



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

Facultad de Ciencias

ANÁLISIS DE LA DIVERSIDAD BACTERIANA NO
CULTIVABLE Y DE LA DIVERSIDAD METABÓLICA
PRESENTE EN UN BANCO METAGENÓMICO DE
UNA MUESTRA DE SUELO.

TESIS

QUE PARA OBTENER EL GRADO ACADÉMICO DE
MAESTRA EN CIENCIAS BIOLÓGICAS
(BIOLOGÍA EXPERIMENTAL)

P R E S E N T A

BIOL. NOEMI SIRENA SÁNCHEZ

DIRECTOR DE TESIS: DR. JOSÉ ADELFO ESCALANTE LOZADA

MÉXICO, D.F.

Febrero, 2009



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Dr. Isidro Ávila Martínez
Director General de Administración Escolar, UNAM
Presente

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día 18 de agosto de 2008, se aprobó el siguiente jurado para el examen de grado de **MAESTRÍA EN CIENCIAS BIOLÓGICAS (BIOLOGÍA EXPERIMENTAL)** de la alumna **SIRENA SÁNCHEZ NOEMÍ** con número de cuenta **94103247** con la tesis titulada **"ANÁLISIS DE LA DIVERSIDAD BACTERIANA NO CULTIVABLE Y DE LA DIVERSIDAD METABÓLICA PRESENTE EN UN BANCO METAGENÓMICO DE UNA MUESTRA DE SUELO"**, realizada bajo la dirección del **DR. JOSÉ ADELFO ESCALANTE LOZADA**:

Presidente: DRA. NATHALIE CABIROL
Vocal: DRA. ESPERANZA MARTÍNEZ ROMERO
Secretario: DR. JOSÉ ADELFO ESCALANTE LOZADA
Suplente: DR. JESÚS CABALLERO MELLADO
Suplente: DR. GUILLERMO GOSSET LAGARDA

Sin otro particular, me es grato enviarle un cordial saludo.

Atentamente
"POR MI RAZA HABLARA EL ESPIRITU"
Cd. Universitaria, D.F., a 10 de febrero de 2009.



Dr. Juan Nuñez Farfán
Coordinador del Programa

Agradecimientos

Al Posgrado en Ciencias Biológicas de la UNAM

Al Consejo Nacional de Ciencia y Tecnología por la beca de maestría otorgada con el número de registro 203627

Para la realización de este proyecto se contó con los siguientes financiamientos:

CONACYT 46052

PAPIIT-UNAM IN213508-2

Agradezco profundamente:

A la **Universidad Nacional Autónoma de México**, por cultivar a la raza y enriquecer el espíritu.

Al **Instituto de Biotecnología**

A los miembros de mi comité tutorial

A los miembros de mi jurado

Al laboratorio Bolívar-Gosset por abrirme sus puertas.

A mi familia por todo.

A todos mis compañer@s por sus comentarios siempre útiles.

A Mechita, no mil, ni un millón, un infinito de agradecimientos por todo lo que nos ayudas en el lab.

A Chivis, por el parote que nos haces todos los días (no sabes como haces falta en vacaciones).

A la Unidad de Síntesis del IBT

A la Unidad de secuencia del IBT

A los miembros de la BSCP por su compañía en los ratos de ocio, y las charlas algunas veces interesantes, algunas veces sin sentido pero igualmente amenas.

A mis viej@s y nuev@s **amig@s**, de quienes no escribiré un solo nombre excusándome en el temor de una visita inesperada de mi siempre fiel Alzheimer y olvidar momentáneamente a alguien igualmente importante.

Dedicatoria

A José Ángel, motor que me impulsa a mover y respirar todos y cada uno de mis días, gracias por compartir conmigo la gran aventura de la Vida que por nada del mundo estaría dispuesta a perderme, gracias por enseñarme lo que se puede ser y hacer por un sueño, cuanto puede significar una lágrima y el gran Valor de una sonrisa sincera.

A mi papi y a mi mami.

A mis hermanas.

A Alex

A todos los que me apoyaron en esta etapa de mi existencia.

Y por qué no...?

...A mí

Índice

Índice de tablas.....	III
Índice de figuras.....	IV
Resumen	1
1. Introducción	2
2. Antecedentes	
2.1 Diversidad bacteriana del suelo	7
2.2 Diversidad metabólica bacteriana	8
2.3 El papel de la diversidad bacteriana en la dinámica de los suelos	11
2.4 Análisis de diversidad microbiana utilizando métodos moleculares no dependientes de cultivo.....	13
2.5 Identificación bacteriana basada en el análisis de secuencias de ADNr ARNr 16S	17
2.5.1 Características generales de los ADNr 16S.....	17
2.6 Aportaciones del análisis de metagenomas	24
3. Justificación, hipótesis y objetivos	26
4. Material y Métodos	
4.1 Toma de muestra para extraer ADN metagenómico de suelo	27
4.2 Extracción de ADN metagenómico	28
4.3 Amplificación de los ADNr 16S por PCR	30
4.4 Clonación del producto de PCR	30
4.5 Identificación de clonas positivas	31
4.6 Análisis de la diversidad de secuencias ADNr 16S presentes	31
4.7 Construcción del banco metagenómico.....	32
4.8 Análisis y caracterización del banco metagenómico	34
5 Resultados	
5.1 Extracción del ADN metagenómico	36
5.2 Análisis de diversidad de secuencias de ADNr 16S	37
5.3 Construcción del banco metagenómico	45
5.4 Análisis y caracterización del banco metagenómico	46

6. Discusión	50
7. Conclusiones	54
8. Perspectivas	55
9. Referencias	56
10. Anexos	62

Índice de tablas

Tabla 1. Ventajas y desventajas de algunos métodos moleculares para estudiar la diversidad microbiana del suelo	16
Tabla 2. Bancos construidos de ADN metagenómico de suelo y sus aportaciones	23
Tabla 3 Cepas y plásmidos utilizados en este trabajo	27
Tabla 4 Tabla comparativa de los métodos de extracción	29
Tabla 5 Mezcla de reacción de la PCR para amplificar ADNr 16S	30
Tabla 6 Reacción de digestión parcial de ADN metagenómico de suelo.....	33
Tabla 7. Mezcla de reacción de ligación de ADN metagenómico de suelo y el vector de clonación de productos de PCR p JET1.2	33
Tabla 8. Mezcla de reacción de digestión del pJET1.2/ MSCL 1-10	34
Tabla 9 Clonas obtenidas del metagenoma de suelo y las secuencias que presentaron mayor identidad depositadas en el GenBank	41
Tabla 10. Índice de diversidad H y E.....	43

Índice de figuras

Figura 1. Árbol filogenético del dominio <i>Bacteria</i>	3
Figura 2. Diferentes aplicaciones del estudio de la metagenómica	5
Figura 3. Construcción de bancos metagenómicos.....	6
Figura 4. Diversidad metabólica de los microorganismos	11
Figura 5. Técnicas moleculares no dependientes de cultivo microbiano utilizados en el análisis de la diversidad bacteriana de suelos	14
Figura 6. Estructura secundaria del ARNr 16S de <i>E. coli</i>	19
Figura 7. Distribución de las secuencias de ADNr 16S depositadas a lo largo de los últimos años	21
Figura 8. Diferentes antibióticos aislados a partir de bancos metagenómicos	25
Figura 9. Estrategia de secuenciación para insertos o fragmentos inespecíficos, primer walking	35
Figura 10 Imagen del ADN extraído.....	36
Figura 11 Electroforesis del ADN metagenómico extraído.....	37
Figura 12, Se muestran algunas de las digestiones que producen diferentes patrones de restricción	38
Figura 13. Gráfica de abundancia genética de los análisis ARDRA.....	39
Figura 14 Árbol filognético construido con el paquete PHYLIP 3.67	44
Figura 15 Digestiones de 10 clonas tomadas al azar	46
Figura 16. Análisis de restricción de las 10 colonias elegidas al azar	47
Figura 17. Mapa de las secuencias de 5 clonas analizadas	49

Resumen

El suelo representa uno de los hábitats más diversos y heterogéneos para las bacterias y arqueas en el planeta, en él existen comunidades bacterianas con mucha mayor diversidad que la que se presenta en la mayoría de los hábitats naturales. Sin embargo, el estudio de la diversidad presente por técnicas microbiológicas dependientes del cultivo ha demostrado que solo es posible obtener hasta el 1% de la diversidad presente. La aplicación de la aproximación metagenómica ha permitido poder acceder a esta diversidad aparentemente no cultivable y ha permitido determinar no tan solo la diversidad microbiana presente –cultivable, como no cultivable- y ha permitido también conocer la diversidad metabólica y genética presente en las muestras analizadas. En este trabajo se analizó la diversidad bacteriana y metabólica presente en el metagenoma de una muestra de suelo de bosque no perturbada aledaña al Instituto de Biotecnología, UNAM y que forma parte de la región del corredor ecológico Chihinautzin, Cuernavaca, Morelos. El análisis de diversidad bacteriana se realizó por secuenciación de los ADNr 16S amplificados directamente del metagenoma, mientras que para el análisis de la diversidad metabólica presente se construyó un banco metagenómico y se secuenciaron varias clonas. El análisis de diversidad bacteriana por análisis de secuencias de ADNr 16S permitió detectar Proteobacterias principalmente del grupo ϵ (como *E. coli*), bacterias no cultivadas y/o sin representantes cultivados, algunas que no poseen similitud significativa con las secuencias depositadas en el NCBI, las menos representadas fueron las Gram positivas. El banco metagenómico construido equivale a poco más de 87 genomas (con relación al tamaño del genoma de *E. coli*). La secuenciación de 5 clonas seleccionadas aleatoriamente de este banco utilizando el método de *primer walking*, permitió identificar los genes correspondientes a una posible proteína reguladora, una endonucleasa dependiente de ATP, una proteína involucrada en la biosíntesis de riboflavina RibF, una GTP difosfoquinasa y CHAP (dominio de GTP sintetasa) de *Sphingomonas*, los cuales pertenecen en su mayoría al grupo de las Proteobacterias.

1. Introducción

Los microorganismos procariontes representan por mucho el más grande reservorio de diversidad genética sobre la tierra. La bacteriología desarrollada alrededor de técnicas de cultivo puro ha sido el eje de estudios desde la taxonomía hasta la genómica. Sin embargo, sabemos ahora con un alto nivel de certeza que el porcentaje de bacterias cultivables en la mayoría de los suelos es muy pobre alrededor de 1% o menos (Rodríguez-Valera, 2004). La actual diversidad de los genes dentro de las especies bacterianas es otra incógnita importante en el estudio de la diversidad procarionte (Boucher, *et al.* 2001).

Se han propuesto diferentes hipótesis para explicar la naturaleza no cultivable (aparentemente el componente mas abundante) de la diversidad procarionte. Una de estas hipótesis propone que en realidad se trata de microorganismos relacionados a especies previamente cultivadas y que la incapacidad de poderlos crecer en laboratorio se debe a su naturaleza recalcitrante. Se ha determinado que la modificación de las condiciones de cultivo han hecho posible revertir este estado no cultivable para algunos microorganismos (Keller y Zengler, 2004). Sin embargo, otras tendencias basadas en el análisis de secuencias de ADNr 16S amplificadas directamente del ambiente, han demostrado la presencia de grupos bacterianos completamente nuevos y sin representantes previamente cultivados. En un reporte en el 2003 existían 26 phyla candidatos que no contienen representantes cultivados, algunos de ellos se han denominado como OP10, OP3, WS3, BRC1, NKB19, OP9 y Termite group1 entre otros (Handelsman, 2004). En la figura 1 se muestra un árbol filogenético del dominio bacteria donde se incluyen phyla definidos y también aquellos que mayoritariamente están integrados por microorganismos no cultivados, además de incluir un gran número de phyla propuestos pero que no poseen miembros o representantes cultivados.

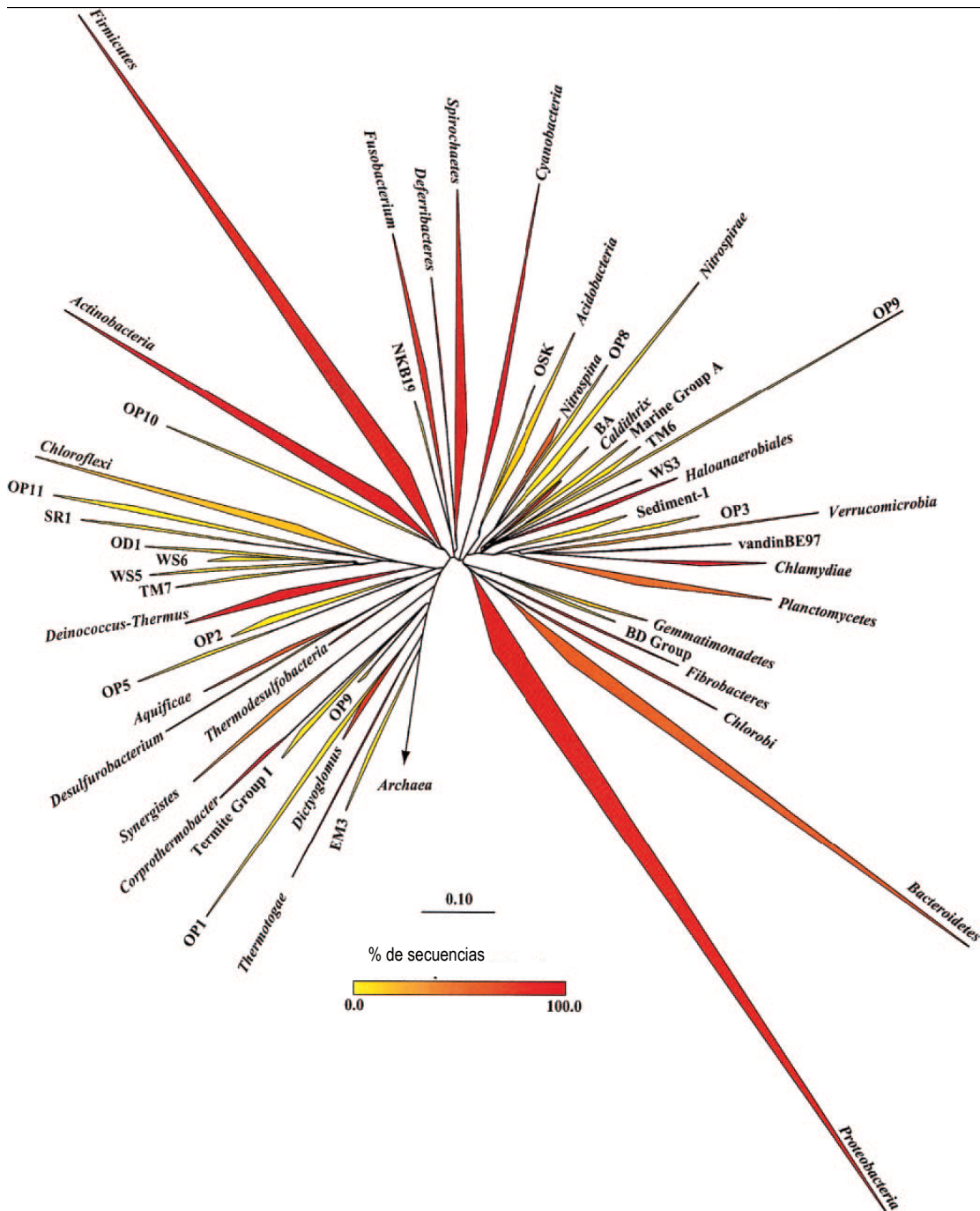


Figura 1. Árbol filogenético del dominio *Bacteria* (en itálicas los phyla establecidos) y phyla candidatos descritos en 2003 en la base de datos ARB (software para análisis de secuencias ambientales [del latín arbor, árbol]). La intensidad del color rojo representa la cantidad de secuencias en cada phylum obtenidas de representantes cultivados. Los phyla candidatos no poseen miembros cultivados (Handelsman, 2004).

Del total de 53 phyla descritos, únicamente 27 poseen representantes cultivados y algunos solo poseen unos cuantos microorganismos aislados y muchos phyla contienen únicamente una especie descrita. Solo cinco phyla como son: *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Firmicutes* y *Proteobacteria* incluyen especies que producen moléculas bioactivas. Esos cinco representan 95% de las especies cultivadas publicadas. Los 22 phyla restantes con miembros cultivados representan únicamente el 5% de todas las especies publicadas (Keller y Zengler, 2004).

Para 2008 existen cerca de 100 phyla descritos mediante análisis de secuencias de fragmentos de ADNr 16s de 103 bases en promedio (región hipervariable V9) obtenidas de muestras de cuatro suelos diferentes (Fulthorpe, *et al.*, 2008). En el anexo 1 se muestra la lista de los diferentes phyla que se encontraron en este trabajo y que están depositadas en <http://greengenes.lbl.gov>

La posibilidad de acceder a esta diversidad bacteriana no cultivable solo es posible por medio de la aplicación metagenómica (figura 2) (Rajendhram y Gunasekaran, 2008). El punto de partida para el análisis metagenómico de una muestra ambiental es la extracción del metagenoma bacteriano (conjunto de genomas microbianos presentes en un ambiente). Existen dos aproximaciones metodológicas para su obtención: 1) Un método de lisis directo que se basa en la lisis de todos los microorganismos presentes en una muestra ambiental por una combinación de métodos físicos, químicos y enzimáticos, para posteriormente purificarlos; y 2) la extracción indirecta, que se basa en una separación diferencial de la fracción bacteriana por homogenización, centrifugación y su posterior lisis (Rajendhram y Gunasekaran, 2008).

Una vez obtenido el metagenoma bacteriano, este se fragmenta en una serie de tamaños discretos para ser clonado en vectores tales como los plásmidos, fósidos, cósmidos o los cromosomas bacterianos artificiales (BAC, por sus siglas en inglés), para ser entonces transformados en cepas hospederas

específicas. Con la obtención del banco metagenómico se tiene un pool que representa varias veces el tamaño de un genoma de referencia (p. ej. *E. coli*).

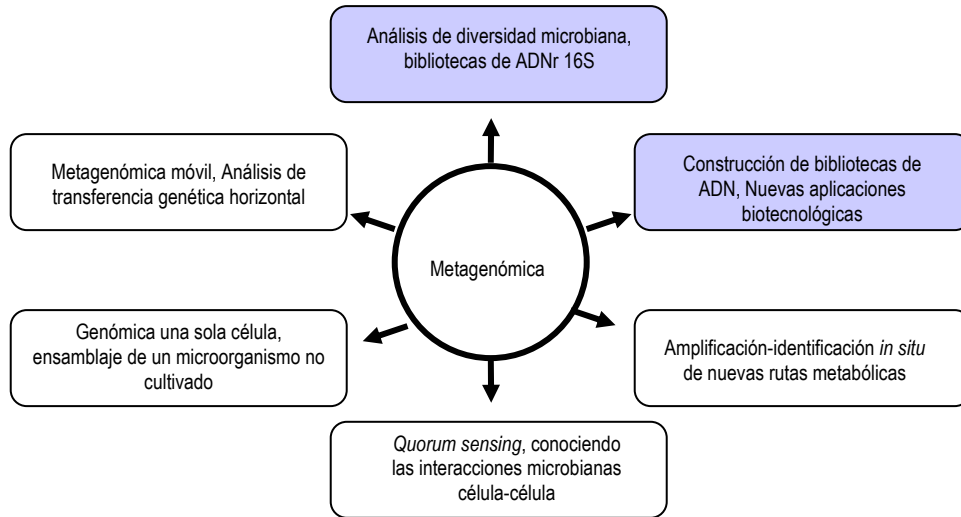


Figura 2. Diferentes aplicaciones del estudio de la metagenómica (Rajendhram y Gunasekaran, 2008). Para los fines de este estudio se realizó un análisis de diversidad microbiana y la construcción de una biblioteca o banco de ADN que en el esquema se muestran en gris.

