Stocki, S., & Schoenfeld, T. W. (2012). Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme. *PLoS One.* 7(6), e38371.
- Muto A. and Osawa S. (1987). The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Nati. Acad. Sci. USA* Vol. 84, pp. 166-169.
- Neveu J, Regard C, DuBow MS. (2011). Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts. *Appl. Microbiol. Biotechnol.* 91(3):635-644.

- Niehaus F., Gabor E., Wieland S., Siegert P., Maurer K. H. and Eck J. (2011). Enzymes for the laundry industries: tapping the vast metagenomic pool of alkaline proteases. *Microbial Biotechnol.* 10: 1-8.
- O'Connor-Sánchez A, Rivera-Domínguez AJ, De los Santos-Briones C, López-Aguiar L, Peña-Ramírez YJ, Prieto-Davo A. (2014). Acidobacteria appear to dominate the microbiome of two sympatric Caribbean Sponges and one Zoanthid. *Biol. Res.* 47: 67 (ISSN 0717-6287) Doi: 10.1186/0717-6287-47-67
- Papineau, D., Walker, J. J., Mojzsis, S. J., Pace, N. R. (2005). Composition and structure of microbial communities from stromatolites of Hamelin Pool in Shark Bay, Western Australia. *Appl. Environ. Microbiol.* 71 (8):4822-4832.
- Peimbert, M., Alcaraz, L. D., Bonilla-Rosso, G., Olmedo-Alvarez, G., García-Oliva, F., Segovia, L. & 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(7), 648-658.
- Perry, E., L. Marin, J. McClain and G. Velázquez (1995). Ring of Cenotes (sinkholes), northwest Yucatán, Mexico: its hydrogeologic characteristics and possible association with the Chicxulub impact crater. *Geology.* 23: 17-20.
- Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*, 8(10), 785-786
- Piel J. (2004). Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* 21:519–38.
- Piel, J., (2011). Approaches to capturing and designing biologically active small molecules produced by uncultured microbes. *Annu. Rev. Microbiol.* 65, 431-453.
- Pope, K.O., Ocampo, A.C., Kinsland, G.L., and Smith, R. (1996). Surface expression of the Chicxulub Crater. *Geology.* 24: 527-530.
- Portillo, M., Sririn, V., Kanoksilapatham, W., Gonzalez, J. (2009). Differential microbial communities in hot spring mats from Western Thailand. *Extremophiles.* 13 (2):321-331.

- Privé F, Newbold CJ, Kaderbhai NN, Girdwood SG, Golyshina OV, Golyshin PN, Scollan ND, Huws SA. (2015). Isolation and characterization of novel lipases/esterases from a bovine rumen metagenome. *Appl. Microbiol. Biotechnol.* 99: 5475-5485.
- Purohit, M. K., & Singh, S. P. (2013). A metagenomic alkaline protease from saline habitat: cloning, over-expression and functional attributes. *Int. J. Biol. Macromol.* 53, 138-143.
- Pushpam PL, Rajesh T, Gunasekaran P. (2011). Identification and characterization of alkaline serine protease from goat skin surface metagenome. *AMB Express.* 1:3.
- Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteases. *Microbiol. Mol. Biol. Rev.* 62(3), 597-635.
- Rappé MS, Giovannoni S.J. (2003). The uncultured microbial majority. *Annu. Rev. Microbiol.* 57:369–394.
- Rawlings, N.D. and Barrett, A.J. (1993). Evolutionary families of peptidases. *Biochem. J.* 290, 205- 218.
- Rawlings, N.D., Barrett, A.J. (2004). Introduction: serine peptides and their clans. In: Barrett, A.J., Rawlings, N.D., Woessner, J.F. (Eds.), *Handbook of Proteolytic Enzymes*, second ed. Elsevier, London, United Kingdom, pp. 1425–1427.
- Rebolledo-Vieyra, M. (2005). Breve historia geológica de la península de Yucatán y el origen de los cenotes: Medina, A. (ed) “Agua, rocas y siglos”, Ed. Pixel, México.
- Rojas-Herrera, R., Zamudio-Maya, M., Arena-Ortiz, L., Pless, R. C., & O'Connor-Sánchez, A. (2011). Microbial diversity, metagenomics and the Yucatán aquifer. *Phyton (Buenos Aires)*. 80, 231-240.
- Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., ... & Tiong, C. L. (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. and Env. Microbiol.* 66 (6):2541-2547.

