



UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS

DEPARTAMENTO DE CIENCIAS BIOLÓGICAS

Trabajo de Tesis Doctoral

***Análisis de las poblaciones de
microorganismos fijadores de
nitrógeno del suelo aplicando
procedimientos metagenómicos***

Tesista: Lic. Calderoli Priscila A.

Director/a: Prof. Dr. Aguilar O.M

Codirector / a: Prof. Dra. Collavino M.M

Marzo 2016

El presente trabajo ha sido realizado en el Instituto de Biotecnología y Biología Molecular CONICET-CCT de la Facultad de Ciencias Exactas, Universidad Nacional de La Plata, bajo la dirección del Prof. Dr. Mario. O. Aguilar y la codirección del Prof. Dra. Mónica M. Collavino.

Mi reconocimiento

Al Instituto de Biotecnología y Biología Molecular (IBBM), a la Facultad de Ciencias Exactas (UNLP) por brindarme el lugar y los equipos para desarrollar este trabajo.

A la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) y al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), por el otorgamiento de las becas que posibilitaron mi dedicación a la investigación científica.

Al Centro de Biotecnología (CeBiTec) de la Universidad de Bielefeld, Alemania, donde fue posible realizar los análisis bioinformáticos del presente trabajo.

Mis agradecimientos

Quiero agradecer enormemente al Dr. Mario Mario. Aguilar y a la Dra. Mónica Collavino, por su confianza y enseñanza durante estos cinco años de trabajo en equipo. Mario gracias por ayudarme y permitirme conocer como es el trabajo científico en otras partes del mundo!! Moni gracias por los infinitos Skype, que a pesar de que Corrientes nos separa más de 1000 km, siempre estuviste ayudándome en todos los momentos que te necesite!!

Al Prof. Dr. Alfred Pühler y a todos los integrantes de su equipo por su colaboración en los análisis bioinformáticos durante mi estancia en Bielefeld.

Al Prof. Dr. Daniel Grasso por su valiosa colaboración en este trabajo de tesis.

A mis queridísimos compañeros y amigos de R4 que trabajar junto a Uds. quedará siempre en mi corazón!!

A las chicas CPAs que desde el primer día que ingresé al IBBM me brindaron una sonrisa!!

A mi familia y amigos que en los momentos más críticos (*lease*: escritura de la tesis) su apoyo logístico fue un pilar.

A vos Fran por tu apoyo INCONDICIONAL y por ayudarme a crecer en mi carrera profesional, gracias amor !!

A TODOS GRACIAS!!!!!!!

<i>Capítulo I</i>	1
<i>Introducción y Objetivos</i>	1
<i>El suelo</i>	2
<i>Diversidad de microorganismos en el suelo</i>	3
<i>Ecología microbiana</i>	5
<i>Herramientas para el estudio de microorganismos ambientales</i>	8
Métodos dependientes de cultivos	8
Métodos independientes de cultivos	9
- Métodos moleculares de huellas genéticas	12
- Clonación y secuenciación de ADN de muestras ambientales	13
Tecnología de pirosecuenciación aplicada al estudio de la diversidad microbiana	14
La Metagenómica.....	17
<i>Ecología de comunidades: el estudio de la biodiversidad</i>	19
<i>Los microorganismos como principales protagonistas de los ciclos biogeoquímicos</i>	20
<i>Dinámica del ciclo del nitrógeno y su importancia ecológica en suelo</i> ..	21
a. Fijación Biológica de nitrógeno	24
b. Mineralización del nitrógeno	26
c. Nitrificación	27
d. Reducción asimilativa y desasimilatoria del nitrito	28
e. Denitrificación	29
<i>La FBN y el ambiente agrícola</i>	30
<i>SoilGene y BIOSPAS: Consorcios argentinos interdisciplinarios para el estudio de la biología del suelo</i>	32
<i>Objetivos</i>	33
Objetivo general.....	33

Objetivos específicos	33
Capítulo II	34
Materiales y Métodos	34
<i>Descripción de los sitios y muestreo</i>	<i>35</i>
1. Suelos para el análisis funcional in silico de los metagenomas de suelo.....	35
2. Suelos para el análisis de perfiles T-RFLP del gen nifH	37
3. Suelos para el estudio de los transcritos del gen nifH en suelo	38
<i>Extracción de ADN, preparación de bibliotecas y secuenciación</i>	<i>39</i>
1. Procedimiento para la obtención de ADN metagenómico de suelo.....	39
2. Preparación de bibliotecas metagenómicas.....	41
3. Procesamiento de secuencias obtenidas por secuenciación masiva al azar	42
<i>Plataforma bioinformática para el análisis de datos metagenómicos de suelo</i>	<i>42</i>
1. Plataforma bioinformática MG-RAST.....	42
2. Plataforma bioinformática MGX	43
<i>Técnica molecular de tipo huella digital T-RFLP</i>	<i>44</i>
1. Amplificación del gen nifH	44
2. Purificación del pool nifH.....	46
3. Digestión del ADN-nifH con endonucleasas	46
4. Purificación de los productos de restricción nifH	47
5. Resolución de los fragmentos de restricción terminales del gen nifH por electroforesis capilar en secuenciador automático	47
6. Análisis de los perfiles T-RFLP con el Programa PeakScanner TM v1.0	48
7. Alineamiento de los perfiles T-RFLP con la herramienta T-Align.....	48
<i>Extracción de ARN de muestras de suelo y RT-PCR nifH</i>	<i>49</i>
1. Generación de amplicones ADNc-nifH para pirosecuenciación	50
2. Procesamiento y asignación taxonómica de las secuencias ADNc-nifH	51

3. Comparación de las base de datos ADN-nifH y ADNc-nifH	52
<i>PCR cuantitativa (qPCR) de grupos específicos de nifH</i>	52
1. Diseño de cebadores específicos para taxones.....	52
2. Condición de reacción de la qPCR	53
<i>Estimación de la diversidad</i>	53
<i>Análisis Multivariados</i>	54
a. Métodos de Ordenación	55
b. Métodos de clasificación.....	56
Capítulo III	58
<i>Análisis in silico de las comunidades microbianas involucradas en el ciclo del nitrógeno en suelos de la Región Pampeana Argentina</i>	58
<i>Introducción</i>	59
- Breve descripción de las plataformas informáticas usadas para análisis de metagenomas de suelo	62
a. Plataforma MG-RAST	62
a.1 Anotación funcional basada en la tecnología de subsistema.....	65
b. Plataforma bioinformática MGX	67
<i>Objetivo general</i>	69
Objetivos específicos	69
<i>Resultados y discusión</i>	70
Bases de datos metagenómicos de suelos: características y tamaño.....	71
Análisis in silico de los atributos funcionales de las bases de datos metagenómicas de suelo y comparación con otros ambientes naturales.....	74
Identificación de secuencias del suelo relacionadas a genes con función asignada del metabolismo nitrogenado.....	81
a. Clasificación de las secuencias de acuerdo a las categorías funcionales COGs.	83

b. Clasificación de las secuencias de acuerdo la base de datos de referencia personalizada	87
<i>La estructura de la comunidad de microorganismos del suelo asociada al metabolismo nitrogenado</i>	92
a. Fijación biológica de nitrógeno molecular (genes nif) y nitrificación (gen amoA).....	92
b. Reducción de nitrito	93
c. Proceso de denitrificación.....	95
<i>Conclusiones</i>	101
<i>Capítulo IV</i>	104
<i>Potencialidad del análisis T-RFLP para el estudio de las poblaciones fijadoras de nitrógeno en suelo</i>	104
<i>Introducción</i>	105
<i>Objetivo general</i>	108
Objetivos específicos	108
<i>Resultados y Discusión</i>	108
<i>Etapas preliminares de optimización de T-RFLP</i>	109
a. Selección de cebadores para la reacción PCR de amplificación del gen nifH.....	109
b. Selección de endonucleasas para la digestión del pool de amplicones de nifH.....	112
c. Ensayos de PCR anidada	113
d. Resolución y alineamiento de los perfiles T-RFLP	114
<i>Evaluación de la reproducibilidad del método T-RFLP</i>	115
a. Variación Intra/Inter gel.....	116
b. Variación introducida en la etapa de PCR	120
c. Variación en el tratamiento enzimática.....	121
d. Variación en el método completo de T-RFLP	122
<i>Análisis comparativo de las poblaciones diazotróficas de suelo mediante T-RFLP</i>	123

a. Composición de los perfiles de T-RFLP del gen <i>nifH</i>	123
b. Relación entre las comunidades diazotróficas de suelo analizadas mediante los perfiles T-RFLP del gen <i>nifH</i>	127
<i>Restricción in silico de secuencias del gen nifH: Asignación taxonómica tentativa de los T-RFs</i>	130
<i>Conclusiones</i>	134
Capítulo V	137
<i>Caracterización de la comunidad diazotrófica activa en suelo mediante el análisis de ADNc-<i>nifH</i></i>	137
<i>Introducción</i>	138
<i>Objetivo general</i>	140
Objetivos específicos.....	140
<i>Resultados y discusión</i>	141
<i>Obtención de ARN de suelo</i>	141
<i>Abundancia y distribución de los OTUs</i>	145
<i>Estimación de la diversidad y de la riqueza de la comunidad diazotrófica activa</i>	145
<i>Patrón de diversidad beta de las poblaciones diazotróficas con distinto manejo agrícola y profundidad</i>	151
<i>Composición de las comunidades diazotróficas activas en suelo</i>	155
a. Asignación taxonómica de las secuencias <i>nifH</i>	155
b. Distribución de los subclusters <i>nifH</i> en las muestras de suelo por tratamiento-profundidad.....	156
c. Clasificación taxonómica de los OTUs más abundantes.....	158
<i>Comparación entre la estructura de diazótrofos revelada por análisis de ADN y ADNc del gen nifH</i>	167
Relación filogenética y distribución de los OTUs <i>nifH</i> dominantes en ADN- <i>nifH</i> y ARN- <i>nifH</i> de suelo.....	170

<i>Cuantificación en suelo de filotipos de interés mediante qPCR del gen nifH</i>	173
<i>Conclusiones</i>	178
<i>Capítulo VI</i>	181
<i>Conclusiones generales y .perspectivas</i>	181
<i>Material Suplementario</i>	188
<i>Referencias Bibliográficas</i>	202

Capítulo I

Introducción y Objetivos

El suelo

El suelo es la capa superficial delgada que recubre la Tierra y actúa como interfase entre la atmósfera y la litosfera. Es una matriz compleja formada por tres componentes importantes: - un agregado de partículas minerales de tamaños varios y diferente composición química; - componentes orgánicos en las diferentes etapas de su descomposición; y - un componente biológico que comprende a bacterias, hongos, levaduras y también a las raíces de plantas. Los minerales están divididos en tres tipos: arenas, arcillas y limos. Los porcentajes de estos tres componentes y la materia orgánica se organizan a nivel micro- y macro- en agregados que determinan la textura y estructura del suelo [1]. Como propone Nogales *et al.* 2005, el tipo y composición de la materia mineral viene dado por las características de las rocas del subsuelo, así como de los procesos edáficos que hayan tenido lugar en su formación [2]. Esta porción inorgánica se halla muy relacionada con la disponibilidad de nutrientes, retención de agua, etc. La materia orgánica procede de la actividad de los distintos organismos vivos del suelo, a su vez su composición y cantidad es variable especialmente en función del tipo de cubierta vegetal. La cantidad de materia orgánica es, sin duda, uno de los mejores indicadores de la calidad agronómica y fertilidad de un suelo.

El origen de suelo es el resultado de cinco factores principalmente: la materia parental, el clima, la biota, la topografía y el tiempo [3]. Todos los factores desempeñan un papel importante en el desarrollo de cualquier tipo de suelo y son altamente interactivos. Las combinaciones de estos factores llevan a la formación de tipos únicos de suelos, pero con horizontes relativamente predecibles que componen el perfil del suelo.

Los contenidos de arena, arcilla, limo y materia orgánica se organizan a nivel micro- y macro- en agregados que determinan la textura y estructura [1]. Dependiendo de la textura del suelos estos se los dividen en: Arenoso arcillosos (35% o más de arcilla y 45% o más de arena), Limo arcillosos (40% de arcilla y 40% o más de limo) y Francos (contienen igual proporción de arena, arcilla y limo).

En este escenario los microorganismos se distribuyen heterogéneamente dentro de microagregados y en macroporosidades fuera de los microagregados [4], [5]. Además, estos se adhieren o adsorben fuertemente en las partículas del suelo (incluyendo partículas de arcilla y materia orgánica) a través de una variedad de mecanismos de unión [6].

A la manera de un órgano, el suelo desempeña múltiples funciones en el proceso de transformación de la materia orgánica y el reciclado de los nutrientes, resultando en uno de los principales proveedores de servicios al ecosistema (Figura I.1) [7]. Estos servicios, no son sólo esenciales para el funcionamiento de los ecosistemas naturales sino que también constituyen un importante recurso para la gestión sostenible de los sistemas agrícolas.

La degradación de los suelos, en parte por el mal uso agronómico de los mismos, en muchas partes del mundo, se refleja en la pérdida de estos servicios para el ecosistema. La tarea de conservar, optimizar y restaurar los servicios ambientales del suelo, constituye uno de los principales desafíos de este siglo.

Diversidad de microorganismos en el suelo

Se considera al suelo como uno de los principales reservorios de diversidad microbiana, y consecuentemente de genes para la degradación o transformación de compuestos orgánicos sobre el planeta [3]. Representa un ambiente sumamente promisorio para el descubrimiento de capacidades metabólicas novedosas y determinantes genéticos que podrían tener aplicación en la industria biotecnológica.

Se ha estimado que un gramo de suelo prístino presenta 2×10^9 células procariotas y kilómetros de hifas fúngicas [8]. En base a los estudios de cinéticas de reasociación de ácidos desoxirribonucleicos (ADN) realizados sobre muestras de ADN extraído y purificado de suelo, han permitido estimar que un gramo de suelo comprende entre 2.000 y 18.000 genomas procariotas distintos [9] un valor que puede ser subestimado ya que tales análisis difícilmente consideren los genomas de especies raras o difícil acceso.

Por lo tanto, estos análisis nos dan una idea preliminar del grado de diversidad de procariotas presentes en un gramo de suelo, la cual posiblemente exceda la diversidad conocida y descrita en el catálogo de procariotas del Centro para la Información Biotecnológica (NCBI) (4987 especies de procariotas listadas en 2015).

A diferencia de los hábitats acuáticos, las superficies de los suelos experimentan cambios dramáticos en el contenido de agua, desde situaciones de saturación hasta extrema sequía. Estos cambios drásticos sobre la población microbiana puede resultar en que una fracción de la comunidad microbiana cesa (muere) durante cada ciclo de estadios de sequía-inundación [10]. En consecuencia, la composición de la comunidad microbiana del suelo fluctúa. Los estudios de ecología microbiana han contribuido a un mejor entendimiento de la estructura de las comunidades microbianas en particular debido al efecto de variantes ambientales - climática, bióticos, xenobióticos-, sin embargo aún existen desafíos metodológicos para el análisis de la diversidad microbiana en ambientes de extrema complejidad, como lo es el suelo. La implementación de técnicas ómicas viene colaborando enormemente en los estudios actuales de la dinámica funcional microbiana en suelos y otros ambientes [11].

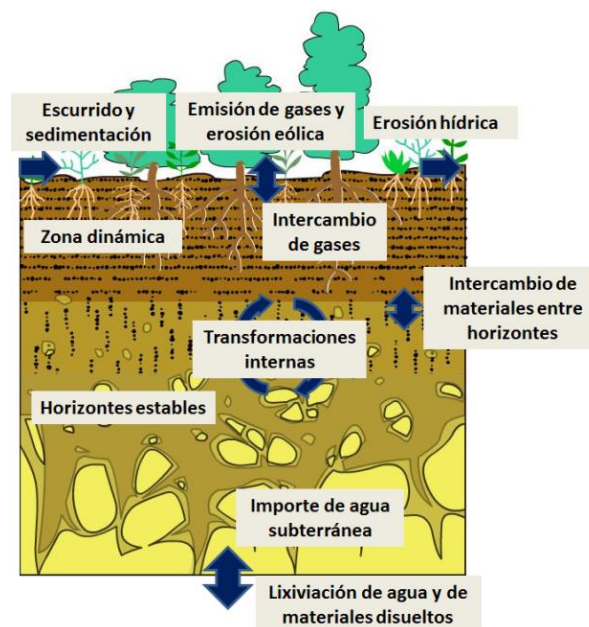


Figura I.1. Procesos y servicios ambientales del suelo. Adaptado de Needelman *et al.* 2013 [12].

La compleja estructura del suelo genera un nivel de heterogeneidad de hábitats donde nutrientes y agua tienen distintos grados de disponibilidad, esto se observa tanto a escala microscópica (microhábitat) como a nivel del paisaje. Las características específicas de los microhábitats, deben ser consideradas en la descripción de la actividad de los microorganismos del suelo, ya que las microregiones dentro de los agregados del suelo controlan la naturaleza y disponibilidad de los recursos de nutrientes, afectando directamente la dinámica de la población microbiana. Es sabido que la gran biodiversidad de microorganismos encontrada en los suelos está ligada a esta amplia heterogeneidad [3]. De hecho, los organismos del suelo no suelen estar distribuidos al azar sino que exhiben patrones predecibles en distintas escalas espaciales [13]. Aún es necesario mucho trabajo de investigación para llegar a comprender los procesos que generan la biodiversidad de microorganismos del suelo y generar predicciones de su distribución en los diferentes suelos.

Ecología microbiana

La ecología microbiana se propone establecer la composición de una comunidad así como también analizar lo que hacen y cómo interactúan entre sí los diferentes componentes de dicha comunidad. Se interesa también por revelar las propiedades emergentes de un ecosistema, aquellas capacidades que no se manifiestan en microorganismos aislados sino en el conjunto de un consorcio [14]. Un entendimiento profundo acerca de la dinámica de las comunidades microbianas es crítico para inferir situaciones y explorar predicciones acerca de cómo la biosfera modula y responde a condiciones ambientales presentes y futuras. Desde el punto de vista del conocimiento básico, el avance en el entendimiento de las comunidades microbianas naturales influirá nuestra capacidad para entender cómo funcionan los ecosistemas.

La biosfera se halla dominada por microorganismos de los cuales se conoce una parte pequeña sobre su diversidad y capacidades metabólicas. La abundancia de microorganismos ha sido estimada en varios ambientes naturales (Tabla I.1) [8]. La gran biodiversidad y diversidad metabólica de los microorganismos está fuertemente relacionada con la gran diversidad de nichos presentes en nuestro planeta. El

crecimiento de los microorganismos depende de los recursos disponibles y de las características fisicoquímicas del hábitat, las cuales definen el nicho ecológico de un microorganismo cubriendo una escala espacial muy pequeña.

Tabla I.1. Abundancia de microorganismo en ambientes naturales.

Ambiente	Abundancia*
Agua de mar	10^6 bacterias/ml
Sedimento superficial	10^8 - 10^9 bacterias/g
Tapíz fotosintético en un ambiente hipersalino	10^{12} bacterias/ml
Suelo de bosque	4×10^7 bacterias/g
Suelo cultivado	10^{12} bacterias/g

Las comunidades microbianas presentan atributos característicos, como la composición, diversidad o la dinámica en el tiempo, útiles para describir y comparar distintos hábitats. La composición de la comunidad, es decir los diferentes taxones o las poblaciones indígenas que la componen, es el resultado de la selección del sistema, tanto en sus componentes bióticos como abióticos. En particular el análisis de una comunidad busca entender cuál es la estructura de los ensamblajes biológicos, cuál es su función y cómo cambian espacial y temporalmente. Según Konopka *et al.* 2009 [14], el análisis de las comunidades microbianas involucra varios elementos relacionados con la determinación de:

- Diversidad taxonómica: referida a inventarios de diversidad (diversidad taxonómica dentro de un hábitat definido espacialmente) en una muestra discreta, entendiendo diversidad como la riqueza de taxones o número de especies y la abundancia relativa de los diferentes taxones en la comunidad.
- Determinación de características funcionales: involucra conocer la actividad de los microorganismos en procesos biogeoquímicos específicos.

- Interacciones microbianas: busca determinar la dinámica del ecosistema y la ecología evolutiva de cada individuo. Las interacciones se pueden originar por competencia por recursos, por asociaciones simbióticas, alelopatía, producción de moléculas señal, transferencia horizontal de genes y co-evolución, entre otros.
- Estabilidad, resistencia y resiliencia: determinan la respuesta del ecosistema a perturbaciones.
- Potencial genómico: Hace referencia al metagenoma de la comunidad. Una de las características de las comunidades microbianas es la capacidad de seguir ofreciendo servicios ecológicos frente a disturbios ambientales.
- Redundancia funcional: que proporciona la capacidad de responder rápidamente a cambios ambientales.

Sobre este último punto, Konopka propone un modelo conceptual de relación entre la redundancia funcional y la estabilidad del ecosistema, es decir la resistencia y resiliencia de un ecosistema frente a cambios ambientales (Figura I.2). A medida que se incorporan nuevas taxa en un ecosistema prístino aumenta la riqueza de la comunidad al igual que el número de funciones ecológicas. Sin embargo, en un determinado momento nuevas taxa no contribuyen con nuevas capacidades metabólicas, pero si incorporan redundancia para ciertas funciones. Sistemas con elevada redundancia son más propensos a retener funciones luego de un disturbio ambiental por la probabilidad que un taxón residente presente mayor resistencia al estrés.

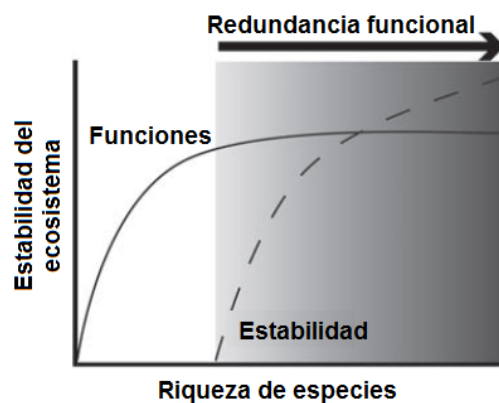


Figura 1.2. Modelo conceptual sobre relación de la redundancia funcional y la estabilidad de un ecosistema. Fuente: Konopka *et al.* 2009 [14]

Las propiedades funcionales de una comunidad definen el comportamiento de la misma, y hacen referencia a los procesos catabólicos que llevan a cabo, a que sustratos procesan, de qué forma interactúa con el ambiente, como responde a perturbaciones, entre otras cuestiones. Existen algunas metodologías (que se describen en mayor detalle en la próxima sección) que permiten el estudio sobre el potencial metabólico de una comunidad. Por ejemplo, en algunos casos existen regiones conservadas en genes funcionales (*dsrA* para la reducción de sulfato, *nifH* para la fijación de nitrógeno) que pueden ser utilizadas para evaluar la abundancia de tales genes en muestras ambientales.

Por tanto, uno de los desafíos actuales en el estudio de comunidades microbianas se trata de la identificación tanto de los microorganismos cultivables, basados en aislamientos y/o el análisis de las propiedades metabólicas de los mismos, como de los microorganismos no cultivables, basados en el análisis del metagenoma (análisis de todo el ADN presente en una muestra ambiental) por medio de técnicas moleculares. A continuación se describen algunas de las herramientas tanto dependientes como independientes de cultivo, usadas para el estudio de las comunidades microbianas en ambientes naturales.

Herramientas para el estudio de microorganismos ambientales

Métodos dependientes de cultivos

Para el cultivo puro de microorganismos presentes en la naturaleza, se han hecho continuos esfuerzos para aumentar la probabilidad de aislamiento de los mismos. En la literatura existe una extensa lista de medios de cultivo diseñados, que incluyen diferentes tipos de fuente de carbono, de vitaminas, de energía, entre otros, al igual que consideraciones sobre los factores fisicoquímicos que afectan el crecimiento, ejemplo temperatura, pH, requerimientos de oxígeno, salinidad, entre otros [15]. Desde 1880 con la introducción de cultivos sólidos en placa de *petri* se ha logrado descubrir una gran diversidad microbiana. A su vez, la utilización de cultivos de enriquecimiento permitió el aislamiento de microorganismos con características deseadas, debido a las condiciones de cultivo selectivas y diferenciales. Por otro lado, el cultivo de comunidades se refiere al aislamiento y cultivo de un consocio microbiano a partir de un proceso de selección artificial, la cual puede resultar en la

producción de una comunidad compleja y componentes diversos optimizados para llevar a cabo una función específica.

Estas metodologías dependientes de cultivo son esenciales para definir especies y para su caracterización genotípica y fenotípica. Algunas de las principales ventajas asociadas a esta metodología reside en su costo económico relativamente bajo y de que el investigador cuenta con el “material” microbiano para su uso en futuros estudios [16]. Mientras que la principal desventaja que presentan las metodologías dependientes de cultivo es la limitación de recuperar la mayor parte de microorganismos presentes en el ambiente. Es sabido que muchos de los microorganismos del ambiente son recalcitrantes al cultivo en condiciones *standard* de laboratorio [17]. Se estima que sólo entre el 0.1-1.0% de las bacterias viviendo en el suelo, pueden ser cultivadas. Para superar estas limitaciones, se han desarrollado métodos independientes de cultivos que se describen a continuación.

Métodos independientes de cultivos

El desarrollo de técnicas para la extracción del ADN o ARN de toda la comunidad microbiana a partir de muestras ambientales y su análisis ha hecho posible un mayor conocimiento de la diversidad microbiana. Las técnicas para la extracción de ADN usualmente son una combinación de métodos físicos y químicos para romper las células tratando de minimizar el daño causado a los ácidos nucleicos. Los métodos para extraer metagenomas bacterianos de diversos ambientes se dividen en dos grupos: lisis directa de las células contenidas en una suspensión amortiguada de la muestra y seguidamente la separación del ADN del substrato y los residuos celulares [18], o la lisis indirecta con la separación primero de las células del substrato para luego ser lisados [19]. El rendimiento de ADN asilado de diferentes tipos de suelo utilizando diversos protocolos puede variar de 1 µg a 500 µg de ADN por gramo de suelo, aproximadamente [8]. Se ha observado un mayor rendimiento de ADN en términos de masa de ADN por masa de suelo mediante la extracción directa [20], [21]; por ejemplo Gabor *et al.* 2003 registró una reducción de 10 a 100 veces en el rendimiento de ADN utilizando la lisis indirecta en comparación con la lisis directa [22]. Para lograr una lisis directa se han probado diferentes tratamientos enzimáticos,

altas temperaturas y detergentes. A su vez, otros métodos utilizan pasos de disrupción mecánica, como *bead-beating*, congelamiento-descongelamiento, etc. [23].

Por otro lado, los métodos de extracción de ADN basado en la separación de células, aunque menos eficiente en términos de la cantidad de ADN recuperado, son menos dañinos para la integridad de los ácidos nucleicos que la lisis directa. La separación de los microorganismos de la matriz del suelo se consigue mediante fuerzas mecánicas leves, procedimientos químicos o la adición de resinas de intercambio catiónico, seguido por una centrifugación por gradiente de densidad o diferencial. Este tipo de lisis permite obtener ADN con una mayor integridad y pureza. Además, el ADN recuperado se encuentra menos contaminado con sustancias asociadas a la muestra ambiental que interfieren con la manipulación enzimática del mismo (por ejemplo ácidos húmicos, que interfieren con la actividad de las enzimas restricción o ADN polimerasas). Otra de las ventajas que presenta este método se relaciona con el tamaño promedio de los fragmentos del ADN aislado, estos suelen ser mayores que los obtenidos con una lisis directa [8].

Por lo tanto sería conveniente la lisis indirecta, en aquellos casos que se desea construir una biblioteca de insertos grandes, por ejemplo. Dentro de las principales desventajas de la lisis indirecta encontramos la baja eficiencia de lisis y los rendimientos de ADN extraído. Delmont *et al.* 2011, en su trabajo donde combinan diferentes métodos de extracción de ADN para recuperar espectros diferentes de la diversidad microbiana en suelo, demuestran que desafortunadamente los diferentes métodos de extracción no proporcionan una submuestra uniforme e imparcial de ADN metagenómico [24]. Además estos autores, resaltan la importancia de la necesidad de considerables esfuerzos y tecnologías para acceder no sólo a *pools* de ADN, sino a todo un metagenoma para estudios de ecología microbiana confiables.

Los estudios ecológicos actuales pretenden caracterizar tanto a los microorganismos cultivables, como a los no cultivables. Es por ello que estos métodos que incluyen el aislamiento de ácidos nucleicos se han convertido en una herramienta importante en la ecología microbiana. El uso de técnicas independientes del cultivo ha permitido revelar no solo una mayor diversidad microbiana en los distintos

ecosistemas, sino grupos funcionales desconocidos previamente. Por ejemplo, fueron los resultados de los estudios sobre las poblaciones de suelo funcionalmente asociadas a la oxidación de amonio, los que nos revelaron a las arqueas como los contribuyentes dominantes a esa función. Las arqueas constituyen una división de microorganismos recalcitrantes a su cultivo *in vitro* [12], [13].

La mayoría de las técnicas moleculares usadas para estudios de diversidad microbiana están basadas en la amplificación de una secuencia blanco. La secuencia más frecuentemente usada es la del gen que codifica para el ARN ribosómico, y específicamente el gen que codifica para la subunidad menor del ribosoma bacteriano, ARNr 16S. La descripción de los microorganismos presentes en el ambiente experimentó un gran avance gracias a la utilización de la información que proporcionan estos genes, y ha permitido establecer firmas moleculares a varios niveles taxonómicos utilizadas como la base de una identificación bacteriana por comparación filogenética [25]. Esta aproximación se ha utilizado para caracterizar la diversidad bacteriana en diferentes tipos de suelos [26]–[28] y sedimentos [29], [30]. Este marcador ha sido extensamente usado dado que: (1) está presente en todos los microorganismos y tiene la misma función en todos ellos; (2) debido a restricciones estructurales, diferentes regiones de la molécula presentan distinto grado de variabilidad en su secuencia, lo que permite realizar comparaciones con diferente nivel de resolución; (3) su transmisión es principalmente vertical ya que se considera que no está sujeto a transferencia génica horizontal entre microorganismos; (4) la longitud de su secuencia tiene un tamaño adecuado para proporcionar suficiente información, y (5) el análisis de la secuencia permite realizar reconstrucciones filogenéticas de los microorganismos [31]. Por otro lado, el reconocimiento de genes biomarcadores funcionales ha permitido estudiar procesos metabólicos específicos en comunidades microbianas, tales como la fijación de nitrógeno, la oxidación de metano, la nitrificación, etc.

La técnica de PCR permite la detección de fragmentos de ADN que se encuentran en poca cantidad en muestras ambientales, no obstante, esta técnica presenta sesgos relacionados con la diferencia en el acceso del iniciador a la cadena molde y la eficiencia en la extensión [32]. Se disponen de varios criterios de análisis

de los amplicones tales como la -técnicas de huella digital (en inglés *Fingerprinting*), -PCR cuantitativa (qPCR),- bibliotecas de productos de amplificación y secuenciación masiva. A continuación se describen las principales características de alguna de ellas.

- ***Métodos moleculares de huellas genéticas***

A pesar que es posible obtener información filogenética precisa de los microorganismos dominantes en un consorcio microbiano mediante el secuenciamiento de clones con insertos de producto de PCR del gen ARNr 16S, el esfuerzo y el costo que involucra limita el análisis de múltiples muestras. Esto a su vez, obstaculiza la investigación de cambios en la estructura de la comunidad microbiana en relación a cambios en las condiciones fisicoquímicos y/o biológicos en el tiempo y espacio, lo cual es esencial cuando se pretende estudiar la funcionalidad de una comunidad. Una alternativa para la investigación de comunidades bacterianas de muestras complejas ha sido el análisis del gen ARNr 16S con métodos moleculares de huellas genéticas como DGGE (del inglés *Denaturing Gradient Gel Electrophoresis*) [33], TGGE (del inglés *Temperature Gradient Gel Electrophoresis*) [34][35] SSCP (del inglés *Single-Strand Conformation Polymorphism*) [36], [37], que permiten el análisis simultaneo de múltiples muestras. Otras metodologías han incorporado sistemas de secuenciación automatizado para la detección láser de fragmentos de ADN marcadas con fluoróforos, por ejemplo T-RFLP (del inglés *Terminal Restriction Fragment Length Polymorphism*) [38], y más recientemente RAPD (del inglés *Randomly Amplified Polymorphic DNA*) [39].

La elevada resolución de estos métodos, y la posibilidad de incorporar un marcador interno de tamaño en cada muestra, son características que deben dotarlos de un grado necesario de reproducibilidad. Además, de su bajo costo y sencillez, publicaciones recientes han demostrado la utilidad de los T-RFLPs en investigaciones ecológicas que requieran seguimiento de comunidades en el tiempo o en tratamientos experimentales [40]. Es importante resaltar que todas las técnicas de tipo huella digital presentan desvíos asociados a la amplificación por PCR suelen presentar una baja cobertura y también baja resolución taxonómica, ya que se obtienen fragmentos del mismo tamaño, que pueden resultar de grupos filogenéticos muy distintos [41].

- ***Clonación y secuenciación de ADN de muestras ambientales***

Uno de los métodos más utilizados en estudios de la diversidad biológica de comunidades naturales es la clonación y secuenciación tradicional vía *Sanger* (o de primera generación) del gen ARNr 16S [42]. En términos generales, la clonación molecular se refiere al proceso de aislar el ADN de interés, insertarlo en un plásmido y obtener múltiples copias de ella en un organismo (generalmente procariota) por acción de la ADN polimerasa (Figura I.3). La composición de la biblioteca de clones dependerá de la especificidad de los iniciadores y de la eficiencia con la cual los genes del ARNr sean amplificados y clonados. Las secuencias obtenidas pueden ser clasificadas taxonómicamente o utilizadas para reconstrucciones filogenéticas. Resulta en una metodología relativamente simple y que además requiere poco equipamiento. La desventaja principal de esta técnica reside en su falta de practicidad, ya que se estima que para poder obtener una representación adecuada de la diversidad microbiana de una muestra de suelo por ejemplo, se deben realizar más de 10.000 reacciones de secuenciación independientes, lo que es equivalente a analizar más de 10.000 clones [43]. La necesidad de una metodología más rápida, fácil y económica, de alto rendimiento y altamente confiable ha impulsado el desarrollo de tecnologías de secuenciación, denominadas segunda generación.

La característica principal de las tecnologías de segunda generación es su capacidad de generar millones de secuencias representantes de una comunidad microbiana sin la necesidad de armar bibliotecas de clones, y dentro de las cuales encontramos a la pirosecuenciación (Figura I.4). Las tecnologías de segunda generación, permitió a muchos grupos de investigación realizar análisis profundos y lograr un mejor entendimiento acerca de las propiedades estructurales de comunidades microbianas en diversos ambientes [44]–[46].

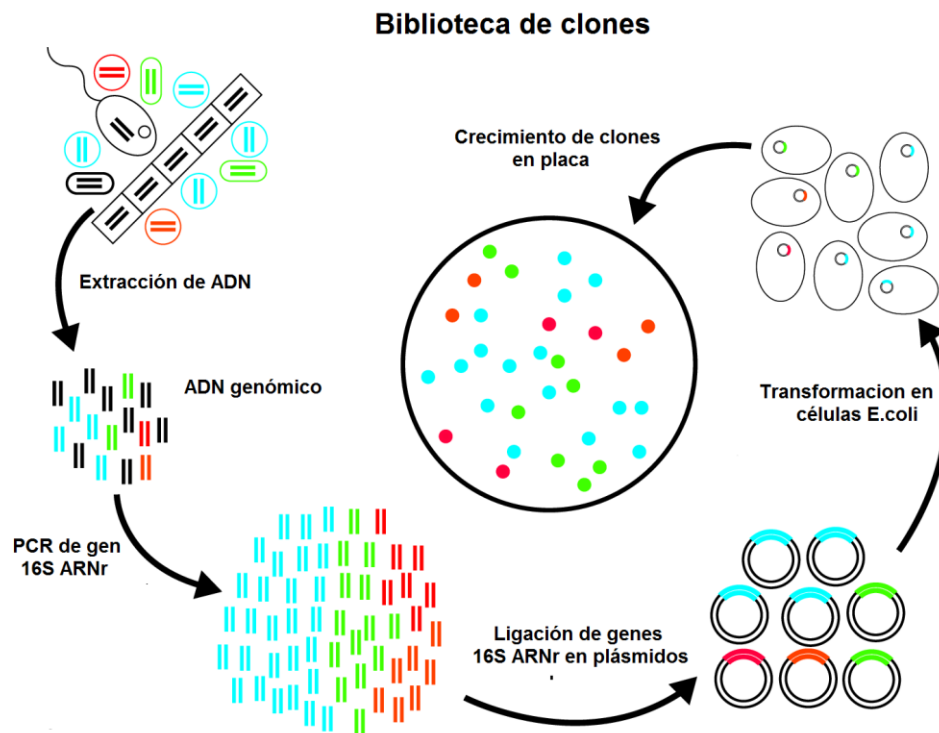


Figura I.3 Construcción de una librería de clones. Representación esquemática de las etapas implicadas en la construcción de una biblioteca de clones a partir del ADN extraído de una muestra ambiental. Fuente: <http://img.docstoccdn.com/thumb/orig/86901294.png>

Tecnología de pirosecuenciación aplicada al estudio de la diversidad microbiana

La pirosecuenciación es un método de secuenciación de ADN basado en la monitorización en tiempo real de la síntesis de ADN. La llegada de los sistemas de secuenciación por síntesis o pirosecuenciación de alto rendimiento dieron solución a los problemas mencionados anteriormente, para el análisis de la diversidad microbiana mediante librería de clones. La principal ventaja de las tecnologías de segunda generación es su capacidad de generar millones de secuencias representantes de una comunidad microbiana sin la necesidad de armar bibliotecas de clones (Figura I.4). El primer equipo de secuenciación basado en esta tecnología, fue el 454 GS FLX, comercializado por Roche. Esta plataforma de secuenciación 454-Roche puede generar hasta 1,3 millones de secuencias de un largo promedio de 400-600 pb en cada corrida de secuenciación, comparado con las 96 secuencias de una placa completa en

el sistema de *Sanger*. Por otro, lado el método permite el etiquetado de ADN generando la posibilidad de secuenciar múltiples muestras a la vez [47].

Brevemente, en el método de 454-Roche, el ADN del ambiente en estudio es fragmentado mecánicamente. Luego copias clonales de los fragmentos son producidas por PCR en emulsión sobre perlas de 28 μm (i.e. cada perla tendrá un solo fragmento original de ADN copiado miles de veces). Posteriormente, las perlas son tratadas con un desnaturalizante para recuperar sólo ADN simple cadena. Se hibridiza después un cebador de secuenciación a cada molécula de ADN presente en las perlas, conocidos como adaptadores. Las secuencias adaptadores están diseñadas para cumplir funciones en los pasos de selección, amplificación y secuenciación. A continuación se produce la pirosecuenciación. Para ello las perlas son depositadas en una picoplaca con pocillos de tamaño tal que quede depositada una perla única por pocillo. La secuenciación se produce por síntesis, cada ciclo consiste en la introducción de uno de los cuatro nucleótidos mediante una polimerasa. A su vez la incorporación es seguida de la adición de luciferina, la cual emite luz. Esta señal lumínica, que es capturada por una cámara, indica la adición del nucleótido en cada pocillo, produciéndose suficientes ciclos de A,G,C,T para completar secuencias de un promedio de 500 pb [48].