A partir del banco metagenómico se pueden aplicar diferentes estrategias para su análisis: detección directa de genes de interés por PCR o por hibridación con sondas específicas; secuenciación masiva del metagenoma o bien por complementación funcional de cepas con mutaciones específicas que son complementadas con el material metagenómico inserto en los vectores utilizados para la construcción del banco. Como resultado de la aplicación de estas aproximaciones ha sido posible la detección y obtención de genes que codifican para una gran diversidad de productos que comprenden desde enzimas hasta metabolitos (debido a la clonación de vías metabólicas completas) (Figura 3)

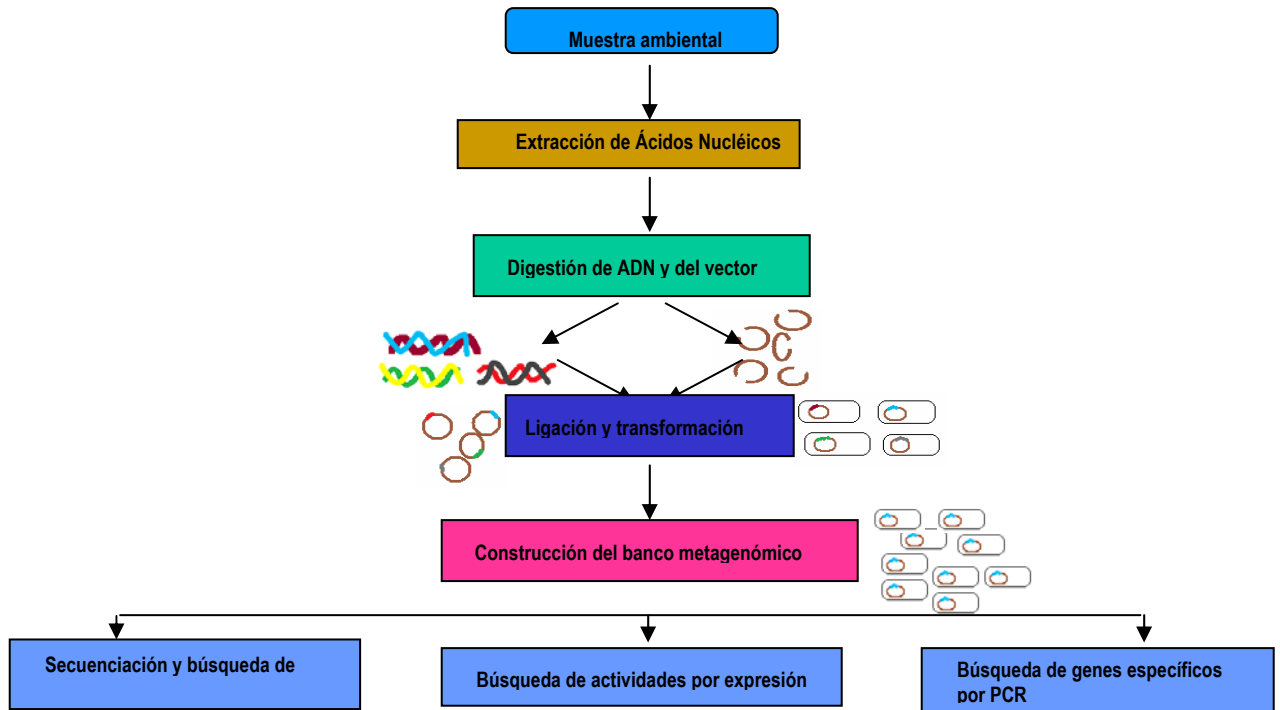


Figura 3. Construcción de bancos metagenómicos de muestras ambientales y estrategias utilizadas para su caracterización (Schloss y Handelsman, 2003; Cowan *et al.*, 2005)

2. Antecedentes

2.1 Diversidad bacteriana del suelo

La diversidad biológica o biodiversidad, se define como las variaciones colectivas a todos los niveles de organización biológicos, desde variaciones genéticas dentro de poblaciones y especies, especies con comunidades, a comunidades que componen un ecosistema. La biodiversidad comúnmente describe tanto la información contenida y como se distribuye en los sistemas biológicos o comunidades naturales (Torsvik y Øvreås, 2007).

El suelo representa uno de los hábitats más diversos y heterogéneos para las bacterias y arqueas en el planeta, en él se presentan comunidades bacterianas con mucha mayor diversidad que la que se presenta en la mayoría de los hábitats naturales. Sumado a la gran diversidad de la vida procarionte, en la mayoría de los suelos la cantidad de microorganismos están en el rango de 10^7 - 10^{10} células por gramo de suelo seco. La evaluación de la diversidad dentro de las comunidades microbianas es uno de los aspectos más fascinantes de la microbiología, especialmente por el hecho de que esta puede ser muy pequeña. Esto se refleja por el número de especies bacterianas que existen sobre la tierra, el cual varía entre 10^4 y 10^9 (Torsvik y Øvreås, 2007).

La abundancia de las células procariontes en el suelo puede ser visualizada por técnicas microscópicas directas. En un trabajo publicado por Hugenholtz y colaboradores (en Nogales, 2005), los autores analizaron la presencia de los diferentes grupos filogenéticos bacterianos en estudios realizados en diferentes ambientes, aquellos grupos microbianos que se encuentran mas frecuentemente en muestras de suelo pertenecen a la Subclase α -Proteobacteria, seguido de microorganismos Gram-positivos de la clase Actinobacteria y dos phyla nuevos, Acidobacteria y Verrumicrobia (Xu, 2006).

Con el uso de aproximaciones metagenómicas y otras herramientas moleculares (bibliotecas genómicas de ARNr 16S) recientemente se ha obtenido evidencia de

que los miembros del Phylum Acidobacteria, el cual contiene comúnmente al menos ocho subgrupos filogenéticos, son ubicuas en gran número de ambientes terrestres, desde los suelos agrícolas hasta los suelos boscosos y de algunos contaminados con metales o radioactividad a suelos prístinos. Una biblioteca genómica de ADN de suelo que es sondeada para secuencias de ADNr 16S pueden contener alrededor de 25%-30% de secuencias afiliadas a acidobacterias, esto se observó en muestras de suelo de pastoreo australianas (van Elsas, *et al.* 2007).

Para calcular una medida de la diversidad total en los sistemas biológicos, se han usado ampliamente los índices de diversidad. Los índices primarios dan medidas del número de taxa o grupos en una comunidad, mientras que un índice secundario o compuesto se basa en dos componentes, riqueza y abundancia. La riqueza describe el número de diferentes taxa en una comunidad y la abundancia describe que tan distribuidos están los individuos dentro de los taxa (la abundancia relativa de los individuos entre los taxa. Índices de diversidad comúnmente usados son el índice de dominancia de Simpson (D), el índice de reciprocidad de Simpson ($1/D$), el índice de series logarítmicas (el cual es sensible a la riqueza de taxones) y el índice de Shanon-Weaver (H'), el cual es una medida de riqueza y abundancia (Odum, 1971).

2.2 Diversidad metabólica bacteriana

Las bacterias exhiben una gran diversidad metabólica. La distribución de estos tipos metabólicos dentro de un grupo de bacterias se ha utilizado tradicionalmente para definir su taxonomía, pero estos rasgos no corresponden a menudo con las clasificaciones genéticas modernas. El metabolismo bacteriano se clasifica con base en tres criterios importantes: el origen de la fuente de carbono, y los donadores de electrones. Un criterio adicional para clasificar a los microorganismos que respiran es el receptor de electrones usado en la respiración.

Según la fuente del carbono, las bacterias se pueden clasificar como heterótrofas cuando usan compuestos orgánicos, o como autótrofas cuando el carbono celular se obtiene mediante la fijación del dióxido de carbono. Las bacterias autótrofas típicas son las cianobacterias fotosintéticas, las bacterias verdes azufrosas y algunas bacterias purpúreas. Pero hay también muchas otras especies quimiolitótrofas, por ejemplo, las bacterias nitrificantes y oxidantes del azufre. Según la fuente de energía, las bacterias pueden ser fotótrofas, es decir, que emplean la luz a través de la fotosíntesis, o quimiótrofas, es decir, que obtienen energía a partir de sustancias químicas que son oxidadas principalmente a expensas del oxígeno o de otros receptores de electrones alternativos (respiración aerobia/anaerobia).

Finalmente, las bacterias se pueden clasificar como litótrofas u organótrofas según utilicen como donadores de electrones compuestos inorgánicos u orgánicos, respectivamente. Los organismos quimiótrofos usan donadores de electrones para la conservación de energía (durante la respiración aerobia/anaerobia o durante la fermentación) y para las reacciones biosintéticas (por ejemplo, para la fijación del dióxido de carbono), mientras que los organismos fotótrofos los utilizan solamente para propósitos biosintéticos. Los organismos que respiran usan compuestos químicos como fuente de energía, tomando electrones del sustrato reducido y transfiriéndolos a un receptor terminal de electrones en una reacción redox. Esta reacción desprende energía que se puede utilizar para sintetizar ATP y conducir el metabolismo. En los organismos aerobios, el oxígeno se utiliza como receptor de electrones. En los organismos anaerobios se utilizan como receptores de electrones otros compuestos inorgánicos tales como los nitratos, sulfatos o el dióxido de carbono. Esto conduce a los procesos ecológicamente importantes de la desnitrificación, reducción del sulfato y acetogénesis, respectivamente.

Otra posibilidad de los quimiótrofos en ausencia de receptores de electrones es la fermentación. En la fermentación, los electrones tomados de los sustratos reducidos se transfieren a intermediarios oxidados para generar

productos de fermentación reducidos (por ejemplo, lactato, etanol, hidrógeno, butirato). La fermentación es posible porque el contenido de energía de los substratos es mayor que el de los productos, lo que permite que los organismos sinteticen ATP y conduzcan su metabolismo. Estos procesos son también importantes en las respuestas biológicas a la contaminación. Por ejemplo, las bacterias reductoras de sulfato son en gran parte responsables de la producción de formas altamente tóxicas de mercurio (metil- y dimetil-mercurio) en el ambiente. Los organismos anaerobios utilizan la fermentación para generar energía, secretando subproductos metabólicos (tales como el etanol en la elaboración de la cerveza) como productos de desecho. Los anaerobios facultativos pueden elegir entre la fermentación y diversos receptores terminales de electrones dependiendo de las condiciones ambientales en las cuales se encuentren.

Las bacterias litótrofas pueden utilizar compuestos inorgánicos como fuente de la energía. Los donantes de electrones inorgánicos más comunes son el hidrógeno, monóxido de carbono, amoníaco (que conduce a la nitrificación), hierro ferroso y otros iones de metales reducidos, y varios compuestos de azufre reducidos. Inusualmente, las bacterias metanótrofas pueden usar gas metano tanto como fuente de electrones como substrato para el anabolismo del carbono. En la fototrofia y quimiolitotrofia aerobias, se utiliza el oxígeno como receptor terminal de electrones, mientras que bajo condiciones anaerobias se utilizan compuestos inorgánicos. La mayoría de los organismos litótrofos son autótrofos, mientras que los organismos organótrofos son heterótrofos (Xu, 2006). La figura 4 muestra la clasificación de los microorganismos de acuerdo al uso de diferentes fuentes de carbono y energía (Madigan, et al., 2008).

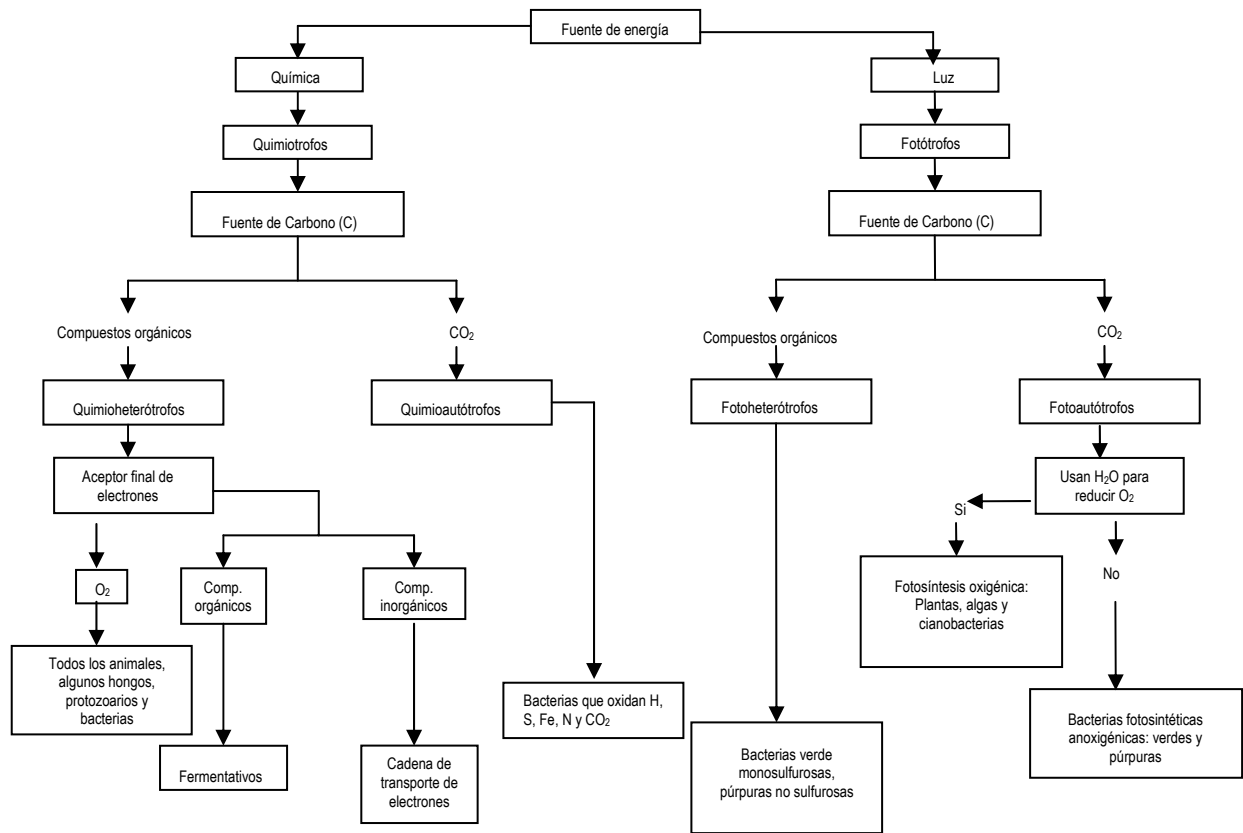


Figura 4. Diversidad metabólica de los microorganismos. Esta clasificación se ha hecho basada en la fuente de energía y carbono que utilizan los microorganismos, así como el tipo de metabolismo que estos presentan (Madigan, et al., 2008).

2.3 El papel de la diversidad bacteriana en la dinámica de los suelos

Las bacterias en el suelo se dividen tradicionalmente en *autóctonas* y *zimógenas*. Los microorganismos autóctonos son habitantes comunes que tienden a mantenerse a niveles relativamente constantes en el medio, a pesar de las fluctuaciones en las condiciones locales, incluyendo los niveles de materia orgánica. Los organismos clasificados como zimógenos pueden únicamente crecer cuando los recursos son abundantes y por lo tanto son oportunistas en términos tróficos, los organismos zimógenos son usualmente copiotróficos, lo cual significa que necesitan acceder a la materia orgánica fácilmente oxidable para su crecimiento (van Elsas, et. al. 2007).

El suelo es un sistema muy complejo que comprende una variedad de microhábitats con diferentes gradientes fisicoquímicos y condiciones ambientales discontinuas. Los microorganismos se adaptan a microhábitats y viven juntos en consorcios e interactuando de una u otra manera con otras partes de la biota del suelo (Torsvik y Øvreast, 2002). Los microorganismos activos en el suelo son responsables en gran parte de los ciclos biogeoquímicos, median el ciclo del nitrógeno proporcionando la única ruta biológica para que el nitrógeno atmosférico entre en los sistemas vivos a través de la fijación de nitrógeno oxidándolo a través de la nitrificación y reciclándolo a la atmósfera por la desnitrificación. De manera similar, éstos juegan un papel central en los ciclos de carbono y azufre y en la transformación oxidativa y reductora de metales como hierro y mercurio (Handelsman, 2004). Por tanto, la identificación de todos los miembros de la comunidad y genes en un ambiente es fundamental para obtener una perspectiva de los sistemas ecológicos (Noguchi, *et al.* 2006).

La diversidad de consorcios microbianos naturales es una fuente rica para nuevas enzimas industriales y compuestos bioactivos, esta diversidad ha sido explorada durante muchos años basada en el cultivo y aislamiento de especies bacterianas. Dichas técnicas son conocidas por su selectividad, pero no permiten aislar un número de especies representativas del grueso de las comunidades bacterianas (Knietsch, *et al.* 2003).

El suelo tiene una basta diversidad filogenética (Lozupone y Knight, 2007), pero el desarrollo de las prácticas industriales en los últimos, 200 años ha dado como resultado, severos daños en el balance ecológico, en la mayoría de los ecosistemas. Como producto de estos disturbios hay un decremento en la biodiversidad, una extinción irreversible de muchas especies sensibles así como la selección de otras que se han adaptado mejor a un ambiente industrializado (Kozdrój y van Elsas, 2001).

2.4 Análisis de diversidad microbiana utilizando métodos moleculares no dependientes de cultivo

La aplicación de herramientas moleculares al estudio de la diversidad microbiana ha permitido la posibilidad de detectar una mayor diversidad bacteriana que aquella observada por técnicas microbiológicas dependientes de cultivo (Riesenfeld, *et al.* 2004).

Existe un amplio rango de métodos disponibles para estudiar la diversidad microbiana del suelo incluyendo a la fracción no cultivada. Cada método tiene sus limitaciones y únicamente proporciona una imagen parcial de un aspecto de la diversidad del suelo (Kirk, *et al.* 2004). El estudio de la diversidad bacteriana ha revolucionado, gracias a la amplificación por PCR de genes (principalmente ADNr 16S) que de alguna forma permiten una identificación directamente de las muestras ambientales. La conclusión mas relevante es que una vasta sección de la diversidad procarionte no está representada en las colecciones cultivadas y por lo tanto sigue siendo básicamente desconocida (Rodríguez-Varela 2004). En la figura 5 se muestra un esquema de las diferentes estrategias no dependientes de cultivo que pueden seguirse para analizar la diversidad bacteriana de los suelos (Escalante-Lozada, *et al.* 2004).

Las técnicas basadas en PCR han sido usadas para contrarrestar las limitaciones de los métodos basados en cultivo, sin embargo, ellas también poseen sus propias limitaciones (Kirk, *et al.* 2004). Las bacterias existen dentro o sobre la superficie de agregados de suelo; la capacidad para separar esas células de los componentes del suelo es vital para estudiar la biodiversidad (Trevors, 1998). Si el método de extracción celular usado es muy gentil, se romperán las paredes de las bacterias Gram-negativas, pero no las Gram-positivas. Y si el método es muy agresivo, podrán lisarse tanto Gram-positivas como Gram-negativas, la desventaja es que el ADN podría fragmentarse (Wintzingerode, *et al.*, 1997).