- Saitou N, Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406-425.
- Salzberg SL, Delcher AL, Kasif S, White O. (1998). Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* 26(2): 544-548. Doi: 10.1093/nar/26.2.544.
- Schmitter-Soto, J. J., Comín, F. A., Escobar-Briones, E., Herrera-Silveira, J., Alcocer, J., Suárez-Morales, E., ... & Steinich, B. (2002). Hydrogeochemical and biological characteristics of cenotes in the Yucatan Peninsula (SE Mexico). *Hydrobiologia*, 467(1-3), 215-228.
- Seizen RJ, Leunissen JA. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* 6:501–523.
- Sharma S, Aneja MK, Mayer J, Schloter M, Munch JC. (2004). RNA fingerprinting of microbial community in the rhizosphere soil of grain legumes. *FEMS Microbiol. Lett.* 240:181–186.
- Shinde, U., and M. Inonye (2000). Intramolecular chaperones: polypeptide extensions that modulate protein folding. *Semin. Cell. Dev. Biol.* 11:35–44.
- Siezen RJ and Leunissen JA. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* 6(3): 501–523. Doi: 10.1002/pro.5560060301.
- Simões, M. F., Antunes, A., Ottoni, C. A., Amini, M. S., Alam, I., Alzubaidy, H. & Bajic, V. B. (2015). Soil and rhizosphere associated fungi in gray mangroves (*Avicennia marina*) from the Red Sea—a metagenomic approach. *Genomics, proteomics & bioinformatics.* 13(5), 310-320.
- Simons, C., & Daniel, R. (2009). Achievements and new knowledge unraveled by metagenomic approaches. *Appl. Microbiol. Biotechnol.* 85:265–276.
- Solovyev V. & Salamov A. (2011). Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies* (Ed. R.W. Li), Nova Science Publishers, p. 61-78.

- Sonnhammer EL, von Heijne G, Krogh A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. 6: 175-182. PubMed: 9783223.
- Steele HL, Jaeger KE, Daniel R, Streit WR. (2009). Advances in recovery of novel biocatalysts from metagenomes. J. Mol. Microbiol. Biotechnol. 16(1-2):25-37.
- Subbian, E., Yabuta, Y. & Shinde, U. P. (2005). Folding pathway mediated by an intramolecular chaperone: intrinsically unstructured propeptide modulates stochastic activation of subtilisin. J. Mol. Biol. 347: 367–383.
- Sudek, S., Lopanik, N. B., Waggoner, L. E., Hildebrand, M., Anderson, C., Liu, H. & Haygood, M. G. (2007). Identification of the putative bryostatin polyketide synthase gene cluster from “*Candidatus Endobugula sertula*”, the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. J. Nat. Prod. 70:67–74.
- Tamura K, Stecher G, Peterson D, Alan Filipinski A, Kuma S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. Doi: 10.1093/molbev/mst197.
- Tang, B., S. Nirasawa, M. Kitaoka, C. Marie-Claire, and K. Hayashi (2003). General function of N-terminal propeptide on assisting protein folding and inhibiting catalytic activity based on observations with a chimeric thermolysin-like protease. Biochem. Biophys. Res. Commun. 301:1093–1098.
- Tang, W. L., & Zhao, H. (2009). Industrial biotechnology: tools and applications. Biotechnol. J. 4(12), 1725-1739.
- Terrón-González, L., Martín-Cabello, G., Ferrer, M., & Santero, E. (2016). Functional Metagenomics of a Biostimulated Petroleum-Contaminated Soil Reveals an Extraordinary Diversity of Extradial Dioxygenases. Appl. Environ. Microb. 82(8), 2467-2478.
- Torsvik, V., Goksøyr, J., and Daae, F. L. (1990). High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. 56(3), 782-787.
- Uchiyama T. and Watanabe K. (2008). Substrate-induced gene expression (SIGEX) screening of metagenome libraries. Nat. Protocols. 3 (7): 1202-1212.

- Venter, J.C., K. Remington, J.F. Heidelberg, A.L. Halern, D. Rusch, J.A. Eisen, D. Wu, I Paulsen, K.E. Nelson, W. Nelson, D.E. Fouts, S. Levy, A.H. Knap, M.W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tilson, C. Pfannkoch, Y. Rogers y H.O. Smith. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*. 304: 66-74.
- Vester JK, Glaring MA, Stougaard P. (2015). An exceptionally cold-adapted alpha-amylase from a metagenomic library of a cold and alkaline environment. *Appl. Microbiol. Biotechnol.* 99: 717-727. Doi: 10.1007/s00253-014-5931-0.
- Wang, G.Y., E. Graziani, B. Waters, W. Pan, X. Li, J. McDermott, G. Meurer, G. Saxena, R.J. Andersen y J. Davies (2000). Novel natural products from soil DNA libraries in a streptomycete host. *Org. Lett.* 2: 2401-2404.
- Waschkowitz, T., Rockstroh, S. and Daniel, R. (2009). Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. *Appl. Environ. Microbiol.* 75, 2506–2516.
- Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL *et al.* (1999). Protein identification and analysis tools in the ExpASY server. *Methods. Mol. Biol.* 112: 531-552. PubMed: 10027275.
- Williamson LL, Borlee BR, Schloss PD, Guan CH, AllenHK, Handelsman J. (2005). Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Appl. Environ. Microbiol.* 71:6335–44.
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods.* 12: 7-8.
- Yeats, C., S. Bentley, and A. Bateman (2003). New knowledge from old: in silico discovery of novel protein domains in *Streptomyces coelicolor*. *BMC Microbiol.* 3:3.
- Zhang Y, Zhao J, Zeng R. (2011). Expression and characterization of a novel mesophilic protease from metagenomic library derived from Antarctic coastal sediment. *Extremophiles.* 15(1):23-29.

## BIBLIOGRAFÍA

---

Zuckerlandl E & Pauling L. (1965). Evolutionary divergence and convergence in proteins. *Evolving Genes and Proteins*. 97. 97-166.