Se utilizan dos enfoques de secuenciación de alto rendimiento para el análisis de las comunidades microbianas: secuenciación en gran escala de un gen (*amplicon sequencing*) y secuenciación masiva al azar (*whole shotgun sequencing*). El primero se basa en amplificación de un gen o de la región de un gen de interés. Por la limitación en la longitud de las secuencias producidas generalmente se amplifica solo una región del gen. A este gen, se lo conoce como biomarcador, y dependiendo de la información que aporta, filogenética o funcional, se lo denomina biomarcador filogenético o biomarcador funcional. Para el estudio de diversidad y composición de comunidades microbianas el gen de ARNr 16S sigue siendo el biomarcador filogenético más utilizado [46], [49], [50], mientras que otros biomarcadores funcionales como *nifH* y *amoA* son ampliamente utilizados para el estudio de la fijación biológica de N_2 y la nitrificación, respectivamente [51], [52].

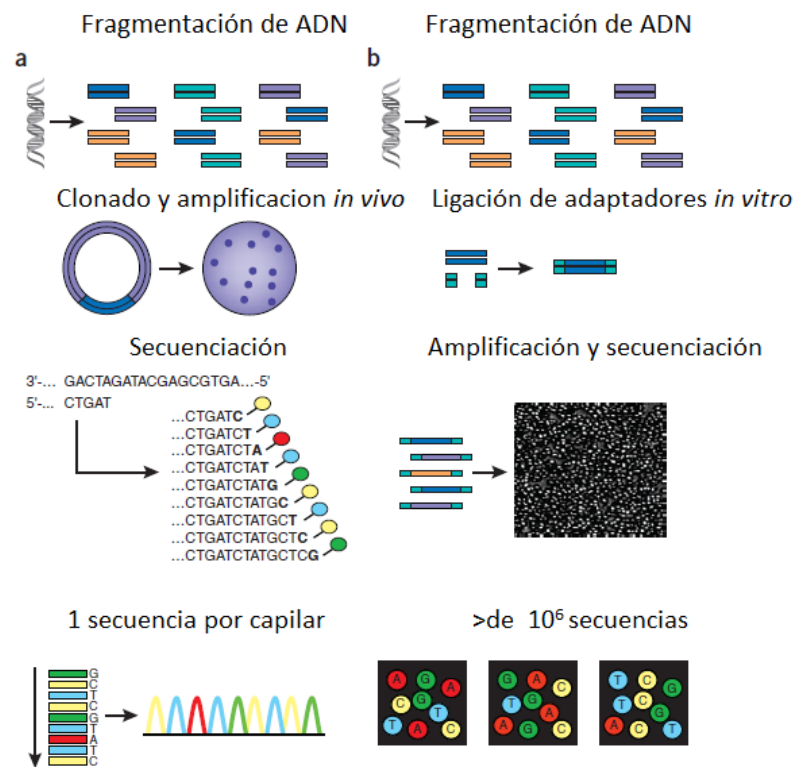


Figura I.4. Comparación entre métodos de secuenciación de primera (a) y segunda (b) generación. Adaptado de Shendure *et al.* 2008 [48].

La secuenciación masiva al azar se basa en la secuenciación directa del ADN genómico extraído de las muestras ambientales, y aporta información a las siguientes preguntas: quiénes están presentes en una determinada muestra ambiental? y qué están haciendo? [53]. El análisis de las secuencias producidas se centra en su anotación funcional. Para ello el primer paso es la identificación de marcos de lectura abiertos (ORF), existen dos métodos para identificar ORFs en secuencias cortas. Existen dos métodos para identificar ORFs. El método más sencillo se basa en búsqueda por homología (ej. BLAST) contra una base de datos [54]. El segundo método utiliza algoritmos de predicción de genes [55]. Los péptidos predichos necesitan ser luego comparados contra una base de datos de referencia para asignarles una anotación funcional.

En resumen, las nuevas tecnologías de secuenciación han traído grandes ventajas al estudio de la microbiología de suelos y a la determinación y análisis de su

diversidad. Varios proyectos internacionales con el objetivo de analizar la biodiversidad de suelos utilizaron herramientas de última generación. Entre ellos se destacan: "TerraGenome" [56], los proyectos llevados a cabo por el Departamento de Energía de los Estados Unidos [57], y proyectos nacionales como "BIOSPAS" [58] y "SoilGene" [59]. Todos ellos se propusieron caracterizar la biodiversidad taxonómica y funcional de distintos tipos de suelos y estudiar los efectos de distintos fenómenos sobre las comunidades microbianas (ej. cambios del clima, tratamiento agronómico, etc.)

La Metagenómica

El análisis metagenómico consiste en el estudio de todos los genomas presentes en un determinado ambiente a partir del ADN extraído directamente de muestras ambientales [60]. Representa una capacidad tecnológica para lograr un conocimiento más amplio y profundo de las estructuras microbianas y potencial metabólico en ambientes naturales. Dentro de este campo se destacan los estudios de biopelículas del drenaje de la mina ácida y el metagenoma del mar del Sargasso, los cuales han permitido la identificación de una gran diversidad de genes putativos nuevos, y hasta el ensamblaje de genomas [61], [62].

Los primeros pasos en la metagenómica de suelos consintieron en la preparación de bibliotecas metagenómicas para acceder a los genomas de los microorganismos (Figura I.5) [63]. El ADN metagenómico extraído es fragmentado por digestión con enzimas de restricción o ruptura mecánica, y luego clonado en vectores. La etapa de búsqueda (*screening*) de los clones es una tarea muy laboriosa ya que para tener una representación robusta del metagenoma, la cantidad de clones a evaluar es muy grande. El *screening* puede ser funcional, expresando las secuencias clonadas en organismos huésped (por ejemplo en *Escherichia coli*); o puede basarse en la inferencia de la función mediante la secuenciación del fragmento de ADN clonado. Una alternativa para estudios metagenómicos resulta en la secuenciación masiva al azar del ADN metagenómico y posterior análisis bioinformático para inferir la composición y potencial metabólico de la comunidad microbiana. Cabe mencionar, que la gran ventaja de la construcción de bibliotecas a diferencia del enfoque *shotgun*,

reside en que los clones pueden preservarse y es posible acceder a estos para análisis posteriores.

Actualmente, con la metagenómica se ha logrado conocer y estudiar en más detalle como los microorganismos, cultivables o desconocidos, se adaptan o interactúan con factores bióticos y abióticos en diversos ambientes [64]–[66]. A la vez mediante las nuevas tecnologías de secuenciación se han descubierto nuevas moléculas con aplicaciones biotecnológicas [67], [68]. Muchos estudios se complementan en algunos casos con el análisis de las propiedades metabólicas de las muestras, utilizando por ejemplo microplacas para determinar la capacidad de utilización de diferentes fuentes de carbono y nitrógeno por los microorganismos presentes (placa EcoPlate^{MR} de BIOLOG), o mediante la determinación de actividades enzimáticas específicas a partir de suspensiones de muestras ambientales [69]

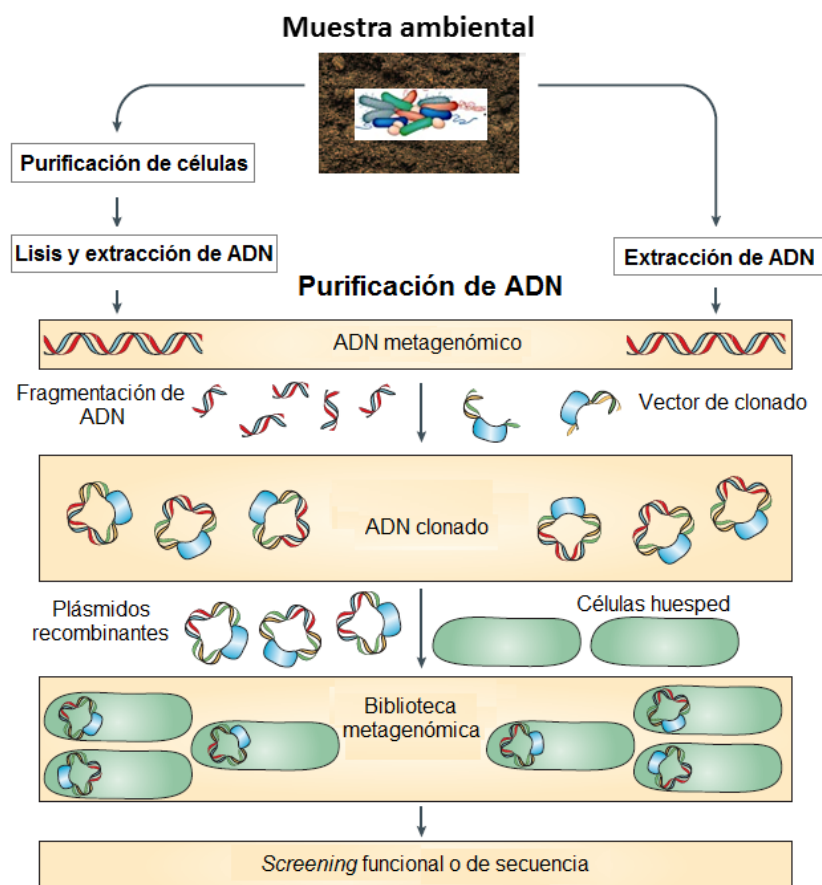


Figura I.5. Construcción de bibliotecas metagenómicas. Adaptado de Daniel *et al.* 2005 [8]

Ecología de comunidades: el estudio de la biodiversidad

La biodiversidad de un ambiente se estudia en el marco de la ecología de comunidades. Por definición una comunidad es el conjunto de especies que coexisten en un determinado lugar y tiempo [70]. Definir especies entre microorganismos no es tarea sencilla [50], por ello para el análisis de comunidades microbianas hablaremos de unidades taxonómicas operativas (OTUs, en inglés *Operational taxonomic units*).

El análisis de biodiversidad basado en amplicones ARNr 16S o de un gen funcional pueden llevarse a cabo mediante distintas estrategias: analizar un perfil de la diversidad de la comunidad y/o analizar su composición taxonómica. Suele utilizarse el 97% de similitud entre secuencias nucleotídicas para definir un OTU (para Bacteria y Arquea). Además, existen distintos métodos para agrupar secuencias en OTUs y se dividen entre: aglomerativos, divisivos y de partición [71], [72]. Por su lado, los métodos de clasificación taxonómica suelen seguir alguna de las siguientes estrategias: - alineamientos pareados contra un *set* de secuencias de referencia con identidad taxonómica conocida (ej. BLAST) [73]; - comparación de la composición de las secuencias con una referencia de entrenamiento utilizando “k-meros”, o “palabras” de una *n* longitud de pares de bases [74] o - métodos basados en filogenia que asignan la clasificación buscando la secuencia más cercana en una filogenia de referencia [75]. En el presente trabajo se utilizaron dos niveles de diversidad para el estudio de las comunidades diazotroficas. Por un lado, la *diversidad alfa* que se refiere a la diversidad de un hábitat o sitio, y por otro lado, la *diversidad beta* que se refiere a la diversidad entre hábitats o sitios [76].

Whittaker definió que la diversidad beta puede medirse a partir de la presencia-ausencia de especies o a partir de los datos cuantitativos de abundancia [77]. Si la diversidad beta se mide entre pares de sitios esta puede variar de par a par. Para analizar este tipo de variaciones se utilizan distancias ecológicas. La ventaja del uso de distancias es que se puede resumir la variación en la composición de especies en un solo valor; la desventaja es que se pierde la información sobre la identidad de las especies [78]. Una distancia ecológica será buena si logra describir las diferencias entre la composición de especies. Para sitios que comparten la mayoría de las especies la distancia será pequeña, mientras que para los sitios con pocas especies en común la distancia será grande. Para el cálculo de las distancias ecológicas existen varias

posibilidades (Figura I.6). Por ejemplo por el tipo de valor que se calcula: hay distancias que se miden en el intervalo entre cero y uno. Los valores de distancia de Bray Curtis y Kulczynski cercanos a cero indican que los sitios comparados tiene una composición similar de especies, mientras que si el valor se acerca a uno implica que la composición varía entre sitios. Otras distancias calculan valores no restringidos, que pueden ser mayores a uno. Este es el caso de la distancia euclidiana y de Hellinger. Otras medidas de distancia tienen en cuenta solo las diferencias en presencia y ausencia de especies, entre ellas: Jaccard y Sorenson [79].

$$\begin{aligned}
 \text{Bray-Curtis: } D &= 1 - 2 \frac{\sum_{i=1}^S \min(a_i, c_i)}{\sum_{i=1}^S (a_i + c_i)} \\
 \text{Kulczynski: } D &= 1 - \frac{1}{2} \left(\frac{\sum_{i=1}^S \min(a_i, c_i)}{\sum_{i=1}^S a_i} + \frac{\sum_{i=1}^S \min(a_i, c_i)}{\sum_{i=1}^S c_i} \right) \\
 \text{Euclidiana: } D &= \sqrt{\sum_{i=1}^S (a_i - c_i)^2} \\
 \text{Hellinger: } D &= \sqrt{\sum_{i=1}^S \left(\sqrt{\frac{a_i}{a_+}} - \sqrt{\frac{c_i}{c_+}} \right)^2} \quad \text{con } a_+ = \sum_{i=1}^S a_i
 \end{aligned}$$

Figura I.6. Fórmulas para el cálculo de distancias ecológicas entre dos sitios hipotéticos A y C. Donde a indica la abundancia de especies en el sitio A y c la abundancia de especies en el sitio C. Adaptado de Kindt *et al.* 2005 [79]

Los microorganismos como principales protagonistas de los ciclos biogeoquímicos

Los ciclos biogeoquímicos son considerados sistemas dinámicos que suceden a través de la biosfera, la disponibilidad de macro/micronutrientes para diferentes formas de vida dependen de los mecanismos de transformación involucrados en esos sistemas. El término “ciclo biogeoquímico” hace referencia claramente a los roles entrelazados de los componentes bióticos y abióticos para proporcionar las moléculas necesarias para el crecimiento y reproducción de los organismos vivos. El concepto comprende también a los microorganismos como principales protagonistas involucrados en la disponibilidad de nutrientes [3].

La extensa historia evolutiva de los microorganismos en el curso de la formación de la tierra como la conocemos hoy, les ha permitido desarrollar la maquinaria metabólica necesaria para reciclar prácticamente la totalidad de la moléculas complejas construidas naturalmente, convirtiéndolas en sustratos para la nutrición de organismos superiores [3].

Un proceso crítico en cualquier ciclo nutricional es la conversión de formas orgánicas de nutrientes provenientes de biomasa muerta (detritos), en formas inorgánicas simples y solubles que pueden ser incorporadas por plantas y otros microorganismos, proceso conocido como mineralización. El requerimiento cuantitativo de los macronutrientes (C, O, H, N, Ca, P, Mg, K, S, Na, Cl) es superior al de los elementos considerados micronutrientes (Fe, Mn, B, Co, Cu, Mo, Zn, I, Se).

Cada elemento tiene un patrón de ciclado resultando en formas asimilables para las plantas y ulteriormente objeto de reciclado por los microorganismos. Disponemos de una importante información y conocimiento de las reacciones bioquímicas involucradas en las transformaciones, como asimismo de los microorganismos y determinantes genéticos de esas capacidades biológicas. Una revisión de estos aspectos, especialmente dirigida al ciclo del nitrógeno (N) es encarada en las secciones que siguen.

Dinámica del ciclo del nitrógeno y su importancia ecológica en suelo

El ciclo del N comprende a diversos procesos llevados a cabo principalmente por microorganismos del suelo [80] y que permiten interconvertir distintas formas de compuestos nitrogenados, contribuyendo a mantener niveles de nutrientes asimilables para las plantas. Ningún otro elemento esencial para la vida toma tantas formas químicas en suelo como el nitrógeno (Tabla I.2). La dinámica de este elemento en la biosfera comprende la fijación biológica de nitrógeno (N_2), la mineralización, la nitrificación, la denitrificación y la oxidación anaeróbica del amonio [3]. El gas dinitrógeno (N_2) representa la forma de nitrógeno más abundante en la biosfera, pero no puede ser utilizada por la mayoría de los organismos, incluyendo las plantas. La fijación biológica del nitrógeno (BNF) es el proceso dominante por el cual el N ingresa inicialmente a los *pools* biológicos del suelo, resultando en la producción de amonio. Además, mediante una reacción de alta energía también se lleva a cabo la

fijación abiótica del nitrógeno molecular en la atmósfera. Comprende la electrificación y reducción fotoquímica donde el oxígeno y el nitrógeno se combinan para formar nitrato. Este es arrastrado por la lluvia a la superficie terrestre y a los cuerpos acuáticos en forma de ácido nítrico (H_2NO_3). El aporte de la fijación abiótica de N_2 a los depósitos de nitrógeno en suelos se ha estimado en 20 Kg N/ha/año comparado con los 100 a 200 Kg N/ha/año que provienen de la BNF [3]. Esto significa que la BNF es la fuente principal de N en suelos y por lo tanto un proceso eminentemente importante para el mantenimiento de la vida en la tierra [81].

Tabla I.2. Formas principales del nitrógeno en suelo y sus estados de oxidación

Nombre	Formula química	Estado de oxidación
Nitrato	NO_3^-	+5
Dióxido de nitrógeno	NO_2	+4
Nitrito	NO_2^-	+3
Oxido nítrico (g)	NO	+2
Oxido nitroso (g)	N_2O	+1
Dinitrógeno (g)	N_2	0
Amoniaco (g)	NH_3	-3
Amonio	NH_4^+	-3
N orgánico	R_{NH_3}	-3

El concepto del ciclo de N formulado por Löhnis en 1913 formalizó la noción acerca de que las distintas formas de nitrógeno son convertidas unas a otras, de una manera ordenada y predecible. Además, propuso la existencia de un equilibrio global, consistente en que el nitrógeno fijado por año debería ser anualmente convertido a N_2 mediante denitrificación [3]. Actualmente los niveles de fijación de nitrógeno - biológico e industrial- han superado significativamente las tasas de denitrificación y es una de las principales razones por las cuales el N se ha convertido en un importante contaminante [82]. Entonces si bien en varios ecosistemas terrestres la disponibilidad de N limita el crecimiento de las plantas y por ende la producción primaria, algunas formas de nitrógeno son nocivas afectando tanto al medio ambiente como a la salud humana. El nitrato (NO_2^-) puede ser arrastrado fácilmente por escorrentía, de modo que no se infiltra en el campo, sino que fluye normalmente hacia terrenos más bajos o cursos superficiales de aguas subterráneas y si alcanza aguas potables plantea un riesgo para la salud humana. Este compuesto puede combinarse con la hemoglobina y bloquear a nivel de respiración el intercambio normal de gases.

Lo descripto provee una justificación para intentar mejorar nuestro conocimiento sobre la comprensión de la transformación del nitrógeno y sobre los microorganismos del suelo que llevan a cabo dichas transformaciones. En la Figura I.7 se resumen los procesos de transformación química que intervienen en el ciclo del nitrógeno. Las etapas mediadas por los microorganismos y relevantes para el reciclado de nutrientes en los sistemas agrícolas incluyen la mineralización, íntimamente ligada a la degradación de materia orgánica, la nitrificación, la cual posibilita la conversión de amonio a nitrato, y las etapas de fijación de N_2 , y desnitrificación que cierran el ciclo global del N. A continuación se describen en detalle los procesos estudiados en el presente trabajo.

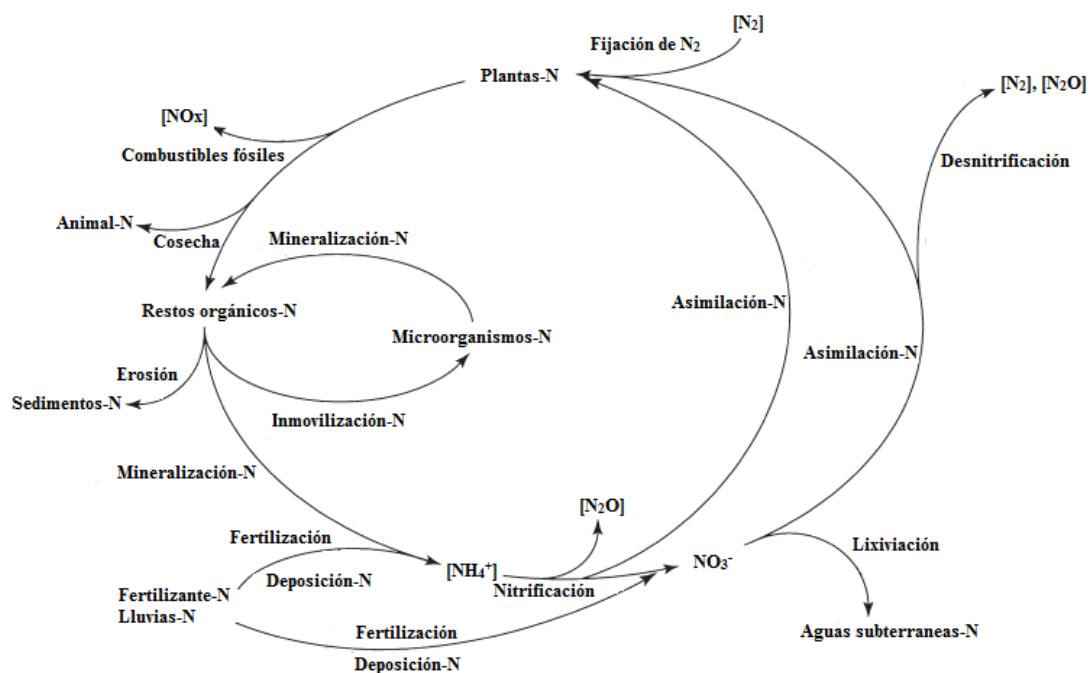


Figura I.7. Ciclo del nitrógeno. Representación esquemática de los principales componentes del ciclo del nitrógeno. Las formas gaseosas de N se encuentran en corchetes. Los procesos de transformación química del nitrógeno estudiados en el presente trabajo se encuentran recuadrados en celeste. Esta figura es una adaptación de su versión original de Paul *et al.* 2006 [3].

a. Fijación Biológica de nitrógeno

La fijación biológica de nitrógeno es la reducción de nitrógeno molecular (N_2) en amonio (NH_4^+), mediado por la enzima nitrogenasa. Dicha reacción enzimática es energéticamente costosa, 16 moléculas de adenosín trifosfato (ATP) y 8 electrones por molécula reducida, resultando en un proceso celular altamente regulado. La nitrogenasa es un complejo enzimático constituido por dos metaloproteínas: el Componente I (dinitrogenasa) presenta un peso molecular de aproximadamente 250 KDa y está constituido por dos heterodímeros codificados por los genes *nifD* y *nifK*. Además, en dicho componente se halla el sitio activo para la reducción de N_2 . El Componente II (dinitrogenasa reductasa) de aproximadamente 70 KDa acopla la hidrólisis de ATP a la transferencia de electrones entre los complejos proteicos. Presenta dos subunidades idénticas codificadas por el gen *nifH*. Las nitrogenasas "convencionales" contienen molibdeno (Mo) en los centros Fe-S de su grupo prostético, mientras que otras nitrogenasas conocidas como "alternativas" reemplazan Mo con vanadio (V) o hierro (Fe), siendo los genes codificantes *vnfH* y *anfH*, respectivamente [52].

El gen *nifH* es el gen más conservado entre los diversos microorganismos fijadores de N_2 . Las relaciones filogenéticas basadas en secuencias *nifH* obtenidas de microorganismos cultivables muestran una buena congruencia con las obtenidas a partir del estudio del gen ARNr 16S [52], [83]. El uso de las secuencias *nifH* en el análisis de poblaciones microbianas, ha permitido caracterizar la diversidad genética de la comunidad fijadora de N_2 independientemente del cultivo. Esta estrategia de análisis se ha aplicado en numerosos ambientes tanto acuáticos como terrestres generando una de las bases de datos de genes no ribosomales más extensa [84]–[88].

La reducción biológica de nitrógeno molecular es llevada a cabo únicamente por microorganismos procariotas. Los organismos diazótrofos utilizan diversas fuentes de energía incluyendo la luz del sol (fotótrofos), elementos y compuestos inorgánicos reducidos (litótrofos), y una gama de diferentes sustratos orgánicos (heterótrofos). Están representados por aerobios obligados y facultativos y anaerobios obligados. Esta diversidad metabólica indica que los diazótrofos pueden contribuir

con la fijación de N a otras formas de vida en una amplia variedad de ambientes. (Tabla I.3)

Tabla I.3. Ejemplos de géneros de bacterias diazotróficas.

AEROBIOS Y ANAEROBIOS FACULTATIVOS DE VIDA LIBRE		
Quimioorganotrofos	Fototrofos	Quimiolitotrofos
Eubacterias:	Cyanobacterias:	Eubacterias:
<i>Azotobacter spp.</i>	<i>Anabaena</i>	<i>Alcaligenes</i>
<i>Klebsiella</i>	<i>Nostoc</i>	<i>Thiobacillus**</i>
<i>Beijerinckya</i>	<i>Synechococcus</i>	
<i>Bacillus polymixa</i>	<i>Gloeocapsa</i>	
<i>Mycobacterium flavum</i>	<i>Dermocarpa</i>	
<i>Azospirillum lipoferum</i>	<i>Xenococcus</i>	
<i>Citrobacter freundii</i>	<i>Myxosarcina</i>	
Methylotrofos **	<i>Pleurocapsa</i>	
<i>Azomonas</i>	<i>Oscillatoria</i>	
<i>Derxia</i>	<i>Calothrix</i>	
<i>Chromobacterium</i>	<i>Microcoleus</i>	
	<i>Nodularia</i>	
	<i>Phormidium</i>	
	<i>Prochlorales</i>	
ANAEROBIOS DE VIDA LIBRE		
Quimioorganotrofos	Fototrofos	Quimiolitotrofos
Eubacterias:	Eubacterias:	Arqueobacterias:
<i>Clostridium spp.</i>	<i>Chromatium</i>	<i>Metanosarcina</i>
<i>Desulfovibrio</i>	<i>Chlorobium</i>	<i>Metanococcus</i>
<i>Desulfotomaculum</i>	<i>Rhodospirillum</i>	
	<i>Rhodospseudomonas</i>	
	<i>Rhodomicrobium</i>	
	<i>Rhodopila</i>	
	<i>Rhodobacter</i>	
	<i>Heliobacterium</i>	
	<i>Heliobacillus</i>	
	<i>Thiocapsa</i>	
SIMBIOTES EN:		
Plantas Leguminosas	Plantas No-Leguminosas	
Eubacterias:	Cianobacterias: *	Eubacterias:
<i>Rhizobium</i>	<i>Anabaena</i>	<i>Frankia (Actinomiceto)</i>
<i>Bradyrhizobium</i>	<i>Nostoc</i>	

* Incluye otras cianobacterias; ** Llevado a cabo solo por algunas especies. Datos tomados de Paul *et al.* 2006 [3]

Los diazotrofos son muy diversos, por ello resulta conveniente categorizarlos según su estilo/estrategia de vida. En el presente trabajo, el término diazotrofos se refiere a todos los organismos que presentan la capacidad genética para llevar a cabo la fijación biológica de nitrógeno. Sin embargo, esto no significa que el organismo en realidad ejecute esta función en el ambiente estudiado. Los diazotrofos simbióticos y de vida libre comprenden los dos grupos comúnmente definidos de estilos de vida. En particular, utilizaremos el término diazotrofos de vida libre o fijadores de N de vida libre para describir a los diazotrofos que habitan en el suelo o en asociación con raíces de plantas, excluyendo a los microorganismos endófitos y simbióticos. Según Kennedy *et al.* [89] hay muchos microorganismos que presentan varios grados de asociación con plantas, desde endófitos verdaderos a cepas que se enriquecen en la rizosfera o el la

superficie de las raíces. Estos últimos, suelen denominarse diazótrofos “asociativos”. Los diazótrofos “no simbióticos” presentan una definición más amplia e incluye a todos los diazótrofos que no presentan una simbiosis específica. Por lo tanto esta definición incluye a todos los diazótrofos que viven en condiciones libres o en asociación con raíces de plantas.

b. Mineralización del nitrógeno

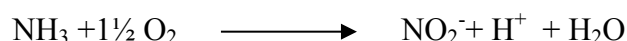
El proceso conocido como mineralización es llevado a cabo por microorganismos liberando nutrientes como producto secundario del consumo de detritos. Mientras que los microorganismos utilizan principalmente los detritos como fuente de energía y carbono (C) para su crecimiento, también presentan la necesidad de nutrientes, tales como el N, para ensamblar proteínas, ácidos nucleicos, y otros componentes celulares [3]. Si el detrito ofrece un excedente de nitrógeno para los organismos descomponedores este será liberado en forma inorgánica al medio mientras el C es consumido. Por el contrario, si el sustrato presenta un déficit de N, los microorganismos precisan capturar N adicional de su alrededor, ya sea incorporando o inmovilizando N en su biomasa (proceso conocido como inmovilización de N). Además, la absorción microbiana de N es afectada por la eficiencia del crecimiento del organismo. Por ejemplo, los hongos presentan una proporción mayor de C:N en sus tejidos en comparación a las bacterias, por lo tanto debido a su baja necesidad de N, pueden crecer más eficientemente en sustratos pobres en nitrógeno [3]

Tradicionalmente se consideró al NH_4^+ como el producto intermedio de la mineralización del N y de hecho se suele encontrar en la literatura referido como amonificación. Recientemente, el hallazgo de que las plantas pueden incorporar formas orgánicas solubles y simples de nutrientes, se ha ampliado la definición de productos de mineralización incluyendo cualquier forma de N soluble y simple que pueda ser incorporado por las plantas [90] Se ha demostrado que las plantas incorporan aminoácidos y otras formas orgánicas de nitrógeno siendo las micorrizas las que desempeñan un papel importante en esta incorporación. Son capaces de llevar a cabo la absorción de aminoácidos, azúcares aminados, péptidos, proteínas, y quitina que luego son utilizados por su planta huésped como fuentes de N.

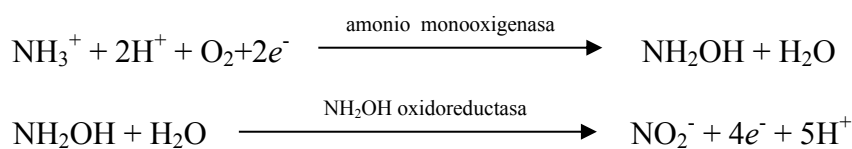
c. Nitrificación

La nitrificación mediada por microorganismos consiste en la conversión de amoníaco a formas de nitrógeno menos reducidas, principalmente nitrito (NO_2^-) y nitrato (NO_3^-). En ambientes con un potencial redox alto, el amoníaco puede ser oxidado a óxidos de nitrógeno y nitrato. Bacterias autotróficas aisladas a finales del 1800, presentaban una ganancia de hasta 440 kJ de energía por mol de amonio (NH_3) oxidado con nitrato como producto final. Actualmente, se conocen microorganismos heterotróficos capaces de llevar a cabo el proceso de nitrificación [91]–[93], no obstante la nitrificación autotrófica parece ser el proceso más importante en suelos [3].

La nitrificación autotrófica ocurre en dos etapas y es llevado a cabo por dos grupos separados de bacterias: las oxidantes de amoníaco y nitrito, respectivamente. En su conjunto a estas bacterias se las conoce como bacterias oxidantes de amonio (AOB) o nitrificantes. Sin embargo, la oxidación de NH_3 no se restringe solamente al dominio Bacteria, se han descubierto representantes de archeas que poseen la capacidad metabólica para oxidar amonio (AOA) [94], [95]. Entonces el proceso de nitrificación comienza con la oxidación de amoníaco a nitrito en condiciones estrictamente aeróbicas y se caracteriza por la siguiente reacción:



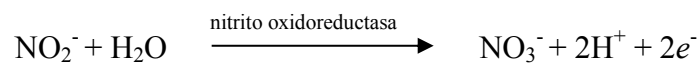
La fase inicial de nitrificación envuelve los siguientes eventos de oxidación:



La primera reacción de oxidación es mediada por la enzima de unión a membrana, amonio monooxigenasa (AMO). Se sintetiza el compuesto intermedio hidroxilamina (NH_2OH) el cual es posteriormente oxidado a nitrato mediante la NH_2OH oxidoreductasa. La enzima AMO también puede oxidar una amplia variedad de compuestos orgánicos no polares y de bajo peso molecular, entre ellos fenol, metanol, metano y otros compuestos halogenados alifáticos como tricloroetileno. Los productos intermedios son oxidados rápidamente, encontrándose en muy bajas

concentraciones en ambientes naturales. Otros compuestos intermedios que pueden resultar durante la oxidación de hidroxilamina a nitrito tales como óxido nítrico (NO) podrían escapar a la atmosfera e influir en la producción fotoquímica de ozono [3]. Esta fase del proceso de nitrificación libera energía ($\Delta G^\circ = -84$ kcal/mol) que es utilizada principalmente por bacterias quimiolitotróficas para la reducción de CO_2 a través del ciclo de Calvin-Benson. Basados en análisis de secuencias de genes ARNr 16S y el gen *amoA*, las bacterias aerobias estrictas del suelo que llevan a cabo la primera etapa de oxidación de amoníaco, se encuentran representadas principalmente por los siguientes géneros: *Nitrosomonas* (β -proteobacteria), *Nitrospira*, (β -proteobacteria) y *Nitrosococcus*, (γ -proteobacteria).

La segunda etapa del proceso de nitrificación conlleva la oxidación directa del nitrito a nitrato mediante la enzima reversible nitrato reductasa.



Las bacterias quimiolitotróficas del suelo del género *Nitrobacter* y de ambientes marinos de los géneros *Nitrococcus*, *Nitrospira* y *Nitrospina* son responsables de catalizar esta fase del proceso de nitrificación.

Los microorganismos nitrificadores desempeñan un rol substancial en la funcionalidad del ecosistema, ya que el aporte predominante de nitrato tiene su origen en la nitrificación *in situ* siguiendo el proceso descrito más arriba. Mientras que un aporte menor proviene de su incorporación en el ambiente mediante lluvias acidas o fertilizantes. El estudio de la variabilidad de la secuencia del gen codificante de la *amoA* es un criterio alternativo usado en estudios de ecología microbiana [96].

d. Reducción asimilativa y desasimilatoria del nitrito

El transporte activo de nitrato al interior de la célula es el primer paso en la ruta de asimilación de esta fuente de nitrógeno, mayoritariamente utilizada por las cianobacterias. Una vez dentro de la célula el nitrato se reduce hasta amonio para utilizarse como nutriente. Dicho proceso de reducción de nitrato es conocido como metabolismo asimilativo de nitrito [97]. El metabolismo asimilativo de nitrito tiene como producto principal al amonio por la acción de la nitrito reductasa (NiR). El

amonio es posteriormente incorporado a esqueletos carbonados mediante el ciclo glutamina sintetasa-glutamato sintasa.

En la respiración anaeróbica de nitrato, la acumulación de nitrito en la célula procariota puede resultar tóxico y para evitarlo, ciertos microorganismos catalizan la reducción de nitrito a amonio, mediante una enzima citoplasmática dependiente de NADH, y codificada por el gen *nirB*. A su vez este proceso regenera el poder reductor NAD^+ . A pesar que el amonio, generado por este proceso puede ser asimilado, el proceso se lo denomina reducción desasimilatoria del nitrito [97].

e. Denitrificación

La reducción disimilativa de nitratos o denitrificación mediada por bacterias anaeróbicas facultativas ocurre en condiciones limitantes de oxígeno y en presencia de nitratos (NO_3^-) y nitritos (NO_2^-), los cuales se utilizan como aceptores finales de la cadena respiratoria. La desnitrificación se lleva a cabo por una amplia gama de bacterias del suelo, incluyendo órgano-, quimio- y fotolitotrofos, fijadoras de N_2 , termófilos, halófilos y distintos patógenos. Se conocen alrededor de 50 géneros de bacterias denitrificantes con más de 125 especies [98], y algunos hongos son capaces de producir N_2O y N_2 a través de dos rutas de denitrificación fúngica en condiciones aeróbicas y anaeróbicas [99]. Los siguientes géneros bacterianos contienen especies que demuestran la habilidad de llevar a cabo el proceso de denitrificación: *Bacillus*, *Chromobacterium*, *Micrococcus*, *Pseudomonas*, *Spirillum*, *Hypomicrobium*, *Achromobacter*, *Moraxella*, *Paracoccus*, *Alcaligenes* y *Aquifex*. Este último género corresponde a bacterias hipertermofílicas y quimiolitotróficas obligadas. En el suelo, la mayoría de bacterias denitrificantes cultivables son anaerobios facultativos pertenecientes principalmente a los géneros, *Pseudomonas* y *Alcaligenes* y, en menor medida *Bacillus*, *Agrobacterium* y *Flavobacterium*.

El retorno del nitrógeno a la atmósfera mediante el proceso de denitrificación completa el ciclo biogeoquímico del nitrógeno. Este proceso acarrea una pérdida de nitrógeno en ambientes naturales, por lo que resulta ser un proceso detrimental para la actividad agrícola [100].

En la Figura I.8 se muestra las etapas del proceso de denitrificación, y para entender la estructura y diversidad de las comunidades implicadas en la reducción de

NO_3^- y su contribución en el proceso de desnitrificación se han utilizado como marcadores moleculares los genes para las enzimas: nitrato reductasa (*napA*, *narG*), nitrito reductasa (*nirK-nirS*), óxido nítrico reductasa (*norB*) y óxido nitroso reductasa (*nosZ*) codificadas por diversos grupos polifilético de bacterias [84], [98], [101]–[104].

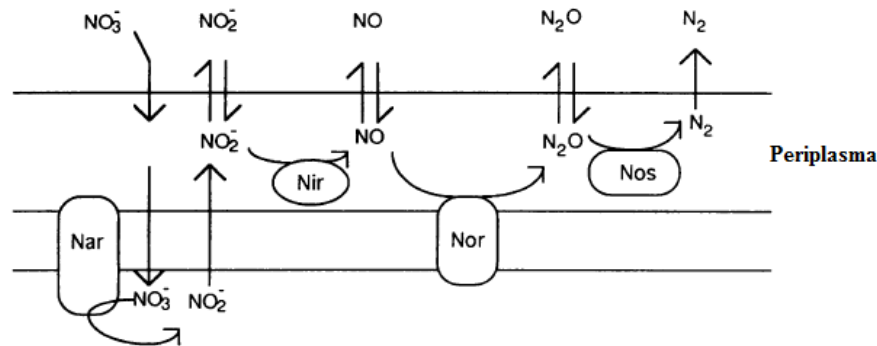


Figura I.8. Proceso de desnitrificación en bacterias Gram negativas. Nar, nitrato reductasa; Nir, nitrito reductasa; Nor, óxido nítrico reductasa y Nos, óxido nitroso reductasa. Figura adaptada de Paul *et al.* 2006 [3]

La FBN y el ambiente agrícola

La actividad del complejo agroindustrial de nuestro país tiene el potencial suficiente para abastecer el mercado interno y además exportar bienes y servicios de clase mundial. La producción primaria de origen agropecuario depende de un creciente nivel de productividad para permanecer competitivo en el futuro inmediato. Una de las limitantes más importantes de dicha producción es la disponibilidad de N en el suelo. El proceso de fijación biológica de nitrógeno (FBN), consistente en la reducción enzimática de nitrógeno molecular a amonio, aporta aproximadamente 140 millones de toneladas de nitrógeno por año constituyéndose en la principal entrada de este elemento a la biosfera [105]. Considerando el costo económico y ecológico del uso de fertilizantes químicos, se pone de manifiesto el rol central de la FBN en la implementación de sistemas productivos y sustentables en el suelo.