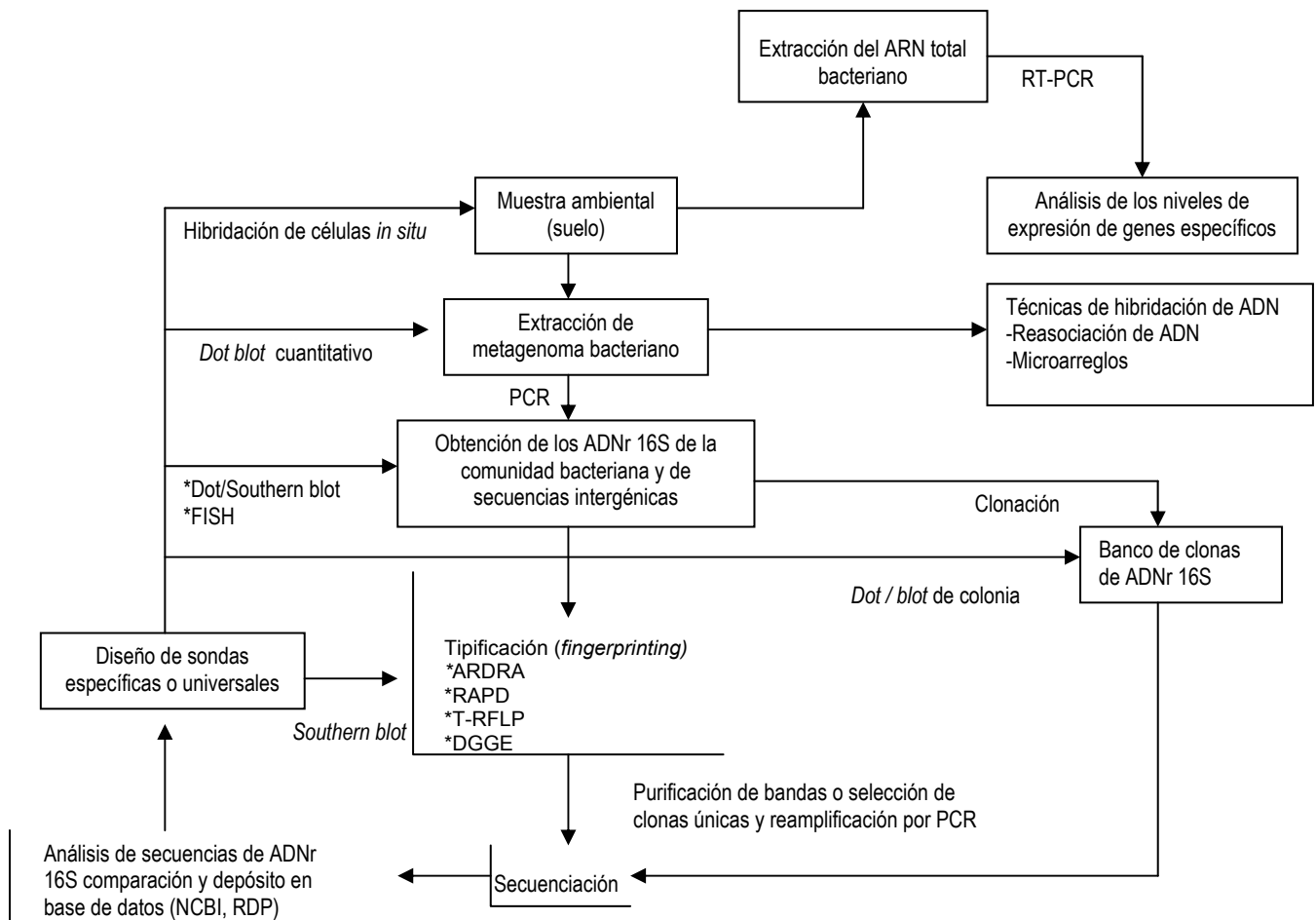


Figura 5. Técnicas moleculares no dependientes de cultivo microbiano utilizados en el análisis de la diversidad bacteriana de suelos (Escalante-Lozada, *et al.* 2004).

El método usado de extracción de ADN o ARN puede sesgar los estudios de diversidad. Los métodos de extracción agresivos como la agitación con perlas pueden fragmentar los ácidos nucleicos también, originando problemas en subsecuentes amplificaciones por PCR (Wintzingerode, *et al.* 1997). Con muestras ambientales, es necesario remover sustancias inhibitorias como ácidos húmicos, los cuales normalmente se coextraen e interfieren con subsecuentes análisis de PCR. Pero, los subsecuentes pasos de purificación pueden originar una pérdida de ADN o ARN, sesgando potencialmente los análisis de diversidad molecular (Kirk, *et al.* 2004).

El análisis de los genes de ARNr 16S proporciona información valiosa acerca de la evolución y diversidad de poblaciones microbianas, pero no revela el

papel de diferentes microorganismos dentro de las poblaciones naturales. Las copias del gen 16S presentes en un microorganismo abarcan aproximadamente 0.05% del genoma procarionte promedio y no puede ser usado para obtener información acerca de la fisiología del organismo del cual se obtuvo (Rodríguez-Varela, 2002).

El gen ADNr 16S presenta una gran heterogeneidad intraespecífica, por lo tanto, en una misma especie pueden existir mas de dos copias del mismo. Diversos estudios han revelado que dichas copias evolucionaron de manera separada, lo cual confiere una desventaja del uso de este gen. Una alternativa al uso del gen 16S (Ueda, *et al.* 1999), puede ser el uso de *rpoB*, el cual se encuentra en una sola copia, es común en todas las bacterias, es conservado y también tiene regiones variables y funciona como un reloj biológico. El gen para la subunidad beta de la ARN polimerasa, gracias a las características antes mencionadas puede ser utilizado para la identificación de especies. Un producto de PCR de este gen puede dar únicamente una banda en un gel de DGGE, haciendo posible distinguir diferentes especies usando DGGE O TGGE sin tener la secuencia de bandas individuales, las medidas de correlación y diversidad pueden ser aplicadas directamente a los patrones de DGGE o TGGE (Dahllöf, *et al.* 2000).

Otra técnica similar es la conocida como técnica Terminal Restriction Fragment Length Polymorphism (T-RFLP), la cual es una mejora del ARDRA, y que mide únicamente los fragmentos terminales de restricción del ADNr 16S, la complejidad del patrón de RFLP es más reducido y donde cada fragmento es representativo de un ribotipo o unidad taxonómica operacional (Kozdrój y van Elsas, 2001).

Las relaciones filogenéticas entre proteínas RecA de mas de 65 especies de bacterias han sido reportadas en diferentes estudios (Eisen 1995). Los árboles son robustos y correlacionan con los árboles generados por ADNr 16S de las bacterias correspondientes, generando la conclusión de que RecA es un marcador genético utilizado para reconstruir filogenias bacterianas (Sandler, *et al.* 1999). En

la tabla 1 se enlistan algunos de los diferentes métodos moleculares usados para estudiar la diversidad bacteriana del suelo (Kirk, *et al.* 2004).

Tabla 1. Ventajas y desventajas de algunos métodos moleculares para estudiar la diversidad microbiana del suelo (Kirk, *et al.* 2004)

Método	Ventajas	Desventajas
Porcentaje de G+C	No está influenciado por el sesgo de la PCR. Incluye todo el ADN extraído.	Requiere grandes cantidades de ADN. Depende de la lisis y eficiencia de extracción
	Incluye miembros raros de la comunidad	Nivel de resolución grueso
Reasociación de ácidos Nucleicos e hibridación	No está influenciado por el sesgo de la PCR.	Poca sensibilidad
	Estudia ARN y ADN Puede ser estudiado <i>in situ</i> Pueden analizarse miles de genes	Necesita un alto número de copias de las secuencias para poder ser detectadas Depende de la lisis y eficiencia de extracción
Microarreglos e hibridación de ADN	Si se usan genes o fragmentos de ADN se incrementa la especificidad	Únicamente se detectan las especies más abundantes Solo es preciso en sistemas de baja diversidad
	Pueden analizarse un gran número de muestras simultáneamente	Sesgo de la PCR, depende de la eficiencia de lisis y extracción Una banda puede representar más de una especie (co-migración)
DGGE o TGGE	Confiable, reproducible y rápido	Únicamente se detectan las especies dominantes
	Detecta cambios estructurales en la comunidad microbiana	Sesgo de la PCR Patrones de bandeo extremadamente complejos
ARDRA o RFLP	Patrones de bandeo más simples que los de RFLP	Depende de la eficiencia de lisis y extracción
	Puede ser automatizado, gran número de copias	Sesgo de la PCR El tipo de <i>Taq</i> puede incrementar la variabilidad
T-RFLP	Altamente reproducible. Compara diferencias en comunidades microbianas	Elección de primers universales
	RISA y ARISA	Perfiles de comunidad altamente reproducibles Bases de datos insuficientes

2.5 Identificación bacteriana basada en el análisis de secuencias de ADNr ARNr 16S.

El campo de la ecología microbiana fue revolucionada por los estudios de Woese en 1990 y Pace en 1997 basados en análisis de secuencias de ADN o ARN, la aplicación de primers universales para la amplificación directa por PCR de genes de ARNr 16S del ADN total de la comunidad, combinado con tecnologías de clonación y secuenciación, ha generado una vasta cantidad de datos y una diversidad procarionte redefinida (Green y Keller, 2006).

El factor más obvio que hace a una molécula una buena medida de relaciones filogenéticas es su tamaño, aunque no es la condición determinante ya que hay aspectos de la organización estructural de la molécula que son importantes, p. ej. que todas las moléculas parecen estar organizadas en unidades estructurales. La diversidad de tasas de cambio dentro de los ARNs es una gran ventaja. Aunque esta permite que la misma molécula sea usada para relaciones globales y locales, esta diluye la cantidad de información usada en alguna distancia. Las posiciones conservadas son esenciales para el reconocimiento de secuencias homólogas, pero no proporciona información filogenética (Olsen y Woese, 1993).

Las razones por las que algunas moléculas son mejores indicadores de relaciones filogenéticas que otros, no siempre es claro. Por ejemplo, el ARN 7S es mas grande que el ARN 5S, pero es menos usado como indicador filogenético a niveles taxonómicos superiores (Kaine, 1990).

2.5.1 Características generales de los ARNr 16S

El ARN ribosomal (ARNr) 16S es la macromolécula más ampliamente utilizada en estudios de filogenia y taxonomía bacterianas. Su aplicación como reloj molecular fue propuesta por Carl Woese a principios de la década de 1970. Los estudios de Woese originaron la división de los procariontes en dos grupos o reinos: *Eubacteria* y *Archaeobacteria*, cuya divergencia es tan profunda como la

encontrada entre ellos y los eucariontes y permitieron establecer las divisiones mayoritarias y subdivisiones entre ambos reinos (Olsen y Woese, 1993).

El ribosoma bacteriano tiene un coeficiente de sedimentación de 70S (expresado en unidades Svedberg) y puede disociarse en dos subunidades, la subunidad grande 50S y la subunidad pequeña 30S. La subunidad 30S contiene al ARNr 16S. En bacterias, los genes que codifican los ARN ribosomales están organizados en operones. Cada operón ribosomal (*rrn*) incluye genes para los ARNr 23s (*rrl*), 16S (*rrs*) y 5S (*rrf*) separado por regiones intergénicas, y contiene además genes para uno o más ARN de transferencia (ARNt) (Rodicio y Mendoza 2004). El número de copias del operón ribosomal por genoma bacteriano varía considerablemente, de 1 a 15, siendo relativamente constante a nivel de especie, género e incluso familia (Klappenbach, *et al.* 2001).

El ARNr 16S es un polirribonucleótido de aproximadamente 1,500 nucleótidos, codificado por el gen *rrs*, también denominado como ADN ribosomal 16S (ADNr 16S), a partir de cuyas secuencias se puede obtener información filogenética y taxonómica. En la figura 6 se puede observar que, como cualquier secuencia de nucleótidos de cadena sencilla, el ARNr se pliega en una estructura secundaria, caracterizada por la presencia de segmentos de doble cadena, alternando con regiones de cadena sencilla (Neefs, *et al.* 1990).

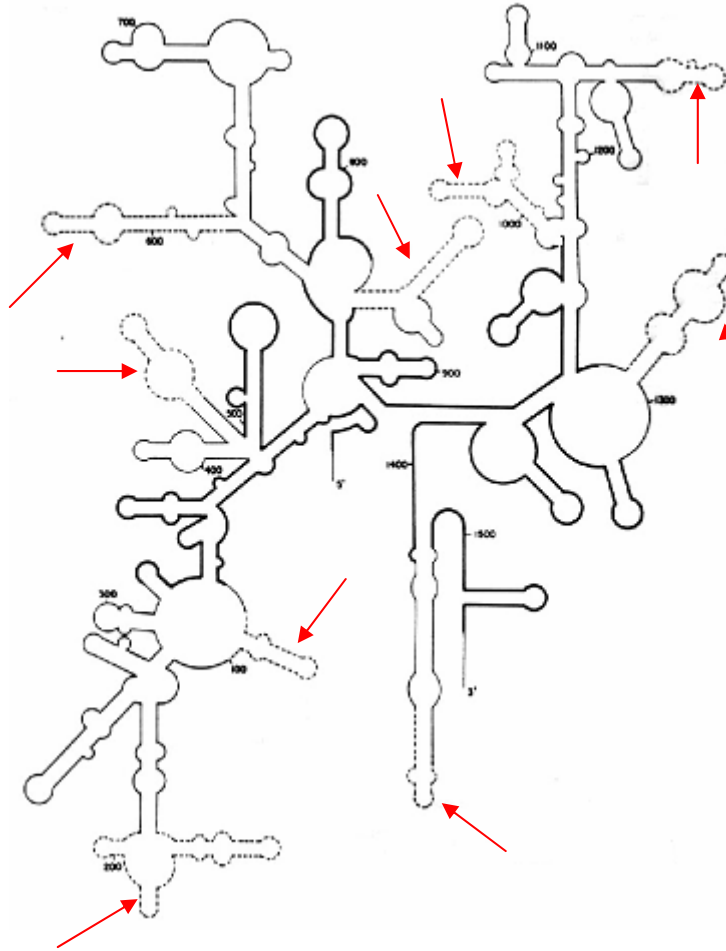


Figura 6. Estructura secundaria del ARNr 16S de *E. coli*: Líneas gruesas: dominios de conservación universal, líneas normales: dominios de conservación intermedia y líneas punteadas y señaladas con flechas rojas: las 9 regiones hipervariables.

Las características de los ARNr 16S por las cuáles han sido elegidos como reloj molecular son las siguientes:

1. Es una molécula muy antigua, presente en todas las bacterias actuales.
2. Su estructura y función han permanecido constantes durante un tiempo muy prolongado, de modo que las alteraciones en la secuencia probablemente reflejan cambios aleatorios.
3. Los cambios han ocurrido de manera suficientemente lenta, para aportar información sobre las relaciones de todos los procariontes. Contienen

suficiente variabilidad para diferenciar a los organismos mas alejados y también a los más próximos.

4. La conservación en la estructura secundaria, puede servir de ayuda en las comparaciones, aportando una base para el alineamiento preciso (Rodicio y Mendoza, 2004).

Existen bases de datos en las que se encuentran depositadas todas las secuencias de ADNr 16S obtenidas tanto de microorganismos cultivables como de microorganismos no cultivados, resultado del análisis de secuencias ambientales (figura 7). De ella, en la base de genes no redundantes del National Center for Biotechnology Information (NCBI) se encuentran hasta noviembre de 2008, 43 657 registros de ADNr 16S totales y 37 619 registros de ADNr 16S bacterianos. Por otro lado, la base de datos ribosomales (RDP) para octubre de 2008 tenia 706 103 registros de ARNr.

El concepto de especies procariontes ha evolucionado y hoy en día involucra una combinación de métodos - también llamado *taxonomía polifásica* - para diferenciar especies procariontes tanto en el contexto genético como en el fenotípico. En términos de criterio genético, la taxonomía polifásica incluye, en particular, la secuenciación de la subunidad pequeña del ARN ribosomal e hibridación genómica (Madigan, *et al.*, 2008). La obtención de la secuencia del ADNr 16S es el primer paso para identificar o determinar la filiación taxonómica de un microorganismo. La secuencia, se compara contra las entradas de ADNr 16S de alguna base de datos y de acuerdo a los resultados de salida, es posible tomar una decisión sobre la identidad del microorganismo o secuencia en cuestión. Los otros criterios para definir una especie con base en el análisis de secuencias dependen de varios factores: el tamaño de la secuencia de ADNr 16S (parcial o completa) y el grado de identidad con la(s) secuencia(s) de referencia (Clarridge III, 2004).

Se ha propuesto que un procarionte cuya secuencia de ADNr 16S difieren mas de 3% de todos los otros organismos (menos de 97% de identidad con las otras

secuencias de la base de datos) deben ser consideradas como una nueva especie. El apoyo a esta propuesta emerge de la observación de que de dos secuencias de ADNr 16S bacteriano que poseen menos de 97% de identidad típicamente hibridan menos del 70% que es el valor mínimo considerado como evidencia de que dos organismos son la misma especie (Madigan, *et al.*, 2008).

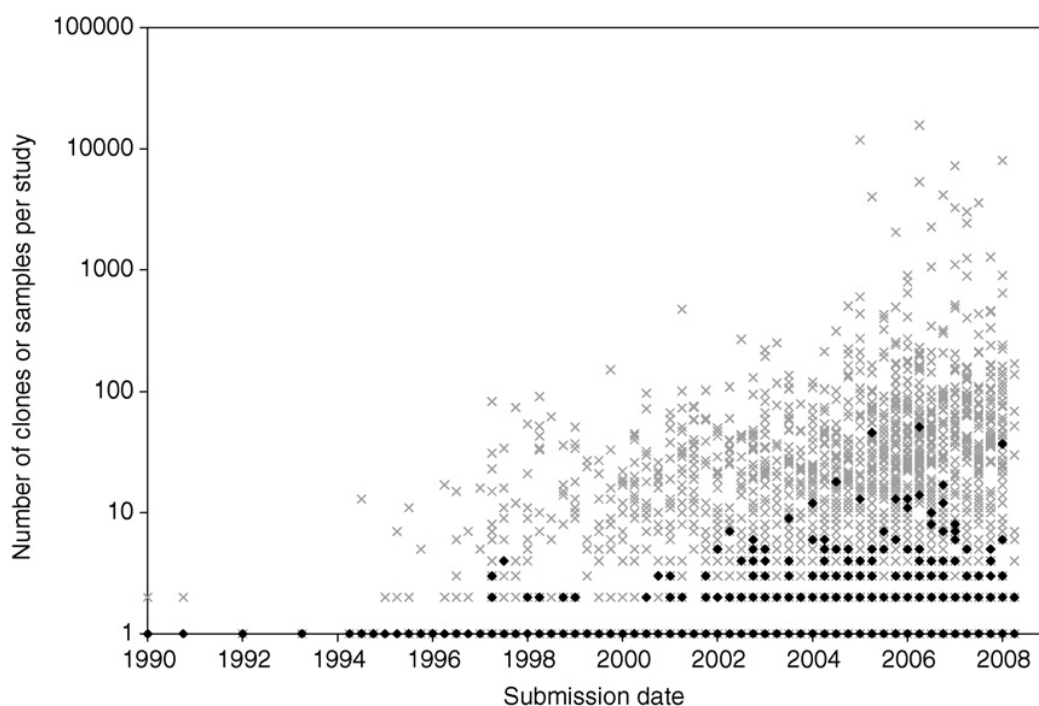


Figura 7. Distribución de las secuencias de ADNr 16S depositadas a lo largo de los últimos años. Las X representan las clonas de ADNr 16S secuenciadas y los diamantes negros representan las muestra analizadas por año teniendo una tendencia de aumento exponencial con en transcurrir del tiempo (Tringe y Hugenholtz, 2008).

Estos criterios permiten entonces tomar una decisión del tipo de primers más adecuado para realizar el análisis del ADNr 16S. Se ha reportado el uso de un par de primers universales que permiten la amplificación del gen completo (aproximadamente 1500 pares de bases) y que ha sido utilizado para la amplificación de una gran variedad de ADNr 16S de cepas de referencia (Weisburg, *et al.* 1991) y a partir de muestras metagenómicas (Escalante, *et al.*

2001). Mientras que se ha reportado también el uso de un par de primers que amplifican un fragmento de 1000 pares de bases correspondiente a las posiciones 500 a 1500 respecto a la secuencia de ADNr 16S de *E. coli*. En este caso, se señala que el amplicón obtenido corresponde a una región que contiene el extremo -3' del ADNr 16S la cual es una región hipervariable, mientras que la secuencia hacia el extremo -5' es mas conservada. Este par de primers fue de gran utilidad para la determinación de los grupos dominantes en una muestra de suelo (Borneman, *et al.* 1996).

Debido a que las estrategias de secuenciación son capaces de leer solamente algunos cientos de nucleótidos, la técnica desarrollada por Fred Sanger en 1982 llamada *shotgun sequense* (Pop, 2004). Esta técnica consiste en fragmentar azarosamente el ADN y clonar en el vector de elección dichos fragmentos, se secuencian cada una de las clonas y se utilizan programas computacionales para comparar los datos, y buscar solapamientos en los datos, de esta forma se arma una secuencia completa de gran tamaño (Staden, 1979),

La construcción y caracterización de bancos genómicos obtenidos a partir de ADN aislado directamente de muestras ambientales es una herramienta reciente y poderosa para el descubrimiento de nuevas enzimas de interés biotecnológico (p. ej. lipasas, esterasas, quitinasas, celulasas, y amilasas) (Gabor, *et al.* 2004). En la tabla 2 se enlistan algunas de las aportaciones del análisis metagenómico, el tamaño promedio de los bancos obtenidos, el tipo de vector utilizado y la cantidad de ADN clonado en dichos bancos (Daniel, 2005).

Tabla 2. Bancos construidos de ADN metagenómico de suelo y sus aportaciones (Daniel, 2005).

Origen	Tipo de vector	No. de clonas	Tamaño promedio del inserto (Kb)	ADN total (Gb)	Genes de interés	Año de construcción
Pradera, cultivos de caña de azúcar	Plásmido	~ 1,500,000	5-8	7.8	Utilización de hidroxibutirato, enzimas lipolíticas y antiportadores	4- 1999
Suelo no cultivado	BAC	3,648;24,576	27;44.5	1.19	Antimicrobianos, resistencia a antibióticos; ARNr 16S, varios biocatalizadores	2000
No especificado	No especificado	---	---	---	Antimicrobianos	2000
No especificado	Cósmido	700,000	No especificado	24.5	Antimicrobianos	2000
No especificado	Cósmido	No especificado	No especificado	---	Pigmentos	2001
Suelo no cultivado	BAC	12,000	37	0.42	Antimicrobianos	2001
No especificado	Cósmido	No especificado	No especificado	---	Acidos grasos enol esterres	2002
Suelo alcalino	Plásmido	100,000	8-12	1.0	Proteasas	2002
Pastizales	Fósmido	25,287	32.5-43.5	0.90	ARNr 16S	2002
Pastizales	Fósmido	55,680	32.5-43.5	2.12	ARNr 16S de Acidobacterias	2003
Campo cultivable	Cósmido	5,000	No especificado	0.18*	Policetido sintasas, actividades diversas	2003
Pradera, cultivos de caña de azúcar, suelo cultivable	Plásmido	583,000 360,000 324,000	4.4;3.8;3.5	4.05	Formación de carbonilos	2003
Suelo arenoso y arcilloso	Fósmido	100,000	30-40	3.50	Policetido sintasa	2004
Suelo boscoso	Fósmido	33,700	35	1.18	Enzimas lipolíticas	2004
No especificado	Cósmido	No especificado	No especificado	---	N-aciltirosinas de cadena larga	2004
Planicie arcillosa con limo	Plásmido	200,000; 58,000; 250,000; 650,000	4.1; 2.7; 3.5; 3.4	4.2	Resistencia a antibióticos	2004
Superficie cubierta de musgo	Plásmido	30,000	3.5	0.11	Enzimas amilolíticas	2004
Campo agrícola	Plásmido	80,000	5.2	0.42	Amidasas	2004

2.7 Aportaciones del análisis de metagenomas

El análisis funcional de bancos de ADN metagenómico ha permitido la identificación y obtención de una gran diversidad de productos y actividades de interés biotecnológico, entre los que se encuentran: antibióticos nuevos (turbomicinaB), lipasas y esterasas (capaces de sintetizar precursores de importancia farmacéutica y agroquímica), quitinasa, proteínas de membrana, genes involucrados en la síntesis de compuestos como el poli-hidroxibutirato (polímero biodegradable), genes que codifican la ruta biosintética de vitaminas, el descubrimiento de nuevas moléculas transductoras de energía, entre otros productos. (Rondon, *et al.* 2000; Handelsman, *et al.* 2002; Gillespie, *et al.* 2003; Schloss, *et al.* 2003; Voget, *et al.* 2003; Brennan, *et al.* 2004; Keller y Zengler, 2004; Venter, *et al.* 2004). En la figura 8 se puede ver la diversidad estructural de los metabolitos aislados mediante las aproximaciones metagenómicas, los cuales poseen actividades como antibióticos contra diferentes microorganismos según estudios reportados (Handelsman, 2004).

En un estudio realizado por Smit y colaboradores en 2001, revela que la diversidad que puede encontrarse en bacterias cultivables es menor que la que se encuentra por métodos moleculares no dependientes de cultivo. Los análisis comparativos de metagenomas en el contexto de referencias disponibles de genomas aislados pueden potencialmente revelar patrones a gran escala de interacciones bioquímicas y correlaciones hábitat-específicos en el ambiente del hospedero que pueden perderse de otra manera. Estudios de microbiomas ambientales, como las biocapas de los drenajes de minas ácidas y muestras del mar de los Sargazos así como estudios de los microbiomas humanos, tales como la microbiota intestinal, son ejemplos de un área en rápida expansión de las aplicaciones del análisis de metagenomas (Markowitz, *et al.* 2006).

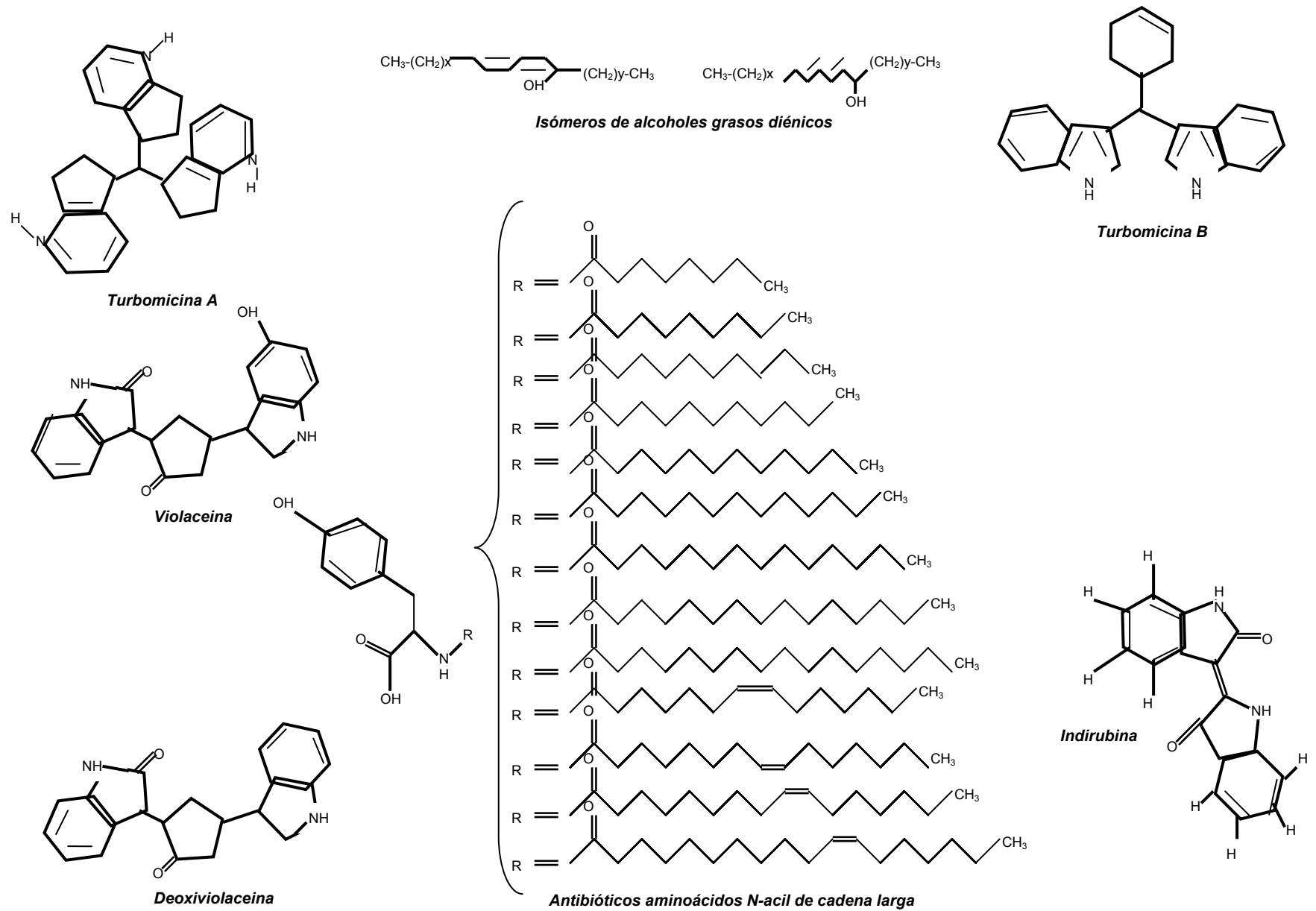


Figura 8. Diferentes antibióticos aislados a partir de bancos metagenómicos (Handelsman, 2004).

3. Justificación, hipótesis y objetivos

Justificación

La creación de bancos metagenómicos permite acceder a la diversidad genómica de microorganismos de un ambiente determinado (en este caso del suelo), incluyendo aquellos que no son cultivados. Dentro de las líneas de investigación que se siguen en nuestro laboratorio, está la de sobreproducir metabolitos aromáticos. Debido a lo anterior, la creación y caracterización de un banco metagenómico, podrá contribuir en el avance del conocimiento de la diversidad microbiana del suelo y por lo tanto será de utilidad en la búsqueda de nuevos genes de interés para los fines que son necesarios en ésta área de trabajo.

Hipótesis

El análisis metagenómico de una muestra de suelo, permitirá abundar en el conocimiento de la diversidad microbiana y metabólica de microorganismos cultivados y no cultivados.

Objetivo general

Generar un banco metagenómico del suelo y analizar la diversidad bacteriana genética y metabólica presente.

Objetivos particulares

- Extraer ADN metagenómico a partir de una muestra de suelo.
- Analizar la diversidad bacteriana presente por secuenciación de ADNr 16S.
- Construir un banco metagenómico de una muestra de suelo.
- Analizar la diversidad genética y metabólica del banco.

4. Materiales y métodos

En la tabla 3 se enlistan los plásmidos y la cepa utilizada en este trabajo.

Tabla 3 Cepas y plásmidos utilizados en este trabajo.

Cepa	Características	Referencia
<i>E.coli</i> TOP10	<i>recA</i> ⁻	Invitrogen
Plásmido		
pJET1.2/Blunt	Permite clonar productos Blunt o con tratamiento especial puede generar extremos adhesivos,	Fermentas
PTOPO Blunt	Permite la clonación de productos de PCR con la selección directa de recombinantes por la interrupción del gen letal <i>ccdB</i> . Tiene el gen <i>kan</i> (Km ^R)	Invitrogen

4.1 Toma de muestra para extraer ADN metagenómico de suelo

El corredor biológico Chichinautzin presenta una variedad de unidades edafológicas a lo largo y ancho del área que cubre. Principalmente en la zona donde fue tomada la muestra se presume que prevalecen el andosol y litosol, derivados de cenizas volcánicas y que poseen alto contenido de humedad (INEGI, 1978). El corredor representa varios tipos de vegetación tales como los bosques de pino con una altura promedio del dosel de 20-30 m, bosques de oyamel con alturas promedio del dosel de 30-40m, los bosques de encino con una altura promedio de 20m, y los pastizales con una altura promedio de 50 cm, entre otros tipos de vegetación.

El corredor biológico es un área de gran biodiversidad, así como también un centro de endemismos, existen especies como el conejo zacatuche, el ajolote de Zempoala, ratón de los volcanes, gallina de monte, el gorrión serrano zacatero, en peligro de extinción y la víbora de cascabel, alrededor de 132 especies de mariposas y 47 especies de libélulas.

Se tomó una muestra de alrededor de 500g de suelo del corredor biológico Chichinautzin, en un punto aledaño al Instituto de Biotecnología de la UNAM donde hay escasa perturbación humana con las siguientes coordenadas

18°54'04.89"N, 99°13'57.07"O para extraer ADN metagenómico, la muestra se extrajo a una profundidad no mayor a 10 cm con una pala de jardín y se almacenó a 4°C.

4.2 Extracción de ADN metagenómico

Para extraer el ADN metagenómico de una muestra de suelo se utilizaron dos métodos distintos:

1) Método de lisis directa: Se utilizó un método basado en extracción directa (Rondon, *et al.* 2000). Se resuspendieron 5 g de suelo en 10 mL de buffer de extracción a pH 8.05 (Tris HCl 100mM, EDTA 100mM, K₂HPO₄ 100mM, NaCl 15M). Para homogeneizar la muestra, se incubó a TA (Temperatura ambiente) durante 2 horas a máxima velocidad de agitación en plataforma. Se centrifugó a 1000xg por 10 min a TA en una centrifuga Sorvall Mach 1.6, se recuperó el sobrenadante y se le adicionó 100 µL de lisozima para romper la pared celular. Se incubó 2 horas a 56°C, se adicionó un volumen de cloroformo y se agitó 30 seg. en vórtex a máxima velocidad, se centrifugó a 5000 rpm durante 5 min. para recuperar la fase acuosa. Al sobrenadante se le adicionó un volumen de isopropanol, se incubó 1 hora a TA y se centrifugó a 12000xg por 20 min se recuperó la pastilla y se lavó con un volumen de etanol frío al 70%, se centrifugó a 12000xg por 20 min y se resuspendió en 500 µL de agua estéril.

El ADN que se obtiene directamente de la extracción contiene gran cantidad de compuestos húmicos que lo contaminan, la separación de estos compuestos se hizo por electroforesis en gel de agarosa al 1.2%, en una cámara horizontal Modelo #MGU-200T de C.B.S. Scientific Co. a 20V por 12 horas, se cortó la fracción de la agarosa donde se encontraba el ADN cromosomal de suelo (banda) y se purificó a partir del gel con el Kit Rapid Gel Extraction System de Marligen, una vez limpio se concentró con el Kit DNA Clean & Concentrator de Zymo Research.

2) Por Kit de extracción: Se utilizó el Power Soil DNA Isolation Kit de laboratorios Mobio siguiendo las especificaciones del fabricante, donde por cada columna de extracción se utilizan 0.25 g de suelo y se obtiene ADN de buena calidad (íntegro y sin rastro de degradación), el kit permite aislar ADN a partir de una muestra de suelo cualquiera, contiene agentes que permiten separar los compuestos húmicos y de esta manera obtener ADN perfectamente limpio (www.mobio.com).

El kit se basa en una homogenización rápida de la muestra de suelo mediante la agitación en tubos que contienen partículas para mejor mezclado. La lisis celular se lleva a cabo física y químicamente. El ADN se captura en una membrana de sílica que se encuentra en una columna de la cual será recuperado. En la tabla 4 se enlistan los pasos de la extracción de ADN del método directo e indirecto y sus diferencias entre ellas.

Tabla 4 Tabla comparativa de los métodos de extracción de ADN.

Proceso	Extracción directa	Característica	Extracción por kit	Característica
Homogenización	Por agitación	Lenta	Por agitación	Rápida
Lisis	Física y enzimática	Menos vigorosa	Física y química	Más vigorosa
Extracción	Con Cloroformo Por centrifugación	Menos eficiente	Por inmovilización en una membrana de sílica.	Más eficiente
Recuperación del ADN	Precipitación con isopropanol	Se obtiene ADN contaminado con compuestos húmicos	Elusión del ADN de una membrana de sílica	Se obtiene ADN limpio
Purificación	Por gel de agarosa		No es necesaria	
rendimiento	Bajo		Alto	

4.3 Amplificación de los ADNr 16S por PCR

El ADN metagenómico extraído mediante la técnica de lisis directa se utilizó como templado para amplificar por PCR los ADNr 16S bacterianos presentes en la muestra en un termociclador (Perkin Elmer modelo 9600). Para este fin se utilizaron primers universales para ADNr 16S de eubacterias (Fw Eu530: 5'-TGACTGACTGAGTGCCAGCAGCCGCGG-3' y R Eu1440: 5'-TGACTGACTGAGGCTACCTTGTTACGACTT-3') (Borneman, *et al.* 1996), los cuales amplifican un fragmento de aproximadamente 1000 pb a partir de la base 500 de la secuencia de ADNr 16S. Las condiciones de amplificación fueron las siguientes: 1 ciclo de desnaturalización a 95°C durante 10 min., 30 ciclos de amplificación: 95°C durante 1 min., 42°C durante 1 min. y 72°C durante 1.5 min. y un ciclo de elongación de 72°C durante 5 min. Se verificó la presencia del producto esperado en un gel de agarosa 1.2%. En la tabla 5 se muestra la mezcla de reacción utilizada y cada uno de los componentes para las PCR que se llevaron a un volumen final de 50µL.

Tabla 5. Mezcla de reacción de la PCR para amplificar ADNr 16S

ADN metagenómico de suelo 40µg/µL	0.8 µg
dNTP's 10mM	0.2 mM
Primer Rv Eu1440	0.5 µM
Primer Fw Eu530	0.5 µM
Buffer TermoPol	1 X
Vent DNA Polimerasa (Fermentas)	1 UI
Agua MQ	41.5µL

4.4 Clonación del producto de PCR

El producto de PCR obtenido se concentró y purificó con el Pure Link PCR purification kit de Invitrogen. Una vez purificado, se tomó una alícuota y se clonó con el kit TOPO Blunt de Invitrogen para clonación de productos de PCR con las especificaciones del fabricante. Se transformaron en 50 µL de células electrocompetentes de la cepa *E. coli* TOP10 incluidas en el kit. Las células se

recuperaron durante 1 hora/37°C en medio SOC, se sembraron en cajas con medio LB Km⁵⁰ y se incubaron 12 horas a 37°C.

4.5 Identificación de clonas positivas con ADNr 16S

Con la finalidad de identificar 100 clonas positivas, fue necesario crecer en medio LB Km⁵⁰ 312 colonias tomadas al azar y se realizaron extracciones de plásmido por lisis alcalina. El plásmido obtenido fue digerido con la enzima *EcoRI* durante 2 horas a 37°C. Para verificar el tamaño del inserto se observaron las digestiones en gel de agarosa al 1.2%.

A partir de los plásmidos obtenidos que se usaron como templado, se amplificó nuevamente por PCR al ADNr 16S y se verificó el producto en gel de agarosa al 1.2%. Para conocer la diversidad presente los ADNr 16S amplificados, estos se sometieron a un ARDRA (Amplified rDNA Restriction Analysis) con la enzima *HaeIII*, las reacciones se incubaron durante 2 horas a 37°C. Para observar los patrones de restricción se realizó una electroforesis de las muestras en un gel de agarosa al 2.5%.

4.6 Análisis de la diversidad de secuencias ADNr 16S presentes

Secuenciación de las clonas únicas

Las clonas que presentaron un patrón de restricción único se secuenciaron por el método Taq FS Dye Terminator Cycle Sequencing Fluorescence-Based Sequencing. Perkin Elmer/Applied Biosystems Modelo 3730, en la unidad de secuencia del Instituto de Biotecnología de la UNAM. Para secuenciar se utilizó el producto de PCR a concentraciones entre 500 y 700 ng, el primer utilizado fue Fw Eu530, a una concentración de 10 pmol para cada clona.