El sistema de siembra directa se ha implementado en forma creciente en los últimos años, cubriendo actualmente más de 70 millones de hectáreas en el mundo y aproximadamente el 70% de la superficie cultivada en la Argentina [106]. En el año

2003 se definió el concepto de Buenas Prácticas Agrícolas (BPA) por parte de la *The Food and Agriculture Organization* (FAO) [107]. El mismo incluye una lista de prácticas agrícolas que fomentan la sostenibilidad ambiental, económica y social de los procesos agropecuarios, promoviendo la producción de alimentos y productos agrícolas seguros y de calidad. Las bases del manejo BPA son las siguientes: 1) Mínimo disturbio mecánico del suelo; 2) Cobertura de suelo permanente; 3) Rotación de cultivos, 4) manejo integral de plagas, malezas y enfermedades. Numerosos estudios demuestran que este sistema de siembra directa bajo BPA disminuye los efectos negativos del uso agrícola tales como la erosión, la pérdida de capacidad de retención de humedad o la disminución del *pool* de materia orgánica del suelo [108]. Sin embargo, a pesar del rol fundamental de los microorganismos en los procesos del suelo, la influencia del uso y manejo del suelo sobre la composición y diversidad de la comunidad microbiana suelo es limitada, particularmente considerando comunidades asociadas a procesos específicos tales como la fijación de nitrógeno [109]–[111].

Varios autores reportaron un aumento de la biomasa total y actividad microbiana en respuesta al incremento de la incorporación de materia orgánica al suelo y a la reducción de aplicación de agroquímicos [112]–[114]. Por otro lado, Donnison *et al.* 2000 encontraron que cambios en el manejo agronómico no provocan efectos significativos en el estado nutricional, en la biomasa y actividad microbiana del suelo; aunque también demostraron que el manejo agronómico afecta significativamente la estructura de la comunidad microbiana, sugiriendo que esto se debe principalmente a cambios en la naturaleza de los cultivos y la cantidad de fertilizante aplicado [115]. En particular, para el caso de la comunidad diazotrófica en suelo se ha demostrado que la estructura y diversidad responde a cambios asociados a la naturaleza del nitrógeno aplicado, al pastoreo, a las enmiendas y a la incorporación de rastrojos [116]–[118]. Orr *et al.* analizaron la diversidad y actividad de las bacterias fijadoras de N_2 en suelos con manejo agronómico orgánico y convencional, encontrando que la rotación de cultivos tiene un efecto significativo sobre la diversidad de la comunidad diazotrófica.

SoilGene y BIOSPAS: Consorcios argentinos interdisciplinarios para el estudio de la biología del suelo

Los consorcios se establecieron en el marco de dos Proyectos de Areas Estratégicas (PAE) financiado por el Ministerio de Ciencia e Innovación Tecnológica de la Argentina. El objetivo de SoilGene consistió en evaluar el impacto de la agricultura sobre ecosistemas agrícolas típicos de la Región Pampeana, con enfoque metagenómico en el estudio de los procesos biológicos y diversidad microbiana de los suelos [59]. El proyecto interdisciplinario denominado "Biología del Suelo y Producción Agropecuaria Sustentable" (BIOSPAS), presentó el objetivo general de desarrollar nuevos indicadores de calidad de suelo en base a sus propiedades químicas, físicas y biológicas [119]. Numerosos trabajos científicos con grandes aportes para el estudio de la biología del suelo han sido publicados en el contexto del desarrollo de ambos consorcios [119]–[123].

Objetivos

Objetivo general

En este trabajo se propone el estudio de las poblaciones microbianas participantes en el metabolismo nitrogenado, en suelos bajo distintas condiciones de uso y prácticas agrícolas de la Argentina, aplicando enfoques metagenómicos

Objetivos específicos

- Analizar la composición de las comunidades microbianas del suelo relacionadas con el ciclo del nitrógeno mediante el análisis *in silico* de bibliotecas metagenómicas..
- Estudiar la población diazotrófica potencial (gen *nifH*) en suelo mediante la técnica de “Polimorfismo en la longitud de los fragmentos de restricción terminal” (T-RFLP, *Terminal-Restriction Fragment Length Polymorphism*).
- Definir la estructura de la población diazotrófica activa en suelos bajo diferentes manejos agronómicos de la región de Pergamino mediante el análisis de ADNc-*nifH*.

Capítulo II

Materiales y Métodos

Descripción de los sitios y muestreo

1. Suelos para el análisis funcional in silico de los metagenomas de suelo

Para el análisis metagenómico se utilizaron suelos representativos de diferentes situaciones de manejo agropecuario, los cuales fueron objeto de estudios multidisciplinarios por otros grupos pertenecientes al consorcio *SoilGene* (PAE n° 37164). Las muestras de suelo se obtuvieron de cuatro lugares diferentes de la región pampeana argentina, el área más productiva y con la más larga historia de manejo agrícola del país [124]. Las regiones de la Pampa Ondulada y Balcarce fueron seleccionados de acuerdo a isoyetas previamente definidas: clima húmedo (1000 - 1200 mm anuales) y semi - húmedos (800 - 1000 mm anuales).

El muestreo en la Pampa Ondulada se llevó a cabo en tres establecimientos productivos: La Estrella, La Negrita y Criadero Klein (Figura II.1) en Junio del año 2010. Los suelos han estado bajo manejo agronómico en los últimos 30 años y bajo labranza cero (siembra directa) en los últimos 16 años, utilizando una rotación trigo/soja-maíz. La última rotación antes del muestreo fue trigo-soja. Se tomaron muestras en lotes agrícolas y en los parques asociados a los cascos en cada establecimiento. Se determinaron así dos tratamientos: suelos cultivados en siembra directa, "CSD", y un control de campos no cultivado, "NC". La utilización de muestras de suelo de parques como tratamiento "referencia" o "control", en análisis de suelos de la región pampeana, es una práctica altamente difundida. Las muestras de suelos no agrícolas se obtuvieron de las zonas de pastos cercanas al casco con herbáceas (no leñosas), como *Cirsium sp*, *Trifolium sp*, *Micropsis sp*, *Festuca sp*, *Dichondra sp*, *Cyperus sp* y *Taraxacum officinale*. Los suelos muestreados fueron clasificados como Argiudoles, de las series Chacabuco y O' Higgins. La temperatura media anual de los sitios fue de 16,0°C a 16,3°C y la precipitación media anual fue de 1027 mm a 1065 mm.

Otro suelo muestreado correspondió al campo Experimental INTA Balcarce, que se encuentra ubicado en la Pampa sureña (Figura II.1). Dicho campo forma parte

de un experimento de labranza (arado de vertedera bajo) y rotación mixta de cultivos de 34 años.

Las parcelas bajo labranza convencional (LC) tuvieron ese manejo desde 1976 y las parcelas bajo siembra directa sufrieron labranza convencional desde 1976 a 1993 y luego siembra directa por 16 años hasta el momento del muestreo en Agosto de 2010. Las rotaciones variaron entre pasturas, trigo/soja-maíz. Las subparcelas muestreadas recibieron una fertilización de 60 Kg N/año y el mismo tipo de rotación. La fuente de nitrógeno utilizada fue urea aplicada al voleo en V6 (sexta hoja) para maíz, en macollaje para el trigo y en R1 (inicio de floración) para soja. La última rotación antes del muestreo fue maíz-soja. También se tomaron muestras en el parque asociado al casco. Los suelos muestreados fueron clasificados como un complejo de Argiudol típico (Serie Mar del Plata) y Paleudol petrocálcico (Serie Balcarce). La temperatura media anual en Balcarce fue de 13.8 (°C) y la precipitación anual media de 875 mm.



Muestra	Latitud	Longitud
La Estrella	-34° 41' 4,2"	-60° 6' 2,4"
La Negrita	-34° 34' 46,2"	-59° 55' 8,4"
C. Klein	-35° 7' 17,4 "	-60° 15' 42,6"
Balcarce	-37° 45' 09",	-58° 17' 47'"
Pergamino	-33° 56' 36"	-60° 33' 57"
Viale	-31° 52' 59,6"	59° 40' 07"

Figura II.1. Localización geográfica de los sitios muestreados en la región pampeana argentina
En la provincia de Buenos Aires los sitios:- La Estrella, - La Negrita, - Criadero Klein, - Balcarce, - Pergamino y en la provincia de Entre Ríos - Viale.

En cada uno de estos sitios se recolectó suelo no rizosférico (*bulk*) tomados de los primeros 20 cm de profundidad con barrenos diseñados para tal fin. El experimento se llevó a cabo con un diseño de bloques al azar, con un arreglo de parcelas divididas por tratamiento agrícola. En los lotes correspondientes a cada uno de los tratamientos se tomaron en un área de aproximadamente 500 m², 20 muestras al azar de los primeros 20 cm. Las submuestras de cada lote luego fueron homogeneizadas en una única muestra, y enviadas inmediatamente al laboratorio a 4°C donde fueron tamizadas a través de malla de 1 mm para una homogeneización profunda, rotura de los agregados y retiro de las raíces y residuos vegetales. Las muestras de suelo se almacenaron a -80 °C inmediatamente después de la homogeneización hasta la extracción de los ácidos nucleicos. Para mayor información acerca de estos suelos ver Rascovan *et al.* 2013 [59].

2. Suelos para el análisis de perfiles T-RFLP del gen *nifH*

Además de los suelos anteriormente descriptos, para el análisis de T-RFLP también se incluyeron suelos de otras regiones de la Pampa Argentina. Se muestrearon campos de la región de Pergamino en Pcia. de Buenos Aires y Viale en Pcia. de Entre Ríos en el marco del proyecto PICT-PAE 2006-11 (Consortio BIOSPAS). Los suelos muestreados fueron clasificados como un complejo de Argiudol típico y Vertisol, respectivamente.

En cada una de las localidades se tomaron muestras de 2 manejos agrícolas contrastantes y un control de campos no cultivados, cada uno por triplicado (submuestras). Los distintos tratamientos agronómicos se definieron según las características de las prácticas agrícolas que se llevan a cabo en ellos:

1. “Buenas prácticas agrícolas (BPA)”: Se caracteriza por un manejo permanente en siembra directa contemplando la rotación intensiva de cultivo, fertilización en reposición, manejo integral de plagas, malezas y enfermedades, el reemplazo de nutrientes, con cultivos de cobertura en las etapas del año donde no se produce.
2. “Pobres prácticas agrícolas (PPA)” en siembra directa: Campos con mínima rotación de cultivos o monocultivo de cultivos (soja) y baja reposición de nutrientes. Estos suelos se caracterizan por pasar gran parte del año sin ningún tipo de cobertura.

3. No Cultivados (NC). Ambiente natural con mínima actividad antrópica, ej: reserva natural, sectores “santuarios” que se conservan en algunos establecimientos, montes, sectores de parques. Estas áreas no han sido cultivadas ni trabajadas al menos por 30 años.

El muestreo se realizó en Febrero del 2011 y dentro de cada sitio-tratamiento se muestrearon tres zonas separadas por un distancia de al menos 50 m. Las muestras consistieron en un *pool* de 16-20 submuestras de suelo *bulk* (fracción 0-20 cm de profundidad) separados al menos por dos metros unas de otras, dispuestos en forma de cuadrícula. Estas submuestras se combinaron en el campo para obtener una única muestra compuesta por zona y se transportaron al laboratorio a 4 °C. Tres días posteriores a la recolección las muestras se tamizaron a través de 4 mm de malla para eliminar raíces, residuos vegetales y ayudar a romper los agregados. Las muestras se almacenaron a -80 °C hasta el momento de su análisis. Para mayor información acerca de estos suelos ver Figuerola *et al.* 2012 [119].

Las mediciones fisicoquímicas fueron realizadas por el laboratorio del Ing. Agr. R. Alvarez para el caso de las muestras de suelo de La Estrella, La negrita, Criadero Klein y Balcarce [120], [125]. Para el caso de las muestras de suelo de Viale y Pergamino, el grupo liderado por el grupo del Ing. Agr. J.A. Galantini [119]. En las Tablas II.1, II.2 y II.3 se detallan respectivamente las abreviaturas, propiedades físicas y químicas de todos los suelos analizados en el presente trabajo.

3. Suelos para el estudio de los transcriptos del gen *nifH* en suelo

El muestreo se realizó en la zona de Pergamino, provincia de Buenos Aires, en donde previamente se analizó la comunidad diazotrófica potencial [121]. Se consideraron lotes bajo siembra directa, con situaciones de manejos diferentes, similares a los descriptos previamente, i) BPA, ii) PPA (soja), y un tercer ambiente NC. Además, de los distintos tratamientos se muestrearon dos profundidades de suelo 0-10 y 10-20 cm.

El muestreo se realizó en Junio del 2012 y al momento del muestreo se encontraban los rastros correspondientes a la cosecha de soja. Dentro de cada

tratamiento se tomaron tres muestras de dos profundidades 0-10 cm y 10-20 cm, separadas por una distancia de al menos 50 m. Cada muestra de suelo no rizosférico se conformó por un *pool* de 4 submuestras. Las muestras fueron transportadas al laboratorio en hielo seco y la homogenización se realizó en el laboratorio. Las muestras se tamizaron a través de una malla de 1 mm para eliminar raíces, residuos vegetales y ayudar a romper los agregados. La extracción de los ácidos nucleicos se realizó el día posterior al muestreo y las muestras de suelo se conservaron -80 °C.

Tabla II.1. Sitios de colección de muestras de suelo y sus abreviaturas usados en este trabajo.

Abreviatura	Lugar de muestreo del suelo	Tratamiento agrícola*
S1 CSD	Establecimiento productivo "La Estrella"	Cultivado bajo siembra directa
S1 NC	Establecimiento productivo "La Estrella"	No cultivado
S2 CSD	Establecimiento productivo "La Negrita"	Cultivado bajo siembra directa
S2 NC	Establecimiento productivo "La Negrita"	No cultivado
S3 CSD	Establecimiento productivo "Criadero Klein"	Cultivado bajo siembra directa
S3 NC	Establecimiento productivo "Criadero Klein"	No cultivado
S4 CSD	Unidad Experimental INTA Balcarce	Cultivado bajo siembra directa
S4 LC	Unidad Experimental INTA Balcarce	Labranza convencional
S4 NC	Unidad Experimental INTA Balcarce	No cultivado
S5 CSD BPA	Pergamino	Cultivado bajo siembra directa y buenas prácticas agrícolas
S5 CSD PPA	Pergamino	Cultivado bajo siembra directa y pobres prácticas agrícolas
S5 NC	Pergamino	No cultivado
S6 CSD BPA	Viale	Cultivado bajo siembra directa y buenas prácticas agrícolas
S6 CSD PPA	Viale	Cultivado bajo siembra directa y pobres prácticas agrícolas
S6 NC	Viale	No cultivado

* BPA, PPA y NC corresponden a suelos con buenas prácticas agrícolas, pobres prácticas agrícolas y no cultivado, respectivamente.

Extracción de ADN, preparación de bibliotecas y secuenciación

1. Procedimiento para la obtención de ADN metagenómico de suelo

Para la extracción del ADN total del suelo se utilizó un método de extracción directa, que presenta la ventaja de que los ácidos nucleicos obtenidos no depende de un paso previo de separación de los microorganismos de la matriz del suelo [1]. Por otro lado, la utilización de *kits* comerciales específicamente diseñados para esta tarea aporta mayor reproducibilidad y estandarización. Básicamente, la mayoría de estos

kits se basan en dos etapas: Comienzan con una ruptura mecánica-química de las células en el suelo, y luego la purificación del ADN con columnas de afinidad.

Tabla II.2. Propiedades físicas y clasificación de los suelos de acuerdo al sitio-tratamiento agrícola. Los valores corresponden al promedio de tres mediciones para los suelos S1 a S4 y para los suelos S5 y S6 al promedio de las mediciones para cada submuestra.

Lugar de muestreo	Parcela muestreada*	Arcilla%	Limo%	Arena%	Textura	Clasificación del suelo
LA ESTRELLA (S1)	S-Directa	19.28	51.00	29.72	Franco	Argiudol típico
	No cultivado	22.28	49.00	28.72		
LA NEGRITA (S2)	S-Directa	24.28	49.00	26.72	Franco limoso	Argiudol típico
	No cultivado	23.28	51.00	25.72		
CRIADERO KLEIN (S3)	S-Directa	14.28	46.00	39.72	Franco	Argiudol típico
	No cultivado	20.28	42.00	37.42		
BALCARCE (S4)	Labranza C.	15.33	33.33	51.33	Franco	Argiudol típico Paleudol petrocálcico
	S-Directa	15.67	40.33	44.00		
	No cultivado	11.33	42.00	46.67		
PERGAMINO (S5)	S-Directa BPA	22.8	60.2	16.9	Franco limoso	Argiudol típico
	S-Directa PPA	21.1	64.3	14.6		
	No cultivado	21.8	62.3	15.9		
VIALE (S6)	S-Directa BPA	47.3	51.1	1.6	Franco arcillo limoso	Vertisol
	S-Directa PPA	40.2	57.8	1.9		
	No cultivado	41.6	56.7	1.7		

* S-Directa y Labranza C. corresponde respectivamente a siembra directa y labranza convencional.

La extracción de ADN de las muestras de los suelos S1 a S4 (10 gramos) se realizó con el *kit* comercial *MaxSoil DNA Isolation Kit* (MO BIO Laboratories, Inc, CA, USA), tales extracciones fueron realizadas por el grupo liderado por el Dr. Martín Vázquez del consorcio *SoilGene* [59]. Para la extracción de ADN de los suelos S5 y S6 (0.25 gramos) utilizamos el *kit* *FastDNA kit for soil* (MP Biomedicals, Solon, OH, USA) siguiendo las instrucciones del fabricante. Para disminuir la co-extracción de compuestos que inhibitorios en las reacciones enzimáticas (ej. ácidos húmicos) durante la extracción de ácidos nucleicos [23], se realizó un paso adicional de purificación durante la extracción, utilizando polivinilpolipirrolidona PVPP. Este compuesto presenta la capacidad de retener compuestos húmicos y disminuir la cantidad de contaminantes en la solución final del ADN. El ADN se eluyó con buffer Tris 10 mM y se almacenó a -20°C para su posterior utilización. La concentración de ADN en cada muestra de suelo se estimó utilizando el equipo NanoDrop 2000

(Thermo Scientific, Waltham, MA, USA) y la integridad del ADN se confirmó en gel de agarosa 0.8 %.

Tabla II.3 Propiedades químicas de los suelos de acuerdo al sitio –tratamiento agrícola. Los valores corresponden al promedio de tres mediciones para los suelos S1 a S4 y para los suelos S5 y S6 al promedio de las mediciones para cada submuestra.

Lugar de muestreo	Parcela muestreada*	pH	Carbón%	N%	P (ppm)	Nitratos (ppm)**	Humedad%
LA ESTRELLA (S1)	S-Directa	6.1	2.6	0.2	8.3	44.3	22.0
	No cultivado	5.4	6.2	0.4	92.3	62.80	26.10
LA NEGRITA (S2)	S-Directa	5.6	2.7	0.2	11.6	43.00	23.7
	No cultivado	6.2	5.8	0.3	81.7	62.4	26.7
CRIADERO KLEIN (S3)	S-Directa	5.4	2.9	0.2	12.2	39.3	21.3
	No cultivado	5.9	3.6	0.22	31.8	52.30	21.80
BALCARCE (S4)	Labranza C.	5.2	5.6	0.2	26.8	11.2	29.0
	S-Directa	5.2	5.3	0.2	25.8	28.4	25.6
	No cultivado	5.6	5.6	0.3	12.4	40.17	34.30
PERGAMINO (S5)	S-Directa BPA	5.6	1.8	0.2	13.3	N/D	19.50
	S-Directa PPA	6.0	1.9	0.1	2.0	N/D	14.98
	No cultivado	5.8	3.0	0.3	4.7	N/D	26.65
VIALE (S6)	S-Directa BPA	6.5	3.6	0.3	20.4	N/D	19.83
	S-Directa PPA	5.9	2.7	0.2	36.3	N/D	9.92
	No cultivado	6.1	3.6	0.3	10.2	N/D	10.58

* S/Directa y Labranza C. corresponde respectivamente a siembra directa y labranza convencional

** N/D corresponde a no disponible.

2. Preparación de bibliotecas metagenómicas

La secuenciación masiva al azar (*whole shotgun sequencing*) de los metagenomas de suelo y la evaluación de la calidad de las lecturas fueron realizadas por el grupo liderado por el Dr. Martín Vázquez [59], [120]. Nosotros accedimos a las secuencias curadas para encarar los análisis bioinformáticos pertinentes. A continuación se describe brevemente la metodología usada en la preparación de las bibliotecas metagenómicas, su secuenciación, y evaluación de la calidad de las lecturas.

Las muestras utilizadas para la preparación de las bibliotecas para secuenciación masiva al azar se obtuvieron de 10 g de suelo y se construyeron a partir de 500 ng de ADN siguiendo las instrucciones del manual de Roche - FLX 454. Las

bibliotecas se amplificaron mediante la amplificación clonal a base de emulsión (emPCR), y se secuenciaron en un secuenciador *Genome Sequencer FLX* (Roche *Applied Science*) de acuerdo con las instrucciones del fabricante, en INDEAR (Rosario, Prov. de Santa Fe, Argentina). Se utilizaron los siguientes manuales de Roche: *Rapid Library Preparation Method Manual* (Octubre de 2010) para preparación de las bibliotecas *shotgun* y *emPCR Method Manual - Lib-A LV* (Enero de 2010) para la captura de las perlas y la amplificación en emulsión. El procesamiento de datos crudos se realizó siguiendo los procedimientos estándar sugeridos por el fabricante. Para mayor información acerca de la preparación de las bibliotecas ver los trabajos de Rascovan *et al.* 2013 y Carbonetto *et al.* 2014 [59], [120]

3. *Procesamiento de secuencias obtenidas por secuenciación masiva al azar*

Los archivos *output* de la secuenciación masiva (extensión SFF) se cargaron en el servidor MG-RAST para el análisis y ejecución de algunos pasos adicionales de filtrado de calidad de las secuencias. Como se describe en el trabajo de Carbonetto *et al.* 2013 [120] se descartaron las lecturas que presentaron dos o más desviaciones estándar respecto del tamaño promedio de las lecturas. Para la remoción de los artefactos de secuenciación (duplicados técnicos) MG-RAST emplea una aproximación tipo k-mer para identificar los 20 caracteres idénticos prefijados en las secuencias [126].

Plataforma bioinformática para el análisis de datos metagenómicos de suelo

1. *Plataforma bioinformática MG-RAST*

El servidor provee un análisis del tipo organigrama que permitió estimar en cada metagenoma la cantidad de categorías funcionales asignadas. Los subsistemas funcionales fueron inferidos mediante una búsqueda por similitud de secuencias contra la base de datos SEED (*Seed Subsystems Annotation*). El valor de corte (*E value*) establecido fue de $1e^{-5}$, con un alineamiento mínimo de 60 aminoácidos y un 80% de identidad. Se adoptó el subsistema funcional con la mayor puntuación para anotar la secuencia incógnita (*query*). Para mayor información acerca de la asignación

de las lecturas a subsistemas funcionales se sugiere ver el apartado “Anotación funcional basada en la tecnología de subsistema” en el Capítulo III. En el entorno MG-RAST se calcularon también las curvas de rarefacción para el nivel de especie utilizando la anotación taxonómica de los subsistemas funcionales (base de datos SEED).

2. *Plataforma bioinformática MGX*

La plataforma bioinformática MGX creada por el Centro de Biotecnología (CeBiTec) de la Universidad de Bielefeld es una facilidad brindada a los grupos de trabajo del CeBiTec. El uso de la plataforma nos fue factible por la colaboración con el grupo alemán dirigido por el Dr. Alfred Puhler en el marco de un Proyecto de Cooperación Bilateral entre MINCyT (Argentina) y DAAD (Alemania). El análisis *in silico* de los metagenomas consistió en una búsqueda por similitud de las secuencias metagenómicas contra la base de datos de referencia construida específicamente para este trabajo y también contra la base de datos de EggNOG (versión 3.0), la cual se compone de 721,801 grupos ortólogos que cubren 1.133 organismos en 41 rangos taxonómicos diferentes [127]. El algoritmo seleccionado fue BlastX versión 2.2.24+, E *value* $1e^{-20}$ y filtro para complejidad de secuencias desactivado ('- F F'). La mejor puntuación se usó para la anotación de la secuencia incógnita.

Para la construcción de una base robusta de datos de referencia, con secuencias de genes codificadores de enzimas claves del metabolismo del nitrógeno, se utilizó el reservorio de secuencias *FunGene* (*Functional gene Pipeline & Repository*) v6.1. Brevemente, *FunGene* consta de secuencias de familias de genes y proteínas (de ciclos biogeoquímicos, biodegradación, resistencia a antibióticos, marcadores filogenéticos, entre otros) de acceso público y actualización mensual [128]. Emplea una estrategia de análisis comparativa basada en modelos para construir la base de datos, permitiendo el estudio de la diversidad funcional y filogenética de familias de genes específicas. Esta estrategia se basa en el uso de HMMER3 (<http://hmmer.org/>) y modelos ocultos de Markov (HMM), la misma herramienta utilizada por Pfam para el estudio de homología de proteínas. Los HMMs se construyen a partir de alineamientos semillas de proteínas. La utilización de estos

modelos permite que se creen grupos de secuencias (obtenidas de GenBank), clasificadas y alineadas.

Los parámetros seleccionados para la creación de las base de datos personalizada para el análisis funcional de los metagenomas fueron: *Score* = 250 o más, largo de secuencia aminoacídica > 200 y *Evalue* > e^{-10} . Se obtuvieron un total de 17.000 secuencias descargadas de *Fungene* (<http://fungene.cme.msu.edu>) y correspondieron a los genes: *-nifH* y *nifD* para el estudio de la fijación de nitrógeno; *-amoA* para la nitrificación, *-narG* y *-napA* para la reducción de nitrato los genes, *-nirA* y *-nirB* para la reducción de nitrito y *-nirK*, *-nirS*, *-norB* y *-nosZ* para la desnitrificación (Tabla II.4)

Para realizar una comparación de las abundancias de los genes funcionales - medida como número de secuencias asignadas- entre los metagenomas de suelos con y sin historia de manejo se normalizaron las abundancias por tamaño de datos con respecto al metagenoma que presentó el menor número de secuencias curadas que correspondió a S4 LC N= 684.967. Por otro lado, para la comparación de las abundancias relativas de cada categoría funcional dentro de un metagenoma se dividió el número de *hits* de cada categoría funcional (ejemplo categoría COG) por el número total de secuencias curadas en cada muestra. Para evaluar las significancia se utilizó la prueba de *Student* en el programa GraphPad Prism 5.

Técnica molecular de tipo huella digital T-RFLP

1. Amplificación del gen nifH

Las reacciones de PCR se hicieron por triplicado y en forma independiente por cada replica de al menos dos muestras de suelo (réplicas biológicas) por cada sitio-tratamiento, y los amplicones de un mismo sitio-tratamiento se combinaron en una sola muestra que fue utilizada en la etapa de purificación del ADN en gel de agarosa.

Para la amplificación por PCR de los genes *nifH* se utilizó el protocolo de PCR anidada descrito por Yeager *et al.* 2004 [129]. Introdujimos una modificación menor que consistió en no diluir el producto de la primera amplificación para su uso como templado de la segunda PCR. Se utilizaron cebadores específicos del gen *nifH* (Tabla

II.5) y se amplificó por PCR anidada una región de 359 pares de bases. Para la primera reacción de amplificación se utilizó el cebador directo 19F y el cebador reverso *nifH3* en una concentración final de 10 μ M. La mezcla de reacción en un volumen final de 25 μ L fue: 200 μ M de dNTP's (Invitrogen), 2.5 mM de $MgCl_2$ (Promega), 10 μ g de albumina sérica bovina o BSA (*albúmina sérica bovina*, Promega), 1.25 U de Go Taq Polimerasa (Promega) y 20 ng de ADN como molde.

Tabla II.4. Genes seleccionados para el análisis funcional *in silico* del metabolismo nitrogenado en suelo. Las secuencias de referencia para cada uno de estos genes se obtuvieron del reservorio de secuencias *FunGene*.

Gen (abreviatura)	Enzima codificada	Proceso metabólico
<i>nifH</i>	Dinitrogenasa reductasa	Fijación de nitrógeno
<i>nifD</i>	Nitrogenasa cadena α	Fijación de nitrógeno
<i>amoA</i>	Amonio monooxigenasa	Nitrificación
<i>nirB</i>	Nitrito reductasa	Reducción desasimilatoria de nitrito
<i>nirA</i>	Nitrito reductasa	Reducción asimilativa de nitrito
<i>napA</i>	Nitrato reductasa periplásmica	Denitrificación
<i>narG</i>	Nitrato reductasa de membrana	Denitrificación
<i>nirK</i>	Cu-nitrito reductasa	Denitrificación
<i>nirS</i>	Citocromo Cd1-nitrito reductasa	Denitrificación
<i>norB</i>	Óxido nítrico reductasa	Denitrificación
<i>nosZ</i>	Óxido nitroso reductasa	Denitrificación

Para la segunda PCR, se utilizó el cebador directo *nifH11* marcado con 6-carboxifluoresceína (5-FAM) y el cebador reverso *nifH22*. La composición y volumen de la mezcla de reacción, fueron idénticos a la descrita anteriormente, excepto que contenía 2.0 mM $MgCl_2$. Para la segunda PCR los parámetros de ciclado (Tabla II.6) fueron idénticos a la primera PCR, excepto que la temperatura de hibridación fue de 55 °C y 32 ciclos de reacción. Las reacciones se llevaron a cabo en un termociclador GenePro BIOER. El análisis del resultado de la PCR se efectuó en geles de agarosa

1% (p/v) preparados con buffer TBE 0.5X y teñidos con bromuro de etidio. La electroforesis fue realizada aplicando 100 V durante 15 min (MiniRun GE-100) y luego los geles fueron visualizados en un transiluminador.

Tabla II.5. Secuencia de los cebadores *nifH* utilizados en la PCR anidada. Los 4 cebadores utilizados fueron degenerados (**Y**= C/T; **R**= A/G; **D**= A/G/T; **N**= A/C/G/T; **W**=A/T; y **I**= inosina). Para la segunda PCR, el cebador *nifH11* fue marcado con fluorescencia.

	Secuencia	Nombre	Posición ^a	Deg. ^b	Ref. ^c
1^{er.}	GCIWTYTAY GGI AAR GGIGG	19F	19-38	128	[130]
PCR	ATR TTRTTNGCNGCRTA	<i>nifH3</i>	494-478	128	[131]
2^{da.}	GAYCCNAAR GCNGACTC	<i>nifH11</i> (5-FAM)	118-134	64	[129]
PCR	ADWGCC ATCATY TCRC	<i>nifH22</i>	460-476	24	[129]

a. La posición es relativa a la *A. vinelandii nifH* (GenBank ACCN # M20568)

b. La degeneración se da como el número de oligonucleótidos que comprenden el cebador.

c. Referencias bibliográficas en el que se describen los cebadores.

2. Purificación del *pool nifH*

Se sembró en un gel preparativo de agarosa 1%, un volumen de 80 µl de cada muestra compuesta y se cortó la banda correspondiente al tamaño *nifH* (359 pb). Posteriormente, para la purificación de los ácidos nucleicos del taco de agarosa se utilizó el *kit* comercial *NucleoSpin® Extract II* (Macherey-Nagel) y los ácidos nucleicos se eluyeron en un volumen final de 30 µl.

3. Digestión del ADN-*nifH* con endonucleasas

Para identificar polimorfismos en la secuencia de gen *nifH*, se realizaron digestiones con las endonucleasas *MspI* y *HaeIII* (con secuencia de reconocimiento de corte de 4 pb). Las condiciones de restricción en un volumen final de 20 µl fueron las siguientes: *Buffer B* (Promega) 1X, *MspI* o *HaeIII* (Promega) 2.5U, BSA (Promega) 1X y 200 ng de ADN. El ADN fue el producto PCR marcado y purificado. La mezcla de reacción fue incubada a 37°C durante toda la noche.

Tabla II.6. PCR del gen *nifH*. Condiciones de ciclado utilizadas para amplificar un fragmento del gen *nifH*.

Condiciones de ciclado		
	Temperatura (°C)	Tiempo
Desnaturalización inicial	95	5 min.
Desnaturalización	94	45 seg.
Hibridación	48 (1 ^{er.} PCR) / 55 (2 ^{da.} PCR)	1 min.
Extensión	72	1 min
Extensión final	72	10 min

4. Purificación de los productos de restricción *nifH*

La inyección de muestras ADN en la electroforesis capilar puede realizarse por dos métodos. Una es la inyección hidrodinámica que requiere una diferencia de presión sobre el capilar; y la otra alternativa es la denominada inyección electrocinética. Esta última tecnología utiliza una combinación de electroforesis y electroendosmosis, por lo que la presencia de iones puede interferir con la toma del ADN, debido a una inyección preferencial de moléculas de mayor carga-tamaño (ej. Iones Cl⁻). Dado que en el presente trabajo la inyección de las muestras fue del tipo electrocinética, fue necesario desalar previamente las muestras conteniendo los productos de digestión. Con este propósito se utilizó el kit comercial *Quiaquick Nucleotide Removal Kit* (Qiagen) siguiendo las instrucciones del fabricante.

5. Resolución de los fragmentos de restricción terminales del gen *nifH* por electroforesis capilar en secuenciador automático

La detección y cuantificación de los fragmentos de restricción terminales (T-RFs) se realizó en un secuenciador automático de 16 capilares modelo ABI3130XL de *Applied Biosystems*. Características de la inyección: El tiempo fue de 15 seg., el voltaje de 1,6 kV; y el tiempo de corrida fue de 30 min. En cada corrida capilar además de las muestras se inyectó el marcador de tamaño GS500 (-250). Las corridas de electroforesis capilar fueron llevadas a cabo en la Unidad de Genómica, Instituto de Biotecnología, INTA Castelar.

6. Análisis de los perfiles T-RFLP con el Programa PeakScannerTM v1.0

Los datos generados a partir de la electroforesis capilar en secuenciador automático se analizaron mediante el software gratuito *PeakScanner*TM v1.0, con el cual fue posible evaluar la calidad de corrida de los perfiles T-RFLP.

Se realizó un análisis preliminar de fragmentos de restricción terminal de ADN, que permite separar una mezcla de fragmentos de restricción de acuerdo a sus tamaños y proporcionar un perfil de la separación (electroferograma). Además, con este programa es posible calcular con precisión los tamaños de los fragmentos al compararlos con un control interno de tamaño inyectado en combinación con las muestras. Por lo tanto, con el programa *PeakScanner*TM v1.0 se visualizó, editó y exportaron los datos de análisis de fragmentos en un formato tab-delimitado los cuales fueron examinados en el entorno de Excel (Microsoft office).

7. Alineamiento de los perfiles T-RFLP con la herramienta T-Align

Se utilizó la herramienta T-Align disponible en la red y de acceso gratuito para comparar los diversos perfiles de T-RFLP y las abundancias de los T-RFs [132]. Los T-RFs se estandarizaron de acuerdo a lo descrito por Smith *et al.* 2005 [132]. En una primera instancia se eliminaron los picos no consistentes entre réplicas técnicas (picos espurios o pseudo T-RFs) generando un perfil único de consenso para cada muestra. Seguidamente, se compararon los perfiles consenso que representaban diferentes comunidades para generar una lista en la cual se detallan si un fragmento de restricción terminal está presente en una muestra particular y su intensidad de fluorescencia relativa. Los T-RFs cuyos tamaño difirieron en 1 pb o menos fueron considerados idénticos y fueron alineados por el algoritmo media móvil (*moving average algorithm*).

Brevemente se describe como fueron alineados los perfiles T-RFLP. Para el alineamiento entre los perfiles réplica de una muestra en particular, se aplica el algoritmo media móvil (*moving average*). En primera instancia se selecciona el T-RF de menor tamaño y se lo marca como “usado”, subsecuentemente se busca en los otros perfiles réplica de la muestra los T-RFs que presenten una tamaño mayor de hasta 1 pb respecto del T-RF marcado. Cada perfil solo puede contribuir con un único T-RF para el alineamiento general, y se determina el tamaño promedio de los T-RFs

identificados de esta manera. Aquellos perfiles réplica que no hayan contribuido con un T-RF en la búsqueda inicial son examinados nuevamente en la búsqueda de T-RFs cuyos tamaños se encuentren dentro de ± 1 pb del tamaño promedio del T-RF previamente definido. Si en esta búsqueda se identifica un nuevo T-RF el proceso de búsqueda se repite utilizando otro tamaño promedio de T-RF. Todos los T-RFs utilizados en este promedio se marcan como “usados”. Sin embargo, si no se encuentra un T-RF en esta búsqueda, se repite el proceso con el T-RF de menor tamaño “no usado” entre todos los perfiles réplica. Como resultado se obtiene un perfil alineado que contiene el tamaño promedio de todos los T-RFs hallados con este algoritmo. Para la construcción del perfil consenso de cada muestra, todos los T-RFs de los perfiles réplica fueron marcados nuevamente como no “usados”. Comenzando con el T-RF de menor tamaño, cada T-RF en todos los perfiles réplica de la muestra se compararon con el primer T-RF del perfil alineado generado previamente. El T-RF más pequeño en cada perfil réplica con un tamaño ± 1 pb respecto del T-RF del perfil alineado se marca como “usado”. Únicamente, si se encuentra un T-RF *match* presente en cada uno de los perfiles réplica, este será retenido en el perfil consenso final y la fluorescencia del pico (su abundancia) de cada T-RF réplica *match* se utilizará para calcular el promedio de la fluorescencia de tal T-RF consenso. Esto resulta en un perfil consenso que contiene únicamente los T-RFs que detectados en todos los perfiles de las réplicas con su correspondiente área de pico promedio. Seguidamente, se normalizan las áreas de los picos, cada valor representa el porcentaje del total de la fluorescencia. Mayor información sobre el algoritmo *moving average* puede encontrarse en el manual de aplicaciones de T-Align en <http://inismor.ucd.ie/~talign/>.

Extracción de ARN de muestras de suelo y RT-PCR nifH

El ARN total fue extraído a partir de 2 g de suelo de dos réplicas biológicas, utilizando el kit comercial *RNA PowerSoil® Total RNA Isolation* de MoBio (MO BIO Laboratories, Inc, CA, USA). En el protocolo usado se introdujeron las siguientes modificaciones: i) se extendió el tiempo de agitación con vórtex a 20 minutos durante la homogenización y lisis celular y ii) la precipitación del ARN se realizó durante toda la noche a -20°C . Con el objetivo de aumentar el rendimiento de ARN por muestra, en

la etapa final de elución del ARN de la columna, se cargaron en una única columna a los dos volúmenes de solución de ARN provenientes de las preparaciones a partir de ambas replicas biológicas. Después de la purificación del ARN, fue tratado con DnasaI *Amplification Grade* (InvitrogenTM, Carlsbad, CA, USA) para eliminar el ADN residual, y su calidad y concentración se evaluaron en una electroforesis en gel de agarosa y por espectrometría UV utilizando el instrumento NanoDrop 2000 (Thermo Scientific, Waltham, MA, USA). El ADNc se sintetizó a partir de 200 ng de ARN con la transcriptasa reversa 5X *SuperScript III Reverse Transcriptase* (InvitrogenTM, Carlsbad, CA), 2 pM *nifH3* [131] y el resto de los reactivos según las instrucciones del fabricante. El volumen final en la reacción fue 20 µl. La temperatura de hibridación fue 55 °C.