Búsqueda de identidad de las secuencias de ADNr 16S

Las secuencias obtenidas tuvieron un tamaño promedio de 750 nucleótidos, aquellas que pertenecen a las clonas únicas, se compararon con las depositadas en la base de datos de secuencias de nucleótidos no redundante del GeneBank

mediante el programa Blast, utilizando las herramientas en línea de la página del National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) para conocer su identidad.

Se realizó un alineamiento múltiple con cada una de las secuencias únicas obtenidas del metagenoma de suelo con secuencias depositadas en el GeneBank, elegidas al azar y las que resultaron con alta identidad en nuestro estudio para dar soporte al análisis. Esto se llevó a cabo con las herramientas de Clustal W (Thompson, *et al.* 1994), a partir del programa BioEdit. 1) se introdujeron las secuencias al Programa BioEdit) se hizo un alineamiento de todas las secuencias en ClustalW, 3) las secuencias alineadas se utilizan para construir un árbol en el paquete PHYLIP.

Para construir un árbol filogenético con las secuencias antes mencionadas, se utilizó el paquete computacional para análisis filogenéticos PHYLIP 3.67 (Phylogenetic Interference Package, Felsenstein 1989). Incluyó las siguientes operaciones 1) BOOTSTRAP con 100 réplicas, 2) se obtuvo una matriz de distancia en DNADIST y 3) la construcción de un árbol filogenético con el programa NEIGHBOR. El árbol que se obtuvo se enraizó a una Arqueobacteria (*Sulfolobus acidocaldarius*) y fue visualizado con el programa MEGA.

4.7 Construcción del banco metagenómico

Para la construcción del banco se utilizó el ADN extraído mediante kit, en primer lugar se realizaron digestiones parciales del ADN metagenómico con la enzima *Sma*I para obtener fragmentos flanqueados por extremos rasos y de un tamaño deseado para clonar en el vector de elección. En la tabla 6 se muestran las mezclas de reacción de las digestiones. La reacción se incubó a 37°C durante 15 min. y se inactivó 20 min. a 60°C. La enzima *Sma*I produce extremos rasos que son necesarios para ligar en el vector de elección. La reacción fue purificada de la siguiente forma: se agregó un volumen de la mezcla de cloroformo:alcohol isoamílico (24:1), recuperándose la fase acuosa, la cual se precipitó con acetato

de sodio y se lavo con etanol. Finalmente se resuspendió en 50 μL de agua estéril y se verificó la concentración en gel de agarosa 1.2%, donde se obtuvo que una concentración final del ADN de 23.6 ng/ μL .

Tabla 6. Reacción de digestión parcial de ADN metagenómico de suelo

ADN metagenómico de suelo	76 μL
Buffer 4 Biolabs 10X	10 μL
<i>Sma</i> I (Biolabs)	10 μL
Agua estéril	4 μL
Volumen final	100 μL

Para la construcción del banco metagenómico se probaron dos condiciones para verificar cual de ellas era la mejor, que lograra capturar la mayor cantidad de ADN y generar el mayor número de colonias. Los fragmentos de ADN obtenidos de la digestión fueron de 10 Kb y 5 Kb. Se eligieron de este tamaño debido a que de esta manera se aumenta la probabilidad de clonar genes completos e incluso operones. La reacción de ligación del inserto y el plásmido pJET1.2 de Fermentas se realizó como se muestra en la tabla 7, la relación de inserto - vector es de 3:1 y se incubó a 22° durante 30 min, como especifica el fabricante (www.fermentas.com/reviewer).

Tabla 7. Mezcla de reacción de ligación de ADN metagenómico de suelo y el vector de clonación de productos de PCR p JET1.2

ADN digerido	8 μL
Buffer 5X	4 μL
pJET 1.2	1 μL
Ligasa	1 μL
Agua estéril	6 μL
Volumen final	20 μL

Las ligaciones fueron purificadas por medio del Pure Link PCR purification kit de Invitrogen, se concentró por evaporación y resuspendió en 20 μL de agua estéril. Se transformaron 2.5 μL de la ligación por cada tubo de células

electrocompetentes de la cepa *E. coli* TOP10 por electroporación, estas células se recuperaron en 250 μL de medio rico SOC durante 1 hora a 300 rpm y 37°C.

Las colonias resultantes fueron evaluadas para determinar la presencia de insertos por digestión de los plásmidos extraídos por lisis alcalina. Finalmente, el total de la ligación (20 μL) fue utilizado para inocular un cultivo de 100 mL en medio LB Ap²⁰⁰ para realizar una extracción masiva de plásmido con el inserto metagenómico y mantener el banco en glicerol a -70°C y mantener así el banco denominado pJET/ MS (plásmido/Metagenoma de Suelo).

4.8 Análisis y caracterización del banco metagenómico

De las colonias crecidas se recuperaron 10 al azar para la extracción de plásmido verificando la presencia y tamaño de los insertos con digestiones con la enzima *BglII* como se muestra en la tabla 8. Estas reacciones se incubaron 2 horas a 37°C, y se analizaron en gel de agarosa al 1%. Algunas reacciones de digestión presentaron más de una banda correspondiente al inserto. El tamaño promedio del inserto se calculó haciendo la suma de los insertos para cada una de las colonias.

Tabla 8. Mezcla de reacción de digestión del pJET1.2/ MSCL 1-10 (Metagenoma de suelo Clona 1-10)

pJET1.2/ MSCL 1-10	5 μL
Buffer 3 Biolabs	2 μL
<i>BglII</i> Fermentas	0.5 μL
Agua estéril	12.5 μL
Volumen final	20 μL

Las 10 clonas fueron digeridas con la enzima *HaeIII* para determinar aquellas que presentaran patrones únicos de restricción.

Secuenciación de clonas únicas:

Para secuenciar los insertos se utilizó la estrategia de *primer walking*, que consiste en secuenciar un fragmento inicial utilizando primers correspondientes a una

secuencia conocida p.ej. los extremos del vector. Al obtener la secuencia de éste se diseña un segundo primer al tomar como referencia la última sección de la primer secuencia obtenida den ambos sentidos (Fw y Rv) y así sucesivamente hasta completar el inserto (Shyamala y Ames, 1989). En la figura 9 se muestra esquemáticamente la estrategia de secuenciación conocida como primer walking.

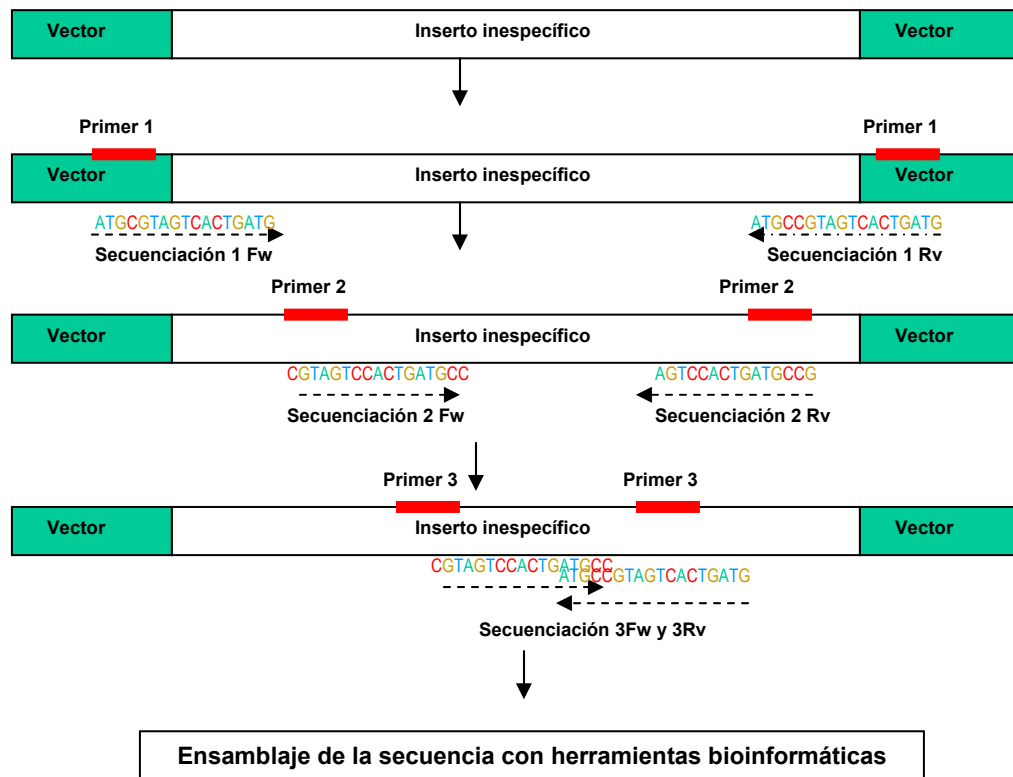


Figura 9. Estrategia de secuenciación para insertos o fragmentos inespecíficos, primer walking.

Análisis de secuencias

El análisis de secuencias se realizó utilizando las herramientas bioinformáticas incluidas en el programa BioEdit: 1) Se introdujeron las secuencias en el programa, 2) se obtuvo la secuencia complementaria reversa y fue la secuencia que se utilizó en el programa BLAST del NCBI para conocer la identidad de las clonas obtenidas en el banco.

5 Resultados

5.1 Extracción del ADN metagenómico

El proceso de extracción y purificación por lisis directa permitió obtener ADN que se encuentra en solución contaminado con compuestos húmicos (Figura 10). Después de la eliminación de éstos compuestos, se concentró el ADN, obteniéndose únicamente 6.4 μg a partir de 5 g de suelo (rendimiento de 1.28 μg de ADN / gr suelo). (Figura 11).

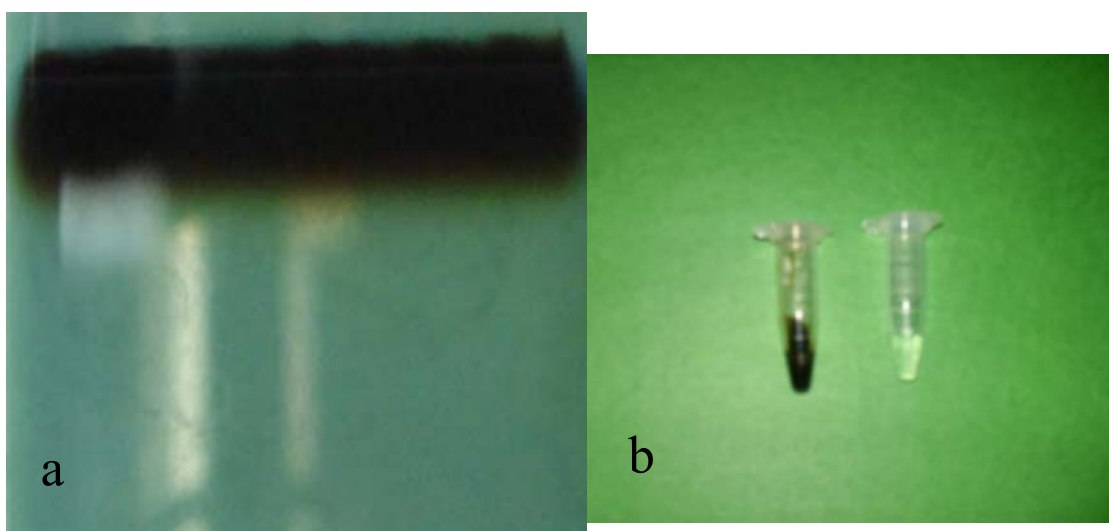


Figura 10. Imagen del ADN extraído. a) Muestra la forma en que se ve el ADN en solución con los compuestos húmicos en el gel de agarosa, b) El tubo oscuro contiene la mezcla de ADN con compuestos húmicos (así se obtiene de la extracción directa) y el tubo claro contiene al ADN purificado de compuestos húmicos mediante electroforesis en gel de agarosa al 0.5%.

Utilizando el kit Power Soil DNA isolation también se obtuvo ADN de buena calidad y en buena concentración como se muestra en la figura 10. A partir de 0.25 gramos de suelo se obtuvieron 1.5 μg de ADN. El análisis por electroforesis en gel de agarosa de ambas muestras de ADN permitió observar una buena calidad (no degradado).

La extracción con el Power Soil DNA isolation kit es muy rápida, característica que puede considerarse como una gran ventaja. Además, el ADN se extrae de una

cantidad muy pequeña de suelo lo cual podría ser una ventaja así como una desventaja, esto se discutirá más adelante.

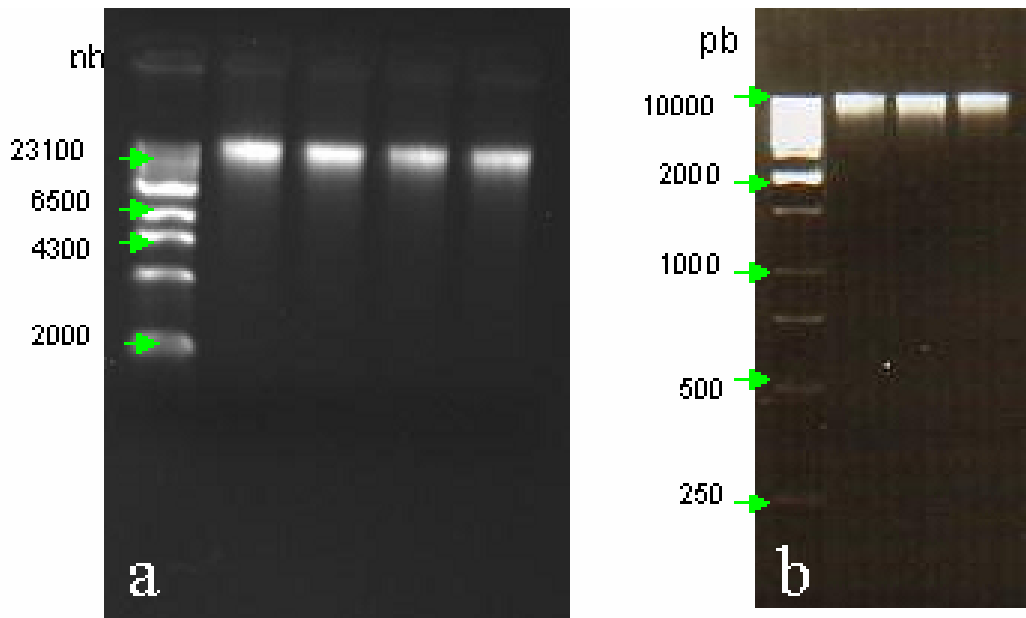


Figura 11. Electroforésis del ADN metagenómico extraído. a) Mediante método manual. b) ADN obtenido mediante el kit de extracción

5.2 Análisis de diversidad de secuencias de ADNr 16S

La amplificación de los ADNr 16S a partir del ADN metagenómico de suelo extraído por el método de lisis directa, permitió la obtención de un producto de PCR de aproximadamente 1000 pb, correspondiente al tamaño esperado para el producto con estos primers. Después de la clonación de los ADNr 16S amplificados y de transformar a la cepa de *E. coli* TOP 10 se obtuvieron 826 colonias, a partir de las cuales se analizaron por extracciones de plásmido por lisis alcalina 325 colonias al azar que fueron necesarias para obtener un total de 100 clonas positivas, lo que significa que solo la tercera parte del total poseen el inserto (indicador de la eficiencia del kit de clonación para productos de PCR TOPO Blunt).

De las 100 colonias obtenidas y sometidas al análisis por ARDRA, se obtuvieron 52 patrones de restricción únicos (algunos de los cuales se muestran en la figura

12), lo cual hace suponer que existe una diversidad bacteriana importante en la muestra de suelo analizada.

La secuenciación del ADNr 16S se realizó con el primer Fw Eu530 iniciando por el extremo 5' y se obtuvieron secuencias de alrededor de 750 pares de bases excluyendo la fracción que corresponde a la región hipervariable y por lo tanto se obtuvo una secuencia de fragmentos conservados (Borneman, *et al.*, 1996).

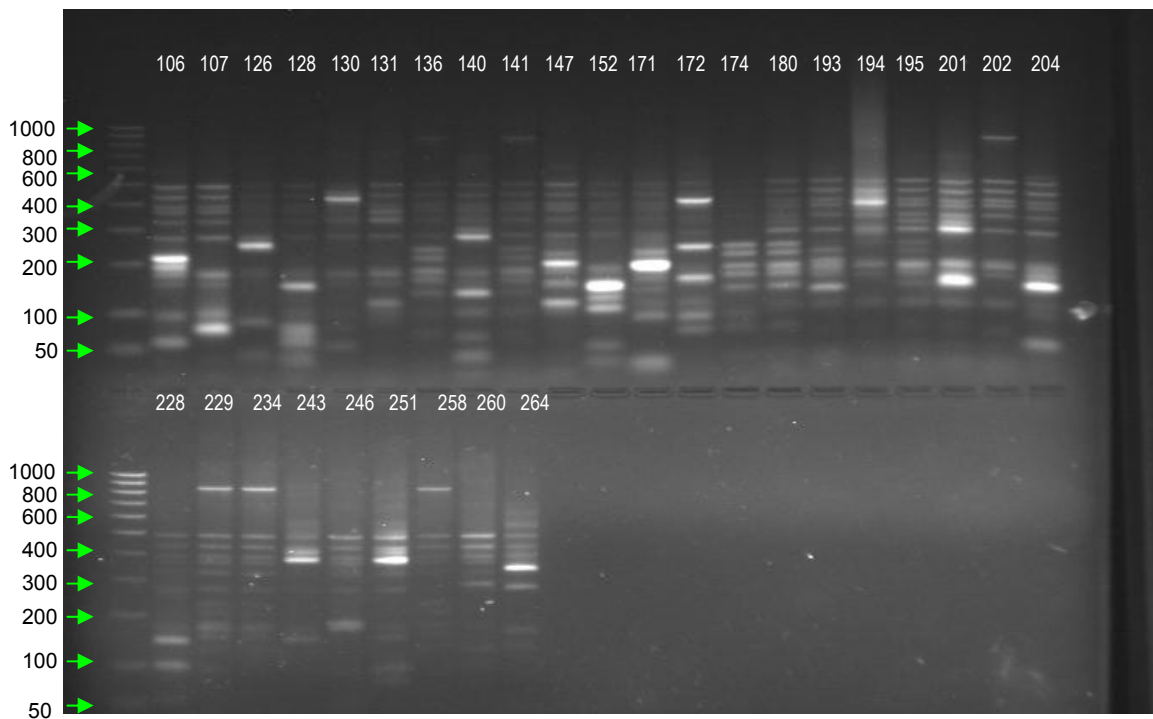


Figura 12, Se muestran algunas de las digestiones que producen diferentes patrones de restricción, para esta electroforesis se utilizó un marcador ladder 100 pares de bases de fermentas

En la gráfica de la figura 13 se muestra la composición genotípica de la muestra de suelo analizada con base en el análisis de ARDRA. Para construir esta gráfica de abundancia genética del análisis por ARDRA los datos que se registraron son los siguientes: la clona *E. coli* MM.1.15 corresponde al 11%, la clona BNC X, 57%, la clona *Escherichia coli* BE33, *Methylobacterium*, *Acidovorax*, *Leuconostoc pseudomesenteroides* y *Sphingomonadaceae* 1% respectivamente. Las clonas sin

similitud significativa representan 17%, las identificadas como Proteobacterias no cultivadas 5%, la clona *Adhaeribacter aquaticus* 3% y Acidobacteria 2%.