1. Generación de amplicones ADNc-*nifH* para pirosecuenciación

En trabajos previos se ha informado que los niveles de expresión *nifH* en suelos son relativamente bajos, lo cual fue tenido en cuenta para este estudio [110], [133]. Con el propósito de aumentar la sensibilidad de detección de secuencias ARNm- *nifH*, se utilizó la RT-PCR propuesta por Yeager *et al.* 2004 [129], consistente en una opción de PCR de tipo anidada. Brevemente, en la primera PCR se utilizó como molde 2 µl de ADNc-*nifH* y los cebadores 19F [134] y *nifH3* [131]. En la segunda PCR se tomaron 2 µl de la primera PCR y los cebadores *nifH11* y *nifH22* [129] modificados con los adaptadores para la pirosecuenciación (remarcado en negritas) TAG-*nifHf* 5' **CACGACGTTGTAACGAC** GAY CCN AAR GCN GAC TC 3' y TAG-*nifHr* 5' **CAGGAAACAGCTATGACC** ADW GCC ATC ATY TCR CC 3'. Se amplificó utilizando 1.25U de GoTaq Polimerasa (Promega, Madison, WI, USA).

Se realizaron ensayos para evaluar en el ARN la posible contaminación con ADN residual, consistentes en reacciones de PCR anidada a sub-muestras de ARN previas al tratamiento con la transcriptasa reversa y además con la reacción RT-PCR a las muestras de ARN tratadas con RNasa A (Promega).

Con el propósito de disminuir el posible sesgo que podría tener lugar en la PCR anidada, se realizaron por triplicado las amplificaciones PCR para cada muestra.

Después de la segunda PCR, los amplicones de las tres repeticiones de una misma muestra fueron combinados en una único *pool* de amplicones. La muestra de ADN-amplicones (80 μ l) fue sembrada en un gel preparativo de agarosa 1% para separar el ADN a partir de la banda correspondiente a *nifH* de 359 pb. Esta etapa de separación fue necesario dado que se observó la co-amplificación de una banda tenue de aproximadamente 900 pb, que fuera descripta anteriormente [135]. Se purificó el ADN *nifH* del taco de agarosa con el *kit* comercial *NucleoSpin® Extract II* (Macherey-Nagel) y el ADN eluido en un volumen final de 30 μ l. Los amplicones de las réplicas fueron combinados en cantidades equimolares en una única muestra antes del marcado con las secuencias identificadoras (*barcodes*), lo que resultó en 6 librerías independientes que fueron secuenciadas con la tecnología GS FLX 454 en la Unidad de Genómica y Bioinformática, INDEAR, Rosario.

2. Procesamiento y asignación taxonómica de las secuencias ADNc-*nifH*

El procesamiento de las secuencias primarias se llevó a cabo utilizando las herramientas de la plataforma Mothur [72]. Se eliminaron de la base de datos a las lecturas que no cumplieran con los siguientes criterios: - tamaño entre 300 y 360 pb; - diferencia en la secuencia (*mismatch*) del *barcode*; - más de dos *mismtach* en región de la secuencia correspondiente a los cebadores; - homopolímeros de más de 8 bp y - un promedio de calidad de lectura menor o igual a 25. Con el mismo software se eliminaran las secuencias que presentaron quimeras mediante el método UCHIME [136]. Los corrimientos en los marcos de lectura (*frameshift*) fueron detectados con el programa *Framebot* (RDP, Ribosomal Database Project) usando el perfil *Hidden Markov Model* (HMM) *Fer4_NifH_fs.hmm* como referencia [52]. Mediante este programa también obtuvimos las secuencias aminoacídicas de *nifH* que fueron agrupadas en unidades taxonómicas operativas (OTUs) usando el programa CD-HIT con un 98% de similitud a nivel de aminoácidos [71]. Se adoptó este umbral de semejanza, considerando el grado alto de similitud de secuencia a nivel de especie que generalmente presentan los genes codificadores de proteínas [137].

Para reducir aún más los potenciales errores asociados a la pirosecuenciación, seleccionamos para los futuros análisis a aquellos OTUs con representación de al menos tres secuencias en la base de datos curada. La abundancia relativa de las

secuencias de los OTUs se normalizó mediante el denominado método *subsampling* descrito en Mothur (<http://www.mothur.org/wiki/Normalize.shared>) previo a los análisis comparativos. Los OTUs compartidos entre los distintos tratamientos y profundidad de suelo se visualizaron con diagramas de Venn, construidos en el entorno Mothur (<http://www.mothur.org/wiki/Venn>). Se alinearon las secuencias representativas de cada OTU s utilizando como referencia el perfil *Hidden Markov Model* (HMM) Fer4_NifH_fs.hmm en el entorno ARB [75]. Los alineamientos fueron examinados manualmente, excluyendo las secuencias que presentaron regiones con un alineamiento ambiguo en los conocidos dominios altamente conservados del gen. Finalmente, la asignación taxonómica y las relaciones filogenéticas de los OTUs fueron realizadas mediante la construcción de árboles filogenéticos con la secuencia aminoacídica representativa de cada OTU e incluyendo también a secuencias *nifH* de la base de referencia de Zehr *et al.* 2003 [138]. Se aplicó el análisis de *Neighbour-joining* (con corrección de Kimura) usando el programa ARB.

3. Comparación de las base de datos ADN-*nifH* y ADNc-*nifH*

Para la comparación de ambas comunidades diazotroficas, la potencial y la activa, se realizó un agrupamiento en OTUs, de las secuencias aminoacídicas de ambas bases de datos *nifH* (ADN y ADNc) con el valor de *cutoff* del 98%. Los OTUs representados con al menos tres secuencias fueron seleccionados para los análisis siguientes usando la metodología descrita en el la sección precedente. Se llevó a cabo la asignación taxonómica de los OTUs mediante arboles de filogenia en el entorno de ARB.

PCR cuantitativa (qPCR) de grupos específicos de *nifH*

1. Diseño de cebadores específicos para taxones

El diseño de los cebadores para el OTU 755 con clasificación taxonómica 1B y para los 6 OTUs agrupados en el *subcluster* 1A (22, 8, 27, 380,488 y 53), se llevó a cabo mediante una primera etapa de comparación de todas las secuencias agrupadas por similitud dentro de cada uno de los OTUs y la identificación de regiones conservadas entre las secuencias, utilizando el programa Jalview. Se adoptó el valor umbral de 95% como el porcentaje de similitud consenso entre las distintas posiciones

nucleotídicas. Se utilizó la herramienta OligoAnalyzer para evaluar los siguientes parámetros: el contenido de GC, la temperatura de *melting*, formación de dímeros y bucles internos (*hairpins*), disponible en el portal IDT DNA (<https://www.idtdna.com/calc/analyzer>). El análisis *in silico* de especificidad de los cebadores se realizó por comparación con la base de datos pública NCBI, mientras que experimentalmente se los evaluó mediante PCR utilizando como templado el ADN de bacterias diazotróficas filogenéticamente alejadas de *Nostoc* y *Geobacter* respectivamente, y las condiciones de ciclado utilizadas en la qPCR que se describen a continuación. Las secuencias de los cebadores utilizados se muestran en la Tabla II.8.

2. Condición de reacción de la qPCR

Se cuantificó la abundancia relativa de estos taxones de *nifH* en el ADN de suelo, mediante qPCR. La composición de la mezcla de reacción qPCR fue: 10 ng de ADN de suelo, 0.8 μ l de cada cebador (5 μ M), 10 μ l de la mezcla 2X SYBER Green iCycler iQ (Bio-Rad), y agua destilada ultra pura hasta un volumen final de 20 μ l. La reacción se llevó a cabo en el termociclador *Applied Biosystem 7500 real-time PCR system* (Applied Biosystems). El programa de ciclado consistió en: 10 min a 95°C, seguido de 40 ciclos que consistieron en -15 seg. a 95°C, -20 seg. a 60°C (grupo 1B)/58 °C (grupo 1A) y -20 seg. a 72 °C. La fluorescencia se midió al final de cada ciclo. Para normalizar los niveles *nifH* en las diferentes muestras se utilizó el nivel de ARNr 16S para lo cual se amplificó con los cebadores universales para bacterias 338F-518R [139]. Para el cálculo de los niveles relativos de *nifH* se utilizó la herramienta GENORM (<http://medgen.ugent.be/~jvdesomp/genorm/>). Todas las reacciones de PCR se corrieron por duplicado. Se analizó la curva de *melting* de cada corrida qPCR para asegurar el desarrollo correcto del ciclo de amplificación.

Estimación de la diversidad

La diversidad de la comunidad diazotrófica activa (α diversidad) en cada una de las muestras tratamiento-profundidad de Pergamino fue evaluada mediante las curvas de rarefacción utilizando el método de re-muestreo sin reemplazo comprendido en el entorno Mothur. Los índices de riqueza Chao1 y de diversidad Shannon-Wiener

(H') se calcularon a partir de la matriz de los OTUs con abundancias normalizadas usando el programa de libre acceso PAST3 [140]. Para el análisis comparativo de la composición entre muestras (β diversidad) se aplicó la medida de disimilitud Bray-Curtis que tiene en cuenta tanto la ausencia como la presencia de OTUs [76] usando el programa PAST3. La distancia *weighted UniFrac* mide la distancia filogenética entre par de OTUs como la fracción del largo de la rama del árbol que lleva al descendiente de una muestra o de la de otra, pero no de ambas.

Tabla II.8. Cebadores utilizados en la qPCR.

Cebador	Blanco	Subcluster <i>nifH</i>	Secuencia (5'–3')	Tamaño de amplicón (pb)
Cyano FW	OTU 755	1B	TGG TAT CAT CAC CGC CAT CAA CTT C	116
Cyano Rev	OTU 755	1B	GAA TTG GCA TAG CGA AAC CAC CG	
Geo Fw	OTUs 22,8,27,380,488 y 53	1A	GCC AAR GCG CAG AAT ACG GT	314*

* Tamaño esperado de amplicón con el cebador reverso *nifH2* [141].

La versión *weighted* del algoritmo tiene en cuenta, además, las abundancias de OTUs en el cálculo de las distancias (<http://unifrac.colorado.edu/>). Las matrices de diversidad beta, tanto basadas en UniFrac como en Bray Curtis, se utilizaron para el Análisis de Coordenadas Principales (PCoA) en PAST3. La significancia estadística de la similitud entre las comunidades fue evaluada con el *test* estadístico *weighted* UniFrac. Para determinar la significancia de diversidad y riqueza entre las distintas muestras tratamiento–profundidad se aplicaron los *tests* no paramétricos ANOSIM [142] del paquete estadístico InfoStat versión 2016 [143].

Análisis Multivariados

En el presente trabajo hemos seleccionado algunos de los métodos del análisis multivariante, los cuales nos permitieron realizar representaciones gráficas,

cuantificar la asociación entre individuos y clasificar respecto a un conjunto de variables observadas, tanto cuantitativas como cualitativas.

A modo de resumen se presentan en la Tabla II.7 los diferentes criterios utilizados en este trabajo para ordenación y clasificación, y dentro de los primeros agrupamos de acuerdo al tipo de variable y medida de distancia asociada. Para llevar a cabo los métodos de análisis multivariante utilizamos el programa PAST3. A continuación se describen los sistemas de análisis y sus características y propiedades usados en el tratamiento de los datos de secuencias.

a. Métodos de Ordenación

Para el análisis de ordenación se utilizó la medida de distancia de Bray-Curtis. Brevemente, en general el objetivo de un análisis de ordenamiento es representar en un número reducido de ejes ortogonales, los datos multivariados provenientes de múltiples variables de una misma muestra tales, como la abundancia de OTUs y la distancia ecológica entre comunidades, [144]. En el gráfico de ordenamiento los puntos que representan a las muestras se ubican en forma tal que las distancias entre ellos representan la distancia ecológica entre las comunidades muestreadas [79].

Por otro lado, así como el análisis de componentes principales es adecuado para variables cuantitativas, cuando las variables son cualitativas resulta más apropiado aplicar un tipo de análisis diferente, tal como el de coordenadas principales [145]. El objetivo del análisis de coordenadas principales es llegar a una representación geométrica de los n individuos a través de una medida de distancia que respete la estructura de similaridad definida por la matriz de similaridad S . El análisis de coordenadas principales no es más que un análisis de componentes principales aplicado a una tabla de datos compatibles con la distancia $d_{ij}^2 = s_{ii} + s_{jj} - 2s_{ij}$. Pero, aquí el sistema de ejes no representa direcciones dado el carácter cualitativo de las variables originales.

Las escalas multidimensionales MDS -en el presente trabajo utilizamos en particular MDS no métricas que abreviaremos NMDS- es un método de ordenación robusto que nos permite representar, la (dis)similitud entre pares de objetos en un espacio métrico de pocas dimensiones, con el mayor grado de confianza posible [146]. Se puede utilizar cualquier de los coeficientes conocidos de disimilitud o

medida de distancia para construir la matriz a utilizar como “*input*” del análisis NMDS.

b. Métodos de clasificación

El análisis de conglomerado jerárquico presenta un enfoque algorítmico para identificar grupos discretos con diversos grados de (dis)similitud en una base de datos representada por una matriz de (dis)similitud. En este trabajo se realizaron los análisis de conglomerado mediante el método de agrupamiento UPGMA aplicando la medida de distancia de Bray-Curtis. Brevemente, los grupos son jerárquicamente organizados a medida que el algoritmo de agrupación avanza y pueden ser representados con un dendrograma. Los dendrogramas indican como se asocian los individuos/parámetros objeto del análisis a medida que disminuye el grado de semejanza entre ellos. Estos árboles generalmente van acompañado por un eje que indica la distancia y/o similaridad a la que se han producido las diferentes uniones \denominado índice de jerarquía. En las publicaciones de Culman *et al.* 2008 y Ramette *et al.* 2007 [146], [147] se encuentra una revisión más amplia de los métodos del análisis multivariante que son utilizados en estudios de ecología microbiana.

Tabla II.7. Métodos del Análisis Multivariado

Métodos de Ordenación	Variables Cuantitativas	Distancia Euclídea	Análisis de Componentes Principales (PCA)
	Variables Cualitativas	Índices de Similitud	Análisis de Coordenadas Principales (PCoA)
	Variables Cualitativas	Indices de Disimilitud	Análisis de Proximidades (NMDS o Escalas Multidimensionales No métricas)
Métodos de Clasificación	Variables Cuantitativas o Cualitativas	Indices de distancia o similitud	Análisis de conglomerados (Clúster) Jerárquico

Capítulo III

**Análisis *in silico* de las
comunidades microbianas
involucradas en el ciclo del
nitrógeno en suelos de la Región
Pampeana Argentina**

Introducción

El avance en el uso de metodologías basadas en la secuenciación de ácidos nucleicos aislados de muestras ambientales, ha generado un importante volumen de datos provocando un gran avance en los estudios de ecología microbiana. Principalmente, estas técnicas se basan en el análisis de marcadores moleculares tales como el perfil de ácidos grasos de fosfolípidos (PLFA) y ácidos nucleicos (ADN y ARN), los cuales se encuentran presentes en todas las células y su extracción no requiere al cultivo. La utilización de estas técnicas permite determinar la composición de las comunidades microbianas en general, y además cuantificar la abundancia de microorganismos específicos. Los estudios pioneros de la complejidad microbiana de suelos, realizados por Torsvik y colaboradores utilizando técnicas de reasociación de ADN, revelaron que la diversidad genética es relativamente muy grande y así el número de genomas bacterianos diferentes estimados en una muestra de suelo forestal es de unos 4.000, en comparación con la diversidad genética de bacterias aisladas de cultivo a partir de la misma muestra (unas 200 veces menor) [148]. En otros estudios sobre muestras de suelo realizados por el mismo autor, la diversidad genómica fue incluso superior (8.000–10.000 genomas diferentes) [9].

El alto nivel de diversidad relativo de la comunidad microbiana que ha sido estimada para el hábitat suelo, ha despertado el interés en laboratorios de todo el mundo en descifrar la estructura y diversidad poblacional de las comunidades microbianas asociadas a este ambiente. Hasta hace pocos años el proceso adoptado para revelar microorganismos, descubrir nuevas especies, y más aún determinar la estructura poblacional de un nicho ambiental, se limitaba al uso de procedimientos experimentales consistente en general, en la incubación de una alícuota de la muestra ambiental en diferentes medios de cultivo tanto generales como específicos. Se disponía de material vivo para encarar su identificación y caracterización morfológica y fisiológica, y eventualmente lograr su identificación taxonómica. En los casos/intentos en que los ensayos de aislamiento y cultivación lograran reunir las condiciones de incubación (ej. temperatura, pH, presencia o ausencia de oxígeno, etc.)

y además se contara con el medio de cultivo que satisficiera los requerimiento para la multiplicación celular. Esta estrategia clásica -relativamente económica- ha sido muy eficiente en el descubrimiento de microorganismos y sus actividades metabólicas, tales como los microorganismos del suelo productores de antibióticos y de otros procesos que han sido derivados a aplicaciones tecnológicas.

Actualmente, con el uso de tecnologías de secuenciación masiva y profunda tales como la pirosecuenciación, se ha acumulado elevada información a partir de los ácidos nucleicos presentes en muestras ambientales. En particular, con la amplificación del gen marcador de filogenia como el codificante para la subunidad ribosomal ARNr 16S o de algún gen marcador de función biogeoquímica específica a partir del ADN ambiental, se ha demostrado la riqueza extraordinaria de tipos filogenéticos y capacidades metabólicas en muchos hábitats microbianos [45], [46], [149], [150]. El estudio global de los genomas de todos los microorganismos presentes en una comunidad se denomina metagenómica [60]. Mediante la secuenciación masiva al azar “*whole genome shotgun sequencing*” (WGS) se ha logrado de una base de datos metagenómicos cuantitativamente sustancioso de varios ambientes tales como suelos, océanos, intestinos de insectos entre otros [62], [64], [120], [151], [152] la cual nos es accesible para su análisis. El volumen de información generado en estos tipos de estudios masivos ofrece la posibilidad de su uso para encarar estudios ecológicos particulares. La literatura muestra diversos reportes que ilustran sobre las contribuciones de aplicar este enfoque al conocimiento de la estructura poblacional y funcionalidad de las comunidades microbianas en ambientes naturales [24], [61], [135], [153].

Uno de los desafíos importantes que aparece en los estudios metagenómicos es aquel asociado con el análisis de un número muy grande de datos. Además, comprende al elevado número de secuencias, la heterogeneidad genómica, secuencias incompletas, lecturas cortas, elevado polimorfismo y abundancias relativas (comunidades simples vs. complejas). El análisis de esos datos, desde los datos primarios hasta las comparaciones de metagenomas, determina la necesidad de disponer de sistemas y herramientas computacionales apropiadas y eficientes, las cuales en su conjunto se las reconoce como plataformas bioinformáticas.

Se encuentran disponibles diversas plataformas bioinformáticas para la carga de datos primarios y procesamiento de metagenomas y muchas son de acceso libre. Cada plataforma presenta sus protocolos de trabajo propios y diversas opciones de análisis bioinformático. Algunas son el producto parcial de proyectos de análisis metagenómicos encarados por consorcios de varios investigadores. En particular, algunas proveen de herramientas para el control de calidad, anotación, análisis comparativo y archivo de datos metagenómicos. A su vez, otras de las herramientas de análisis facilitan las reconstrucciones filogenéticas y metabólicas en los metagenomas. A título ilustrativo se puede mencionar a las siguientes plataformas: - MG-RAST [126], , -IMG/M [154], -CAMERA [155] -MEGAN [156], -MGX (facilidad computacional diseñada por de la Universidad de Bielefeld, Alemania), entre otras.

Considerando el papel protagónico de las comunidades microbianas en las transformaciones biogeoquímicas, Konopka *et al.* 2009 sostiene que avanzar hacia un entendimiento más profundo de su dinámica es crítico para refinar las predicciones sobre cómo la biosfera modula y responde a variaciones ambientales [14]. A un nivel más básico, la comprensión de la estructura y funcionalidad de las comunidades microbianas naturales contribuiría a una mejoría en nuestro conocimiento sobre el funcionamiento de los ecosistemas, aspecto que representa uno de los mayores desafíos en la ecología microbiana del siglo XXI [157].

En resumen, con la metagenómica se ha estudiado la diversidad taxonómica y metabólica en diversos suelos sujetos a efectos de distintos fenómenos sobre las comunidades microbianas, entre ellos: cambios de clima, manejos agrícolas, efecto invernadero, etc. [59], [152], [158], [159]. En la actualidad se encuentran en desarrollo varios proyectos internacionales analizando la biodiversidad de suelos, mediante la utilización de herramientas de última generación. El proyecto llevado a cabo por el consorcio *SoilGene*, presentó dos objetivos principales 1- evaluar el impacto de la agricultura sobre ecosistemas agrícolas típicos de la Región Pampeana, con un enfoque en el estudio de los procesos biológicos y diversidad microbiana de suelos, y 2- generar una base de datos metagenómicos de suelos de Argentina y generar un set de datos para la bioprospección de enzimas de interés biotecnológico.

Dichos datos metagenómicos fueron utilizados en esta tesis para realizar el análisis *in silico* de las comunidades microbianas involucradas en el ciclo del nitrógeno en suelos bajo distinto uso de la Región Pampeana Argentina.

- **Breve descripción de las plataformas informáticas usadas para análisis de metagenomas de suelo**

a. Plataforma MG-RAST

El servidor MG-RAST v.3.1 (*Metagenomics Rapid Annotation using Subsystem Technology*) es un servidor de acceso público que ha sido creado con el objetivo de proveer un sistema de alto rendimiento para el análisis de datos metagenómicos resultantes de secuenciación masiva (Meyer *et al.* 2008). Este servidor, se basa en el concepto de subsistemas y los datos establecidos por los equipos FIG (*Fellowship for Interpretation of Genomes*) y NMPDR (*The National Microbial Pathogen Database Resource*) lanzando la primera versión de la plataforma bioinformática en el año 2007. Este sitio/servidor se ha enriquecido en el curso de años y el desarrollo de proyectos, y actualmente, posee más de 12.000 usuarios registrados, y alrededor de 189.835 metagenomas. Además, y muy importante, esta plataforma permite a los usuarios la carga de sus propios datos de metagenomas para un análisis automatizado (control de calidad, anotación, análisis comparativo, etc.). También la plataforma MG-RAST permite el almacenamiento de datos metagenómicos.

A continuación se describe brevemente las etapas comprendidas en el proceso de un análisis clásico usando MG-RAST. Como se muestra en la Figura III.1, en primera instancia se realiza el procesamiento de las lecturas resultantes de la secuenciación, eliminándose aquellas secuencias que no cumplan con determinados niveles de calidad y también se descartan otras que representan artefactos generados en la secuenciación. Cabe mencionar que en el presente trabajo nos referiremos a la curación de datos, como el procedimiento mediante el cual las secuencias generadas por secuenciación masiva son sometidos a pasos de filtrado de calidad (largo de secuencia, presencia de homopolímeros, ambigüedad de bases, etc.) obteniéndose una base de datos altamente confiable denominada base de datos curada.

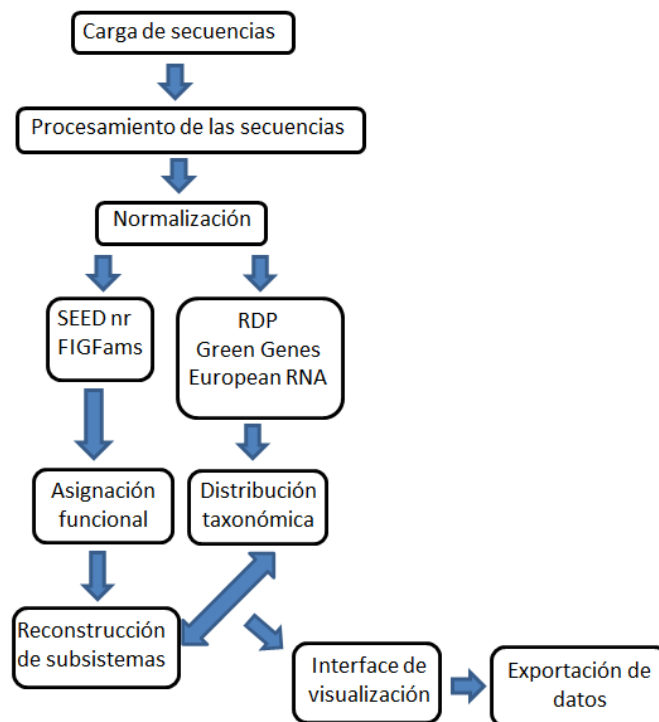


Figura III.1. Esquema del flujo de trabajo implementado en la plataforma MG-RAST. Se muestran las principales etapas de procesamiento de datos. Esta figura es una adaptación de su versión original Meyer *et al.* 2008 [126].

El MG-RAST presenta por defecto valores establecidos para los diversos parámetros de curado de los datos, los cuales a su vez pueden ser modificados por el usuario. Seguidamente, el servidor desarrolla en varios pasos diferentes tareas tales como la normalización, en donde se remueven secuencias exactamente iguales (secuencias duplicadas) generando la rotulación de cada lectura con IDs internos únicos, y encara la búsqueda de potenciales genes codificadores de proteínas (conocido en inglés como PEGs *potential protein codings genes*). Esto último se lleva a cabo a través del empleo del algoritmo BLAT contra la base de datos no redundante del MG-RAST denominada M5NR. Esta base de datos no redundante representa a su vez la integración de otras bases de datos que experimentan actualizaciones continuadas, tales como: GO (*Gene Ontology*), Greengenes (*16S rRNA Gene Database*), JGI (*Joint Genome Institute*), KEGG (*Kyoto Encyclopedia of Gene and Genomes*), NCBI (*National Center for Biotechnology Information*), RDP (*Ribosomal Database Project*), SEED (*Seed Subsystems Annotation*), SILVA (*SILVA rRNA*

Database Project), UniProt (*UniProt Knowledgebase*), VBI (*Virginia Bioinformatics Institute*) y eggNOG (*Non-supervised Orthologous Groups*).

Es posible asumir como un resultado, que muchos de los genes en los metagenomas no encontrarán correspondencia alguna en la base de datos, ya que los mismos no se encuentran aún anotados en alguna de las diversas bases de datos (principalmente por la limitación de cultivo de varios microorganismos), y por lo tanto genes putativos no podrían ser hallados o identificados por búsquedas basadas en homología de secuencias contra bases de datos de referencia. Además, se disponen de otros métodos para predecir ORFs (*Open Reading Frame*) comprendidos en diferentes herramientas como MetaGene [160] y GeneMark [161] que se basan en modelos probabilísticos para evaluar las probabilidades de ocurrencia de regiones codificantes en ciertas regiones de las secuencias de ADN. Se identifican dos grandes limitaciones para la predicción de genes en la colección de datos primarios/lecturas generadas por la pirosecuenciación. Por un lado, el tamaño promedio de los genes en los microorganismo es de 950 pb [160], mientras que el tamaño promedio de lectura alcanzado por la plataforma Roche 454 es aproximadamente de 400 pb, además pudiendo estar representado solo un fragmento del gen.

Por otro lado, las lecturas presentan una alta frecuencia de error de secuenciación de hasta un 3% [162], lo cual puede provocar un corrimiento en los marcos de lectura (*frameshifts*). Por ejemplo, los errores de secuenciación en las lecturas del tipo 454, tienden a ocurrir en las regiones de homopolímeros, resultando en frecuentes inserciones y deleciones. Más aún, se ha demostrado que los métodos de predicción de genes basados en ORFs son afectados mayormente por errores de secuenciación que causan *frameshifts* [162]. Por lo tanto, muchas de las herramientas existentes para predicciones de genes presentaron un decremento en el nivel de su desempeño a medida que aumentó la tasa de error de secuenciación asociadas a la pirosecuenciación. En este contexto, Rho y colaboradores desarrollaron el algoritmo *FragGeneScan* para superar las limitaciones relacionadas a las tasas de error de secuenciación y tamaño de las lecturas, y así aumentar la capacidad de recuperación de secuencias [55]. Se trata de una herramienta basada en un modelo probabilístico que combina modelos de errores de secuenciación y el uso de codones preferenciales

(*codons usage*). *FragGeneScan* es parte de las herramientas implementadas en la plataforma MG-RAST.

En paralelo, a las predicciones de similitud realizadas vía BLAT contra bases de datos de proteínas, las lecturas también se comparan contra bases de datos ribosomales que incluyen a *Greengenes* [163], RDP-II [164] y Silva [165]. El MG-RAST permite realizar la reconstrucción filogenética de la muestra y en combinación con las clasificaciones funcionales, describe modelos metabólicos para el metagenoma sugiriendo caminos y flujos metabólicos, reacciones y enzimas. Además, es posible encarar comparaciones de la composición y capacidades metabólicas entre metagenomas. Esta plataforma bioinformática presenta una interfase de visualización de los análisis filogenéticos y funcionales muy agradables, desde gráficos de barras, *heatmap*, o del tipo Krona. Estos últimos son muy interesantes ya que permiten visualizar de forma dinámica categorías funcionales de distinta jerarquización para un proceso metabólico específico por ejemplo.

a.1 Anotación funcional basada en la tecnología de subsistema

Inicialmente el servidor RAST fue planeado y diseñado para ser usado por la comunidad NMPDR para identificar polimorfismos genéticos correlacionados con la patogenicidad, la resistencia a las drogas, la morbilidad y la ineffectividad [166], pero rápidamente su utilización por varias comunidades científicas fue inminente. El NMPDR fue fundado por el Instituto Nacional de Alergias y Enfermedades Infecciosas (NIAID) de Estados Unidos y es tanto un reservorio central para una extensa variedad de datos científicos de microorganismos patogénicos así como también un centro de recursos bioinformáticos con herramientas de *software* que ofrecen soporte y para el análisis de datos genómicos. La plataforma RAST proporciona una anotación funcional lo más precisa posible en el contexto de un análisis comprensivo y comparativo. Para ello, se integran datos genómicos y funcionales de cientos de genomas completos y parciales de bacterias, archeas y organismos eucariotas, todos los cuales reunidos en los denominados “genomas de soporte”, proveen un contexto filogenético rico para su uso en el análisis de los datos.

Para la correcta interpretación de los datos genómicos (objeto de análisis) es imprescindible lograr una anotación funcional precisa y consistente. La anotación empleada por el NMPDR se realiza en un entorno denominado SEED (<http://www.theseed.org>), el cual organiza los datos genómicos en forma de subsistemas y no sobre la base de organismo por organismo. El concepto de subsistema consiste en una integración bidimensional de funciones biológicas con secuencias genómicas, los cuales se representan en tablas, donde las columnas corresponden a los roles funcionales, las filas a los genomas u organismos y en las celdas se hallan los genes responsables de cada función (ver Figura Suplementaria III.1). Los roles funcionales son en una primera instancia asignados manualmente a un grupo de proteínas, y se lleva a cabo con un número limitado de genomas microbianos. Esto resulta en el denominado subsistema poblado. Se realizan análisis comparativos y de co-localización génica (acoplamiento funcional) para extender los subsistemas a otros genomas, lo que conduce al establecimiento/definición de grupos de proteínas que desempeñan el mismo rol funcional en una variedad de organismos.

Cuando se combinan los subsistemas funcionales y un análisis de conglomerado, es posible identificar genes que codifican funciones conocidas en un organismo, pero que se llevan a cabo por proteínas divergentes las cuales no son reconocidas automáticamente por similitud de secuencias. Para asegurar la integridad de las anotaciones y evitar ediciones el recurso bioinformático NMPDR no permite introducir cambios, es del tipo solo lectura. Contrariamente, el entorno SEED presenta cierto grado de apertura ya que los enlaces públicos se encuentran disponibles para la comunidad científica para la anotación de genes y para la construcción de nuevos subsistemas. Resumiendo, un subsistema se puede describir como un grupo de roles funcionales que juntos conforman/constituyen un proceso biológico específico o alternativamente un complejo estructural. Un subsistema puede ser asimilado al concepto de vía metabólica (*pathway*). Por lo tanto, al igual que la glucólisis se compone de un conjunto de roles funcionales individuales (glucoquinasa, 6-glucosa isomerasa, fosfofructoquinasa, etc.), un complejo como el ribosoma o un sistema de transporte, puede ser considerado también una colección de roles funcionales [167].

La inclusión o incorporación de un gen a un subsistema está determinada por el rol funcional que le fuera asignado, y cada proteína incluida en dicho subsistema tiene una función, por ejemplo puede ser enzimático, de señalización, de regulación, estructural, etc. Una proteína resultará incluida en un subsistema únicamente si su anotación es coincidente con el nombre de un rol funcional ya incluido en tal subsistema. La colección de subsistemas representa la reconstrucción metabólica de un organismo.

En la plataforma MG-RAST la anotación de los datos genómicos se realiza a través de la curación de los subsistemas, y la precisión depende de la calidad de las anotaciones realizadas manualmente por curadores expertos del FIG sobre la base de los genomas de referencia disponibles en las bases de datos mundiales. Los curadores utilizan la literatura y los resultados experimentales recientemente publicados, para mantener actualizada las bases de datos, resultando así en una labor permanente e interactiva. De los subsistemas curados se extrae un conjunto de familias de proteínas (FIGfams del inglés *Fellowship for Interpretation of Genomes Families*) que presenta similitud en toda la extensión de la secuencia. Dentro de este esquema, se asume que las proteínas están asociadas a un rol funcional o más, y todos los miembros de una misma familia FIGfam tienen el mismo rol funcional aunque pertenezcan a variantes divergentes. Los FIGfams son datos de acceso público, y es el componente central para la anotación automatizada del servidor MG-RAST.

b. Plataforma bioinformática MGX

La plataforma bioinformática MGX fue creada por el Centro de Biotecnología (CeBiTec) de la Universidad de Bielefeld para facilitar los estudios genómicos y metagenómicos de los grupos de trabajo del centro de investigaciones CeBiTec. Nos fue posible el uso de esta plataforma informática dentro del marco de la colaboración con el grupo alemán dirigido por el Dr. Alfred Puhler en el marco de un Proyecto de Cooperación Bilateral entre MINCyT (Argentina) y DAAD (Alemania). El análisis funcional *in silico* de los metagenomas que fuera realizado en este contexto, consistió en la comparación de las secuencias metagenómicas desconocidas (*query*) mediante el algoritmo BlastX versión 2.2.24+ contra una base de datos de referencia personalizada y la base de datos eggNOG. La base de datos EggNOG V.3

se compone de 721.801 grupos ortólogos que cubren 1.133 organismos en 41 rangos taxonómicos diferentes [127].

Para la construcción de una base de datos de referencia robusta que define el rasgo distintivo de este análisis complementario de los genomas, se utilizó el reservorio de secuencias *FunGene* (*Functional gene Pipeline & Repository*) v6.1. Nos referiremos a esta base de datos como base de datos personalizada. La base de datos apuntaba claramente a la inclusión de secuencias de genes codificadores de enzimas claves del metabolismo del nitrógeno. En general, *FunGene* se trata de un reservorio de secuencias de familias de genes y proteínas (de ciclos biogeoquímicos, biodegradación, resistencia a antibióticos, marcadores filogenéticos, entre otros) de acceso público y actualización mensual, y resulta de gran ayuda para estudios de genómica funcional, especialmente del medio ambiente [128]. La estrategia de análisis comparativo se basa en el uso de HMMER3 (<http://hmmer.org/>) y modelos ocultos de Markov (HMM), las mismas herramientas utilizadas por Pfam para el estudio de homología de proteínas [168]. Los HMMs se construyen a partir de alineamientos “semillas” de proteínas, y la utilización de estos modelos permite crear grupos de secuencias (obtenidas de *GenBank*), clasificadas y alineadas.

Una de las ventajas que presentó construir una base de datos de referencia personalizada, consistió en la exclusión de secuencias parálogas las cuales no resultan beneficiosas y no corresponden a secuencias codificantes del gen de interés. Contrariamente, cuando la búsqueda de secuencias de genes de interés se realiza en base a similitud contra el NCBI *dataset* en el entorno de análisis de la plataforma MG-RAST, se observó que muchos de los *hits* (Adoptaremos la palabra hit para hacer referencias a la cantidad de lecturas que fueron asignadas a una función de interés) recuperados eran parálogos, por ejemplo en el caso del gen *nifH* se correspondían con el gen *bchl*. En los proyectos de metagenómica que implican el procesamiento y análisis de un alto número de secuencias, encontrar *hits* no verdaderos del gen de interés puede resultar en una sobreestimación de la abundancia de tales genes. En muchos casos no se alcanza a validar los datos de *hits* hallados por tratarse de un volumen de datos considerable.

Objetivo general

Evaluar la base de datos de secuencias de ADN de suelo para la construcción *in silico* de la estructura poblacional de microorganismos del suelo relacionadas con el ciclo del nitrógeno.

Objetivos específicos

- Relevar las bases de datos metagenómicos de suelo: características y tamaño, describiendo además las principales categorías funcionales jerárquicas presentes en los metagenomas de suelo. Comparar los resultados con aquellos obtenidos en el análisis *in silico* de metagenomas de otros ambientes naturales.
- Identificar secuencias codificantes de genes funcionales del metabolismo nitrogenado.
- Realizar la asignación taxonómica de las secuencias codificantes de genes funcionales del ciclo del N.

Resultados y discusión

Algunas consideraciones experimentales sobre la obtención de metagenomas de suelos argentinos

En este capítulo se utilizaron datos metagenómicos de diversos suelos resultantes de estudios multidisciplinarios en el marco del proyecto PAE n° 37164 llevados a cabo por el consorcio SoilGene (ANPCyT). Las muestras de suelo se obtuvieron de cuatro sitios de la Región Pampeana Argentina definidos por dos isoyetas. En total se analizaron nueve metagenomas de suelo comprendiendo distintas localidades y tratamientos agronómicos. Por un lado, tres campos experimentales en la Pampa ondulada (La Estrella: S1, La Negrita S2, Criadero Klein: S3) con un clima húmedo y precipitación anual de 1000 a 1200 mm. Por otro lado, del campo experimental de la estación -INTA Balcarce (S4) con un clima semi-húmedo y precipitación anual de 800-1000 mm. Además de cada sitio, se muestrearon parcelas con diferentes manejos agronómicos de suelo: siembra directa (CSD), Labranza convencional (LC, solo para el sitio de Balcarce) y suelos sin uso agronómico ni perturbaciones antropogénicas (NC).