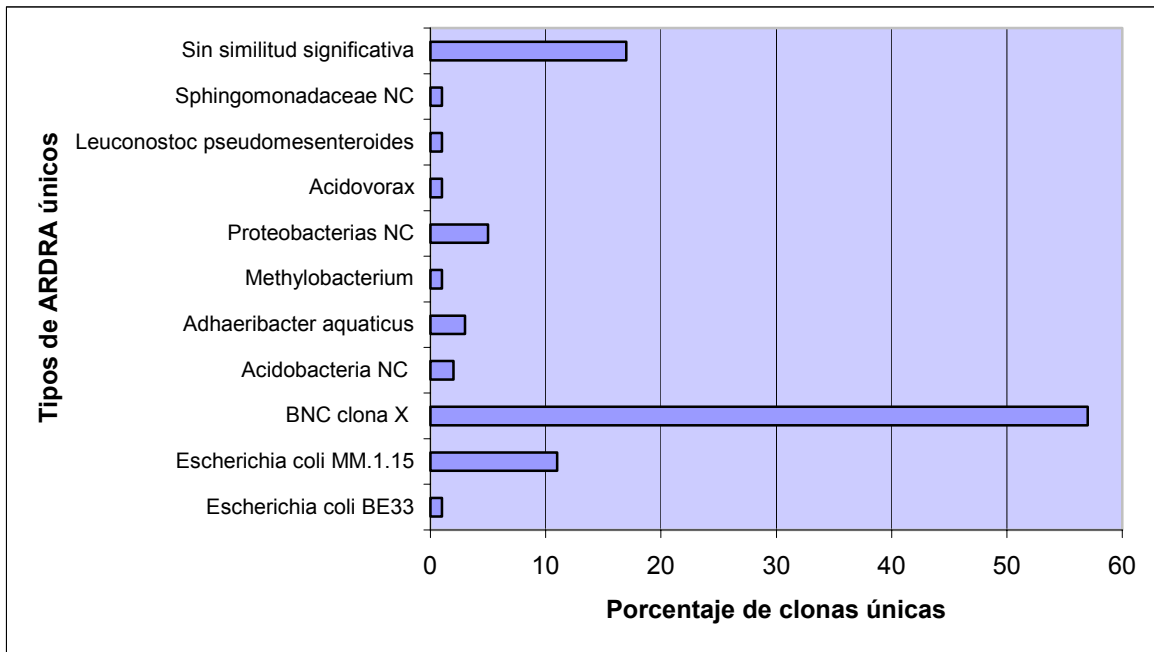


Figura 13. Gráfica de abundancia genética de los análisis ARDRA. Aquellas sin similitud significativa son las que en el estudio se identificaron con los números 33, 48, 106, 140, 152, 171, 228, 240, 243, 284 y 296, las de la familia Sphingomonadaceae y es un organismo no cultivado que se identificó con el número de clona 282, la clona identificada con *Leuconostoc pseudomesenteroides* corresponde al número 254, *Acidovorax* se identificó con la clona 128, las Proteobacterias no cultivadas corresponden a las clonas 280, 290, 291, un tipo único que se identificó con *Methylobacterium* es 301, *Adhaeribacter aquaticus* se identificó con tipo ARDRA único 297 que está representado 3 veces en la población, las Acidobacterias se identificaron como clonas 39 y 281, las bacterias no cultivadas (BNC) pertenecen a las clonas 2, 35, 68, 72, 74, 86, 93, 172, 180, 201, 202, 264, 279, 283, 285, 286, 287, 288, 289, 293, 294, 295, 298, 299, 300 y 302, de los cuales algunos están representados en más de una ocasión. *Escherichia coli* MM.1.15 se identificó en cuatro ocasiones con los números 50, 100, 193 y 195 y la cepa *Escherichia coli* BE33 con el tipo único 30.

Identificación de las clonas únicas

De los 52 perfiles de clonas únicas que se analizaron en la base de datos de secuencias nucleotídicas no redundantes del NCBI, 26 secuencias presentaron el mayor porcentaje de similitud con microorganismos no cultivados, 6 tienen identidad con microorganismos que poseen representantes cultivados

(Sphingomonadaceae, α -Proteobacterias, β -Proteobacterias, Acidothermaceae y Acidobacteria), 5 secuencias que presentaban entre 98 y 99% de similitud con diversas cepas de *E. coli* (cepa MM.1.15 y *E. coli* BE 33), 1 coincidió en 93% con *Leuconostoc mesenteroides*, 1 de ellas coincidió en 97% con *Adhaeribacter aquaticus*, 1 de ellas tuvo similitud de 85% con *Methylobacterium* sp., 1 tuvo similitud de 93% con *Acidovorax*, para 11 tipos únicos ARDRA no se encontró similitud significativa, aunque algunos de ellos están representados más de una vez en la población. Cabe mencionar que este análisis se hizo utilizando un fragmento de los ADNr 16S de aproximadamente 750 bases iniciando a partir de la posición 500 del gen.

Tabla 9 Clonas obtenidas del metagenoma de suelo y las secuencias que presentaron mayor identidad depositadas en el GenBank (BNC: bacteria no cultivada, S: suelo, P: Proteobacteria).

Clona	Identidad	No. de clonas	% de similitud con la secuencia más relacionada en GenBank	No. de acceso GenBank
2	BNC clona FCPS411 gen ribosomal 16S	4	96%	EF516709.1
30	<i>Escherichia coli</i> cepa BE33 gen ribosomal 16S	1	57%	EF560782.1
33	No se encontró similitud significativa.	1		
35	BNC clona S22-33 gen ribosomal 16S	5	98%	EU669630.1
39	Acidobacteria NC clone F9AGEub99 subunidad pequeña del ARN ribosomal	1	98%	EF040888.2
48	No se encontró similitud significativa	2		
50	<i>Escherichia coli</i> cepa MM.1.15 gen ribosomal 16S, secuencia parcial	4	99%	EF550554.1
68	BNC clona RP_1aaa03h10 16S gen ribosomal 16S, secuencia parcial	2	99%	EU470687.1
72	BNC clona RPSD_1aaa01d10 gen ribosomal 16S, secuencia parcial	2	97%	EU472816.1
74	BNC clona Elev_16S_1270 gen ribosomal 16S, secuencia parcial	4	96%	EF019903.1
86	BNC clona FCPN449 gen ribosomal 16S, secuencia completa	3	89%	EF515878.1
93	BNC clona RP_1aaa03h10 gen ribosomal 16S, secuencia parcial	2	98%	EU470687.1
100	<i>Escherichia coli</i> cepaMM.1.15 gen ribosomal 16S, secuencia parcial	4	98%	EF550554.1
106	No se encontró similitud significativa	1		
128	Acidovorax sp. JS42, genoma completo	1	93%	CP000539.1
140	No se encontró similitud significativa	3		
152	No se encontró similitud significativa	1		
171	No se encontró similitud significativa	1		
172	BNC clona FCPO549 gen ribosomal 16S, secuencia completa	4	95%	EF516035.1
180	BNC clona AFEL3_aao14b07 gen ribosomal 16S, secuencia parcial	2	98%	EU465922.1
193	<i>Escherichia coli</i> cepa MM.1.15 gen ribosomal 16S, secuencia parcial	2	99%	EF550554.1
195	<i>Escherichia coli</i> cepa MM.1.15 gen ribosomal 16S, secuencia parcial	1	98%	EF550554.1
201	BNC clona Z009 gen ribosomal 16S, secuencia parcial	1	92%	EU151840.1
202	BNC clona RP_3aaa04f05 gen ribosomal 16S, secuencia parcial	2	99%	EU778511.1
228	No se encontró similitud significativa	4		
240	No se encontró similitud significativa	2		
243	No se encontró similitud significativa	3		
254	<i>Leuconostoc pseudomesenteroides</i> gen ARN r 16S	1	93%	AB326299.1
264	BNC clona 11A-10 gen ribosomal 16S, secuencia parcial	1	96%	DQ906875.1
279	BNC clona WHEATSIP-EL87 gen ribosomal 16S, secuencia parcial	3	98%	EF066600.1
280	Beta proteobacteria no cultivada clona OS-C14 gen ribosomal 16S, secuencia parcial	2	97%	EF612410.1
281	Acidothermaceae NC clona Amb_16S_1471 gen ribosomal 16S, secuencia parcial	1	99%	EF018957.1
282	Sphingomonadaceae NC clona Amb_16S_1187 gen ribosomal 16S, secuencia parcial	1	97%	EF018731.1
283	BNC clona FCPO549 gen ribosomal 16S, secuencia completa	4	98%	EF516035.1
284	No se encontró similitud significativa	1		
285	BNC gen ribosomal 16S, secuencia parcial, clona B1_36R	4	98%	AM157280.1
286	BNC clona DCN-I-22 gen ribosomal 16S, secuencia parcial	1	98%	DQ660846.1
287	BNC clona AKAU3874 gen ribosomal 16S, secuencia parcial	1	98%	DQ125760.1
288	BNC gen ribosomal 16S, clona JG35+U3-JT12	1	98%	AM084890.1
289	BNC clona FFCH572 gen ribosomal 16S, secuencia parcial	1	89%	EU132428.1
290	PBNC Elev_16S_1557 gen ribosomal 16S	1	88%	EF020136.1
291	APBNC clona BDC2_D06 gen ribosomal 16S, secuencia parcial	2	97%	AY690263.1
293	BNC clona FFCH4996 gen ribosomal 16S, secuencia parcial	2	97%	EU134062.1
294	BNC clona FFCH12324 gen ribosomal 16S, secuencia parcial	1	96%	EU133825.1
295	BNC clona 342B gen ribosomal 16S, secuencia parcial	1	94%	AY571850.1
296	No se encontró similitud significativa	1		
297	<i>Adhaeribacter aquaticus</i> gen ribosomal 16S, cepa tipo MBRG1.5	3	97%	AJ626894.2
298	BNC clona F4-04 gen ribosomal 16S, secuencia parcial	1	91%	EU025069.1
299	BNC clona FFCH10554 gen ribosomal 16S, secuencia parcial	1	99%	EU133813.1
300	BNC clona A15 gen ribosomal 16S, secuencia parcial	1	98%	EF589982.2
301	<i>Methylobacterium</i> sp. 4-46, genoma completo	1	85%	CP000943.1
302	BNC gen ribosomal 16S, secuencia parcial, clona: BS054	1	97%	AB240240.1

Cálculo de la diversidad de diversidad H' para la muestra de suelo

Los resultados de la identidad obtenida para las secuencias de las clonas analizadas, fueron utilizados para construir un árbol filogenético y poder observar esquemáticamente la distancia que existe entre nuestras secuencias y las de referencia del NCBI. La mayoría de secuencias de ADNr 16S obtenidas a partir de la muestra de suelo analizada 50% están relacionadas con el grupo de las Proteobacterias y comprende tanto a las α , β , γ , δ y ϵ , siendo mayoritario el grupo de las α ; 24% están agrupadas como bacterias no cultivadas, (BNC); 2%, fueron identificadas como Acidobacterias y una minoría mantienen parentesco con microorganismos Gram-positivos 9% (Figura 14).

A partir de los datos de identidad obtenidos para las clonas analizadas y con la finalidad de comparar la diversidad genotípica de la muestra analizada se determinó el índice de diversidad Shannon-Weaver (H') y el valor de equidad (Evenness, E). Existen diferentes índices de diversidad reportados en la literatura como el de Simpson D, la H de Brillouin de los cuales, uno de los más reportados es el índice H' (Borneman, *et al.* 1996; Escalante, *et al.* 2001; Martín, 2002; Baek, *et al.* 2007) El índice H' se calculó de acuerdo a la siguiente expresión:

$H' = -\sum (p_i) \times \ln(p_i)$ donde:

n_i = # de individuos de la especie.

N = # total de individuos de todas las especies.

S = # total de especies (riqueza).

$p_i = n_i/N$

El valor de E se calculó de acuerdo a la siguiente expresión:

$E = H' / H_{\max}$ donde:

$H_{\max} = \ln S$

El valor de H' (1.42) y el valor de E (0.59), se compararon con valores reportados para otros estudios de diversidad basados en el análisis de secuencias de ADNr

16S amplificadas a partir de muestras de suelos y/o sedimentos. El valor obtenido en este trabajo indica la diversidad detectada (Tabla 10).

Tabla 10. Índice de diversidad H' y E de la muestra de suelo analizada y comparación con otros valores reportados en la literatura.

Valor de H' (índice de Shannon Weaver)			
Aislado	Número	p_i	H
<i>Escherichia coli</i> BE33	1	0.01	-0.046052
<i>Escherichia coli</i> MM.1.15	11	0.11	-0.2428
BNC clona X	57	0.57	-0.320408
Acidobacteria NC	2	0.02	-0.07824
<i>Adhaeribacter aquaticus</i>	3	0.03	-0.105197
Methylobacterium	1	0.01	-0.046052
Proteobacterias NC	5	0.05	-0.149787
Acidovorax	1	0.01	-0.046052
<i>Leuconostoc pseudomesenteroides</i>	1	0.01	-0.046052
Sphingomonadaceae NC	1	0.01	-0.046052
Sin similitud significativa	17	0.17	-0.301233
Total	100		
Índice de Shannon (H') =			1.427923
Equidad (Evenes=E)			0.59549

Algunos valores de E y H' reportados			
Muestra analizada	H'	E	Referencia
Suelo agrícola para cultivo de trigo	1.34	0.98	Smit, <i>et al.</i> 2001.
Suelo de pastizales mejorados	2.04	NR	Martin, 2002.
Suelo de pastizales no mejorado	2.01	NR	Martin, 2002.

Figura 14 Árbol filogenético construido con el paquete PHYLIP 3.67 de las secuencias parciales de ADNr 16S de la muestra de suelo que se nombran únicamente por su número, las que poseen el número de acceso, son las secuencias de referencia del NCBI y las que se identifican por nombre fueron introducidas también como referencia para dar soporte a la topología del árbol. El árbol fue visualizado y editado con el programa MEGA. Las ramas que están marcadas en color azul pertenecen a microorganismos “identificados” como no cultivados, en café se muestran las clonas identificadas con *E. coli* y probables emparentadas entre ellas (que pertenecen también a las proteobacterias), en amarillo se muestran las proteobacterias, la que se encuentra en el extremo superior es *Sulfolobus acidocaldarius* una arqueobacteria utilizada para enraizar el árbol, en rojo se muestran las bacterias Gram positivas y las líneas punteadas pertenecen a las clonas que no poseen similitud significativa con las depositadas en el NCBI.

5.3 Construcción del banco metagenómico

Clonación del ADN metagenómico

Para la construcción del banco metagenómico se probaron dos condiciones para verificar cual era la mejor de ellas que lograra capturar la mayor cantidad de ADN.

1) Se partió de 505ng ADN de un tamaño de alrededor de 10Kb obteniéndose 85,552 Unidades Formadoras de Colonias (UFC)

2) Para ADN de tamaño aproximado de 5 Kb a partir de 252 ng se obtuvieron 14,496 UFC. Se decidió que la prueba 1 sea tomada como la mejor condición ya que se obtuvieron considerablemente más colonias, lo cual permite incrementar el tamaño del banco.

Para fines prácticos se consideró un tamaño promedio de 5 Kb (debido a la baja resolución de los geles de agarosa en el que se verificó el tamaño) para los insertos analizados.

1) por 2.5 μ L de reacción de ligación se obtuvieron 10 944 UFC

2) El volumen final de la reacción de ligación fue de 20 μ L (ya que se hicieron 8 ligaciones de 2.5 μ L)

3) Por 20 μ L de reacción de ligación se obtuvieron 87 552 UFC totales, si se considera un tamaño promedio de 5 Kb para el genoma de *E. coli*, entonces el banco obtenido equivale a:

5 Kb (tamaño promedio de inserto) X 87 552 (UFC totales) = 437 760 Kb

437 760 Kb totales/ 5000 Kb genoma promedio = 87.55 genomas

lo que equivale a 0.43 Gb de ADN.

5.4 Análisis y caracterización del banco metagenómico

En la figura 15 se presentan las digestiones correspondientes a las dos pruebas de clonación para la construcción del banco, éstas digestiones se realizaron con la enzima *Bgl*II para determinar la presencia de insertos en las clonas elegidas. Para hacer la caracterización, se eligió como un buen banco el construido con fragmentos de 10 Kb, ya que al poseer insertos de mayor tamaño la probabilidad de haber capturado genes o incluso pequeños operones o vías metabólicas pequeñas aumenta.

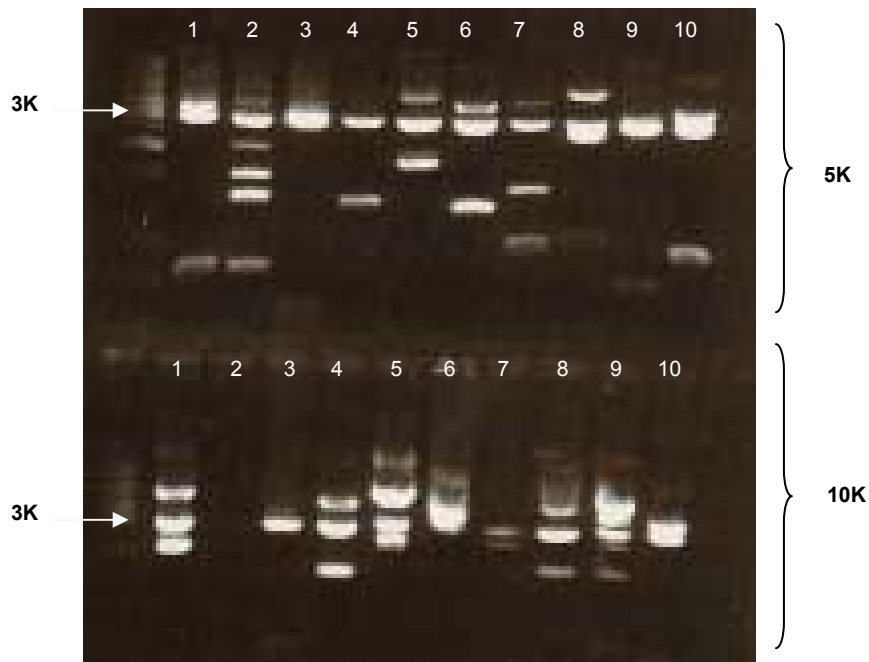


Figura 15 Digestiones de 10 clonas tomadas al azar de cada una de las transformaciones para determinar la eficiencia de ligación y transformación. El corchete que marca 5Kb corresponde a la

digestión de los inserto de ese tamaño, el que marca 10Kb corresponde a la digestión de insertos de ese tamaño y la marca de 3K en ambos casos indica el tamaño del vector.

El análisis de restricción de las minipreps de las clonas MSCL1-10 (Metagenoma de suelo clonas 1-10), permitió identificar patrones de restricción únicos (Figura 16).

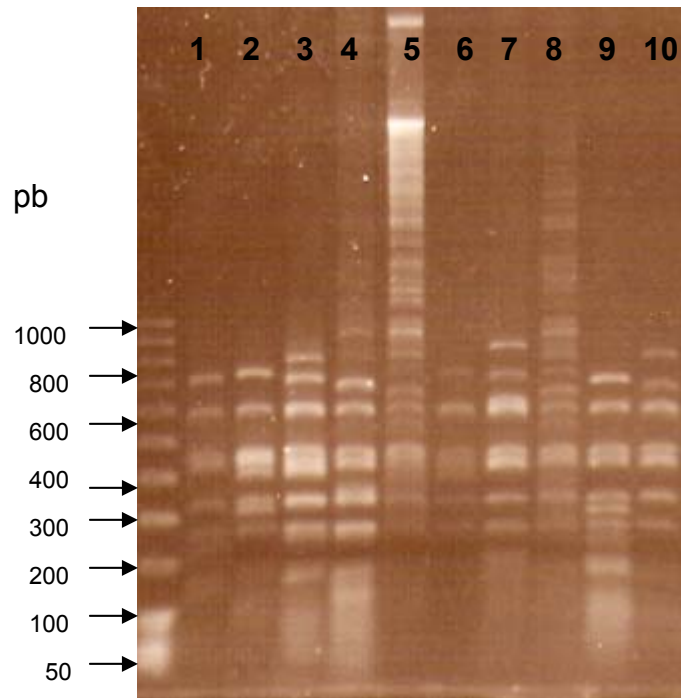


Figura 16. Análisis de restricción de las 10 colonias elegidas al azar, donde se muestran los patrones de restricción diferentes entre ellos.