El proyecto generó un volumen muy importante de datos metagenómicos mediante secuenciación masiva al azar (WGS) y del amplicón del gen ARN 16S usando la tecnología 454-FLX. Estos datos fueron denominados *Pampa dataset* [59], [120]. En el presente trabajo hemos hecho uso de una parte de los mismos, principalmente aquellos datos generados por WGS, al cual por conveniencia práctica también llamaremos *Pampa dataset*. Los datos metagenómicos generados por el consorcio *SoilGene* fueron depositados en el servidor MG-RAST para el procesamiento de las secuencias. Además, en la sección III de Materiales y Métodos, titulada “*Procesamiento de secuencias obtenidas por secuenciación masiva y paralela*” se describen los pasos de filtrado de calidad de secuencias realizados por Carbonetto y colaboradores [120]. Aquí, se adoptó el uso del término base de datos curada para hacer referencia a la base de datos de alta calidad de secuencias de Carbonetto *et al.* 2014

La secuenciación masiva al azar generó un total de 10.445.170 lecturas que luego de los pasos de filtrado se redujo en un 20% a 8.592.774 [120]. El análisis taxonómico realizado previamente por Rascovan *et al.* 2013 demostró que la contaminación con ADN alóctono (ejemplo de plantas, humano, etc.) en estas bibliotecas metagenómicas fue mínimo o inexistente [59]. El 95% de las secuencias fueron identificadas como Bacteria, 0.6% como *Archaea* y 1% como Eucariota, mientras que 3.4% de las secuencias no pudieron ser clasificadas. Dentro de los Eucariontes, 42% de lecturas fueron clasificadas como *Viridiplantae* (plantas), 27% como *Fungi*, 12% como metazoos, 6% como diatomeas y 13% de las secuencias no pudieron ser clasificadas.

Concluyendo, las bibliotecas metagenómicas objeto del análisis bioinformático que se detalla en las secciones siguientes, fueron previamente curadas, y se caracterizan por una buena proporción de secuencias procarióticas y una muy baja contaminación con ADN alóctono.

Bases de datos metagenómicos de suelos: características y tamaño

El Pampa *dataset* comprendió un total 10.445.170 de secuencias y alrededor de 1 millón de secuencias para cada uno de los metagenomas de suelo analizados, como fuese descrito previamente. El número de secuencias que se obtuvo para cada sitio y el largo promedio de las mismas se muestra en la Tabla III.1, entre otras características. El contenido de GC (%) presentó los siguientes valores en las bases de datos: S1 CSD 64.6 ± 5.6 ; S1 NC 65.6 ± 5.4 ; S2 CSD 64.5 ± 7.2 ; S2 NC 64.9 ± 7.1 ; S3 CSD 64.6 ± 6.9 ; S3 NC 64.8 ± 7.1 ; S4 LC 65.2 ± 7.1 ; S4 CSD 64.1 ± 7.2 y S4 NC 65.2 ± 6.8 . Por lo tanto, no hubo una variación significativa en el contenido GC entre los metagenomas de suelo.

Para realizar una comparación de las abundancias relativas de los genes funcionales entre los distintos metagenomas de suelo se normalizó por tamaño de datos con respecto al metagenoma que presentó menor número de secuencias curadas que correspondió a S4 LC N= 684.967. Por otro lado, para la comparación de las abundancias relativas de cada categoría funcional dentro de un metagenoma se dividió

el número de *hits* de cada categoría funcional (ejemplo categoría COG) por el número total de secuencias curadas en cada muestra.

Tabla III.1. El Pampa *dataset* se compone de nueve metagenomas de suelo. Se obtuvo un total de 10.445.170 secuencias equivalente a 1.5 Gb. En la tabla muestra el número de pares de bases (N° pb), número de lecturas (N° de lecturas), y largo promedio de las mismas en cada muestra.

Abreviatura*	Sitio	N° pb	N° de lecturas	Largo de lecturas (promedio)
S1 CSD	La Estrella	557.052.142	1.340.378	415.593
S1 NC	La Estrella	461.058.570	1.166.656	395.197
S2 CSD	La Negrita	578.379.033	1.375.442	420.504
S2 NC	La Negrita	455.626.322	1.129.130	403.520
S3 CSD	C. Klein	598.035.949	1.374.775	435.006
S3 NC	C. Klein	561.993.912	1.314.527	427.526
S4 LC	Balcarce	380.148.362	901.747	421.569
S4 CSD	Balcarce	391.398.392	923.803	423.682
S4 NC	Balcarce	385.446.435	918.712	419.551

* SD corresponde a cultivado en siembra directa, LC a labranza convencional y NC a no cultivado

Con el propósito de evaluar los niveles de riqueza en el inventario del Pampa *dataset*, se realizaron curvas de rarefacción. Este parámetro permite estimar el nivel de cobertura de especies/taxones alcanzado con el número de secuencias logradas. La Figura III.2 muestra las curvas de rarefacción para metagenomas de distintos suelos generadas en el entorno del MG-RAST para el nivel de especie utilizando la anotación taxonómica de los subsistemas funcionales (base de datos SEED). Las curvas de rarefacción del Pampa *dataset* de más de 1.5 Gb de bases no presentaron saturación, reflejando el alto grado de diversidad del ambiente suelo. La muestra identificada como S4 LC con el número menor de lecturas, se utilizó como valor de corte (N=901.747) para comparar el número de especies reveladas de acuerdo al

análisis de rarefacción en cada una de las bibliotecas. A un N=901.704 de lecturas las bases de datos de S1 CSD y S2 CSD revelaron el mayor número de especies (4960 y 4971, respectivamente) y S4 LC presentó el valor mínimo (4389). Sin embargo, para todos los suelos analizados, la secuenciación *shotgun* alcanzó a revelar un valor sustancial de especies (más de 4000 especies) y comparable entre todas las muestras. Tringe *et al.* 2005 observó que las curvas de rarefacción a partir de la secuenciación masiva de clones de ARNr 16S bacteriano de suelo agrícolas tampoco alcanzaron saturación a diferencia de las curvas de muestras de aguas profundas que si lo hicieron [64]. En otros proyectos de WGS donde evaluaron la diversidad de comunidades microbianas en muestras ambientales, se estimó por ejemplo para el mar de los Sargazos un total de 1.800 especies para 1.6 Gb de lecturas [62] mientras que en los suelos agrícolas de Minnesota en Estados Unidos un total de 3000 especies para 1 Gb de lecturas [64], [151].

Por otro lado, los datos publicados en trabajos recientes, donde se ha utilizado la plataforma de secuenciación 454 GS-FLX para estudios metagenómicos de suelo con diferentes enfoques, permite realizar comparaciones de tamaño con nuestra base de datos. Por ejemplo Souza *et al.* 2013 obtuvieron 4.079.233 lecturas de suelos agrícolas de Brasil (labranza convencional y siembra directa) [169], Fierer *et al.* 2012 un total de 1.350.000 de lecturas de suelos sujetos a distintos gradientes de fertilización de nitrógeno en EEUU [65], y Delmont *et al.* 2012 analizaron 12.575.129 lecturas de suelos con pastizales naturales de Rothamsted Research en Reino Unido [152] (para mayor información sobre dichos suelos ver Tabla Suplementaria III.1 en la sección Material Suplementario). Estos datos permiten concluir que los estudios que se describirán en las siguientes secciones presentan un nivel cuantitativo y cualitativo de datos comparable a las de otros proyectos metagenómicos de suelo realizados en distintas partes del mundo y en los cuales también se encaró la identificación de atributos funcionales en las comunidades microbianas.

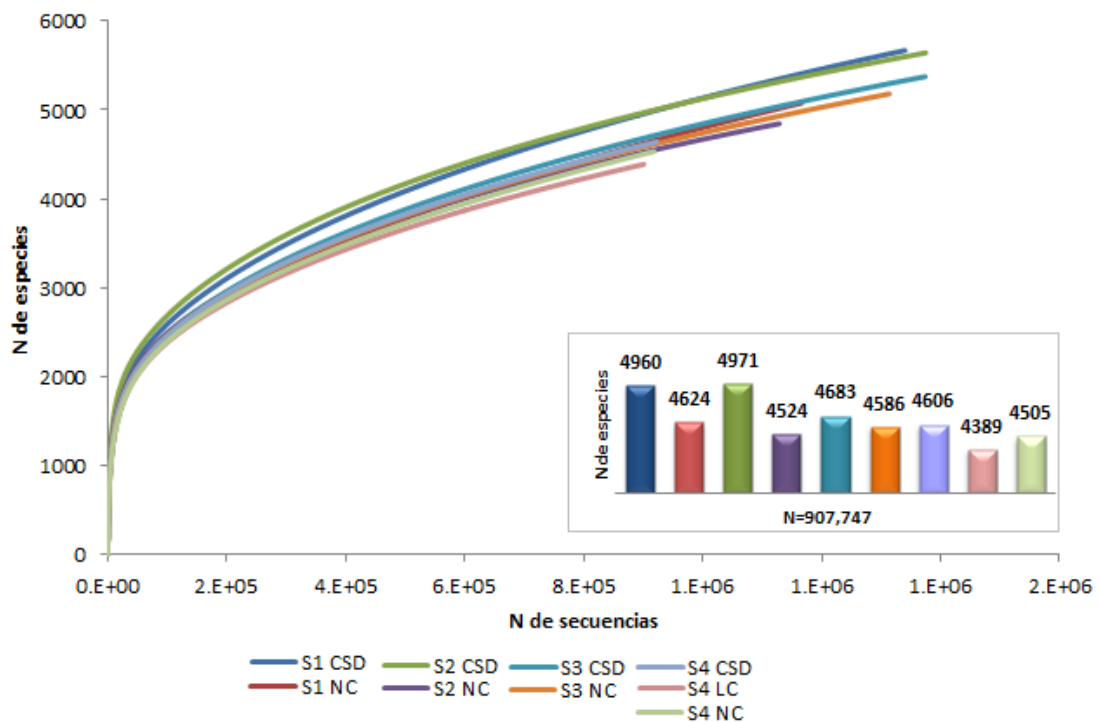


Figura III.2. Curvas de rarefacción para los metagenomas de suelo. Las curvas de rarefacción fueron calculadas a partir de la tabla de abundancia de especies en el entorno de MG-RAST. Las curvas representan el número promedio de distintas especies anotadas para cada submuestra de la base de datos completa.

Análisis in silico de los atributos funcionales de las bases de datos metagenómicas de suelo y comparación con otros ambientes naturales

En un análisis preliminar utilizando las diferentes bases de datos disponibles en M5NR se encontró que entre el 50% y el 58% de las secuencias para todo el Pampa *dataset* que presentaron un marco abierto de lectura pudieron ser anotadas como una proteína con función conocida, y aproximadamente un 90% de las mismas asignables a categorías funcionales (Figura III.3). En este punto se concluyó que estos valores fueron promisorios/robustos para encarar la asignación funcional específica.

En la siguiente etapa del análisis *in silico*, y con el propósito de explorar el potencial metabólico revelado a partir del examen en los metagenomas de suelo, se llevó a cabo la asignación de las secuencias en categorías de subsistemas mediante búsquedas de similitud por BlastX entre las lecturas curadas y la base de datos SEED.

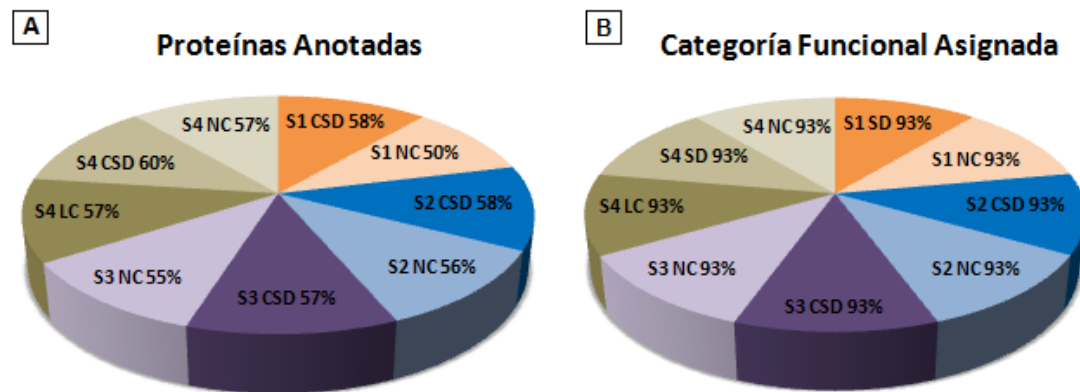


Figura III.3. Los gráficos ilustran el porcentaje de proteínas anotadas y la asignación a categorías funcionales en los metagenomas de suelo. (A) Para la anotación proteica, el MG-RAST utiliza el programa *FragGeneScan* para predecir potenciales genes codificadores de proteínas y luego el algoritmo BLAT contra la base de datos no redundante M5NR. (B) Para la asignación de una función proteica se utilizaron diferentes bases de datos como por ejemplo *Seed Subsystems Annotation*, *Non-supervised Orthologous Groups*, *Cluster of Orthologous Groups* y *KEGG Orthology*.

Dicho intento permitió asignar el 52% del total de secuencias del Pampa *dataset* y se encontraron 28 subsistemas jerárquicos (nivel 1 de clasificación) cuya proporción relativa fue variada (Figura III.4). Cabe mencionar que la forma de presentar los resultados de las clasificaciones funcionales y taxonómicas usando el MG-RAST es dinámica, es decir permite al usuario filtrar dinámicamente diferentes niveles de clasificación. A modo de ejemplo, en el nivel 1 de clasificación se selecciona la categoría jerárquica “aminoácidos y derivados”, en el nivel 2 “alanina, serina y glicina”, en el nivel 3 biosíntesis de alanina” y finalmente se puede filtrar por función (ejemplo cisteína desulfurasa).

Los subsistemas encontrados que presentaron el mayor número de secuencias asignadas en todos los metagenomas son: subsistemas basados en agrupaciones ($15 \pm 0.7\%$), metabolismo de carbohidratos ($11 \pm 0.1\%$), metabolismo de aminoácidos y derivados ($9 \pm 0.1\%$), metabolismo de proteínas ($8 \pm 0.1\%$), cofactores, vitaminas y grupos prostéticos ($7 \pm 0.05\%$) y misceláneos ($8 \pm 0.1\%$). El subsistema basado en agrupaciones consiste en grupos de genes que se encuentran próximos entre sí en los genomas de diversas taxa y cuyos atributos funcionales aún no se conocen bien. Entre los roles funcionales que contiene se hallan los proteosomas, ribosomas y clústeres relacionados a la recombinación, entre otros [152]. El subsistema misceláneo

comprende los siguientes roles funcionales: factores de elongación de la traducción (LepA), ARN polimerasa subunidad alfa, proteína RecA, entre otros [170].

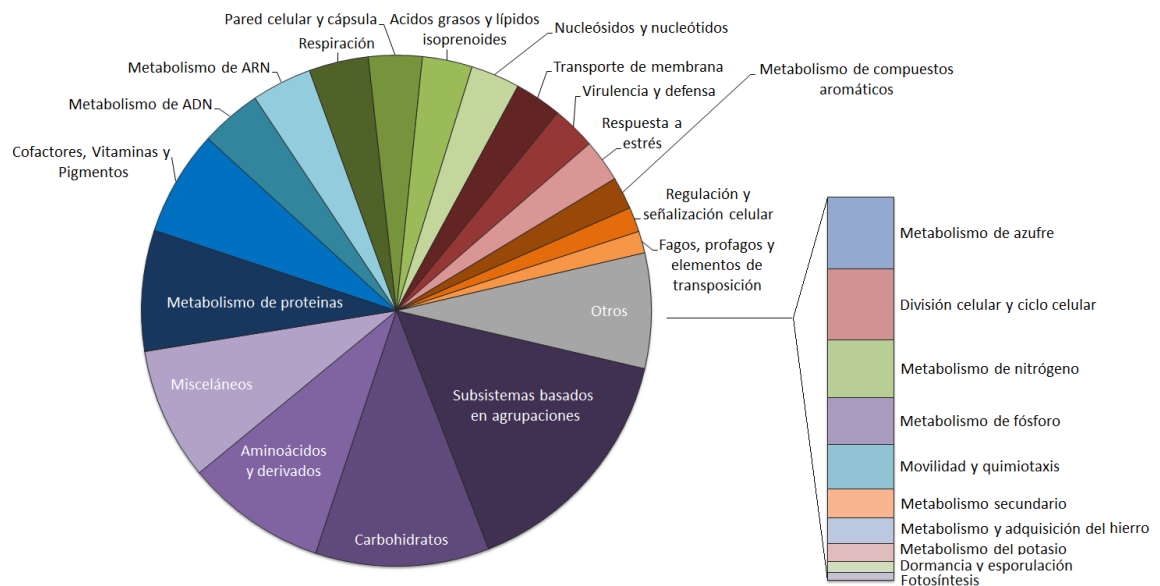


Figura III.4. Proporción relativa de subsistemas representados en el Pampa dataset. La asignación de subsistemas se realizó con un valor de corte de E de 10^{-5} y la base de datos de referencia utilizada fue SEED.

Otros subsistemas tales como metabolismo y adquisición de hierro ($0.5 \pm 0.05\%$), fotosíntesis ($0.15 \pm 0.1\%$), metabolismo del potasio ($0.34 \pm 0.03\%$), dormancia y esporulación (0.2%) se hallaron en menor proporción (Figura III.4). La distribución relativa del resto de los subsistemas para todo el Pampa dataset se encuentra en la Tabla Suplementaria III.2. Los subsistemas descritos como predominantes en el Pampa dataset se encontraron también como tales en diversos metagenomas, sugiriendo la ubicuidad y el papel significativo de dichas categorías funcionales en los microorganismos.

En particular, los subsistemas metabolismo de carbohidratos, metabolismo de proteínas, aminoácidos y derivado y misceláneos presentaron la mayor cantidad de lecturas asignadas en otros suelos [65], [152], [171], lodos activados [170] y en comunidades microbianas marinas [172]. El subsistemas basado en agrupaciones aparece mayoritariamente en los metagenomas de suelo [152], [153], [171].

Contrario a nuestras expectativas, se encontró una baja representación del subsistema metabolismo de nitrógeno, representando entre 0.8 al 1% del total de las secuencias asignadas en todos los metagenomas de suelo. En línea con nuestros resultados, Delmont *et al.* 2012 [152] y Fierer *et al.* 2012 [173] hallaron en sus análisis de metagenomas de suelo que el porcentaje de secuencias correspondientes al metabolismo nitrogenado no superó el 1% del total de las secuencias. Como ha sido discutido en varios trabajos, ciertas categorías metabólicas como fijación de nitrógeno y nitrificación no están distribuidas de manera ubicuas en los microorganismos [24], [174], [175]. Para que los mismas sean recuperadas de bases de datos metagenómico podría procurarse una secuenciación más profunda, o un previo enriquecimiento de una comunidad en particular [53]. En este contexto, Yu *et al.* 2012 evaluaron la estructura de la comunidad microbiana y el potencial metabólico de ADN metagenómico de lodos activados en diferentes profundidades de secuenciación (0.5 Gbp, 1.0 Gbp y 2.4 Gbp), encontrando que con una mayor profundidad de secuenciación (2.4 Gbp) se logró aumentar significativamente el número de secuencias asociadas a categorías funcionales del metabolismo nitrogenado [170]. Cabe mencionar que dicha cantidad de datos se logró con la plataforma de Illumina Hi-seq2000. Esto sugirió que para revelar la presencia de componentes “raros” en la base de datos metagenómicos, la secuenciación debería tener una cobertura de al menos 2 Gb.

La comparación de los patrones de secuencias asignadas aplicando un criterio más fino de clasificación de subsistemas funcionales (nivel 3) en el entorno MG-RAST, resultó en que los subsistemas ciclo de serina-glioxilato y señalización de AMP cíclico en bacterias presentaron la mayor representación en el Pampa *dataset* con un valor de 17.399 y 14.603 lecturas asignadas, respectivamente. El subsistema ciclo de serina-glioxilato también ha sido reportado recientemente por Pacchionni *et al.* 2014 como uno de los subsistemas de mayor predominancia en suelos de dos regiones del estado de Rio Grande en Brasil [159]. Estos autores sugieren que este ciclo representa un camino metabólico alternativo por el cual el carbono podría ser fijado en suelo, y estaría involucrado en la degradación de materia orgánica. Asimismo, este ciclo es una vía alternativa para la asimilación de C1 (mono carbono)

en bacterias metilótrofas, que comprende algunas especies de *Mycobacterium* [176]. Se ha señalado al metanol (CH₄O) como un compuesto muy abundante en suelo resultante de la degradación de lignina y pectina [177]. Por otro lado, el AMP cíclico es un mensajero secundario muy importante en eucariotas y procariotas. En bacterias es un regulador celular universal de la energía/metabolismo, así como también un componente de la señalización celular. Este subsistema que aparece ampliamente representado en nuestros suelos, también fue descrito como predominante en otros estudios similares con suelos de Puerto Rico, Italia y Reino Unido [152]. Por lo tanto, la amplia representación de ciertas categorías funcionales en el suelo resulta suficientemente robusta para ser reveladas a una escala global, indicando que ciertos procesos metabólicos son propios del hábitat suelo y sugiriendo además que dichas funciones son importante para los microorganismos del suelo.

En otra etapa del análisis se encaró la distribución de los subsistemas en los distintos sitios y uso agrícola (CDS y NC). El resultado se muestra en la Figura III.5 y se encontró que los subsistemas mostraron leves variaciones entre los distintos metagenomas de suelo analizados. Con el propósito de reforzar el resultado observado en nuestros datos, e identificar asociaciones en base al potencial metabólico de las comunidades microbianas de diferentes ambientes, se realizó un análisis de conglomerado mediante el método de agrupamiento UPGMA utilizando una matriz de similitud basada en la distancia de Bray-Curtis. Para la construcción de la matriz se utilizaron los valores de distribución relativa (en porcentaje del total de secuencias anotadas) de los subsistemas (Tabla Suplementaria III.2).

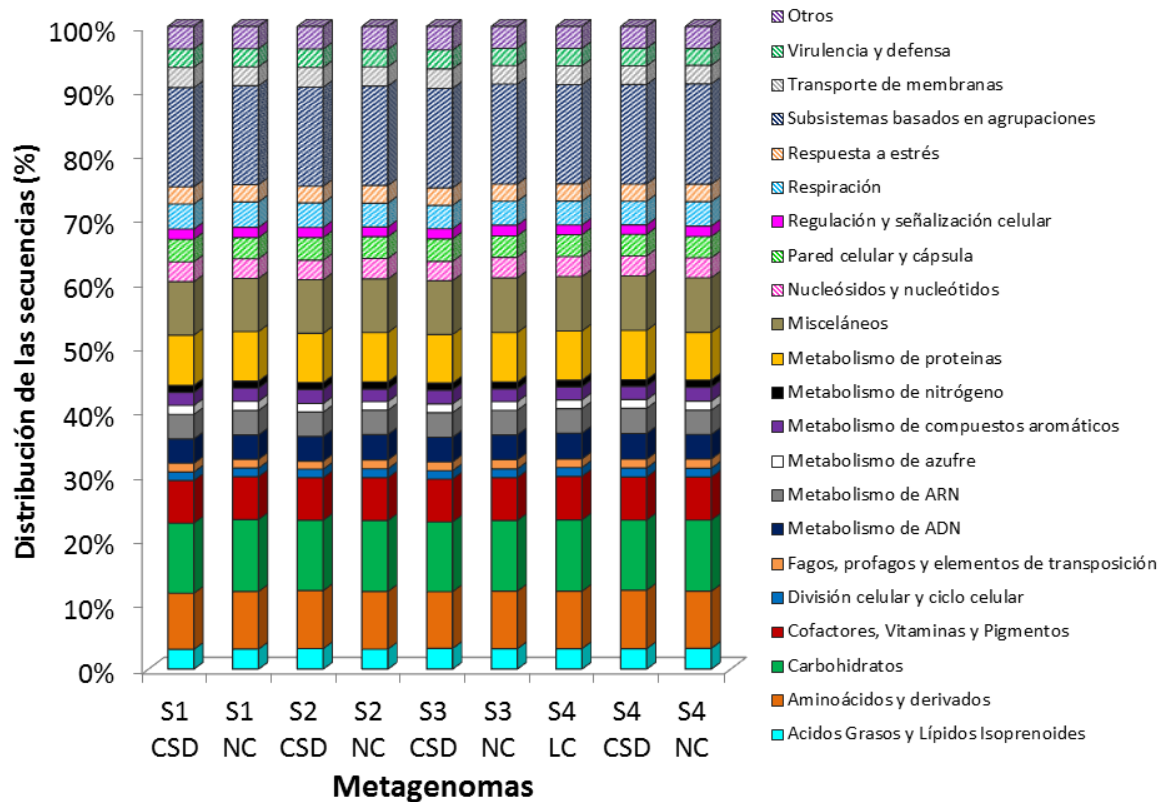


Figura III.5 Proporción relativa (%) de los principales subsistemas funcionales detectados en los metagenomas de suelo de la región pampeana argentina. La asignación de subsistemas se realizó con un valor de corte de E de 10^{-5} y la base de datos de referencia utilizada fue SEED.

Como se muestra en la Figura III. 6 se formaron dos grupos jerárquicos a partir del análisis de conglomerado, indicando que la naturaleza del bioma es el principal conductor de agrupamiento a un valor de *bootstrapping* del 100%. Por otra parte y con el mismo valor de *bootstrapping*, los suelos cultivados y no cultivados de los sitios S2, S3 y S4 formaron dos subgrupos distintos. Asimismo, para el sitio S4 los suelos cultivados se agruparon con respecto al suelo no cultivado, pero a un valor de *bootstrapping* menor al 50%. Este análisis exploratorio refleja que el ambiente suelo y el acuático presentan atributos funcionales únicos y propios, y a su vez que diferencias del potencial metabólico podrían reflejarse para suelos con distinto uso agrícola. En línea con esta observación, Carbonetto *et al*, 2014 utilizando otras categorías funcionales, como GOGs, revelaron que para los sitios S2 y S3 el uso del suelo era un factor determinante en la agrupación de los suelos [120].

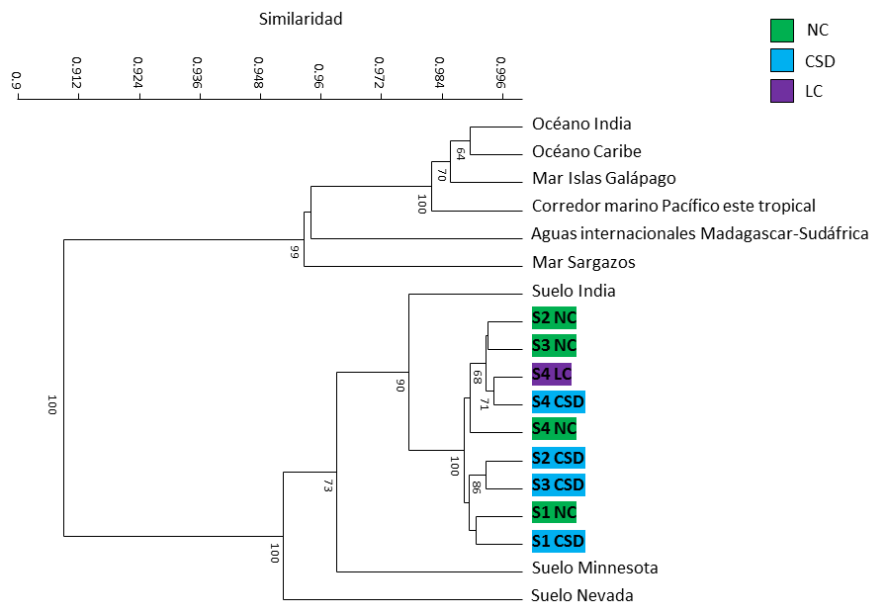


Figura III.6. Dendrograma de similitud UPGMA generado a partir de las abundancias relativas de los subsistemas en diversos metagenomas acuáticos y de suelo. Para el análisis de conglomerado se aplicó la medida de distancia Bray Curtis. Los datos metagenómicos públicos se obtuvieron del reservorio del MG-RAST bajo los siguientes ID: 4441091.3 (Suelo Minnesota de EEUU), 4451105.3 (Suelo Nevada de EEUU), 4461011.3 (Suelo de India), 4441571.3 (Mar Sargazos), 4441611.3 (Océano Indio) 4441614.3 (Aguas internacionales Madagascar-Sudáfrica), 4441594.3 (Mar Islas Galápagos), 4441587.3 (Océano Caribe), 4441661.3 (Corredor marino Pacífico este tropical). Se muestran valores de *bootstrapping* mayores al 50%.

Con el objetivo de evaluar los principales subsistemas que se encuentran dirigiendo el ordenamiento de los metagenomas analizados, se llevó a cabo un análisis de componentes principal (PCA). El resultado de este análisis se muestra en la Figura III.7. La componente principal 1 representa el 76% de la variabilidad y separó los metagenomas de acuerdo al ambiente, suelo y agua. Algunas clasificaciones funcionales aparecieron más representadas en suelo en comparación con los ambientes acuáticos como los subsistemas metabolismo de aminoácidos y derivados (1), metabolismo de carbohidratos (2), transporte de membrana (11), metabolismo de compuestos aromáticos (12) y virulencia y defensa (28). Contrariamente, los subsistemas cofactores, vitaminas y pigmentos (6), misceláneo (13), metabolismo de nucleótidos y nucleósidos (16), fagos, profagos y elementos transposición (17), fotosíntesis (19) y metabolismo de proteínas parecen predominar en los metagenomas

de ambientes acuáticos. Coincidentemente los perfiles de distribución de subsistemas, coincide con lo observado por Delmont *et al.* 2012 al examinar los subsistemas funcionales en metagenomas de suelo, océanos y heces de humano. Revelaron que los subsistemas -metabolismo de proteínas, -metabolismo de nucleótidos y nucleósidos y -fotosíntesis también se asociaban principalmente con los ambientes acuáticos. [152]. A su vez, hallaron que los subsistemas - virulencia y defensa, y transporte de membrana se encontraban mayormente representados en el suelo. Por lo tanto, a partir de los resultados del análisis comparativo con otros ambientes, se pone de manifiesto que los suelos muestran una representación de ciertos subsistemas o aún más general, de ciertas capacidades metabólicas potenciales mayores que en otros ambientes.

Finalmente, se utilizaron las categorías de subsistema (nivel 1 de clasificación) con el fin de evaluar su distribución entre los suelos cultivados y no cultivados de los sitios S1, S2 y S3, para lo cual se usó la prueba estadística Fisher. Como se muestra en la Figura III.8 se pudo revelar que 4 subsistemas presentaron una distribución significativa diferente entre los suelos cultivados y no cultivados. Estos correspondieron a: - división celular y ciclo celular ($p=0.03$), -movilidad y quimiotaxis ($p=0.0003$), - pared celular y capsula ($p\leq 0.002$) y - virulencia y defensa ($p=0.02$). Excepto el subsistema división celular y ciclo celular, los tres restantes presentaron una asignación mayor de secuencias en los suelos cultivados respecto de los no cultivados. Por lo tanto, en general las capacidades metabólicas jerárquicas de estos suelos son comparables entre los distintos usos.

Identificación de secuencias del suelo relacionadas a genes con función asignada del metabolismo nitrogenado

En la sección precedente se describieron los atributos funcionales de las bases de datos metagenómicas de suelo y se realizó una comparación con otros ambientes naturales. En esta sección, nos focalizamos en examinar mediante el análisis *in silico*, la representación de secuencias relacionadas a genes con función asociada con el metabolismo nitrogenado de los suelos. Se realizó la anotación funcional mediante BlastX en la plataforma bioinformática MGX. Además, se utilizaron dos bases de datos de referencia: eggNOG v3.0 [127] y otra construida especialmente para este estudio según se describe en la sección “Plataforma bioinformática MGX” en el

capítulo Materiales y Métodos). Particularmente, nos concentramos en ciertos procesos metabólicos del ciclo del nitrógeno (N) para los cuales además, se cuentan con genes marcadores de tales procesos (Tabla III.2).

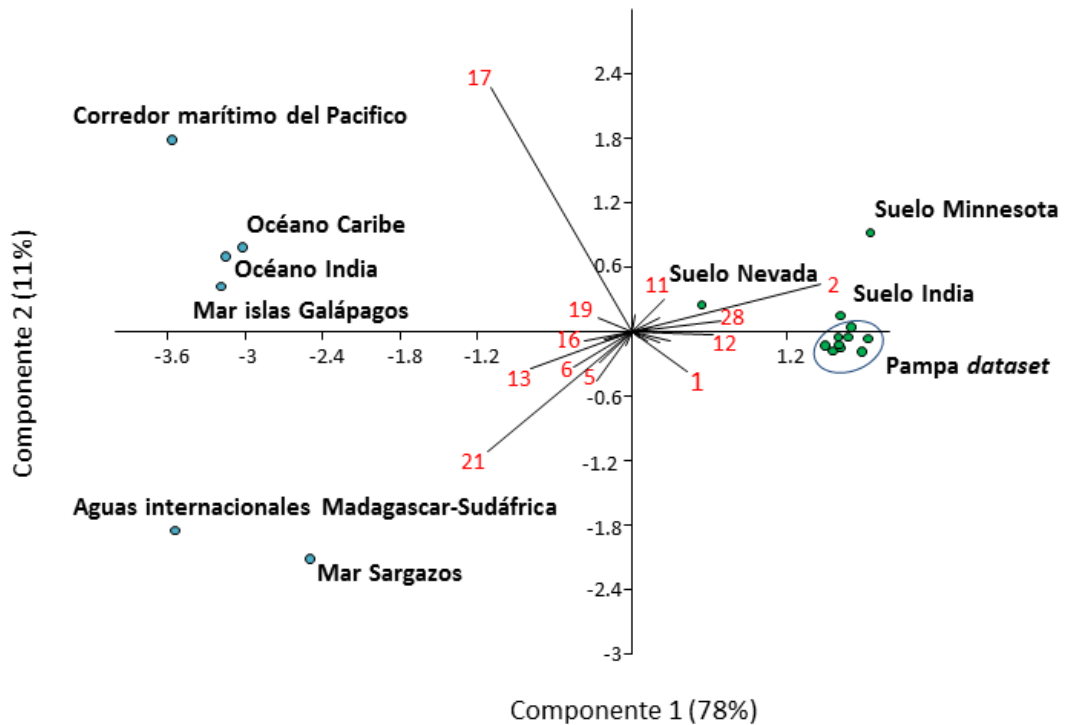


Figura III.7. Análisis de PCA de metagenomas de suelo, mares y océanos utilizando la abundancia relativa de los diversos subsistemas. Solo se muestra el número referencia de los subsistemas cuya magnitud y dirección fueron relevantes en este análisis. Los colores hacen referencia a los distintos ambientes (celestes para los suelos, y verde para los ambientes acuáticos). Los datos metagenómicos públicos se obtuvieron del reservorio del MG-RAST y coinciden con los de la Figura III.6 Los números en cada vector se refieren respectivamente a los subsistemas considerados en el análisis (ver Tabla Suplementaria III.3).

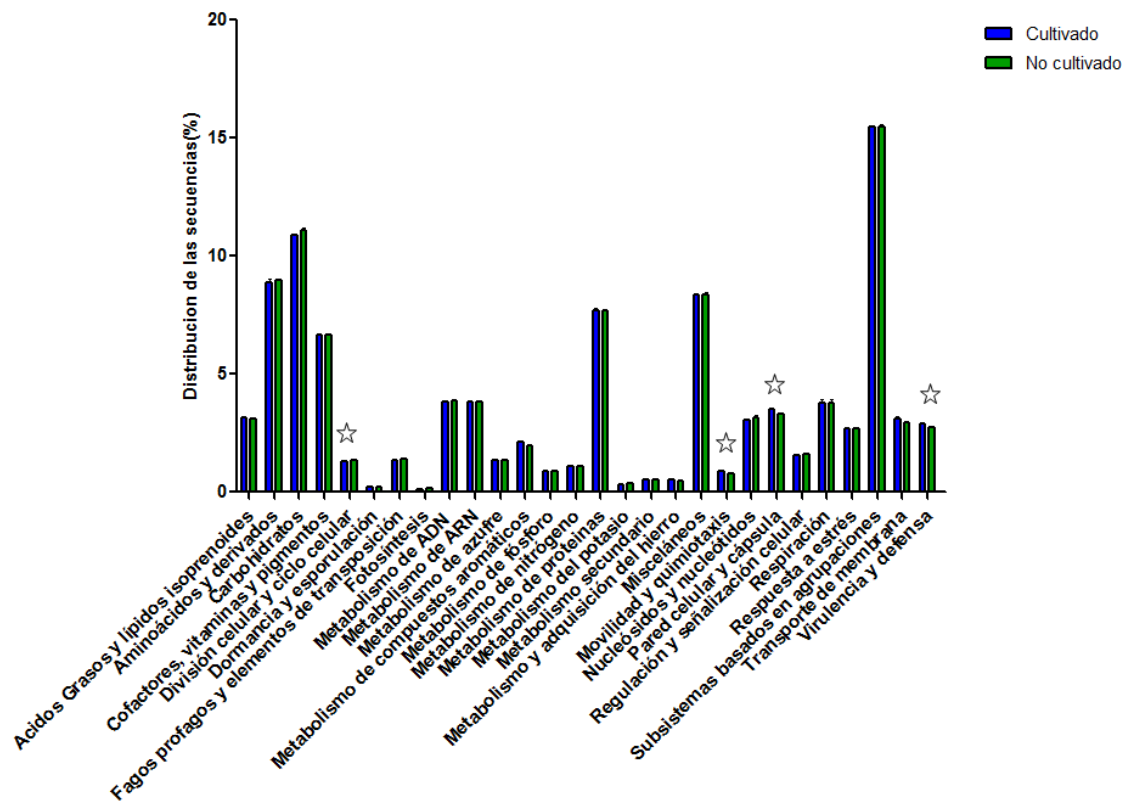


Figura III.8. Comparación de distribución relativa de subsistemas entre metagenomas de suelos cultivados y no cultivados de los sitios S1, S2 y S3. La estrella indica diferencias significativas con un valor de $p \leq 0.05$. Las barras corresponden al porcentaje del total de secuencias anotadas.

a. Clasificación de las secuencias de acuerdo a las categorías funcionales COGs.

La abreviación COG hace referencia a clúster de grupos ortólogos, y cada COG se define como el grupo de tres o más proteínas que son ortólogas. Han sido desarrollados mediante la comparación, todos contra todos, de secuencias de proteínas codificadas en genomas completos secuenciados depositados en el NCBI [127]. En esta sección se describe el resultado del análisis bioinformático basado en la clasificación por COGs. Como se muestra en la Tabla III.3, más del 50% de las secuencias de los metagenomas de suelo resultaron clasificadas en genes codificantes.

Tabla III.2. Genes seleccionados para el análisis funcional *in silico* del metabolismo nitrogenado en suelo.

Gen (abreviatura)	Enzima codificada	Proceso metabólico
<i>nifH</i>	Dinitrogenasa reductasa	Fijación biológica de nitrógeno
<i>nifD</i>	Nitrogenasa cadena α	
<i>amoA</i>	Amonio monooxigenasa	Nitrificación
<i>nirA</i>	Nitrito reductasa	Reducción asimilativa de nitrito
<i>nirB</i>	Nitrito reductasa	Reducción desasimilatoria de nitrito
<i>napA</i>	Nitrato reductasa de membrana	Denitrificación
<i>narG</i>	Nitrato reductasa periplásmica	
<i>nirS</i>	Citocromo Cd1-nitrito reductasa	
<i>nirK</i>	Cu-nitrito reductasa	
<i>norB</i>	Óxido nítrico reductasa	
<i>nosZ</i>	Óxido nitroso reductasa	

Conforme a lo sugerido por Delmont *et al.* 2012 este nivel de porcentaje de clasificación es aceptable para encarar el análisis funcional de los procesos metabólicos jerárquicos de las células procariotas [152]. Encontramos que las representaciones de las categorías jerárquicas COGs en el Pampa *dataset* correspondieron a:

- Metabolismo con el 44% (incluye las categorías de producción de energía, aminoácidos, nucleótidos, carbohidratos, coenzimas, lípidos y metabolitos secundarios, entre otros).
- Procesos de señalización con el 20% de (incluye las categorías de división, movilidad y envoltura celular, modificaciones postranscripcionales, mecanismo de transducción de señales, mecanismos de defensa, entre otros).

- Pobre caracterización con el 20% de (incluye las categorías de predicción de función general y función no conocida).
- Almacenamiento de información y procesamiento con el 16% (incluye las categorías de procesamiento y modificación de ARN, transcripción, replicación, recombinación y reparación, entre otras.).
- **Tabla III.3.** Se muestra para cada metagenoma la proporción (%) de secuencias asignadas en COGs.

Sitio	Abreviatura	(%)
Establecimiento productivo "La Estrella"	S1 CSD	53
	S1 NC	56
Establecimiento productivo "La Negrita"	S2 CSD	55
	S2 NC	57
Establecimiento productivo "Criadero Klein"	S3 CSD	59
	S3 NC	58
Unidad Experimental INTA Balcarce	S4 CSD	58
	S4 LC	59
	S4 NC	57

A un nivel más fino de clasificación, seleccionamos aquellos COGs que presentaron una asignación mayor o igual a 2000 secuencias (Tabla III.4). Hallamos una importante representación en el Pampa *dataset* de funciones tales como: sistema de transporte tipo ABC (componente permeasa), glicosiltransferasa, histidina quinasa, Acil CoA-deshidrogenasa, entre otras. Muchas de las funciones abundantes reveladas en nuestros metagenomas corresponden a las categorías jerárquicas de COGs "predicción de función general". Mientras que otras corresponden a las categorías mecanismos de transducción de señales, mecanismos de defensa, metabolismo y transporte de lípidos y replicación, recombinación y reparación. Sessitsch *et al.* 2012

y Carbonetto *et al.* 2014 también han descripto a estas categorías jerárquicas de COGs como predominantes el ambiente suelo [120], [135].

Tabla III 4. Categorías jerárquicas de COGs y sus funciones asociada más abundantes en los metagenomas de suelo. Se muestran aquellos COGs que superaron un número de secuencia asignada ≥ 2000 Abreviaturas CSD: cultivado en siembra directa, LC: labranza convencional y NC: no cultivado.

Categorías jerárquicas de COGs	Función	S1	S1	S2	S2	S3	S3	S4	S4	
		CSD	NC	CSD	NC	CSD	NC	LC	CSD	NC
Biogénesis de pared celular/membrana/envoltura	Glicosiltransferasa	2423	2607	2551	2583	2736	2850	2904	2886	2746
Mecanismos de defensa	Sistema de transporte tipo ABC (componente permeasa)	4066	3524	3626	3226	3961	3264	3586	3621	2936
Mecanismos de transducción de señales	Histidina quinasa	6715	6316	7226	6251	7210	6494	6836	6773	6350
Metabolismo y transporte de lípidos	Acil CoA-deshidrogenasa	2117	2522	2273	2475	2462	2458	2563	2493	2519
Predicción de función general	Permeasas de la superfamilia principal de transportadores facilitadores (MFS)	4063	4847	4175	4498	4806	4473	5058	5015	5044
Predicción de función general	Deshidrogenasas con diferentes especificidades	3383	3834	3462	3579	3834	3759	3858	3789	4111
Predicción de función general	Hidrolasas o aciltransferasas predichas (superfamilia de alfa/beta hidrolasas)	2347	2688	2175	2515	2580	2839	2619	2700	3094
Replicación, recombinación y reparación	Serina/treonina proteína quinasa	4353	3986	4227	3548	4310	4338	3882	4156	4051

Una vez realizado el análisis preliminar descrito en los párrafos precedentes para validar los datos, enfocamos nuestra atención sobre los COGs relacionados a procesos específicos del metabolismo de nitrógeno. Se hallaron 19 COGs relacionados a procesos específicos del metabolismo de nitrógeno que se muestran en la Tabla III.5. Estos procesos presentaron un bajo porcentaje de representación de entre el 0.15% - 0.20% de las clasificaciones por COGs. Específicamente 2.734 secuencias pudieron ser clasificadas de acuerdo a los COGs de interés en el presente estudio (marcados con asterisco en la Tabla III.5). Los COGs relacionados a *nifH* (COG 1348) y *nifD/nifK* (COG 2710) presentaron una asignación de 37 secuencias en total en el Pampa *dataset*; mientras que la *amoA* (COG 3180) un total de 127 secuencias. Con respecto a las reductasas de nitrito *nirA* (COG 2146) y *nirB* (COG 1251) se rescataron 946 y 668 secuencias, respectivamente. Para los 3 COGs relacionados al proceso de denitrificación, se rescataron 407, 372 y 178 secuencias para la reductasa de nitrato *napA* (COG 2013), *norB* (COG 3256) y *nosZ* (COG 4263), respectivamente. En la Figura III.9 se muestra la abundancia relativa de estas categorías COGs en el Pampa *dataset* (A) y en cada uno de los metagenomas de suelos (B).

Tabla III.5. Categorías de COGs relacionadas al metabolismo del nitrógeno reveladas en el Pampa *dataset*. El identificador de COG corresponde a un código numérico mediante el cual se hace referencia a una función metabólica en la base de datos eggNOG.

Proceso metabólico	Identificador de COG	Función	Gen
	COG 1348*	Nitrogenasa subunidad ATPasa	<i>nifH</i>
	COG 5000	Transducción de señal histidina kinasa involucrada en el la fijación de nitrógeno y regulación del metabolismo	
Fijación de nitrógeno	COG 347	Proteínas PII reguladoras del nitrógeno	
	COG 2710*	Cadena alfa y beta de la nitrogenasa convencional	<i>nifD/nifK</i>
Nitrificación	COG 3180*	Amonio monooxigenasa putativa	<i>amoA</i>
	COG 2146*	Ferredoxina-Nitrito Reductasa	<i>nirA</i>
	COG 1251*	Nitrito reductasa NAD(P)H	<i>nirB</i>
	COG 3303	Nitrito reductasa dependiente de formiato, citocromo periplásmico c552	<i>nrfA</i>
Reducción de nitrato y nitrito	COG 1140	Nitrato reductasa subunidad beta	<i>narH</i>
	COG 2181	Nitrato reductasa subunidad gama	<i>narI</i>
	COG 2180	Nitrato reductasa subunidad delta	<i>narJ</i>
	COG 3043	Nitrato reductasa citocromo c	<i>napB</i>
	COG 4459	Nitrato reductasa periplásmica, componente NapE	<i>napE</i>
	COG 5013*	Nitrato reductasa subunidad alfa	<i>napA</i>
Desnitrificación	COG 4548	Oxido nítrico reductasa, proteína de activación	<i>norD/norE</i>
	COG 4362	Oxido nítrico sintetasa, dominio oxigenasa	
	COG 3256*	Oxido nítrico reductasa, subunidad mayor	<i>norB</i>
	COG 3420	Proteína accesoria de óxido nitroso oxidasa	<i>nosD</i>
	COG 4263*	Oxido nitroso reductasa	<i>nosZ</i>

*COGs utilizados en la identificación de secuencias con función asignada a genes del metabolismo de N.

Este análisis reveló que en el Pampa *dataset* las categorías de COGs relacionadas a los procesos de reducción de nitrato/nitrito y desnitrificación presentaron una abundancia relativa mayor, del 59% y 35% respectivamente, comparada con la fijación biológica de nitrógeno (<1%) y la nitrificación (5%). Por otra parte, el número total de secuencias asignadas a estos COGs fue bajo en los metagenomas de suelo, considerando que en general los metagenomas presentaron un millón de secuencias.

b. Clasificación de las secuencias de acuerdo la base de datos de referencia personalizada

El *BlastX* contra la base de datos de referencia personalizada permitió recuperar un total de 4.618 secuencias correspondientes a genes de interés asociados al metabolismo nitrogenado. Se adoptó el término *hit* para hacer referencia al número de secuencias asignadas a una función en particular. Como se muestra en la Figura III.10 A, se obtuvo una mayor proporción de *hits* para la reducción de nitrito (21%) y la desnitrificación (77%) en el Pampa *dataset*, comprado con el proceso de fijación biológica de nitrógeno y nitrificación representados por <1% del total. Es tan baja la

representación de los genes *nifH* (3 hits), *nifD* (8 hits), *amoA* (2 hits) y *nirS* (13 hits) en los metagenomas de suelo que en la gráfica de barras no se llegan a visualizar (Figura III.10 B).

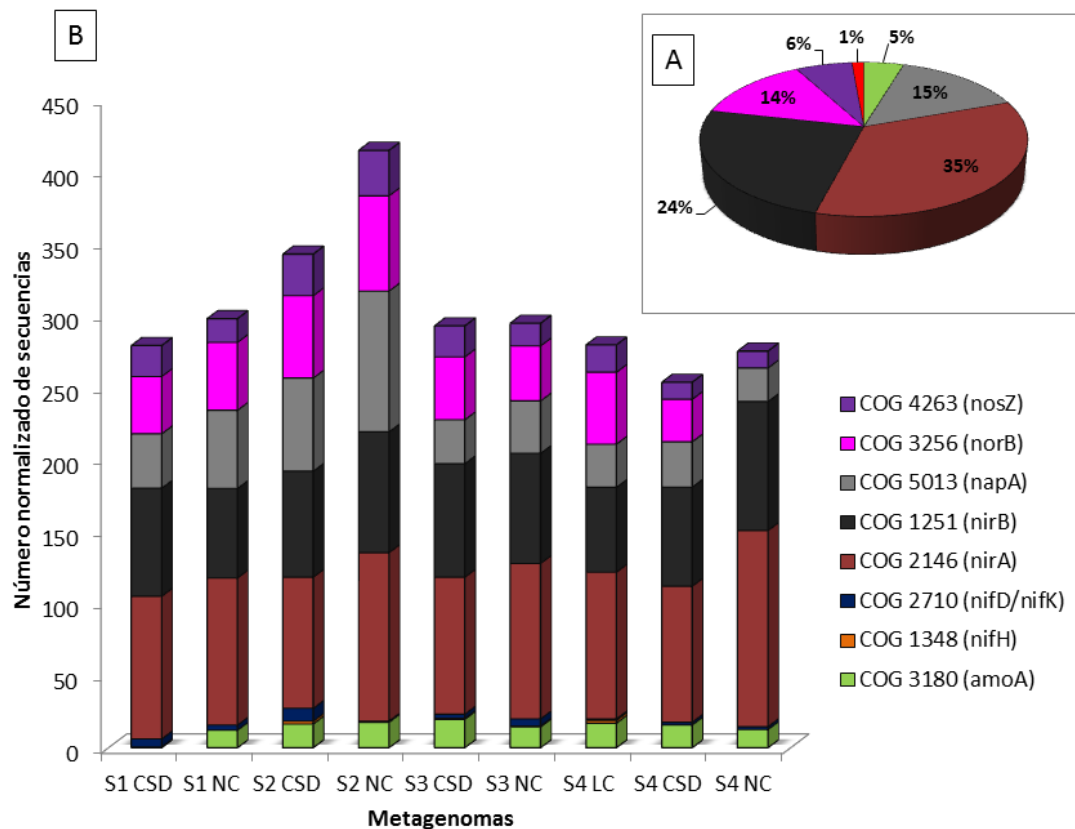


Figura III.9. Abundancia relativa de las categorías de COGs involucradas en los procesos del metabolismo del nitrógeno. En A se muestra la proporción de los COGs en el Pampa dataset y la porción de color rojo representa los COGs <1% del total (COG 2710 y COG 1348). En B se muestra la abundancia relativa de los COGs en cada metagenoma. En total se asignaron 280 secuencias en S1 CSD; 298 secuencias en S1 NC; 343 secuencias en S2 CSD; 415 secuencias en S2 NC; 239 secuencias en S3 CSD; 295 secuencias en S3 NC; 280 secuencias en S4 LC; 254 secuencias en S4 CSD y 276 secuencias en S4 NC.

Al comparar el total de secuencias asignadas para los genes de interés en los suelos cultivado y no cultivado, se observó que el gen *nirB* (codificante para la nitrito reductasa) fue el único que presentó una proporción significativa mayor ($p=0.02$) en los suelos cultivados (Figura III.11). En el presente trabajo seleccionamos al gen

marcador *nirB* como representa del proceso de reducción desasimilatoria de nitrito a amonio. Este mecanismo permite a las células la detoxificación de nitrito, el cual es un potente mutagénico y altamente toxico para los organismos [97].

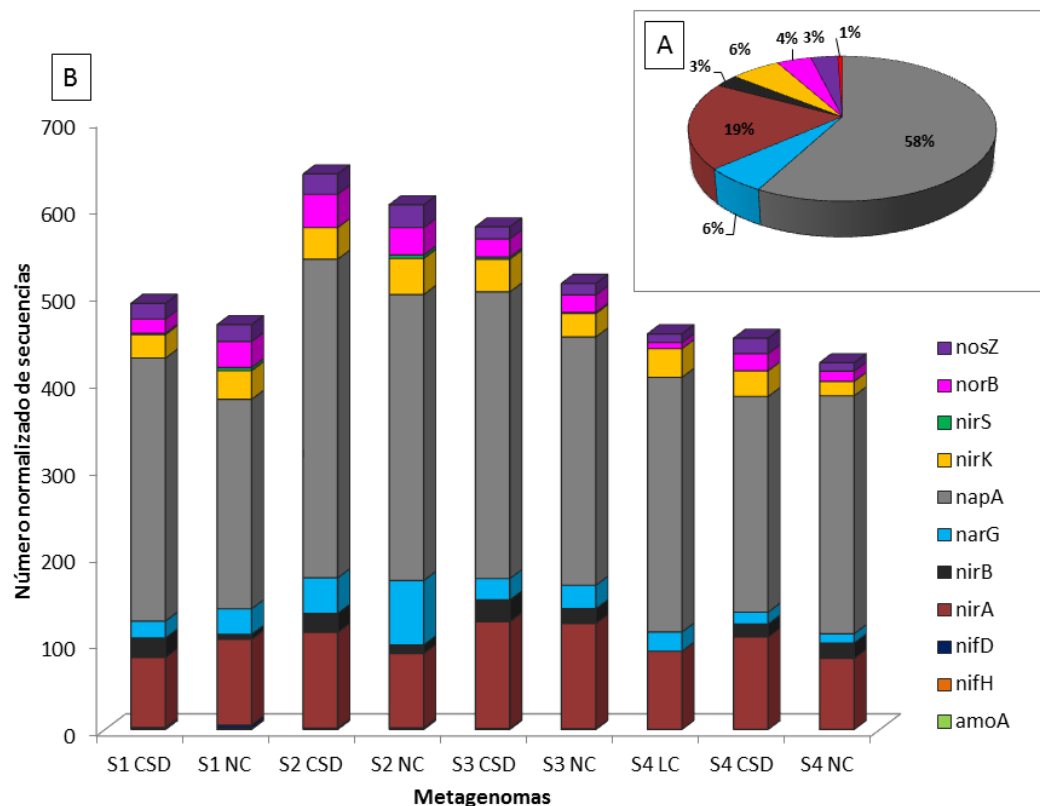


Figura III.10. Abundancia relativa de genes involucrados en el metabolismo nitrogenado en los metagenomas de suelo. En A se muestra la proporción de los genes marcadores del ciclo del nitrógeno en el Pampa *dataset*, y la porción en rojo hace referencia a los genes *nifH*, *nifD*, *amoA* y *nirS* con una representación <1% del total. En B se muestra la abundancia relativa de los genes marcadores en cada metagenoma. El total de secuencias asignadas fue de 490 secuencias en S1 CSD; 466 secuencias en S1 NC; 639 secuencias en S2 CSD; 633 secuencias en S2 NC; 580 secuencias en S3 CSD; 513 secuencias en S3 NC; 455 secuencias en S4 LC; 450 secuencias en S4 CSD y 420 secuencias en S4 NC.

Resumiendo, en esta búsqueda en el Pampa *dataset* de secuencias con función asignada a genes del metabolismo de nitrógeno, encontramos que 2.734 y 4.616 secuencias fueron clasificadas de acuerdo a la base de datos eggNOG y la base de datos personalizada, respectivamente. Cabe mencionar que la información brindada por ambas bases de datos es distinta. Por un lado, el concepto de clúster de grupos

ortólogos implica aquellas proteínas que derivan de una secuencia única ancestral, existiendo en el último ancestro común de las especies que están siendo comparadas y cuya divergencia se debe a la especiación [127].

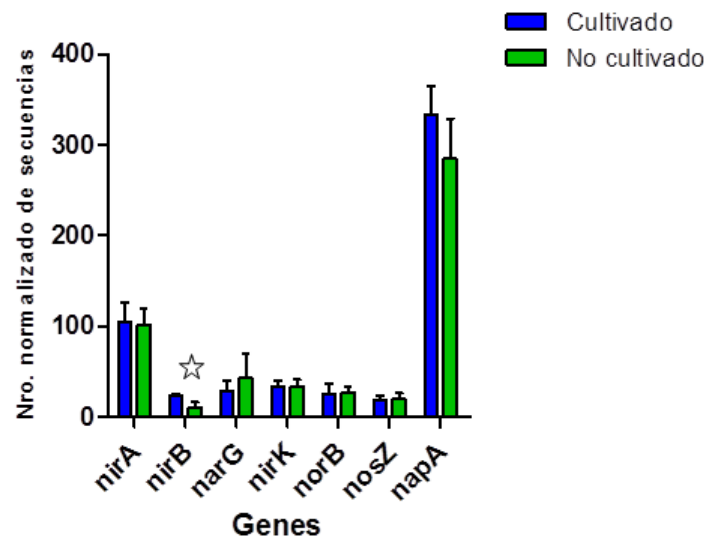


Figura III.11. Comparación de la abundancia de genes marcadores de procesos metabólicos del ciclo del N en los suelos cultivados y no cultivados de los sitios S1, S2 y S3. Se aplicó la prueba de Fisher y la estrella indica diferencias significativas con un valor de $p \leq 0.05$.

El uso de grupos ortólogos puede dar lugar a una herramienta poderosa para el análisis funcional y la anotación de secuencias metagenómicas ya que las proteínas ortólogas tienden a presentar funciones equivalentes [178]. Por otro lado, la otra base de datos construida para este estudio no solo permite recuperar las secuencias codificantes de genes de interés asociados al ciclo del N, sino que también inferir la filogenia de la secuencia incógnita (*query*) a través de la clasificación taxonómica del mejor *hit* del blast. Debido a que en la búsqueda de secuencias mediante blast los parámetros fueron los mismos, el mayor número de secuencias recuperadas utilizando la base de datos personalizada en comparación con la base de datos eggNOG, posiblemente corresponda a que la base de datos personalizada presenta un mayor número de secuencias específicas de los genes de interés del ciclo del N.

Coincidentemente, utilizando las dos bases de datos se obtuvo una baja representación de secuencias asignadas a genes marcadores como *nifH*, *nifD* y *amoA*.

No obstante el análisis funcional metagenómico para estudiar procesos del metabolismo del N en suelo tales como, la reducción de nitrito y denitrificación resultó en una recuperación de datos apreciable.

Ha sido reportado en el trabajo de Sessitsch *et al.* 2012 una baja proporción de secuencias metagenómicas asignadas a los genes *nifH/nifD* en suelo comparado con metagenomas de otros ambientes como sedimentos marinos, intestino de termitas y endófitos de arroz [135]. En el trabajo de Pfister y colaboradores se comparó la distribución de los procesos metabólicos del ciclo del N en varios metagenomas de suelo y marinos, encontrando que la reducción asimilativa de nitrato y nitrito resultaron con mayor cantidad de lecturas asignadas [179]. Por lo tanto, el proceso de reducción asimilativa de nitrato y nitrito, cuyos productos de reducción son respectivamente NO_2^- y NH_4^+ , parecieran estar ampliamente representado en varios biomas. Este proceso metabólico presenta el importante rol de proveer moléculas precursoras para la biosíntesis de compuestos celulares nitrogenados tales como proteínas y ácidos nucleicos esenciales para la viabilidad celular.

Se describió que el proceso de denitrificación presenta una distribución relativamente mayor en comparación con la fijación de nitrógeno y nitrificación en suelo [179]. Por otra parte, el proceso de denitrificación fue predominante (79%) en metagenomas de lodos activados comparado con la amonificación (17%), fijación de nitrógeno (3%) y nitrificación (1%) [170]. La denitrificación es un proceso importante de respiración anaeróbica en condiciones limitantes de oxígeno (O_2) llevado a cabo por una amplia gama de bacterias del suelo (organotrofas, quimio- y fotolitotrofos, termófilas, halófilas, etc.). Se han clasificado más de 50 géneros con al menos 125 especies denitrificantes [98]. En general, los microorganismos denitrificantes constituyen entre el 0.1 al 5% de la población cultivable total del suelo, y hasta el 20% del total de la biomasa microbiana [180] indicando una alta proporción de los mismos. Estos datos podrían explicar que en la identificación de secuencias del suelo relacionadas a genes con función asignada a procesos del ciclo del N, la denitrificación haya sido uno de los procesos que presentó la mayor proporción relativa de *hits*.

Es importante mencionar que estos resultados no infieren que la expresión o actividad de los genes marcadores de la reducción asimilativa de nitrito y la denitrificación se encuentren necesariamente aumentadas en comparación con el resto de los genes marcadores, pero una mayor abundancia de secuencias recuperadas para estos genes si indican la ubicuidad de los mismos en los microorganismos del suelo. Además, la presencia de un alto número de secuencias de genes estratégicos ya ha sido relacionada con formas de vida de microorganismos copiotróficos y oligotróficos [120], [158].

La estructura de la comunidad de microorganismos del suelo asociada al metabolismo nitrogenado

En esta sección se describe la estructura de la comunidad de microorganismos del suelo asociada a los distintos procesos metabólicos analizados del ciclo del N. Para ello se utilizó la clasificación taxonómica del mejor *hit* del blast contra la base de datos personalizada.

a. Fijación biológica de nitrógeno molecular (genes *nif*) y nitrificación (gen *amoA*)

Las secuencias con la asignación funcional *nifH* correspondieron al filo *Proteobacteria* y los géneros hallados fueron *Rubrivivax*, *Anaeromyxobacter* y *Bradyrhizobium*. Con respecto al gen *nifD*, nuevamente el filo *Proteobacteria* se encontró en los 5 metagenomas de suelo que presentaron *hit* para el gen *nifD*. Además, en los suelos S1 CSD y S2 CSD se halló el filo *Cyanobacteria* (género *Nostoc*) y en S2 CSD también el filo *Firmicutes* (género *Desulfotomaculum*). A nivel de géneros para el filo *Proteobacteria* se encontraron los siguientes: *Geobacter*, *Burkholderia*, *Methylocystis*. El proceso de la nitrificación evaluado a través del gen marcador *amoA*, presentó 2 *hits* y correspondieron al filo *Proteobacteria* y al género *Nitrosospira*.

Por lo tanto, si bien la representación de estos procesos metabólicos fue escasa en el Pampa *dataset*, se encontró que el filo *Proteobacteria* fue predominante para ambos procesos en suelo. Este filo ha sido reportado como uno de los más predominantes en la comunidad diazotrófica del suelo, y comprende una amplia

variedad de géneros bacterianos con estilo de vida y capacidades metabólicas diversas [83], [181]. La mayoría de los géneros revelados para genes *nif*, corresponden a microorganismos fijadores de N en vida libre en suelo [52], [182]. Para el caso del gen *amoA* el género *Nitrospira* ha sido identificado como un grupo importante de bacterias aerobias estrictas del suelo que llevan a cabo la primera etapa de oxidación de amoníaco [51], [183].

b. Reducción de nitrito

- Reducción asimilativa de nitrito

El producto final del proceso de reducción asimilativa del nitrito es el NH_4^+ , el cual es utilizado por los microorganismos mediante la acción de la glutamina sintetasa para la biosíntesis de compuestos nitrogenados. Este proceso analizado mediante *nirA* correspondió al 19% del total de los genes marcadores estudiados en el Pampa *dataset*. En la Figura III.12 se muestran los 10 fila asociados a al gen *nirA*. Los fila *Actinobacteria*, *Aquificae* y *Proteobacteria* se hallaron presentes en todos los metagenomas de suelo con una representación porcentual de secuencias del $18 \pm 6.7\%$, $32 \pm 7.2\%$ y $33 \pm 10\%$, respectivamente. Otro filo como *Choloroflexi* fue detectado únicamente en S2 CSD y el resto de los fila se hallaron distribuidos entre los metagenomas de suelo con proporciones diversas.

Se comparó la distribución porcentual de los fila presentes en los suelos CSD y NC de los sitios S1, S2 y S3 y se observó que el filo *Actinobacteria* se encontró mayormente representado ($p=0.02$) en los suelos no cultivados; mientras que las *Cianobacterias* ($p=0.01$) lo hicieron en los suelos cultivados.

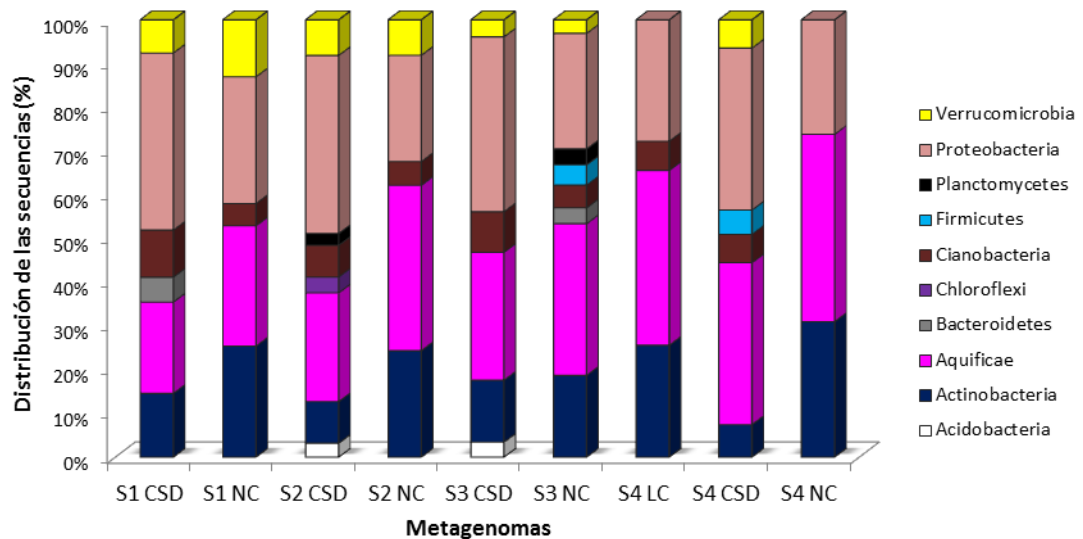


Figura III.12. Distribución porcentual de los filos asignados al gen *nirA* en los metagenomas de suelo. La afiliación taxonómica de las secuencias codificantes para la nitrito reductasa se realizó mediante la anotación del mejor *hit* del Blast con un valor de corte de E de $1e^{-20}$.

En el estudio metagenómico del potencial metabólico y filogenético en suelos agrícolas realizado por Fierer *et al.* 2011, propusieron al filo *Actinobacteria* como una taxa copiotrófica, es decir, aquella que prospera en condiciones de elevada disponibilidad de nutrientes lábiles [158]. Los organismos copiotróficos suelen ser estrategias-r, con altas tasas de crecimiento frente a la disponibilidad de nutrientes. Los suelos bajo manejo agronómico podrían clasificarse copiotróficos frente a los suelos no cultivados, debido a la presencia de fertilización nitrogenada y de residuos orgánicos luego de cada cosecha. En línea con esto, se ha demostrado que la fertilización y los residuos orgánicos han aumentado los niveles de materia orgánica lábil en suelos de la Región Pampeana [184]. No obstante, en nuestros datos, encontramos a este filo representado mayormente en los suelos no cultivados. Este grupo de bacterias es muy interesante desde el punto de vista agronómico ya que presentan varias funciones que favorecen la fertilidad de los suelos. Debido a su estructura filamentosa y a su producción de polisacáridos extracelulares estas bacterias pueden unir partículas de suelo entre sí y de esa forma contribuir a mantener la matriz del suelo [185], [186]

- Reducción desasimilatoria de nitrito

El proceso de reducción desasimilatorio de nitrito, permite la detoxificación del nitrito que se acumula en las células que respiran anaeróbicamente el nitrato [97]. Este proceso está catalizado por la nitrito reductasa codificada por el gen *nirB* que en el Pampa *dataset* representó el 3% del total de los genes analizados. Como se muestra en la Figura III.13 se hallaron 5 fila para el gen *nirB* distribuido en los diferentes metagenomas de suelo, excepto para el sitio S4 LC. El filo *Proteobacteria* predominó entre los metagenomas de suelo ($75 \pm 21\%$) y en algunos suelos también se hallaron los filo *Verrucomicrobia*, *Firmicutes*, *Actinobacteria* y *Planctomycetes*. No se encontraron diferencias significativas en la predominancia del filo *Proteobacteria* entre los suelos cultivados y no cultivados de los sitios S1, S2 y S3.

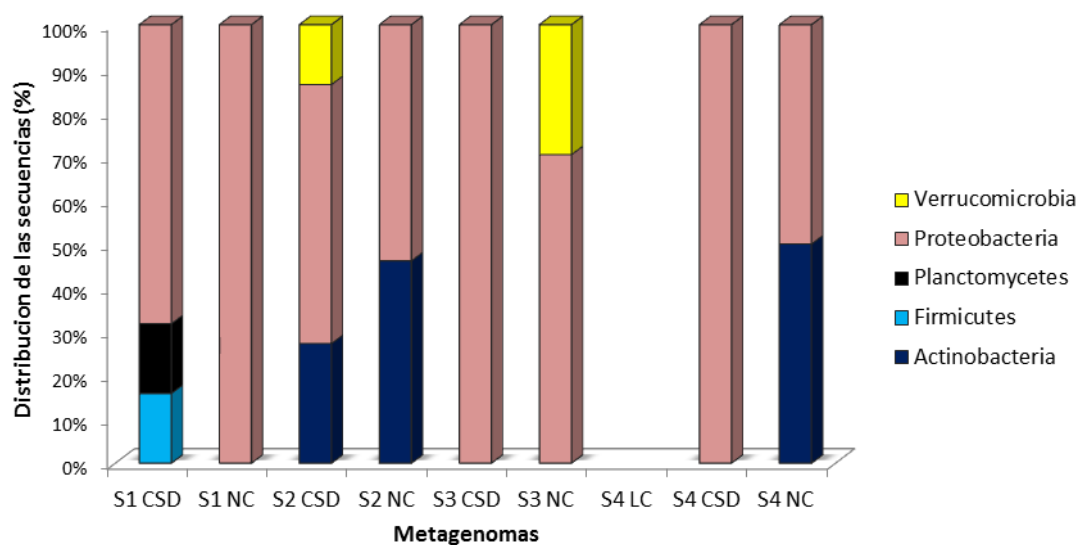


Figura III.13. Distribución porcentual de los filo asignados al gen *nirB* en los metagenomas de suelo. La afiliación taxonómica de las secuencias codificantes para la nitrito reductasa se realizó mediante la anotación del mejor *hit* del *Blast* con un valor de corte de E de $1e^{-20}$.

c. Proceso de denitrificación

El proceso de denitrificación correspondió al 77% del total de los procesos metabólicos asociados al metabolismo de N. Se hallaron 16 filo y un grupo correspondiente a bacteria no cultivable asociadas a las diferentes etapas de la denitrificación evaluadas el Pampa *dataset*. A continuación se describe la inferencia

de la estructura de la comunidad de microorganismos del suelo asociada a los distintos procesos reductivos que comprende la denitrificación.

- **Nitrato reductasa: *narG* y *napA***

El gen *narG* codificante para la enzima nitrato reductasa de unión a membrana representó en el Pampa *dataset* el 6% del total de los genes evaluados. Como se muestra en la Figura III.14 A se encontró que el filo *Proteobacteria* fue predominante en todos los metagenomas de suelo ($67 \pm 15\%$), y en algunos suelos como S4 CSD y S4 NC correspondió al único filo revelado para *narG*. Además, también se halló un grupo de bacterias no cultivables y el filo *Actinobacteria* representados en la estructura de la comunidad microbiana asociada a la reducción disimilativa del nitrato. Con respecto a la distribución de los fila entre los suelos bajo distinto tratamiento agronómico, no se hallaron diferencias significativas para el gen *narG*.

Para a la nitrato reductasa periplásmica (gen *napA*), la cual presentó el mayor porcentaje (42%) de secuencias asignadas respecto de los otros genes analizados, se encontraron 9 fila asociados a este gen (Figura III.14 B). Particularmente, 5 fila los cuales fueron abundantes también se hallaron distribuidos en todos los metagenomas de suelo. Tal fue el caso de *Caldiserica* ($5.4 \pm 1.2\%$), *Euryarchaeota* del dominio Archea ($15 \pm 3.6\%$), *Firmicutes* ($38 \pm 4.9\%$), *Proteobacteria* ($33 \pm 3.9\%$) y *Verrucomicrobia* ($3.1 \pm 0.6\%$).

Ha sido reportada la ocurrencia de *napA* en una amplia gama de bacterias heterótrofas y archeas [187]–[189]. Sin embargo, la información sobre la contribución relativa de la actividad de archeas en la reducción global de nitratos en ecosistemas naturales es aún limitada. El filo *Euryarchaeota* representó uno de los tres más abundantes ($15 \pm 3.6\%$) en los metagenomas aquí examinados, señalándolo como un componente importante de la población de microorganismos asociados a *napA*.

El filo *Deinococcus thermus*, comprende a bacterias quimiorganotrofas resistentes al calor. Fueron aisladas a partir de una variedad de ambientes (suelo, heces y carne), y necesita un medio complejo de crecimiento *in vitro*. Se especula que estas bacterias aeróbicas son propensas a vivir en hábitats orgánicos ricos, como las heces o contenido intestinal [190]. Se encontró que este grupo de bacterias asociadas a

la reducción de nitrato fue detectado únicamente en los suelos cultivados de los sitios S1 (2.5%), S2 (2%) y S3 (1.3%).

Bacteroidetes fue otro filo revelado únicamente en suelos agrícolas con una distribución porcentual del 1.7% en S1, 1.2 en S2 y 1.8 en S3. Este filo comprende un grupo grande de bacterias gram negativas y anaeróbicas con amplia distribución en el medio ambiente, incluyendo suelos, sedimentos marinos y el tracto digestivo de los animales [191]. Ya ha sido reportado que varias especies pertenecientes al filo contienen genes implicados en la denitrificación indicando una posible participación en el ciclo del N en [192]. Por otra parte, Fierer *et al.* 2011 en su trabajo con enfoque metagenómico sobre el potencial metabólico y filogenético en suelos agrícolas identificó a este grupo de bacterias como un posible taxa copiotrófico putativa [158]. Por lo tanto, se podría inferir que la presencia de *Bacteroidetes* exclusivamente en los suelos cultivados analizados aquí se relacionaría con estrategias de vida copiotróficas.

Se encontró que el filo *Proteobacteria* fue el único que presentó una distribución significativamente mayor ($p=0.03$) en los suelos cultivados respecto de los no cultivados de los sitios S1, S2 y S3 para el gen *napA*. Este filo ha sido postulado como una posible taxa copiotrófico [158]. Además, Campbell *et al.* 2010 encontró al filo *Proteobacteria* en mayor abundancia en suelos fertilizados [50], un patrón también observado por Wessen *et al.* 2010 en suelos agronómicos [193].

- **Nitrito reductasa: *nirS* y *nirK***

Con el 0.2% del total de las secuencias asignadas, el gen *nirS* se halló pobremente representado en el Pampa *dataset* al igual que los genes *amoA*, y *nif*. Los *hits* se hallaron distribuidos en los sitios S1 (CSD y NC), S2 NC y en S3 (CSD y NC) y el filo asociado a este gen fue *Proteobacteria* y un grupo de bacterias no cultivables.

Por otro lado, el gen *nirK* presentó una abundancia del 6% respecto de los otros genes evaluados, y como se muestra en la Figura III.15, se reveló que este gen se encontró asociado a 7 fila. En particular, 4 fila se encontraron presentes en todos los metagenomas de suelo, tal es el caso del grupo predominante *Proteobacteria* ($70 \pm 5.8\%$), *Bacteroidetes* ($9.4 \pm 3.9\%$), *Spirochaetes* ($9.1 \pm 2.9\%$) y *Chloroflexi* ($8.1 \pm 3.1\%$). También se reveló un grupo de bacterias no cultivables presente en el sitio S1

CSD (5.7%), S1 NC (6%) y en S2 NC (4%); y el filo *Euryarchaeota* (1.4%), únicamente representado en el sitio S2 CSD.

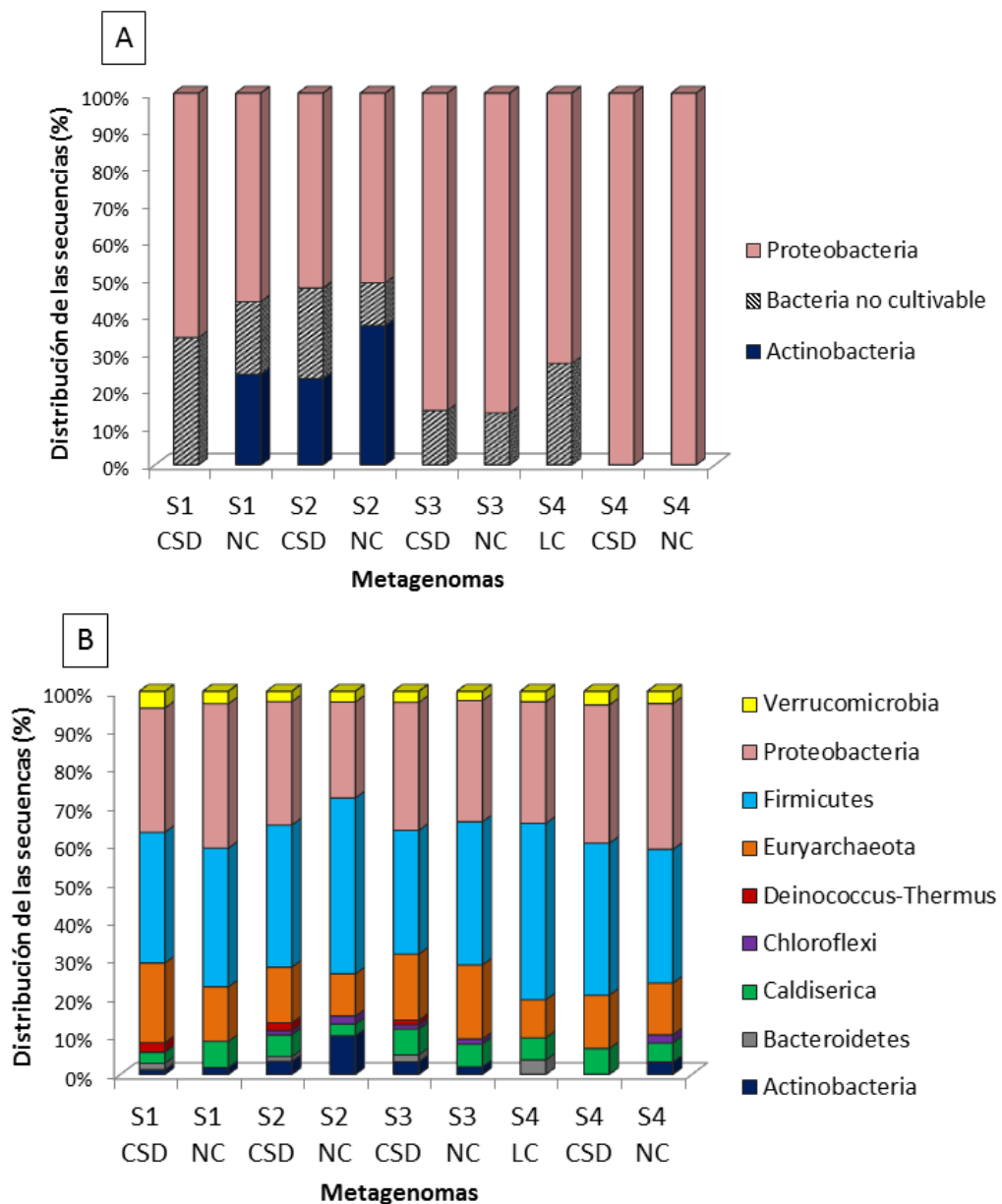


Figura III.14. Distribución porcentual de los fila asignados al gen *narG* (A) y *napA* (B) en los metagenomas de suelo. La afiliación taxonómica de las secuencias codificantes para las nitrato reductasas se realizó mediante la anotación del mejor *hit* del *Blast* con un valor de corte de E de $1e^{-20}$.