Secuenciación de clonas únicas por el método de primer walking

De acuerdo al tamaño promedio estimado para las clonas seleccionadas para secuenciarlas según lo que se muestra en la figura 15 (alrededor de 2 Kb), la estrategia más adecuada para su secuenciación es el método de *primer walking*. Inicialmente se secuenciaron los extremos 5' y 3' de cada inserto metagenómico clonado utilizando los primers incluidos en el kit de clonación pJET1.2, que corresponden a 39 y 36 bases del sitio de clonación. A partir de las secuencias obtenidas se diseñan los siguientes primers para “caminar” sobre el inserto

metagenómico y así sucesivamente hasta completar la secuencia total. De este modo se secuenciaron 5 de las clonas metagenómicas analizadas y por restricción (Figura 16): MSCL1, MSCL2, MSCL4, MSCL5 y MSCL8. La secuencia completa fue entonces analizada en la página del NCBI para hacer un Blast X y de esta manera se encontró la proteína con la que nuestras secuencias problema mantienen la mayor identidad, también en el NCBI se hizo una búsqueda de ORF's (del Inglés Open Reading Frame), cada uno de ellos se utilizó para hacer Blast y buscar alguna otra proteína o gen de interés y finalmente se representó esquemáticamente solo las identidades mayores al 60%.

En la figura 17 están representadas las secuencias de las cinco clonas analizadas y en las que se muestra, el tamaño de la secuencia, los extremos flanqueados por un fragmento de vector, las puntas de flecha rojas indican cada uno de los ORF's encontrados y la dirección en que se traducen (la mayoría no se identificó con alguna proteína) y las flechas azules representan el fragmento de la secuencia que pertenece o tiene similitud con proteínas específicas como en el caso de la colonia MSCL1, donde se encontró un fragmento que pertenece a una posible proteína reguladora de *Blastopirellula marina* DSM 3645–planctomycetes- (número de acceso gb|EAQ79990.1) con 86% de similitud, para la clona MSCL2 que posee identidad con una Endonucleasa dependiente de ATP de *Comamonas testosteroni* con número de acceso gb|EAVI7825.1. 61% de identidad, en el caso de la clona MSCL4 y la mayor similitud la tiene con una proteína involucrada en la biosíntesis de riboflavina RibF de *Salinispora arenicola* CNS-205 -Actinobacteria- con número de acceso gb|ABV97238.1 con quien mantiene 88 % de identidad, para MSCL5 quien tiene identidad con una GTP difosfocinasa de *Salinispora tropica* CNB-440 con número de acceso gb|ABP55721.1 con 73 % de identidad y finalmente para MSCL8 se ve en la figura 21 la similitud de 76% con CHAP (dominio de GSP sintetasa) de *Sphingomonas* sp –Alphaproteobacteria- con número de acceso ZP_01302380.

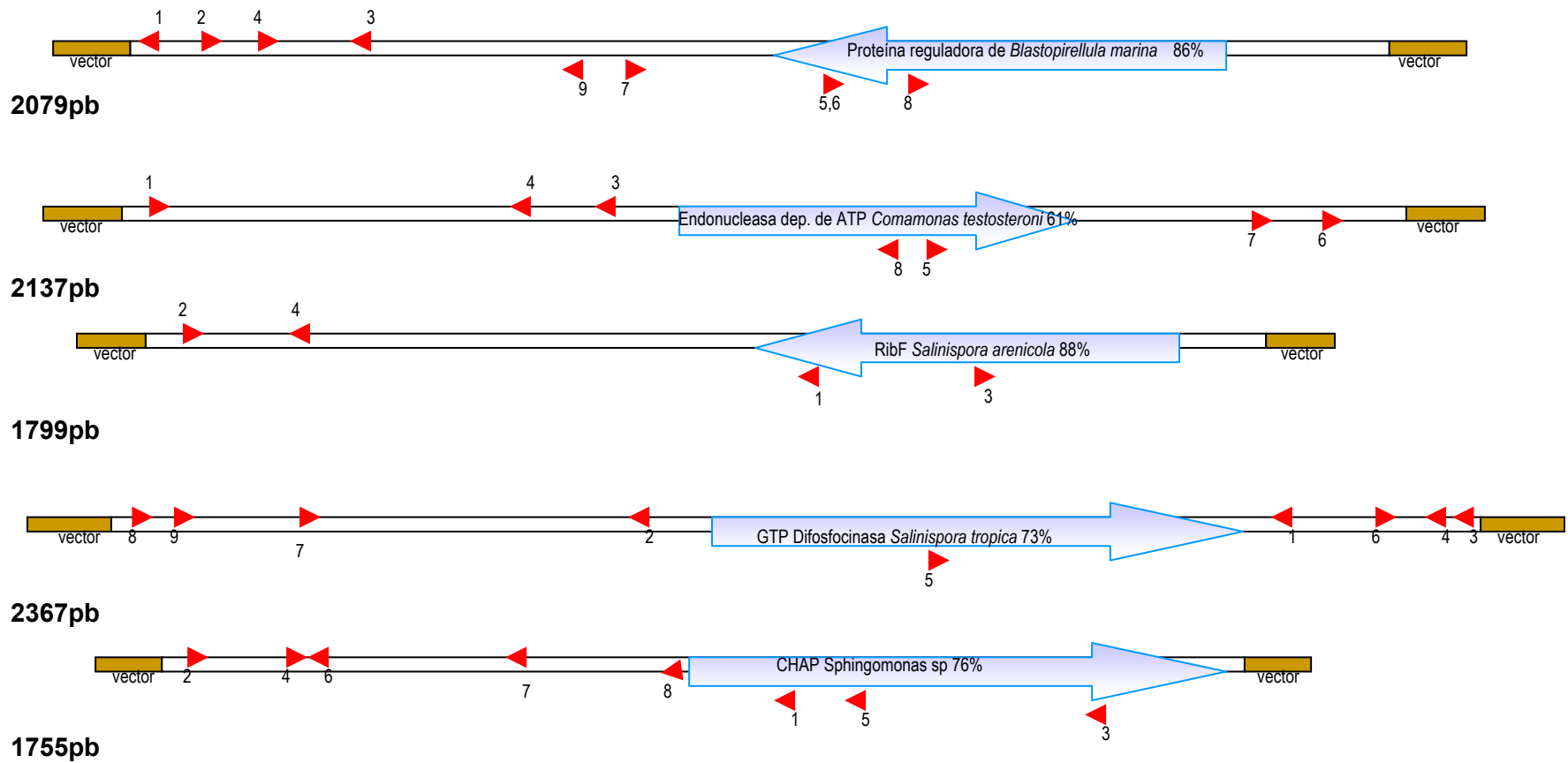


Figura 17. Mapa de las secuencias de los plásmidos analizados, los extremos en café representan un fragmento del vector, las flechas rojas indican diferentes ORF's y la dirección de transcripción, las flechas azules indican el fragmento que abarca la proteína encontrada, el porcentaje de identidad con la clona metagenómica y la dirección en que se traduce.

6. Discusión

Extracción del ADN metagenómico

Al utilizar ambos métodos de extracción (manual y por kit) se evaluó primero el manual y pudimos determinar que la desventaja que presenta este método, es que se necesita extraer el ADN partiendo de una gran cantidad de suelo, para obtener una concentración y volumen aceptable, ya que por cada paso va disminuyendo considerablemente la concentración. La eliminación de compuestos húmicos mediante electroforesis en gel de agarosa implica grandes pérdidas y no necesariamente garantiza la disponibilidad del ADN para usos subsecuentes.

La extracción por kit podría generar un gran sesgo ya que para una extracción se utiliza una cantidad tan pequeña como 0.25 gramos de suelo, esto podría influir en una pérdida de diversidad de especies de la cual se obtenga su ADN. La característica más notable de este método es la cantidad de ADN que se obtiene a partir de esa cantidad de suelo ya que se obtiene 4 veces más por gramo de suelo que en una extracción directa por lisis.

A partir de los resultados obtenidos de la clonación de ADNr 16S, indica que alrededor del 30% de las transformantes es exitoso (este fenómeno puede deberse a que el ADN aún contenga algunos elementos contaminantes que reduzcan considerablemente la eficiencia de ligación), se ha comprobado que en un proceso de extracción donde se involucre una ruptura mecánica de las células, la diversidad que se recupera es mayor en comparación con aquellas cuyo tratamiento es químico (Niemi, et al. 2001).

Análisis de diversidad de secuencias de ADNr 16S

La cuantificación de la riqueza de especies dentro de ambientes complejos es un paso importante en el conocimiento de la dinámica de la comunidad en condiciones naturales y las repercusiones de actividades antropogénicas sobre el funcionamiento del ecosistema (Youssef y Elshahed, 2008). La muestra de suelo

utilizada en este trabajo fue tomada de un área con casi nula actividad humana. Sin embargo, la riqueza de especies que se obtuvo o que se pudo identificar es relativamente escasa a pesar de que aparentemente los ARDRA reflejaban algo diferente, el 65% de los ADNr 16S obtenidos pertenece a microorganismos ambientales no cultivados, 17% no tuvo similitud con ningún organismo reportado en el GenBank, lo que deja un 82% de microorganismos desconocidos y solo 18 % de microorganismos cultivados y reportados como tal en la base de datos (como es el caso de *E. coli*, que es la más abundante).

Al calcular matemáticamente la diversidad presente en nuestra muestra de suelo el valor de H' fue de 1.427923, a medida que este valor se acerca a cero la diversidad presente es menor, y cuando es igual a cero significa que la población está conformada por únicamente una especie (Liu, *et al.* 2008). El valor máximo de H' solo se alcanza cuando las especies de una población están uniformemente distribuidas. Es decir, cuando el valor de E es igual a 1, para el caso de nuestra población este valor fue de 0.59549, lo que nos indica que la diversidad encontrada en esta muestra de suelo es mediana. Si comparamos con los valores obtenidos por Liu y colaboradores, quienes analizaron 4 muestras de suelo y encontraron que la mas alta diversidad era de $H'= 2.63$ para una muestra no contaminada y el menor valor que encontraron fue de $H' = 2.07$ para una muestra contaminada con TCE o tricloroetileno (28mg/Kg). Para otros estudios se han reportado valores de $H'=1.34$ y $E=0.98$ (Smit, *et al.* 2001), que utilizaron muestras de un suelo agrícola y Martin (2002), para suelo de pastizales mejorados y no mejorados reporta valores de $H'=2.04$ y $H'=2.01$. El valor de E es una medida de la uniformidad de la población y las poblaciones son mas uniformes a medida que este se acerca a 1, la muestra analizada en este estudio es medianamente uniforme, y lo reportado por Smit, *et al.* (2001), puede calificarse como una muestra altamente uniforme.

La diversidad y abundancia de los microorganismos también depende en gran parte del tipo de suelo del que se haya tomado la muestra y el método de

extracción utilizado. Los métodos de lisis directa dejan acceder a grandes poblaciones microbianas difíciles de cultivar, gracias a sus grandes rendimientos y los métodos de extracción indirecta de ADN son conocidos por acceder a una diversidad menos rica. Sin embargo, Stach y colaboradores en 2001 hicieron un estudio de diversos métodos de extracción y al compararlos concluyeron que altos rendimientos de ADN en la extracción y purificación, no necesariamente se correlaciona positivamente con la riqueza de especies (Rajendhran y Gunasekaran, 2008). La diversidad incrementa a medida que incrementan la riqueza y abundancia, en la mayoría de los casos la riqueza es alta, si la abundancia y riqueza son altas en un área o muestra determinadas (Borneman, *et al.* 1996).

La especie más abundante identificada en el estudio resultó ser *E. coli*, a pesar de que la muestra fue tomada en una zona sin actividad humana, esta abundancia puede deberse a que existen en esta área diferentes especies de mamíferos principalmente herbívoros y aves ya que se ha demostrado que diferentes especies de ungulados, primates, roedores y algunas especies de aves pueden tener en su tracto digestivo poblaciones de este microorganismo (Souza, *et al.*, 1999).

En el caso de nuestro análisis la baja incidencia de microorganismos puede ser debida al método de lisis utilizada en esa extracción, hay que recordar que para el análisis de diversidad de ADNr 16S se utilizó ADN extraído mediante un método manual de lisis directa. La región hipervariable del gen ADNr 16S está comprendida en el extremo 3' de la molécula (Borneman, *et al.* 1996) y muy seguramente fue excluida al hacer el proceso de secuenciación de esta, ya que esta se realizó utilizando el primer Fw Eu530 que amplifica inicialmente la región 5', por lo tanto los datos presentados fueron obtenidos con la región conservada. Existen estudios de diversidad bacteriana de suelo, donde el grupo predominante es el de Proteobacterias que concuerda con los resultados obtenidos.

Construcción del banco metagenómico

El tamaño de los insertos completos de las clonas secuenciadas, es menor al estimado casi por la mitad, y al utilizar el programa BLAST (Altschul, *et al.* 1997) en el NCBI parece haber una captura de solo un gen por cada clona. La elección del vector de clonación juega un papel crucial en la construcción de un banco, así como también la estrategia que se use para extracción y purificación de ADN ya que de ello depende la integridad y concentración que posea el ADN con el que se inicia. El tamaño del banco se encuentra dentro del rango, según el tipo de vector que se utilizó y el tipo de ambiente del cual se extrajo la muestra de ADN.

Caracterización del banco de ADN metagenómico

Para las cinco clonas que se analizaron donde se encontró una posible proteína reguladora, una endonucleasa dependiente de ATP, una proteína involucrada en la biosíntesis de riboflavina RibF, una GTP difosfocinasa y finalmente una CHAP (dominio de GTP sintetasa). Cada una pertenece a un microorganismo diferente y ninguno se encuentra representado en el estudio de ADNr 16S.

Se debe tener especial cuidado en la elección de la cepa hospedera ya que para la expresión de genes heterólogos es fundamental el fondo genético de la misma, de tal manera que para los fines de búsqueda de funciones específica es necesario el conocimiento de la genética de la cepa y la regulación de los genes en cuestión.

7. Conclusiones

- La diversidad bacteriana de la muestra de suelo analizada y determinada por secuenciación parcial de los ADNr 16S estuvo conformada por Proteobacterias, bacterias no cultivadas, algunas que no poseen similitud significativa con las secuencias depositadas en el NCBI, las menos representadas fueron las Gram positivas. El valor de H indica que esta diversidad es mediana si se compara con el valor de H'_{Max} .
- E indica que se trata de una población medianamente uniforme.
- El banco metagenómico que se obtuvo posee un tamaño equivalente a poco más de 87 genomas de *E. coli*. Se considera como un banco de buen tamaño respecto a lo reportado con anterioridad en la literatura (Daniel, 2005) y puede ser utilizado para la búsqueda de genes o funciones de interés.
- La cantidad de ADN clonado en este banco es de 0.43 Gb.
- La secuenciación de 5 clonas metagenómicas permitió identificar una posible proteína reguladora, una Endonucleasa dependiente de ATP, una proteína involucrada en la biosíntesis de riboflavina RibF, una GTP difosfocinasa y CHAP de Sphingomonas.

8. Perspectivas

- Realizar búsqueda de funciones específicas con fenotipos bacterianos particulares (p.ej. resistencia a antibióticos).
- Buscar funciones o genes determinados para hacer más dirigido el estudio hacia enzimas o metabolitos de interés (p. ej. DAHPS y otras involucradas en la biosíntesis de compuestos aromáticos).
- Ampliar las opciones de cepas con diferentes fondos genéticos para la construcción de bancos por ejemplo con bacterias Gram-positivas.

Referencias

- Baek, K., Yoon, B., Kim, B., Cho, D., Lee, I., Oh, H., and Kim, H. 2007, Monitoring of microbial diversity and activity during bioremediation of Crude Oil-Contaminated Soil with Different Treatments. *J. Microbiol. Biotechnol.* 17:67-73.
- Borneman, J., Skroch, P. W., O'Sullivan, K., Palus, J.A., Rumjanek, N.G., Jansen, J. L. and Triplett, E. W. 1996, Molecular microbial diversity of an agricultural soil in Wisconsin. *Appl. and Environ. Microbiol.* 62:1935-1943
- Boucher, Y., Nesbo, C.L., and Doolittle, W. F. 2001. Microbial genomes: dealing with diversity. *Curr. Opin. Microbiol.* 4:285-289.
- Brennan, Y., Callen, W., Christoffersen, L., Dupree, L., Goubet, F., Healey, S., Hernández, M., Keller, M., Li, K., Palackal, N., Sittenfeld, A., Tamayo, G., Wells, S., Hazlewood, G., Nathur, E., Short, J., Robertson, D., and Steer, D. 2004. Unusual microbial xylanases from insect guts. *Appl. Environ. Microbiol.* 70:3609-3617.
- Clarridge III, J. E. 2004, Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol. Rev.* 17:840-862.
- Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R and Wittwer, P. 2005. Metagenomic gene discovery: past, present and future. *Trends Biotechnol.* 23:321-329.
- Dahlöf, I., Baillie, H., and Kjelleberg, S. 2000, *rpoB*-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. *Appl. Environ. Microbiol.* 66:3376-3380.
- Daniel, R., 2005. The metagenomics of soil. *Nature.* 3:470-8.
- Eisen, J. A. 1995 The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of RecAs and 16S rRNAs from the same species. *J. Mol. Evol.* 41:1105-1123.
- Escalante A., Wachter C. Farrés A. 2001. Lactic acid bacterial diversity in the traditional Mexican fermented dough pozol as determined by 16S rDNA sequence analysis. *Int. J. Food Microbiol.* 64: 21-31.

- Esclante-Lozada, A., Gosset-Lagarda, G., Martínez-Jiménez, A., y Bolívar-Zapata, F. 2004. Diversidad bacteriana del suelo: métodos de estudio no dependientes del cultivo microbiano e implicaciones biotecnológicas. *Agrociencia*, 38:583-592.
- Felsenstein, J. 1989, PHYLIP- Phylogeny Interference Package. *Cladistics*. 5:164-166.
- Fulthorpe, R. R., Roesch, L. FW., Riva, A. and Triplett, E. W. 2008. Distantly sampled soils carry few species in common. *ISME J*. 2:901-910.
- Gabor, E., de Vries, E., and Janssen, D. 2004. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environ. Microbiol*. 6: 948-958.
- Gillespie, D., Brady, S., Bettermann, A., Cianciotto, N., Liles, M., Rondon, M., Clardy, J., Goodman, R., y Handelsman, J. 2003, Isolation of turbomycin A and B fom metagenomic library soil microbial DNA. *Appl. Environ. Microbiol*, 68:4301-4306.
- Green, B., and Keller, M. 2006. Capturing the uncultivated majority. *Curr. Opin. Biotechnol*. 17:236-240.
- Handelsman, J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. and Mol. Bio.l Rev*. 68:669-85
- Handelsman, J. 2004, Soils the metagenomic approach, en *Microbial diversity and Bioprospecting*, Capítulo 11. Ed. Bull, A. 109-119p. En línea.
- Handelsman, J. 2005. Metagenomics or Megagenomics. *Nat. Rev. Microbiol*, 3:457-458.
- Handelsman, J, Liles, M., Mann, D., Riesenfeld, C. and Goodman, R. M. 2002. Cloning the metagenome: culture independent access to the diversity and functions of the uncultivable microbial world. En: *Methods in microbiology*, vol 33. *Functional Genomics*. Wren, B. Y Dorrel, N. (eds). Academic Press. Amsterdam, The Netherlands. pp 241-255.
- Instituto Nacional de Estsdística, Geografía e Informática, 1978. Centro de estudios del territorio nacional (CETENAL), carta edafológica 1:50,000.
- Kaine, B. P. 1990. Structure of Archaeobacterial 7S RNA molecule. *Mol. Gen. Genet*. 221:315-321.