Al igual que para el caso del gen *napA*, el perfil de los fila más abundantes fueron similares en todos los sitios indicando que la estructura microbiana asociada al gen *nirK* resulta comparable a través de los mismos. Desde otra perspectiva, el

criterio de análisis basado en la detección *in silico* de funciones a partir de secuencias de metagenomas es insuficiente para revelar diferencias y particularidades de los suelos con distinto uso.

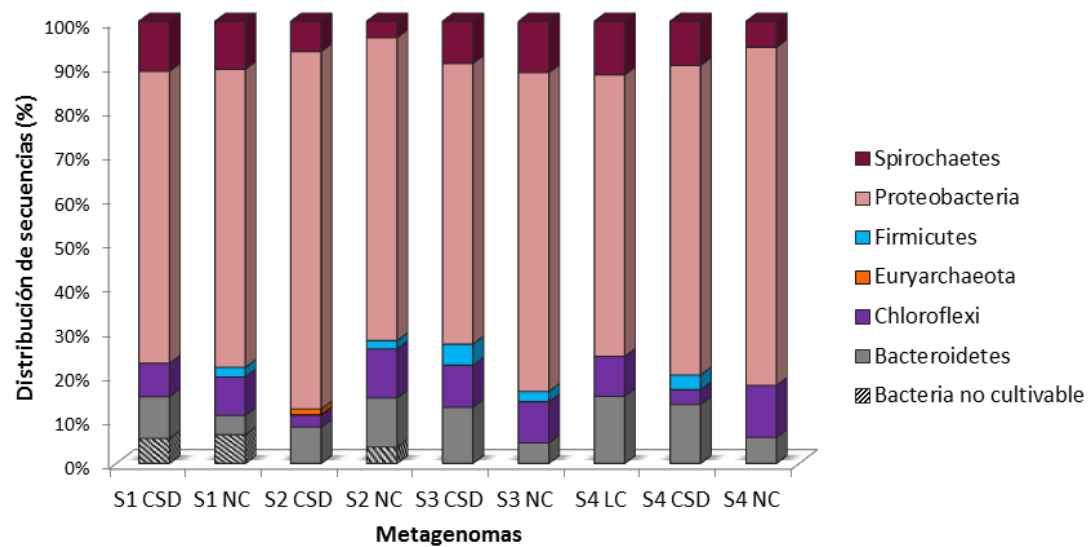


Figura III.15. Distribución porcentual de los filas asignados al gen *nirK* en los metagenomas de suelo. La afiliación taxonómica de las secuencias codificantes para la nitrito reductasa se realizó mediante la anotación del mejor *hit* del Blast con un valor de corte de E de $1e^{-20}$.

- ***Oxido nítrico reductasa: norB***

El gen *norB* codificante de la subunidad mayor de la óxido nítrico reductasa, la cual cataliza la reacción de reducción $\text{NO} \rightarrow \text{N}_2\text{O}$, presentó el 4% del total de secuencias asignadas a los genes de interés evaluados. Como se muestra en la Figura III.16 las secuencias se distribuyeron en 4 filas y a un grupo que se correspondió a bacterias no cultivables. Dentro de los 4 filas, el grupo de *Proteobacteria* y *Acidobacteria* se encontraron distribuidos entre todos los metagenomas de suelo. El grupo de *Proteobacteria* presentó un rango de distribución de entre 33 y 64% en los diferentes sitios, excepto para S1 CSD que fue del 9%. La representación de *Acidobacteria* fue de entre el 15 al 57% con la mayor predominancia en S4 LC. También se reveló el filo *Actinobacteria* en la mayoría de los suelos evaluados. La mayor representación de porcentajes de este filo correspondió al 31%, 25% y 17% respectivamente en los sitios S1 CSD, S4 CSD y S2 NC. Mientras que en los otros sitios la distribución fue menor de entre el 5 y 9%. Además, un grupo correspondiente

a bacterias no cultivables fue encontrado en muchos de los metagenomas. La distribución de este grupo varió entre el 8 y 28%. No se encontraron diferencias significativas con respecto a la de distribución de fila entre los suelos cultivados y no cultivados.

En este estudio, el filo *Acidobacteria* se puso en evidencia solo cuando el marcador de búsqueda fue el gen *norB* y (Figura III.16). Ward *et al.* 2009 a partir de la evaluación de la presencia del gen *norB* en algunas cepas que habitan el suelo reportaron la capacidad del filo *Acidobacteria* para la reducción del óxido nitroso [194]

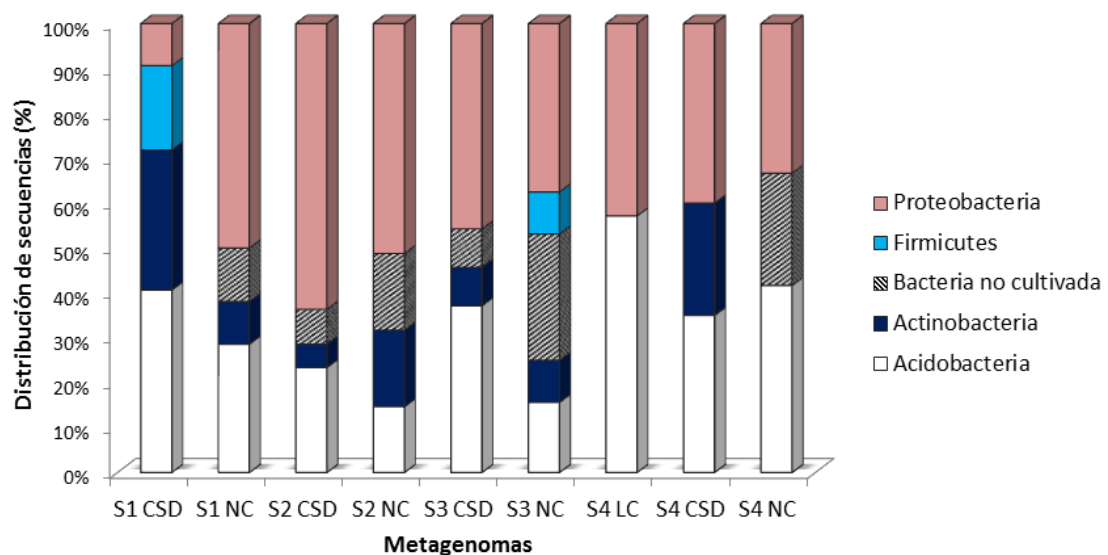


Figura III.16. Distribución porcentual de los filia asignados al gen *norB* en los metagenomas de suelo. La afiliación taxonómica de las secuencias codificantes para la óxido nitroso reductasa se realizó mediante la anotación del mejor *hit* del blast con un valor de corte de E de $1e^{-20}$.

- *Oxido nitroso reductasa: nosZ*

El último paso de la reducción disimilativa en la desnitrificación, $N_2O \rightarrow N_2$, esta mediada por la enzima óxido nitroso reductasa, codificada por el gen *nosZ*. Representó el 3% del total de secuencias asignadas a los genes evaluados y como se muestra en la Figura III.17 la diversidad correspondió a 9 filia. El filo *Verrucomicrobia* y el grupo de bacterias no cultivables se encontraron presentes en todos los metagenomas de suelo. Este último grupo presentó una predominancia considerable del $30 \pm 8.4\%$ en todos los sitios. La mayor abundancia del filo

Verrucomicrobia fue detectada en los sitios S2 NC y S3 CSD con una distribución porcentual del 26.5 y 13%, respectivamente; mientras que en el resto de los metagenomas la abundancia varió entre el 3.5 y el 10%. Otros fila revelados en varios de los suelos fueron *Bacteroidetes*, *Aquificae*, y *Proteobacteria*. Los grupos fotosintéticos anoxigénicos *Chlorobi* y *Chloroflexi* se hallaron en seis metagenomas de suelo, y el mayor porcentaje se reveló en el sitio S4 CSD. En cuanto a los fila *Euryarchaeota* y *Gemmatimonadetes* estos se encontraron en cinco de los sitios analizados, y su distribución varió del 4.3 al 10% y 2.1 al 21.7%. La búsqueda *in silico* de secuencias *nosZ* reveló una apreciable diversidad de microorganismos portadores del gen *nosZ*, aunque ninguno de los grupos taxonómicos predominó significativamente en alguno de los tratamientos (cultivado y no cultivado).

Conclusiones

En este capítulo se encaró el análisis *in silico* de una base de datos de suelo de 1.5 Gb, pertenecientes a la principal región de actividad agrícola de la Argentina. En una primera instancia mediante la clasificación de las secuencias en subsistemas funcionales se comparó el potencial metabólico de los suelos y ambientes acuáticos de diferentes regiones del mundo revelándose que el suelo a escala global presenta ciertos atributos funcionales característicos del ambiente. Esto se observó tanto en los análisis de ordenación como de conglomerado, en donde el bioma fue el principal conductor de agrupamiento de los metagenomas.

Asimismo, se realizó una búsqueda de secuencias con funciones asignadas a procesos metabólicos claves del metabolismo nitrogenado en suelo. Se utilizaron dos bases de referencia para la búsqueda mediante BlastX, una correspondió a eggNOG y la otra fue una base de datos construida especialmente para este estudio, con el propósito de aumentar las chances de recuperación de secuencias. Concluimos, que los procesos metabólicos como la FBN y la nitrificación resultaron representados en una baja proporción comparado a otros procesos, como la reducción asimilativa de nitrito y la denitrificación. Nuestros resultados sugieren que para evaluar mediante secuenciación masiva al azar los procesos tales como fijación biológica de nitrógeno y nitrificación, llevados a cabo por componentes de la comunidad microbiana

considerado por varios autores como componentes “raros” del suelo [195]–[197], es necesario una secuenciación con mayor cobertura. El trabajo reciente de Yu *et al.* 2012 provee bases para esta especulación [170]. Estos autores evaluaron diferentes niveles de secuenciación (0.5 Gbp, 1.0 Gbp y 2.4 Gbp) de ADN de lodos activados, encontrando que con una importante profundidad de secuenciación (2.4 Gbp) se logró aumentar significativamente el número de secuencias recuperadas y asociadas a categorías funcionales del metabolismo nitrogenado, incluida la fijación de nitrógeno y la nitrificación. La profundidad de secuenciación lograda les permitió revelar y describir tales procesos metabólicos, los cuales quedan excluidos o ignorados en proyectos de secuenciación “superficial” con una secuenciación de un millón de lecturas por biblioteca. Concluimos que esta observación es un hallazgo interesante a tener en cuenta en futuros proyectos dirigidos a revelar capacidades metabólicas de poblaciones con baja representación. Una alternativa para el estudio de procesos metabólicos como la FBN y la nitrificación sería la pirosecuenciación dirigida a los genes marcadores de tales procesos [121], [198], [199]

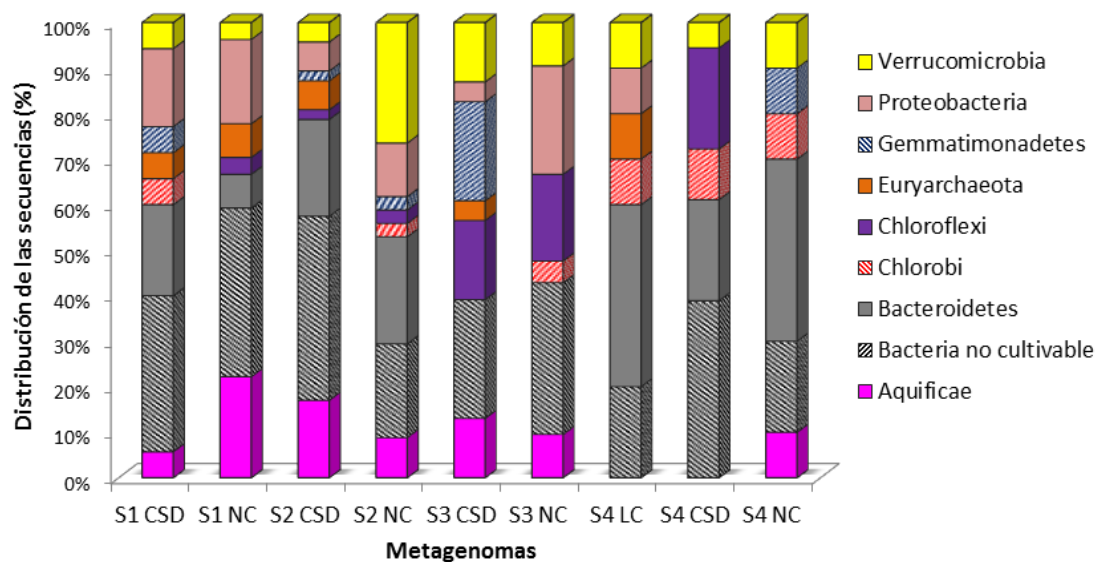


Figura III.17. Distribución porcentual de los filas asignados al gen *nosZ* en los metagenomas de suelo. La afiliación taxonómica de las secuencias codificantes para la óxido nítrico reductasa se realizó mediante la anotación del mejor *hit* del blast con un valor de corte de E de $1e^{-20}$

Hemos mostrado la presencia de diversos taxones involucrados en las distintas etapas del ciclo del nitrógeno, muchos comúnmente distribuidos en suelo y otros

como *Aquificae* y *Deinococcus-Thermus* que suelen habitar ambientes con condiciones extremas de temperatura. Asimismo, se encontró al filo *Euryarchaeota* del dominio Archaea representado en los procesos metabólicos de reducción de nitrato, óxido nítrico y óxido nitroso. Dada la importancia del nitrógeno en la nutrición vegetal, particularmente en los casos de cultivos de importancia económica, el conocimiento de las comunidades potencialmente participantes en esos procesos es clave para evaluar la sustentabilidad de los suelos. Nuestros resultados aparecen como un indicador importante para encarar futuros análisis funcionales con el propósito de caracterizar los suelos en sus potencialidades de llevar a cabo los procesos biogeoquímicos, teniendo en cuenta que este ambiente es uno de los más desafiantes por su alta diversidad microbiana.

Capítulo IV

**Potencialidad del análisis T-RFLP
para el estudio de las poblaciones
fijadoras de nitrógeno en suelo**

Introducción

Muchos de los microorganismos presentes en los diferentes ecosistemas son recalcitrantes al cultivo. Las investigaciones sobre la estructura y la diversidad genética de comunidades microbianas han provocado el continuo desarrollo y uso de métodos independientes del cultivo, principalmente moleculares, tales como los que involucran la reacción en cadena de la polimerasa (PCR) de genes ribosomales (ARNr) [17]. Si bien, es posible obtener información filogenética de los microorganismos mediante el secuenciamiento del gen ARNr 16S, el esfuerzo y el costo limitan el análisis de numerosas muestras. Esto a su vez, limita encarar el estudio de la dinámica de la comunidad microbiana en relación a cambios en las condiciones fisicoquímicas y/o biológicas en el tiempo y espacio, lo cual es esencial cuando se pretende estudiar la funcionalidad de una comunidad.

Ciertas técnicas moleculares de tipo huella digital (*fingerprint*) para el análisis de la estructura y dinámica de comunidades microbianas complejas permiten el análisis simultáneo de múltiples muestras; como por ejemplo: electroforesis en gel con gradiente de desnaturalización DGGE (*Denaturing Gradient Gel Electrophoresis*) [33], electroforesis en gel con gradiente de temperatura TGGE (*Temperature Gradient Gel Electrophoresis*) [34], [35] y polimorfismo de conformación de cadena simple SSCP (*Single-Strand Conformation Polymorphism*) [36], [37]. Debido a que tales métodos pueden ser automatizados, resultan convenientes para el procesamiento de varias muestras de una manera reproducible [41]. En ese contexto, otras metodologías nuevas incorporaron sistemas de secuenciación automatizado con detección láser de fragmentos de ADN marcados con fluoróforos. Entre las diversas técnicas, encontramos por ejemplo, polimorfismo de longitud de fragmentos de restricción terminales T-RFLP (*Terminal Restriction Fragment Length Polymorphism*) [38], análisis de la heterogeneidad de genes amplificados por PCR [200], y más recientemente ADN polimórfico amplificado al azar RAPD (*Randomly Amplified Polymorphic DNA*) [39].

En particular, la técnica de T-RFLP se basa en la variación de tamaño de los fragmentos terminales de restricción (T-RFs) 5' y/o 3' marcados por fluorescencia debido al polimorfismo de la secuencia correspondiente a sitios de reconocimiento de

endonucleasas. Estos T-RFs se resuelven por medio de electroforesis en geles de alta resolución en secuenciadores de ADN y se detectan mediante excitación láser (Figura IV.1). Esta técnica fue originalmente desarrollada en 1996 por Avannis-Aghajani *et al.*, para la identificación de diferentes variantes de micobacterias a partir de muestras de agua y de asilamientos clínicos [201]. Posteriormente, Liu *et al.* 1997 demostró el potencial de la técnica T-RFLP para caracterizar la diversidad microbiana en hábitats con comunidades complejas (lodos activados, lodos de biorreactor, intestino de termitas y arena de acuífero) y la obtención de distintos perfiles T-RFLP permitió comparar rápidamente la estructura de las comunidades microbianas en los diferentes ambientes [38].

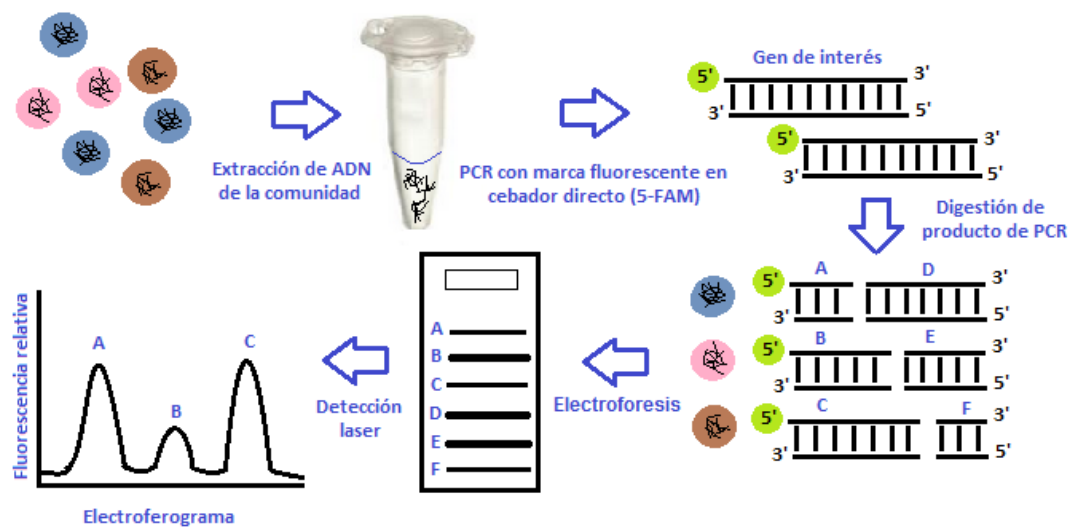


Figura IV.1 Esquema de las principales etapas del ensayo de T-RFLP. El método se basa en la amplificación de una secuencia, ya sea genes ribosomales o funcionales, usando uno o ambos cebadores marcados con un fluoróforo, y la posterior incubación del producto de amplificación con endonucleasas de corte frecuente. Luego, el producto de digestión se mezcla con un marcador interno de tamaño el cual está marcado con un fluoróforo distintivo y finalmente la mezcla de los fragmentos se resuelven mediante electroforesis en gel o capilar. La detección de los fragmentos terminales marcados se realiza en un secuenciador automático con detector láser, generándose un perfil gráfico de los T-RFs denominado, electroferograma.

En los últimos años esta técnica ha sido ampliamente utilizada para caracterizar perfiles poblacionales de bacterias [202]–[206] y de archaeas [207]–[211] en hábitats naturales, así como también la dinámica o evolución de las comunidades microbianas naturales en una escala tiempo/espacio [183], [212]–[214]. Además, la técnica se aplicó para el estudio de diversos genes funcionales [183], [215], [216] como los involucrados en la fijación de nitrógeno [217], [218] y la oxidación de amonio [183], [219]. Algunos autores sostienen que la metodología T-RFLP puede ser utilizada para un análisis semi cuantitativo de las proporciones relativas de los miembros dominantes en una comunidad microbiana [38], [220]. En el trabajo de Tan *et al.* 2003 se utilizó la técnica T-RFLP para reflejar el efecto del agregado de fertilizante nitrogenado en la composición de la comunidad diazotrófica asociada a las raíces de variantes de arroz en una manera semi cuantitativa. Mostraron que luego del agregado de fertilizante ciertos T-RFs presentaban una abundancia menor, y además otros T-RFs se encontraron ausentes comparado con el perfil de T-RFLP de plantas controles [218].

Sobre las limitaciones del T-RFLP, se puede mencionar la baja resolución, ya que existe cierta probabilidad de que varios filotipos se encuentren representados por un solo tamaño de fragmento terminal, es decir aparece solo un pico en el electroferograma, lo que puede llevar a subestimar la diversidad de especies [221]. [221]. Como el método examina amplicones, resulta que los desvíos asociados a la amplificación de secuencias particulares debido a una preferencia en la hibridación de los cebadores, o una mayor incidencia en la síntesis de quimeras por aumento del número de ciclos de PCR reducen la calidad de las predicciones cuantitativas mediante T-RFLP. Por lo tanto, resulta importante reducir el intervalo de variabilidad y/o desvíos asociados en los métodos basados en PCR y definir claramente las condiciones experimentales óptimas.

El T-RFLP es una herramienta de complejidad técnica relativamente rápida e independiente del cultivo de microorganismos, lo que resulta en un método simple y accesible para su aplicación en muchos laboratorios. Asimismo, el método presenta alta sensibilidad, detecta genomas que se encuentran en proporciones tan bajas como el 1% del total de la comunidad [221]. Estas características, sumado a su demostrada capacidad para describir estructuras microbianas de diversos ambientes incluyendo

suelo: [217], [218], [221]–[223], microbiota de intestino humano [38], comunidades endófitas [135], [224], [225] y lodos activados [226], [227] entre otros, justificaron en este trabajo, el uso de T-RFLP como otro criterio de análisis de la comunidad diazotrófica potencial.

Objetivo general

Estudiar la población diazotrófica del suelo mediante la técnica de “Polimorfismo de longitud de fragmentos de restricción terminales” (del inglés *Terminal Restriction Fragment Length Polymorphism*).

Objetivos específicos

- Determinar los perfiles T-RFLP del gen *nifH* de distintas muestras de suelo argentino y compararlos mediante análisis de conglomerados.
- Realizar una asignación taxonómica tentativa de los fragmentos de restricción terminal empíricos mediante un análisis comparativo de restricción *in silico*.

Resultados y Discusión

Algunas consideraciones experimentales

Se analizaron 15 muestras de suelo (Tabla IV.1) y las amplificaciones (por triplicado) se llevaron a cabo de forma independiente a partir del ADN metagenómico de dos réplicas biológicas de suelo. Los amplicones de un mismo sitio-tratamiento se combinaron en una sola muestra (*pool nifH*). Se realizó separadamente la incubación con las enzimas de restricción *MspI* y *HaeIII*. Por conveniencia denominaremos perfil T-RFLP al registro de los fragmentos de tamaños diferentes marcados con fluorescencia, los cuales son detectados como picos de emisión. El conjunto de los picos después de la corrida en el instrumento, conforman para cada muestra el llamado electroferograma.

Tabla IV.1. Sitios analizadas por T-RFLP. Se muestran los sitios de colección y sus respectivas abreviaturas. BPA corresponden a suelos cultivados en siembra directa con Buenas Prácticas Agrícolas, PPA a suelos cultivados en siembra directa con Pobres Prácticas Agrícolas y NC a suelos No Cultivados con pasturas silvestres naturales.

Abreviatura	Lugar de muestreo del suelo	Tratamiento agrícola
S1 CSD	Establecimiento productivo "La Estrella"	Cultivado bajo Siembra Directa
S1 NC	Establecimiento productivo "La Estrella"	No Cultivado
S2 CSD	Establecimiento productivo "La Negrita"	Cultivado bajo Siembra Directa
S2 NC	Establecimiento productivo "La Negrita"	No Cultivado
S3 CSD	Establecimiento productivo "Criadero Klein"	Cultivado bajo Siembra Directa
S3 NC	Establecimiento productivo "Criadero Klein"	No Cultivado
S4 CSD	Unidad Experimental INTA Balcarce	Cultivado bajo Siembra Directa
S4 LC	Unidad Experimental INTA Balcarce	Labranza Convencional
S4 NC	Unidad Experimental INTA Balcarce	No Cultivado
S5 CSD BPA	Pergamino	Cultivado bajo Siembra directa y Buenas Prácticas Agrícolas
S5 CSD PPA	Pergamino	Cultivado bajo Siembra directa y Pobres Prácticas Agrícolas
S5 NC	Pergamino	No Cultivado
S6 CSD BPA	Viale	Cultivado bajo siembra directa y Buenas Prácticas Agrícolas
S6 CSD PPA	Viale	Cultivado bajo siembra directa y Pobres Prácticas Agrícolas
S6 NC	Viale	No Cultivado

Etapas preliminares de optimización de T-RFLP

*a. Selección de cebadores para la reacción PCR de amplificación del gen *nifH**

Según Schütte *et al.* 2008 los cebadores elegidos para un análisis de T-RFLP deben ser específicos y lo suficientemente degenerados para permitir amplificar solo el grupo taxonómico de interés y cubrir la diversidad poblacional existente dentro del grupo [228]. En el trabajo de Poly *et al.* 2001 se estudió la diversidad de secuencias del gen *nifH* en suelo mediante la técnica de polimorfismos en la longitud de los fragmentos de restricción (RFLP), y propusieron tres parámetros esenciales para tener en cuenta en la selección de cebadores *nifH*: - los cebadores deben ser capaces de amplificar secuencias *nifH* de un amplio rango de bacterias fijadoras de N₂, - la amplificación debe ser específica y, - el tamaño del producto de amplificación por PCR debe ser lo suficientemente grande para permitir obtener perfiles de fragmentos de restricción representativos de la diversidad de fijadores de nitrógeno [222].

Gaby y Buckley *et al.* 2012 analizaron varios cebadores universales y otros específicos de grupos usados comúnmente para la amplificación del gen *nifH* [229]. Por una lado, los autores evaluaron *in silico* la capacidad de amplificación encontrando que solo 15 de los 51 cebadores universales analizados presentaron la potencialidad para amplificar más del 90% de las secuencias *nifH* anotadas en la base de datos pública, el resto solo amplificaba menos del 50% de las mismas. Por otro lado, evaluaron empíricamente la capacidad de amplificación de estos 15 cebadores universales a partir de muestras de suelo y de una colección de cepas fijadores de nitrógeno. Hallaron que solo 3 pares de cebadores amplificaron satisfactoriamente genes *nifH*, es decir generaron el producto de PCR del tamaño esperado de todas las cepas fijadores de nitrógeno y muestras de suelo evaluadas, y no hubo amplificaciones en los controles negativos.

Dado que el análisis empírico de los cebadores para la amplificación de *nifH* en el estudio realizado por Gaby y Buckley *et al.* 2012 fue exhaustivo, se utilizó dicho trabajo como referencia para seleccionar los cebadores *nifH* (Tabla IV.2) y llevar a cabo análisis exploratorios empíricos en nuestras muestras de suelo (Tabla IV.1). Los cebadores evaluados aquí también han sido ampliamente utilizados en muestras ambientales y en especial en el suelo [87], [129], [230]–[238].

Tabla IV.2. Pares de cebadores evaluados para la amplificación del gen *nifH* en el presente capítulo.

Par de cebadores	Posición ^a	Tamaño ^b	Referencia
PolF/PolR	115-476	362	[222]
<i>nifH3/nifH4</i>	22-494	473	[131]
<i>nifH1/nifH2</i>	115-476	362	[141]
19F/<i>nifH3</i>	19-494	476	[130], [131]
<i>nifH11/nifH22</i>	118-476	359	[129]

a. Posición del amplicón en *nifH* relativo a *A. vinelandii* (Genbank ACCN# M20568)

b. Tamaño del amplicón en bp esperado de *nifH*.

Se evaluaron los cebadores *nifH1/nifH2*, *nifH3/nifH4*, 19F/*nifH3* y *nifH11/nifH22* mediante la reacción PCR de tipo anidada como ya fuese descrito previamente [129], [131]. Utilizando la combinación de los cebadores *nifH3/nifH4* (1^{er}. PCR) - *nifH1/nifH2* (2^{da}. PCR no se detectó amplificación en algunas muestras de

suelo ensayadas (Tabla IV.1), incluso en aquellos ensayos en que se intentó optimizar variables que mejoran el desempeño de la PCR como la temperatura de hibridación y la concentración de $MgCl_2$.

Los cebadores (*nifH3/nifH4*) fueron diseñados para amplificar el gen *nifH* a partir de ambientes acuáticos y el par *nifH1/nifH2* para la amplificación de secuencias *nifH* de cianobacterias (*Trichodesmium thiebautii*) [141], lo que podría explicar el desempeño limitado que se observó para ciertas muestras de suelo.

Con respecto al uso combinado de los cebadores 19F/*nifH* (1^{er}. PCR) y *nifH11/nifH22* (2^{da}. PCR) se logró la amplificación del gen *nifH* en todas las muestras de suelo evaluadas, así como también cuando se emplearon los cebadores PolF/PolR. Ha sido demostrado por Poly *et al.* 2001 que los cebadores PolF/PolR no solo amplifican satisfactoriamente secuencias *nifH* de 19 cepas de referencia, sino también permiten la amplificación a partir de ADN de muestras de suelo [222]. Estos cebadores que continúan siendo muy utilizados en estudios del gen *nifH* han permitido revelar una amplia diversidad de secuencias *nifH* del suelo [121], [230], [235].

Contrariamente, los estudios *in silico* de Gaby y Buckley *et al.* 2012 muestran que el par cebadores PolF/PolR cubre potencialmente solo el 25% de la diversidad *nifH* de la base de datos utilizada por los autores [229]. Esta baja proporción aparente puede estar relacionada con una evaluación parcial de los cebadores en cepas de referencia que incluyeran representantes del cluster 1A (*Deltaproteobacteria*), cluster III (*Deltaproteobacteria* y *Verrumicrobia*) y Cianobacterias. Poly *et al.* 2001 solo testearon cepas de referencia que consistieron en representantes de *Alpha-*, *Beta-*, y *Gammaproteobacteria*, *Firmicutes* y *Actinobacteria*. Si bien experimentalmente estos cebadores amplifican satisfactoriamente secuencias *nifH* a partir de muestras de ADN metagenómico de suelo, se decidió no utilizarlos en el ensayo de T-RFLP debido a su relativa baja cobertura para ciertos de grupos de diazótrofos ampliamente distribuidos en suelo [229]. No obstante, el juego de cebadores 19F/*nifH3* presentó un 90% de cobertura *in silico* sobre los diversos grupos taxonómicos que poseen *nifH* [229].

Entonces, si bien en una PCR anidada se potencian los desvíos asociados a esta metodología, cuando se combina con condiciones de reacción de rigurosidad baja permite amplificar una diversidad más amplia de templados. Además, se procuró seleccionar aquellos cebadores *nifH* que mejor se ajustaron a los criterios descriptos por Poly *et al.* 2001, relacionados a la especificidad, a la cobertura de la diversidad de secuencias *nifH* y al tamaño de producto de amplificación apropiado para una posterior digestión enzimática. Es por ello, que se decidió utilizar la PCR anidada con el juego de cebadores 19F/*nifH*3 y *nifH*11/*nifH*22 para el estudio del gen *nifH* mediante T-RFLP.

b. Selección de endonucleasas para la digestión del pool de amplicones de *nifH*

En la detección de polimorfismos en ADN comúnmente se utilizan enzimas que reconocen sitios de cuatro pares de bases dada la mayor probabilidad de hallar estos sitios en una secuencia dada. También, varios investigadores han demostrado el aumento de capacidad de discriminación y resolución del método al incorporar varias endonucleasas [38], [41], [222], [239]. Las enzimas *MspI* y *HaeIII* han sido ampliamente utilizadas en el análisis T-RFLP o RFLP de *nifH* en muestras ambientales [130], [217], [222], [223], [230], [238], [240], [241].

Engebretson y Moyer *et al.* 2003 evaluaron la capacidad de 18 enzimas de restricción para resolver T-RFs únicos a partir de comunidades modelos, sugiriendo que solo cuatro de las enzimas entre ellas *MspI* presentaron el mejor desempeño [242]. También concluyeron que para aquellas comunidades con más de 50 OTUs (Unidades Taxonómicas Operativas) ninguna de las enzimas podía resolver más del 70% del total de OTUs. Si bien los suelos presentan una alta diversidad de secuencias *nifH* [83] y un elevado número de OTUs relacionado a la comunidad diazotrófica [121], se consideró que *MspI* y *HaeIII* por sus capacidades ya demostradas de discriminar filotipos y resolver grupos filogenéticos de diazótrofos eran buenas candidatas para el análisis de T-RFLP. Además, estas enzimas no generan cortes en los cebadores seleccionados para la amplificación del gen *nifH*.

c. Ensayos de PCR anidada

En este trabajo se utilizó la reacción de PCR anidada descrita en el trabajo de Yeager *et al.*, 2004 donde el tamaño de amplicón *nifH* obtenido en todas las muestras de suelo analizadas fue el esperado de 359 pb [129]. En la electroforesis en gel de agarosa se observó la banda de 359 pb y además otra banda inespecífica de aproximadamente 900 pb (Figura IV.2). La co-amplificación de la banda inespecífica resultó tenue, y ha sido previamente descrita en un trabajo de Fürnkranz *et al.* 2008 donde utilizaron los mismos cebadores para la amplificación de *nifH* [238].

El trozo de agarosa correspondiente al fragmento de 359 pb (*nifH*) fue aislado del gel y el ADN purificado fue incubado, en reacciones separadas, con las endonucleasas *MspI* y *HaeIII*. A modo de ejemplo, en la Figura IV.3, se muestra un gel analítico con las bandas correspondientes a la muestra de ADN-*nifH* después de la etapa de purificación para el caso de los suelos de Pergamino (S5) y Viale (S6). Se obtuvo un porcentaje de recuperación de ADN-*nifH* de más del 75% entre las diversas muestras.

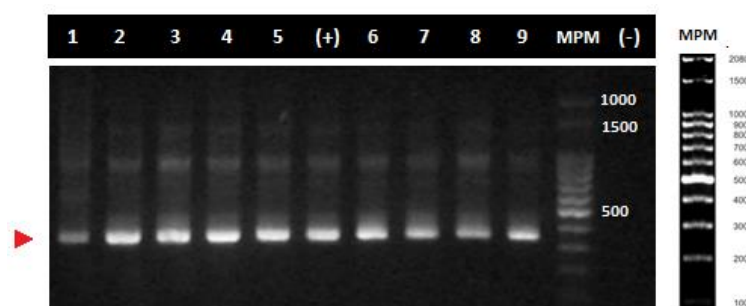


Figura IV.2. PCR anidada de *nifH*. Productos resultantes de la amplificación por PCR anidada del gen *nifH* a partir de ADN metagenómico de suelo. Calles 1-5: S1, S2, S3, S4 BPA y S4 LC; Calles 6-9: S1, S2, S3 y S4 NC. (+) y (-) corresponden al control positivos (resina de *Azotobacter vinelandii*) y negativo de la PCR, respectivamente. En el margen izquierdo de la figura, la punta de flecha roja indica el fragmento de 359 pb de *nifH*. En el margen derecho se muestra los tamaños del marcador de peso molecular (MPM) de 100pb (PB-L Productos Bio-Logicos ®).

En la Tabla Suplementaria IV.1 se detalla el rendimiento y pureza de ADN-*nifH* medida por espectroscopia UV. El rendimiento de ADN-*nifH* presentó valores en el rango de 1 a 2 $\mu\text{g g}^{-1}$ de suelo. La pureza evaluada mediante la relación de los valores

de absorbancia ($A_{260/280}$ y $A_{260/230}$) indicó bajo contenido de contaminantes como ARN, proteínas y ácidos húmicos. Por lo tanto, fue factible utilizar dicho ADN-*nifH* para el análisis de T-RFLP

d. Resolución y alineamiento de los perfiles T-RFLP

Los fragmentos de restricción terminal presentando la marca fluorescente (FAM) en el extremo 5', se identificaron en la mezcla de los fragmentos resultantes de la incubación con las enzimas de restricción *MspI* y *HaeIII* mediante electroforesis capilar. Se adopta la abreviatura T-RFs para hacer referencia a los fragmentos terminales con la marca fluorescente. Es esperable que la incubación de ADN de muestras ambientales con las enzimas de lugar a una mezcla compleja de fragmentos proveniente de diferentes genomas, y esto genere en particular una diversidad de tamaños de fragmentos terminales (Figura IV.4).