- Keller, M. and Zengler, K. 2004. Tapping into microbial diversity. *Nature Reviews*. 2:141-150.
- Kirk, J. L., Beaudette, L. A., Hart, M., Moutoglis, P., Klironomos, J. N., Lee, H., and Trevors, J. T. 2004. Methods of studying soil microbial diversity. *J. Microbiol. Meth.* 58:168-188.
- Klappenbach, J. A., Saxman, P., Cole, J., y Schmidt, T. M. 2001.: rrndb: the Ribosomal RNA operon copy number database. *Nucleic Acids Research* 29: 181-184
- Knietsch, A., Waschowitz, T., Bowien, S., Henne, A., and Daniel, R. 2003. Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* 5:46-56.
- Kozdrój, J., and van Elsas, J. D. 2001. Structural diversity of microorganisms in chemically perturbed soil assessed by molecular and cytochemical approaches. *J. Microbiol Methods.* 43:197-212.
- Lozupone, C. A. and Knight, R. 2007. Global patterns in bacterial diversity. *PNAS*, 104:11436-11440.
- Madigan, M.T., Martinko, J. M., Dunlap, P. V. and Clarck, D. P. 2008. *Brock Biology of Microorganisms* (12th. Ed). USA. Prentice Hall.
- Markowitz, V., Ivanova, N., Palaniappan, K., Szeto, E., Korzeniewski, F., Lykidis, A., Anderson, I., Mavrommatis, K., Kunin, V., García Martin, H., Dubchak, I., Hugenholtz, P. and Kyrpides, N. 2006. An experimental metagenome data management and analysis system. *Bioinformatics.* 22: e359-e367.
- Martin, A. P. 2002. Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Appl and Environ Microbiol.* 68:3673-3682.
- Neefs, J. M., van de Peer, Y., Hendriks, L., de Wachter, R. 1990. Compilation of small ribosomal subunit RNA sequences. *Nucl. Acids. Res.* 18:2237-2330.
- Niemi, R. M., 2001. Extraction and purification of DNA in rizosphere soil samples for PCR-DGGE analysis of bacterial consortia. *J. Microbiol. Methods.* 45:155-165.

- Nogales, B., 2005. La Microbiología del suelo en la era de la biología molecular: descubriendo la punta del Iceberg. *Ecosistemas*. 2005/2.
- Noguchi, H., Park, J., and Takagi, T. 2006. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucl. Acids Res.* 34:5623-5630.
- Odum, E. P., *Fundamentals of Ecology*, 3rd. ed., Philadelphia. W.B. Saunders Co, , PA. 1971.
- Olsen, G. J. and Woese, C. R., 1993. Ribosomal RNA: a key to phylogeny. *FASEB J.* 7:113-123.
- Page, R. D. M., 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comp. Appl. Biosci.* 12: 357-358.
- Pop, M. 2004. Shotgun sequence assembly. *Adv. In comp.* 60:193-248
- Rajendran, J., Gunasekaran, P. 2008. Strategies for accessing soil metagenome for desired applications. *Biotechnology Advances.* 26:576-590.
- Riesenfeld, C. S., Goodman, R. M. and Handelsman. J. 2004. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ. Microbiol.* 6:981-989
- Riesenfeld, C. S., Schloss, P. D., and Handelsman J. 2004. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* 38:525-552.
- Rodicio, M. R. and Mendoza, M. C., 2004. Identificación bacteriana mediante secuenciación del ARNr 16S: fundamento, metodología y aplicaciones en microbiología clínica. *Enferm. Infecc. Microbiol. Clin.* 22:238-45.
- Rodríguez-Varela, F. 2002, Approaches to prokaryotic biodiversity: a population genetics perspective. *Environ. Microbiol.* 4:628-633.
- Rodríguez-Varela, F., 2004. Environmental genomics, the big picture?. *FEMS Microbiol. lett.* 231:153-158
- Rondon, M., Goodman, M and Handelsman J. 1999. The earth's bounty: assessing and accessing the soil microbial diversity. *Trends Biotechnol.* 17:403-409.
- Rondon, M., Augist, P., Bettermann, A., Brady, S., Grossman, T., Loiacono, k., Lynch, B., MacNeil, I., Minor, C., Lai Tiong, C., Gilman, M., Osburne, M., Clardy, J., Handelsman, J. and Goodman, R. 2000. Cloning the soil metagenome: a

- strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66:2541-2547.
- Sandler, S. J., Hugenholtz, P., Schleper, C., DeLong, E. F., Pace, N. R. and Clark. A. 1999. diversity of *radA* genes from cultured and uncultured *Archaea*: comparative analysis of putative RadA proteins and their use as a phylogenetic marker. *J. Bacteriol.* 181:907-915.
- Schloss, P. D., y Handelsman J. 2003. Biotechnological prospects from metagenomics. *Curr. opinion Biotech.* 14:303-310.
- Shyamala, V. and Ames, F. L. 1989, Genome walking by single specific primer polymerase chain reaction: SSP-PCR. *Gene.* 84:1-8.
- Smit, E., Leeflang, P., Commans, S., van den Broek, J., van Mil, S and Wernars K. 2001, Diversity and seasonal fluctutations of the dominant members of the bacterial soil community in a wheat field as determined by cultivation and molecular methods. *Appl. Environ. Microbiol.* 67:2284-2291.
- Souza, V., Rocha, M., Valera, A. and Eguiarte, L. E. 1999. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. *Appl. Environ. Microbiol.* 65: 3373-3385.
- Staden, R. , 1979. A strategy of DNA sequencing employing computer program. *Nucl. Acid. Res.* 6:2601-10.
- .Thompson, J. D., Higgins, D. G. and Gibson, T. J. , 1994, Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl. Ac. Res.*
- Torsvik, V., and Øvreas, L. 2002. Microbial diversity function in soil: from genes to ecosystems. *Curr. Opinion microbiol.* 5:240-245.
- Torsvik, V., and Øvreas, L. 2007. Modern soil microbiology, cap. 2, Van Elsas, J., Jansson, J., and Trevors, J. eds. USA CRC Press.. 646 pp.
- Trevors, J. T. 1998. Molecular evolution in Bacteria: cell division. *Rev. Microbiol.* 29:237-245.
- Tringe SG, Hugenholtz P. 2008. A renaissance for the pioneering 16S rRNA gene. *Curr. Opin. in Microbiol.* 11:442–446.

- Ueda, K., Seki, T., Kudo, T., Yoshida, T. and Kataoka, M. 1999, Two distinct mechanisms cause heterogeneity of 16S rRNA. *J. Bacteriol.* 181:78-82.
- van Elsas J. D., Jansson, J. K., and Trevors, J. T. , *Modern soil microbiology*, 2007, capitulo 4, Eds. van Elsas, J D., Torsvik, V., Hartmann, A., Ovreas, L., and Janet K. Jansson. New York. CRC Press Taylor & Francis Group.
- Venter, J. C., Remington, K., Heidenberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu D., Paulsen, I., Nelson, w., Fouts., D. E., Levy, S., Knap A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsosns, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H. and. Smith, H. O. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Sci.* 304:66-74.
- Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K. H. and Streit. W. R. 2003. Prospecting for novel biocatalysts in a soil metagenome. *Appl. Environ. Microbiol.* 69:6235-6242.
- Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 1991. 16S Ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* 173: 697-703.
- Wintzingerode, F. V., Gobel, U. B., Sackerbrandt, E., 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* 21:213-229.
- Woese, C. R., Kandler, O. and Wheelis, M. L. 1990. Towards a natural system of organisms: Proposal for the domains *Archaea*, *Bacteria*, and *Eukarya*. *Proc. Natl Acad. Sci.* 87:4576-4579.
- Xu, Jianping, 2006. Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol. Ecol.* 15:1713-1731.
- Youssef, N. H. and Elshahed, M. S., 2008. Species richness in soil bacterial communities: A proposed approach to overcome sample size bias. *J. Microbiol. Meth.* 75:86-91.

<http://www.ncbi.nlm.nih.gov/BLAST/>

<http://www.mobio.com>

www.fermentas.com/reviewer

www.rdp.cme.msu.edu

<http://greengenes.lbl.gov>

Anexos

Anexo 1. Listado de todos los phyla encontrados por diferentes autores y los depositados en las do bases de datos (NCBI y RDP).

Phyla	Hugenholtz	Pace	NCBI	RDP	Ludwig
AC1	Hugenholtz	Pace	NCBI		
Acidobacteria	Hugenholtz	Pace	NCBI	RDP	Ludwig
Actinobacteria	Hugenholtz	Pace	NCBI	RDP	Ludwig
AD3	Hugenholtz	(Chloroflexi)			
Aquificae	Hugenholtz	Pace	NCBI	RDP	Ludwig
Bacteroidetes	Hugenholtz	Pace	NCBI	RDP	Ludwig
BD1-5	(OD2)	Pace	(Proteobacteria)		
BRC1	Hugenholtz	Pace	NCBI	RDP	Ludwig
Caldicellulosiruptor	Hugenholtz		(Firmicutes)	(Firmicutes)	(Thermoanaerobacteria)
Caldithrix	Hugenholtz	(CD7_NC3,CD8,KSB1_N C5,marine group A)	(Deferribacteres)	(Firmicutes)	Ludwig
Carboxydocella	Hugenholtz		(Firmicutes)	(Firmicutes)	(Firmicutes)
CD11	Hugenholtz	(Nitrospira)	(Nitrospira)		(Nitrospira)
CD12	Hugenholtz	Pace		(Firmicutes)	
CD13	(WS3)	Pace	(WS3)	(WS3)	(WS3)
CD14	(WS3)	Pace			
CD15	Hugenholtz	Pace	(Proteobacteria)	(Proteobacteria)	
CD18	(OD2)	Pace	(OP11)	(OP11)	(OP11)
CD19	(OD2)	Pace	(Proteobacteria)		(OD1)
CD2	(DSS1, WS2)	Pace		(Cyanobacteria)	(Cyanobacteria)
CD3	(Firmicutes)	Pace	(Firmicutes)	(Firmicutes)	(Firmicutes)
CD4		Pace			(Firmicutes, WS3)
CD5	(WS4)	Pace			(Deinococcus-Thermus)
CD7_NC3	(Caldithrix)	Pace	(Proteobacteria)	(Fimicutes)	(Caldithrix)
CD8	(Caldithrix)	Pace		(Fimicutes)	(Bacteroidetess, Chlorobi)
Cenibacterium	(Proteobacteria)		NCBI	(Proteobacteria)	
Chlamydia	Hugenholtz	Pace	NCBI	RDP	Ludwig
Chlorobi	Hugenholtz	Pace	NCBI	RDP	Ludwig
Chloroflexi	Hugenholtz	Pace	NCBI	RDP	Ludwig
Chrysiogenetes			NCBI	RDP	
Coprothermobacter	Hugenholtz	Pace	(Firmicutes)	(Aquificae)	Ludwig
Crenarchaeota	Hugenholtz	Pace	NCBI		Ludwig
Cyanobacteria	Hugenholtz	Pace	NCBI	RDP	Ludwig
Deep Cyanobacteria	(Cyanobacteria, Spirochates)	Pace	(Fimicutes, Cyanobacteria)	(Cyanobacteria)	(Cyanobacteria)

Deferribacteres	Hugenholtz	Pace	NCBI	RDP	Ludwig
Dehalococcoides		(Chloroflexi)	(Chloroflexi)	RDP	(Thermomicrobia)
Deinococcus-Thermus	Hugenholtz	Pace	NCBI	RDP	Ludwig
Denitromonas	(Proteobacteria)	(Proteobacteria)	NCBI	(Proteobacteria)	(Proteobacteria)
Desulfurobacterium	(Aquificae)	Pace	(Aquificae)	(Aquificae)	(Aquificae)
Dictyoglomi	Hugenholtz	Pace	NCBI	RDP	(Thermoanaerobacter)
DSS1	Hugenholtz	(CD2)		(Firmicutes)	(Cyanobacteria)
EM19	Hugenholtz	Pace		(Aquificae)	Ludwig
Entotheonella_CD10	Hugenholtz	Pace	(Proteobacteria)	(Proteobacteria)	(Proteobacteria)
Euryarchaeota	Hugenholtz	Pace	NCBI	(Aquificae)	Ludwig
Ferribacter	Hugenholtz	(Deferribacteres)	NCBI	(Deferribacteres)	(Firmicutes)
Fibrobacteres	Hugenholtz	Pace	NCBI	RDP	Ludwig
Firmicutes	Hugenholtz	Pace	NCBI	RDP	Ludwig
Fusobacteria	Hugenholtz	Pace	NCBI	RDP	Ludwig
Gemmatimonadales_Gem mimonas_NC9	Hugenholtz	Pace	NCBI	RDP	Ludwig
Guaymas 1	(Proteobacteria)	Pace		(Proteobacteria)	(Proteobacteria)
Haloanaerobiales	Hugenholtz	Pace	(Firmicutes)	(Firmicutes)	Ludwig
KB1	(NKB19,OP1)	(NKB19,OP1)	NCBI		(OP1)
Korarchaeota	Hugenholtz	Pace	NCBI		Ludwig
KSB1_NC5	(Caldithrix)	Pace	NCBI	(Firmicutes)	(Caldithrix)
LC1	Hugenholtz			(Firmicutes)	(Spirocheates)
Lentisphaerae	Hugenholtz	Pace	NCBI		(Victivalliaceae)
Magnetics	(Proteobacteria)	(Proteobacteria)	(Proteobacteria)	(Proteobacteria)	Ludwig
MgA_SAR406	Hugenholtz	Pace	(Proteobacteria)	(Firmicutes,Proto ebacteria))	(Caldithrix)
Moorella	Hugenholtz		(Firmicutes)	(Firmicutes)	(Firmicutes)
MS1	(Gemmatimonadales)	Pace	(Proteobacteria, Actinobacteria)	(Proteobacteria)	(Gemmatimonadales)
Nanoarchaeota	(Euryarchaeota)		NCBI		(Euryarchaeota)
Nanobacterium	(Proteobacteria)	(Proteobacteria)	NCBI	(Proteobacteria)	(Proteobacteria)
Natronoanaerobium	Hugenholtz	(Firmicutes)	(Firmicutes)	(Firmicutes)	(Firmicutes)
Natronohydrobacter	(Proteobacteria)	(Proteobacteria)	NCBI	(Proteobacteria)	(Proteobacteria)
NB01	Hugenholtz	(Proteobacteria)			
NC10	Hugenholtz	Pace		(Firmicutes)	(Nitrospira)
Nitrospina	Hugenholtz	Pace	(Proteobacteria)	RDP	Ludwig
Nitrospira	Hugenholtz	Pace	NCBI	RDP	Ludwig
NKB19	Hugenholtz	Pace	NCBI		
OD1	Hugenholtz	Pace	NCBI	(OP11)	Ludwig
OD2	Hugenholtz	(BD1-5, CD18, CD19)	(OP11,Proteobacteria)	(OP11)	(OD1)
ODP	Hugenholtz			(Firmicutes)	
OP1	Hugenholtz	Pace	NCBI	(OP10, OP11)	Ludwig

OP10	Hugenholtz	Pace	NCBI	RDP	Ludwig
OP11	Hugenholtz	Pace	NCBI	RDP	Ludwig
OP12	(Symbiobacteria)	Pace	(Firmicutes)	(Firmicutes)	(Firmicutes)
OP2		Pace	NCBI		Ludwig
OP3	Hugenholtz	Pace	NCBI		Ludwig
OP4	(Spirocheates)	(Spirocheates)	NCBI		(Spirocheates)
OP5	Hugenholtz	Pace	NCBI	(Firmicutes, Nitrospira)	Ludwig
OP8	Hugenholtz	Pace	NCBI	(Acidobacteria)	Ludwig
OP9_JS1_OPB46	Hugenholtz	Pace	NCBI	(Firmicutes)	Ludwig
OPA2	(Crenarchaeota)	Pace	(Crenarchaeota)		(Crenarchaeota)
OPA3		Pace	(Crenarchaeota)		(Crenarchaeota)
OPB7_EM3	Hugenholtz	Pace	(Thermotogae)		Ludwig
OS-K	Hugenholtz	Pace	NCBI		(Acidobacteria)
OS-L	Hugenholtz	Pace			(OP10)
Planctomycetes	Hugenholtz	Pace	NCBI	RDP	Ludwig
Poribacteria	Hugenholtz	Pace	NCBI		
Proteobacteria	Hugenholtz	Pace	NCBI	RDP	Ludwig
SAM		(Chloroflexi)	NCBI		(Thermomicrobia)
SB1_ZB1	(Chlorobi)	Pace	NCBI		(Chlorobi)
SBR1093_1959	Hugenholtz	Pace	NCBI	(Bacteriodes)	(Bacteriodes, Proteobacteria)
SC3	Hugenholtz	Pace			
SC4	Hugenholtz	Pace			(OP10)
Smegmaraptor	(Proteobacteria)	(Proteobacteria)	NCBI	(Proteobacteria)	(Proteobacteria)
SPAM	Hugenholtz	Pace	NCBI		(Nitrospira)
Sphaerobacter_Thermomicrobia	(Chloroflexi)	(Chloroflexi)	(Chloroflexi)	RDP	Ludwig
Spirochaetes	Hugenholtz	Pace	NCBI	RDP	Ludwig
SR1	Hugenholtz	Pace	NCBI		
Sturt1	Hugenholtz	(Firmicutes)	(Proteobacteria)	(Proteobacteria)	(Proteobacteria)
Sulfobacilli	Hugenholtz	Pace	(Firmicutes)	(Firmicutes)	(Firmicutes)
Symbiobacteria	Hugenholtz	(OP12, Sulfobacillus)	(Actinobacteria)	(Firmicutes)	(Firmicutes)
Synergistes_Desulfohalobium et rel	Hugenholtz	Pace	(Deferribacteres, Firmicutes)	(Firmicutes)	Ludwig
TG1	Hugenholtz	Pace	NCBI	(Actinobacteria, Firmicutes, Proteobacteria)	(OP5)
TG2			NCBI		
Thermoacetogenium	Hugenholtz		(Firmicutes)	(Firmicutes)	(Firmicutes)
Thermoaerobacter	Hugenholtz		(Firmicutes)	(Firmicutes)	(Firmicutes)
Thermoanaerobacter	Hugenholtz	Pace	(Firmicutes)	(Firmicutes)	Ludwig
Thermobaculum	(Chloroflexi)	(Chloroflexi)	NCBI	(Thermomicrobia)	
Thermodesulfobacteria	Hugenholtz	Pace	NCBI	RDP	Ludwig

Thermolithobacter			NCBI	(Fimicutes)	
Thermoprotei	Hugenholtz	Pace	(Crenarchaeota)	(Aquificae)	(Crenarchaeota)
Thermotogae	Hugenholtz	Pace	NCBI	RDP	Ludwig
Thermovenabulum	Hugenholtz	(Fimicutes)	(Fimicutes)	(Fimicutes)	(Thermoanaerobacter)
TM6	Hugenholtz	(Proteobacteria)	NCBI	(Proteobacteria)	Ludwig
TM7	Hugenholtz	Pace	NCBI	RDP	Ludwig
Verrucomicrobia	Hugenholtz	Pace	NCBI	RDP	Ludwig
Victivalliaceae	(Lentisphaerae)	(Lentisphaerae, Proteobacteria)	(Lentisphaerae, Spirocheates, Verrucomicrobia)		Ludwig
WS1	Hugenholtz	Pace	NCBI	(Fimicutes)	Ludwig
WS2	Hugenholtz	(CD2)	NCBI		(Cyanobacteria)
WS3	Hugenholtz	Pace	NCBI	RDP	Ludwig
WS4	Hugenholtz	(CD5)	NCBI		(Deinococcus-Thermus)
WS5	Hugenholtz	Pace	NCBI		
WS6	Hugenholtz	Pace	NCBI		Ludwig
YNPFFA	Hugenholtz	Pace	(Crenarchaeota)		(Crenarchaeota)
ZB3	Hugenholtz	Pace	(Fimicutes)	(Fimicutes)	(Cyanobacteria)