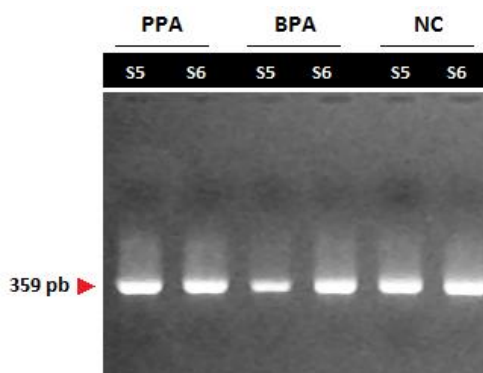


Figura IV.3. Purificación de los amplicones *nifH*. Foto de un gel analítico de agarosa 1% mostrando la banda obtenida después de la purificación de ADN *nifH*, indicado por la flecha de punta roja en el margen derecho de la figura. El ADN fue purificado del taco de agarosa a partir del gel, utilizando el kit comercial NucleoSpin® Extract II (Macherey-Nagel). Las abreviaturas S5 y S6 corresponden a los sitios de Pergamino y Viale, respectivamente. BPA y PPA corresponde a los manejos buenas prácticas agrícolas y pobres prácticas agrícolas, respectivamente. NC a suelo no cultivado.

La utilización del programa *PeakScanner*TM permitió editar y evaluar manualmente la calidad de los perfiles T-RFLP para su posterior alineamiento mediante la herramienta bioinformática T-Align. Los datos de los perfiles T-RFLP se dispusieron en una matriz, donde cada columna correspondió a una muestra y en cada

fila se halló la proporción del área de fluorescencia de los T-RFs (Tabla Suplementaria IV.5).

La base de datos de T-RFs para encarar los diferentes análisis, comprendió 22 T-RFs revelados a partir de la incubación de los amplicones *nifH* con *MspI*, mientras que con *HaeIII* resultó en 33 T-RFs. Aquellos T-RFs revelados por la incubación con *MspI* o *HaeIII* se referirán como T-RF pb_*MspI* y T-RF pb_*HaeIII*, respectivamente.

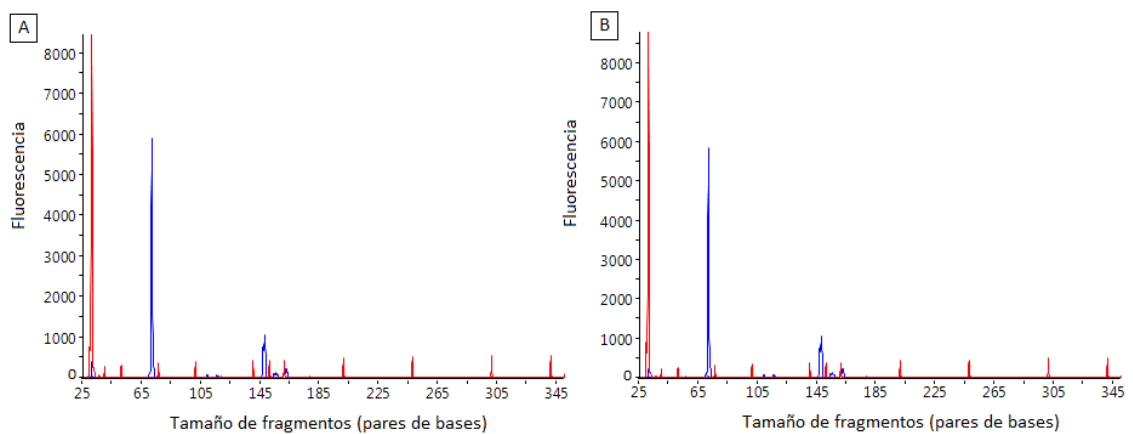


Figura IV.4. Ejemplo de perfiles T-RFLP observados con la enzima de restricción *MspI* en *nifH* amplificado directamente de muestras de suelo. (A) y (B) son réplicas técnicas de la electroforesis capilar, dos submuestras de la mezcla de digestión se analizaron en la misma corrida electroforética en el secuenciador automático. Los picos azules corresponden a los fragmentos de la muestra y los picos rojos al marcador interno de tamaño LIZ-GS500 (-250).

Evaluación de la reproducibilidad del método T-RFLP

Es importante en cualquier metodología molecular que genera perfiles de comunidades microbianas, evaluar el grado de reproducibilidad del procedimiento para así generar confianza y credibilidad a las conclusiones realizadas sobre las comparaciones de similitud entre las comunidades. Además, el uso de T-RFLP como una técnica semi-cuantitativa así como describió Liu *et al.* 1997 requiere eliminar o definir cuantitativamente cualquier variación entre réplicas [38].

A continuación se describen las variables utilizadas para evaluar la reproducibilidad de la metodología de T-RFLP, en el análisis comparativo de la

estructura de las comunidades diazotroficas en muestras de suelos provenientes de la Región Pampeana Argentina. Los parámetros seleccionados consistieron en la presencia/ausencia de los T-RFs y/o en la altura de los picos. Estos análisis permitieron además establecer ciertos criterios de estandarización a la hora de procesar y evaluar los perfiles T-RFLP generados.

a. Variación Intra/Inter gel

Para investigar la variación debida a la técnica electroforética, intra e inter gel, los productos de amplificación por PCR (200 ng de templado) se incubaron con la enzima *MspI* y se dividió el producto de digestión en tres submuestras (a, b y b'). Por un lado, se intentó evaluar la variación intra gel corriendo simultáneamente dos submuestras (a y b) en la misma corrida electroforética, llamados aquí “duplicados técnicos”. Algunas duplicados presentaron perfiles similares como se muestra en la Tabla IV.3, mientras que en otras muestras (ej. S2 NC, S4 LC y S5 NC) las réplicas presentaron T-RFs adicionales con bajo nivel de fluorescencia, no más de 250 UFR (unidades relativas de fluorescencia) como se muestra en la Tabla IV.4.

Por otro lado, con el propósito de evaluar la variación inter gel se corrió la submuestra b' en un gel distinto y el perfil T-RFLP fue comparado con el obtenido para la submuestra a' previamente descripta. Como se muestra en la Tabla IV.3, la mayoría de los T-RFs fueron compartidos entre las submuestras, y en el caso de las muestras S1 CSD, S4 LC y S5 NC se reveló al menos un fragmento de restricción terminal distinto entre las réplicas. Aquí también los valores de URF correspondientes a la altura de tales T-RFs fueron bajos y variaron entre 80 a 210 unidades (Tabla IV.5), representando menos del 4% del total de la fluorescencia en dichas muestras.

Entonces si la variación se debió al sistema de detección del equipo, a la carga de las muestras o a una diferencia en la cantidad inicial de templado de ADN, sería de esperar que los T-RFs que variaron tanto por su presencia como por su ausencia correspondieron a los que presentaron picos con una señal relativamente muy baja. Teniendo en cuenta esta consideración, todos los T-RFs que presentaron valores altos de URF (> 300) estuvieron presentes en los perfiles de ambas réplicas intra e inter gel (Tablas IV.4 y IV.5).

Tabla IV.3. Reproducibilidad de los perfiles de T-RFLP generados a partir de la incubación de los amplicones *nifH* con la enzima *MspI*. Número de T-RFs, indica el número de fragmentos de restricción terminal observado en cada submuestra (a y b).

Variable	Muestra	Número de T-RFs*		Total T-RFs compartidos
		a	b	
INTRA GEL	S1 CSD	6	6	6
	S2 NC	11	12	11
	S3 NC	7	7	7
	S4 LC	10	8	8
	S5 NC	7	6	6
INTER GEL	S1 CSD	6	7	6
	S2 NC	11	11	11
	S3 NC	7	7	7
	S4 LC	10	8	8
	S5 NC	7	5	5

También se evaluó la variabilidad en términos de altura de los picos intra e inter gel. Para ello, se calculó el porcentaje de la desviación estándar (SD %) del promedio de la altura de los picos presentes en los perfiles T-RFLP de ambas réplicas. En la Tabla IV.4 se muestran los valores de SD de los T-RFs intra gel, indicando que la mayoría presentó una variación menor al 15%, excepto el fragmento más pequeño (30 pb) el cual presentó valores de variación elevados (43 al 57%) comparados con los otros T-RFs en todos los perfiles analizados. Con respecto a los valores de SD obtenidos para los T-RFs analizados en geles distintos, se observó también que los picos más pequeños presentaron la mayor variación, alcanzando valores de hasta un 43%. Para el resto de los T-RFs se observaron valores de SD menores al 14%.

Tabla IV.4. Variación intra gel de la altura de los picos en los perfiles T-RFLP generados a partir de la incubación de los amplicones *nifH* con la enzima *MspI*.

Muestra	Tamaño (pb)	Altura *		Altura (\pm)**	SD (%)
		a	b		
S1 CSD	30	393	209	301 \pm 130.1	43.2
	71	5886	5834	5860 \pm 36.8	0.6
	146	1034	1046	1040 \pm 8.5	0.8
	154	107	114	110.5 \pm 4.9	4.5
	160	115	122	118.5 \pm 4.9	4.2
	161	212	216	214 \pm 2.8	1.3
S2 NC	30	185	370	277.5 \pm 130.8	47.1
	69	1550	1532	1541 \pm 12.7	0.8
	109	116	118	117 \pm 1.4	1.2
	111	110	106	108 \pm 2.8	2.6
	115	616	626	621 \pm 7.1	1.1
	140	101	103	102 \pm 1.4	1.4
	145		243		
	147	1994	2014	2004 \pm 14.1	0.7
	156	1253	1281	1267 \pm 19.8	1.6
	160	629	650	639.5 \pm 14.8	2.3
	162	589	590	589.5 \pm 0.7	0.1
	177	165	170	167.5 \pm 3.5	2.1
S3 NC	30	187	370	278.5 \pm 129.4	46.5
	71	589	727	658 \pm 97.6	14.8
	114	179	215	197 \pm 25.5	12.9
	145	329	391	360 \pm 43.8	12.2
	147	1982	2376	2179 \pm 278.6	12.8
	156	204	249	226.5 \pm 31.8	14
	162	606	722	664 \pm 82	12.4
S4 LC	30	434	186	310 \pm 175.4	56.6
	71	1318	1066	1192 \pm 178.2	14.9
	115	334	259	317 \pm 24	7.6
	144	351	276	335.5 \pm 21.9	6.5
	147	2811	2288	2549.5 \pm 369.8	14.5
	154	344	283	313.5 \pm 43.1	13.8
	157	754	614	684 \pm 99	14.5
	159	138	105	141.5 \pm 4.9	3.5
	161	109			
	162	103			
	177	251	277	264 \pm 18.4	7
S5 NC	30	330	263	296.5 \pm 47.4	16
	70	355	367	361 \pm 8.5	2.4
	114	147	149	148 \pm 1.4	1
	115		210		
	147	3826	3971	3898.5 \pm 102.5	2.6
	160	80			
	161	481	512	496.5 \pm 21.9	4.4

* a y b corresponden a los duplicados técnicos. ** Altura promedio del pico.

Tabla IV.5. Variación inter gel de la altura de los picos en los perfiles T-RFLP generados a partir de la incubación de los amplicones *nifH* con la enzima *MspI*.

Muestra	Tamaño (pb)	Altura *		Altura (\pm)**	SD (%)
		a	b'		
S1 BPA	30	393	325	359 \pm 48.1	13.4
	69		115		
	71	5886	5700	5793 \pm 131.5	2.3
	147	747	789	768 \pm 29.7	3.9
	154	107	105	106 \pm 1.4	1.3
	160	115	160	123.5 \pm 12	9.7
	161	212	231	221.5 \pm 13.4	6.1
S2 NC	30	185	141	163 \pm 31.1	19.1
	69	1550	1081	1565.5 \pm 21.9	1.4
	109	116	154	120 \pm 5.7	4.7
	111	110	119	114.5 \pm 6.4	5.6
	115	616	620	618 \pm 2.8	0.5
	140	101	120	110.5 \pm 13.4	12.2
	147	1994	2050	2022 \pm 39.6	2
	156	1253	1245	1249 \pm 5.7	0.5
	160	629	629		
	162	589	685	637 \pm 67.9	10.7
	177	165	136	150.5 \pm 20.5	14
S3 NC	30	187	274	230.5 \pm 61.5	26.7
	71	589	823	606 \pm 24	4
	114	179	170	174.5 \pm 6.4	3.6
	145	329	336	332.5 \pm 4.9	1.5
	147	1982	3706	1844 \pm 195.2	10.6
	156	204	235	219.5 \pm 21.9	10
	162	606	935	670.5 \pm 91.2	13.6
S4 LC	30	434	231	332.5 \pm 143.5	43.2
	71	1318	1748	1433 \pm 162.6	11.3
	115	334	210		5.3
	144	351	247	349 \pm 2.8	0.8
	147	2811	2784.5	2797.75 \pm 18.7	0.7
	154	344	314	329 \pm 21.2	6.4
	157	754	500	702 \pm 73.5	10.5
	159	138			
	161	109			
S4 NC	30	330	263	296.5 \pm 47.4	16
	70	355	367	361 \pm 8.5	2.4
	114	147	149	148 \pm 1.4	1
	115		210		
	147	3826	3971	3898.5 \pm 102.5	2.6
	160	80			
	161	481	512	496.5 \pm 21.9	4.4

Por lo tanto, se encontró que la variación inherente a la técnica electroforética, intra e inter gel, es baja, y se limita principalmente a la presencia/ausencia de T-RFs de pequeño tamaño (en pb) o de baja señal. Considerando este resultado y dado que el marcador interno de tamaño utilizado en las corridas electroforéticas presenta un rango de fragmentos comprendido entre 35 y 500 pb, en los análisis que preceden se excluyeron aquellos T-RFs menores a 35 pb y se consideraron aquellos T-RFs con una altura de pico \geq a 300 UFR.

En el trabajo de Osborn *et al.* 2000 evaluaron la robustez del método T-RFLP para el estudio de la estructura y dinámica de comunidades microbianas a partir de muestras de ADN de suelos, observando que los valores de SD (%) del promedio de la altura de los picos en su mayoría fueron \leq al 11% entre T-RFs en réplicas intra e inter gel [41]. Además, revelaron que el porcentaje de desviación de los fragmentos de menor tamaño (43 y 46 pb) alcanzaba valores de SD de hasta el 18% debido a la cercanía a los límites de detección del secuenciador [40].

b. Variación introducida en la etapa de PCR

Se evaluó la variabilidad generada en la etapa de amplificación por PCR analizando dos replicas (a y b) que fueron amplificadas en forma separadas usando submuestras del mismo molde de ADN metagenómico. Esto se ensayó con cinco muestras de suelo (S1 CSD, S2 NC, S3 NC, S4 LC y S4 NC). Los amplicones fueron incubados separadamente con las endonucleasas *MspI* y *HaeIII* generando respectivamente dos perfiles T-RFLP. Los T-RFs con altos valores UFR (mayor a 250 UFR) fueron reproducibles en ambos perfiles T-RFLP de cada muestra. Por el contrario, algunos T-RFs que presentaron bajos valores de UFR ($<$ al 3% del total de UFR) no estuvieron presentes en las réplicas para ambas enzimas (Tablas Suplementarias IV.2 y IV.3).

Los perfiles T-RFLP se compararon aplicando la prueba estadística *t-Student* ($p \leq 0.05$) a los valores de fluorescencia correspondiente a la integración de todos los picos del electroferograma, y no se consideraron aquellos T-RFs con una altura menor a 250 UFR. En la Figura IV.5 se muestra los valores promedio de fluorescencia total y los desvíos correspondientes para cada muestra y réplica. Ninguno de los perfiles T-RFLPs generados a partir de la incubación de los amplicones *nifH* con *MspI* y *HaeIII* presentaron diferencias significativas entre las réplicas. Los valores de *p* para *MspI* fueron los siguientes: S1 CSD: 0.09, S2 NC: 0.26, S3 NC: 0.08, S4 LC: 0.17 y S4 NC: 0.11 y en el caso de *HaeIII* S1 CSD: 0.73, S2 NC: 0.66, S3 NC: 0.80, S4 LC: 0.56 y S4 NC: 0.99.

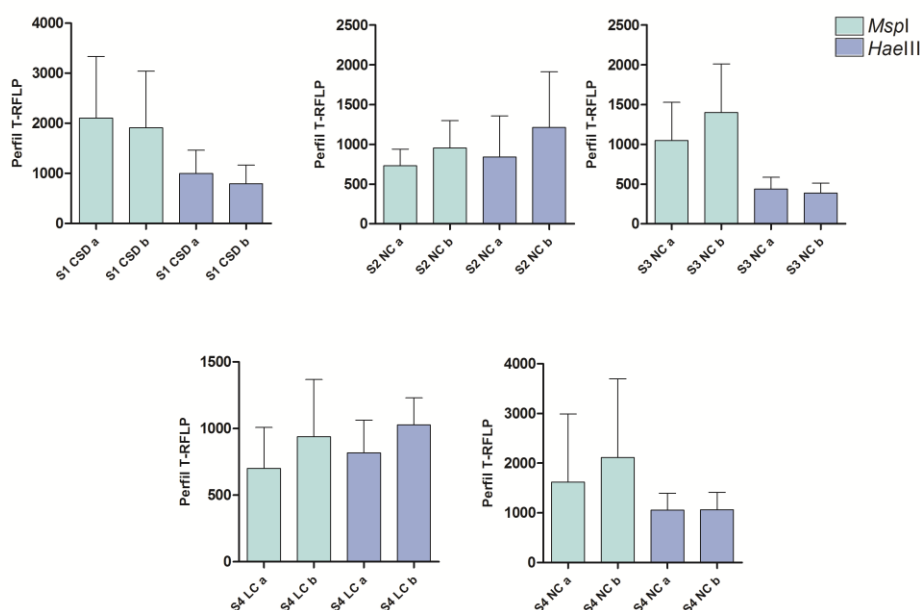


Figura IV.5. Reproducibilidad de la etapa de PCR. Se muestran en distinto color los perfiles T-RFLP de las muestras de suelos S1 CSD, S2 NC, S3 NC, S4 LC y S4 NC que se obtuvieron a partir de incubación de los amplicones *nifH* (a y b) con las endonucleasas *MspI* y *HaeIII*. En el eje Y se representa el valor de fluorescencia correspondiente a la integración de todos los picos del electroferograma.

c. Variación en el tratamiento enzimática

Los amplicones *nifH* de 2 muestras de ADN de suelo (S3 CSD y S4 NC) fueron digeridos en duplicado (a y b) con *MspI* o *HaeIII* y la mezcla analizada en la

misma corrida electroforética con el propósito de evaluar la variabilidad debido a distintas reacciones de digestión enzimática. Como se muestra en la Figura Suplementaria IV.1, en el caso de los dos tratamientos, el resultado de los perfiles T-RFLP en los duplicados fue comparable sin diferencias significativas entre ellos y los valores de p fueron los siguientes: S3 CSD 0.11 y S4 NC 0.35 para *MspI*; S3 CSD 0.32 y S4 NC 0.67 para *HaeIII*. Además, todos los T-RFs fueron reproducibles en ambos duplicados de digestión con ambas enzimas de restricción (Tabla Suplementaria IV.4 y Figura Suplementaria IV.1).

d. Variación en el método completo de T-RFLP

El ensayo de T-RFLP, desde la amplificación por PCR hasta la resolución de los T-RFs por electroforesis capilar, se realizó por duplicado a partir de 2 submuestras del mismo ADN metagenómico de 15 muestras de suelo (Tabla IV.1). Se observó que la presencia y/o ausencia de T-RFs en una réplica se correspondió con picos de baja abundancia (alturas de picos menores a 300 unidades relativas de fluorescencia) y representación menor al 1% del total del área de fluorescencia. Se calculó la similitud de los perfiles T-RFLP basada en la medida de Bray-Curtis que tiene en cuenta tanto la ausencia y presencia de OTUs como su abundancia [76]. Los valores de similitud entre los perfiles T-RFLP de la misma muestra oscilaron entre 0.92 y 0.99 para la digestión con *MspI* y entre 0.80 a 0.92 para la digestión con *HaeIII*.

En el trabajo de Li *et al.* 2007, donde evaluaron la reproducibilidad del método de T-RFLP para la estudiar la microbiota del intestino humano, revelaron que aquellos T-RFs con una representación menor al 1% del total del área de fluorescencia del perfil de la comunidad, no era significativamente distinto al ruido del fondo inherente al método [243]. Asimismo, en otros trabajos también utilizaron dicho valor de corte, para excluir de los análisis los T-RFs con un área relativa de fluorescencia <1% del total [240], [244].

En resumen, en esta sección “Evaluación de la reproducibilidad del método T-RFLP” se mostraron los resultados de los análisis de reproducibilidad donde se evaluaron varios de los parámetros que pueden influenciar significativamente la variación de los perfiles de T-RFLP. De los parámetros analizados, ninguno presentó un efecto significativo en la variación de los perfiles T-RFLP lo que demuestra una

alta reproducibilidad del método. Aun así, se detectaron factores que se pueden corregir para optimizar la aplicación del mismo. Se observaron los mayores valores de variación de corridas (inter/intra gel) en fragmentos de tamaño comprendido entre 30 y 35 pb. A partir de estos análisis exploratorios de reproducibilidad se concluyó excluir, una vez alineados los perfiles T-RFLP, aquellos T-RFs cuyo tamaño fue menor a 35 pb. También se reparó en la importancia de que el perfil T-RFLP consenso de cada muestra se genere a partir de la corrida de al menos dos replicas técnicas. Como se describió, hubo ciertos T-RFS revelados o ausentes en determinado perfil, y si solo se considerase tal perfil se podría sobreestimar o subestimar el número total de T-RFs presentes en una comunidad. Por lo tanto, en el presente trabajo cada muestra fue analizada en duplicado en la corrida electroforética y solo aquellos T-RFs reproducibles en ambas replicas técnicas fueron considerados “verdaderos” y luego alineados. Por último, una vez alineados los T-RFs y sus abundancias normalizadas, los T-RFs que presentaron una baja proporción, ya sea picos con una altura <300 URF o representando <1% del total del área de fluorescencia, fueron excluidos de los análisis posteriores considerando que podrían corresponder a picos espuriosos o al ruido de fondo del método.

Análisis comparativo de las poblaciones diazotróficas de suelo mediante T-RFLP.

a. Composición de los perfiles de T-RFLP del gen *nifH*

En esta sección se describe el análisis comparativo de los perfiles T-RFLP del gen *nifH* obtenidos a partir del ADN metagenómico de 15 muestras de suelo (Tabla IV.1). Con el propósito de mostrar la distribución relativa de los T-RFs en las distintas muestras de suelo, se realizó una representación gráfica (*heatmap*) de los valores cuantitativos de los T-RFs (Figura IV.6).

En la Figura IV.6 se observó que la digestión con las enzimas de restricción *MspI* y *HaeIII* permitió detectar 22 y 33 T-RFs, respectivamente. Estos T-RFs presentaron un tamaño comprendido entre 69 pb a 338 pb para *MspI* y 37 pb a 356 pb para *HaeIII*. La cantidad de fragmentos de restricción terminal encontrados en los diferentes sitios varió de 11 a 16 fragmentos. El número de los fragmentos de restricción terminales refleja la diversidad y heterogeneidad de la población

diazotrófica en una muestra. En general, en la mayoría de los sitios, los suelos cultivados presentaron un mayor número de T-RFs comparado con los suelos no cultivados, excepto en los sitios S3 y en S6. En particular, en el suelo no cultivado de S6 se encontró el mayor número de T-RFs, lo que podría estar indicando una mayor variedad de secuencias *nifH* en este suelo. Coincidentemente, mediante secuenciación masiva y profunda de genes *nifH* de las mismas muestras de Viale, se halló que el suelo no cultivado presentaba los mayores valores de diversidad [121].

A pesar de la técnica de T-RFLP pone de manifiesto principalmente las secuencias *nifH* de los microorganismos más abundantes, y que un T-RF no necesariamente representa un taxón particular (ya que el fragmento puede ser generado por distintos taxa que tengan un mismo sitio de restricción), la caracterización de una muestra a partir del análisis de la diversidad de fragmentos sí permite hacer inferencias indirectas sobre la diversidad de la muestra, como ha sido reportado en análisis de comunidades de suelos por Noguez *et al.* 2005 [245], por Fierer y Jackson *et al.* 2006 [196] y por Reed y Martin *et al.* 2013 [246]. Así, el patrón de fragmentos obtenidos por T-RFLPs para estas muestras de suelo indica una mayor diversidad de taxa fijadores de N₂ en los suelos cultivados de los sitios S1, S2, S4 y S5, en S3 y S6 por el contrario, la comunidad diazotrófica del suelo no perturbado fue la más diversa.

Se observaron rasgos que merecen remarcarse en cuanto a la distribución de los T-RFs asociados con la región geográfica de los sitios, o con el uso (cultivado/no cultivado) del suelo y/o manejos agrícolas (BPA, PPA y LC). Algunos T-RFs fueron detectados en la mayoría de los suelos analizados. Los T-RFs de 161 pb_ *MspI* y 154 bp_ *HaeIII* estuvieron presentes en 14 y 13 de las 15 muestras totales, respectivamente. La proporción del área de fluorescencia varió del 1 al 12% para el T-RF 161 pb_ *MspI* y del 9 al 34% para 154 pb_ *HaeIII*. Estos T-RFs generalistas podrían describir la presencia de taxones ampliamente distribuidos en el hábitat suelo. Por otra parte, la fluorescencia del T-RF 154 pb_ *HaeIII* fue en general mayor en los suelos cultivados (CSD), con respecto de los suelos no cultivados (NC).

Otros T-RFs se asociaron particularmente con algún tipo de uso del suelo, como los T-RFs de 146 pb_ *MspI* y 154 pb_ *MspI* revelados en suelos cultivados. En particular, el T-RF 146 pb_ *MspI* exhibió una elevada proporción de fluorescencia (del 7% al 43%) en los perfiles T-RFLP. Con respecto a los distintos tratamientos bajo siembra directa, no se encontraron T-RFs específicos de los tratamientos BPA o PPA. Contrariamente, para el suelo bajo el manejo de labranza convencional (LC) se revelaron varios T-RFs propios, tales como los T-RFs 144 pb_ *MspI*, 157 pb_ *MspI*, 159 pb_ *MspI*, 151 pb_ *HaeIII* y 153 pb_ *HaeIII* y en particular, este último presentó una elevada proporción del área de fluorescencia (18%).

Por otro lado, se encontraron algunos T-RFs específicos de sitio. Los T-RFs 37 pb_ *HaeIII*, 61 pb_ *HaeIII*, 65 pb_ *HaeIII* y 71 pb_ *HaeIII* se encontraron únicamente en el sitio S6 (Viale), independientemente del uso o práctica agrícola. El T-RF de 65 pb_ *HaeIII* presentó una elevada proporción relativa del área de fluorescencia (10% al 23%) mientras que los otros representaron <4% (Figura V. 8).

Concluyendo, se observaron T-RFs con una distribución diferencial asociada con el sitio, uso del suelo, o la práctica agrícola, y que además presentaron una proporción considerable dentro de la comunidad fijadora de N₂ libre, los cuales podrían ser útiles para el desarrollo de marcadores o indicadores biológicos simples y eficientes de calidad de suelo.

En el trabajo de Fűrnkranz *et al.* 2008 donde analizaron los perfiles T-RFLP de las bacterias de la filósfera de varias plantas superiores, encontraron que muchos T-RFs se compartían entre los distintos perfiles, pero que otros eran característicos de cada planta huésped [238]. Por otro lado, Tan *et al.* 2003 mediante T-RFLP estudió el efecto de la fertilización nitrogenada en la estructura de la comunidad diazotrófica asociada a raíces del cultivo de arroz, revelando que la aplicación de fertilizantes provoca un efecto marcado en el *pool* de genes *nifH* asociado a las raíces. Comparado con el arroz sin fertilizar 8 a 10 T-RFs no se detectan en los cultivos fertilizados, y solo se encontró un T-RF que además presentó un tamaño diferente [218].

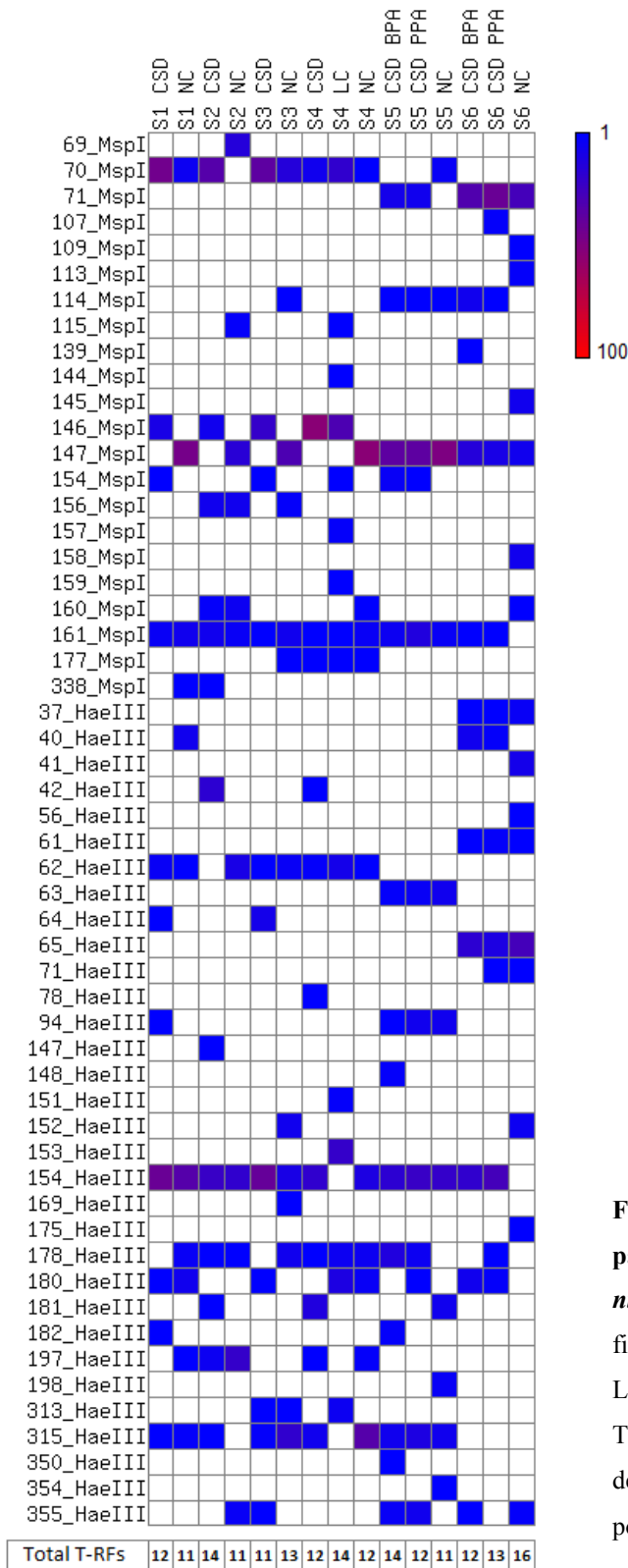


Figura IV.6. Heatmap de los tRFs revelados a partir de la digestión con *MspI* y *HaeIII* del *pool nifH* en varios suelos. En el margen izquierdo de la figura se muestran los tamaños en pb de los T-RFs. Las celdas en color blanco significan ausencia del T-RF en dicha muestra. La proporción relativa (%) del área de fluorescencia de cada tRF se representó por la gama de colores de azul (min) a rojo (máx.).

Dichos estudios, al igual que los datos revelados aquí, ponen de manifiesto que la técnica de T-RFLP permite detectar cambios en los perfiles poblaciones asociados a distintas variables (uso del suelo, aplicación de fertilizantes, especies de plantas, entre otras), y esos cambios se reflejan en la presencia de determinado T-RF y/o en su abundancia.

b. Relación entre las comunidades diazotróficas de suelo analizadas mediante los perfiles T-RFLP del gen *nifH*

Con el propósito de identificar asociaciones entre las diferentes comunidades diazotróficas de los suelos analizados, se realizó un análisis de conglomerado mediante el método de agrupamiento UPGMA sobre matrices de similitud basadas en la de distancia de Bray-Curtis. Para la construcción de la matrices se utilizaron los datos de abundancia de los T-RFs consensos cuya área de fluorescencia fue \geq al 1% del total para (Tabla Suplementaria IV.5).

Como se observa en la Figura IV.7 se formaron tres grandes grupos a partir del análisis de clúster, referido como Grupo 1, 2 y 3, respectivamente. En el grupo 1 se agruparon la mayoría de los suelos cultivados (S1 a S4), y dentro de este grupo se separó el suelo bajo labranza convencional (LC) de los suelos cultivados bajo siembra directa (CSD). En el Grupo 2 se encontraron los suelos no cultivados (NC) de los sitios S1 a S5 y en un subgrupo separado los suelos cultivados de S5. Los suelos del sitio 6 conformaron el Grupo 3, en el cual también se separaron con un alto valor de *bootstrapping* los ambientes cultivados de los no cultivados. Concluyendo, el análisis de los perfiles T-FRLP del gen *nifH* nos muestra que en los suelos S1 a S4 la composición de la comunidad diazotrófica parece estar afectada mayormente por el uso del suelo. En contraste, en los suelos de Pergamino (S5) y Viale (S6) se observa que la localización geográfica fue el principal conductor de agrupamiento de las comunidades y en segunda instancia el uso del mismo.

En el trabajo de Poly *et al.* 2001 donde se utilizó la técnica de RFLP para estudiar la diversidad de genes *nifH* de bacterias diazotróficas de vida libre en suelos cultivados y suelos bajo pastura permanente, se encontró que los perfiles RFLP de las comunidades diazotróficas de los suelos bajo distinto uso, eran notablemente

diferentes entre sí [222]. También en suelos amazónicos la composición y diversidad de la comunidad bacteriana total analizadas por el método de T-RFLP del gen ARNr 16S es afectada por los diferentes sistemas de uso del suelo [247].

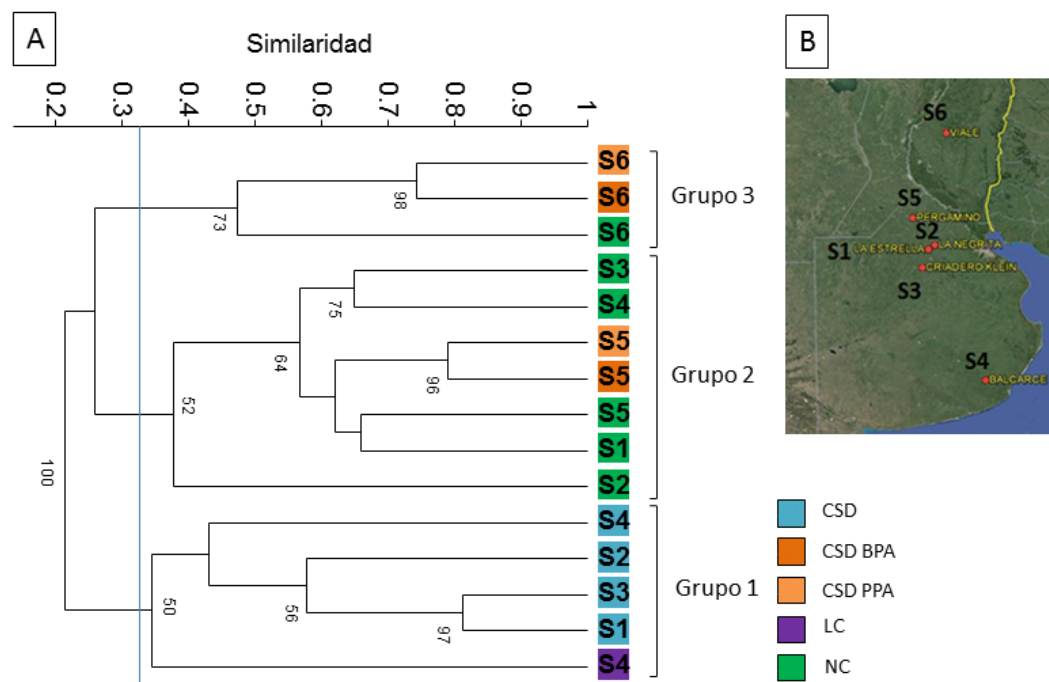


Figura IV.7. Análisis de agrupamiento basado el coeficiente de similitud Bray-Curtis utilizando las abundancias normalizadas de los T-RFS *nifH*. (A) En el dendrograma UPGMA de perfiles T-RFLP del gen *nifH*, los colores hacen referencia a los distintos usos (NC y CSD) y manejos agrícolas (BPA, PPA y LC) del suelo. La escala de la barra representa los valores de porcentaje de similitud. Solo se muestran valores de *bootstrapping* \geq a 50. (B) Mapa con la localización de los sitios de la Región Pampeana Argentina donde se realizó el muestreo de los suelos analizados. La línea vertical celeste en el dendrograma, indica el valor de corte de similitud, a partir del cual se dividen los tres grupos principales.

Por otro lado, en el trabajo de Girvan *et al.* 2003 determinaron mediante el análisis de perfiles T-RFLP del gen ARNr 16S en suelos cultivados, que el tipo de suelo fue el factor determinante de la composición de la comunidad bacteriana total en suelo [248]. Ramond *et al.* 2013 emplearon la técnica de T-RFLP para estudiar la estructura de la comunidad bacteriana rizosférica asociada al cultivo de sorgo en varios sitios, revelando que la ubicación geográfica fue una de las variables principales además, de ciertas variables químicas del suelo, del agrupamiento de las

comunidades microbianas asociadas a la rizosfera de sorgo [249]. Por lo tanto, al igual que para los resultados aquí expuestos, los manejos agronómicos, el tipo de suelo y la región geográfica pueden ser factores determinantes en la agrupación de perfiles T-RFLP de la comunidad bacteriana total, o de grupos específicos (diazótrofos).

A fin de identificar los principales T-RFs que se encuentran dirigiendo el ordenamiento de los suelos analizados, se llevó a cabo un análisis de componentes principal (PCA). El resultado de este análisis se muestra en la Figura IV.8. La componente principal 1 representa el 46% de la variabilidad y separó los suelos cultivados de S1 a S4 de los no cultivados. Se observó que esta separación es resultado principalmente de 5 T-RFS, tres asociados a los suelos CSD, 70 pb_*MspI*, 146 pb_*MspI*, y en menor grado 154 pb_*HaeIII*; y dos asociados a los suelos NC, 147 pb_*MspI* y 315 pb_*HaeIII* (Figura IV. 6 y Tabla Suplementaria IV.5). Asimismo, la ocurrencia de los T-RFs de 71 pb_*MspI* y 65 pb_*HaeIII* separa los suelos de Viale (S6). En particular, el T-RF de 65 pb_*HaeIII* se reveló solo en Viale, y presentó una proporción mayor en el suelo NC (23%) en comparación con los suelos cultivados (Figura IV.6).

Con el objeto de analizar el efecto de distintos parámetros químicos del suelo (Tabla Suplementaria IV.6) en el ordenamiento de las comunidades diazotróficas se llevó a cabo un análisis de escalamiento multidimensional no-métrico (NMDS). En este tipo de análisis un valor de *stress* ideal no supera el valor de 0,1 y es confiable hasta un valor de 0,3 [79]. Como se muestra en la Figura IV. 9. el valor de *stress* obtenido usando dos dimensiones fue suficiente para una representación adecuada de las relaciones entre los datos originales. En este análisis exploratorio se observó que el gradiente humedad ejerció un efecto en la separación de los suelos tanto a nivel horizontal como vertical. En particular, la agrupación de los suelos del sitio Viale (S6) se relaciona con bajos valores de humedad y mayores valores de pH. Con respecto a la agrupación de los suelos cultivados en siembra directa de los sitios S1 a S4, esta podría asociarse principalmente con niveles bajos de N total, mientras que los suelos no cultivados S1 a S4 presentan mayores valores de humedad y N total.

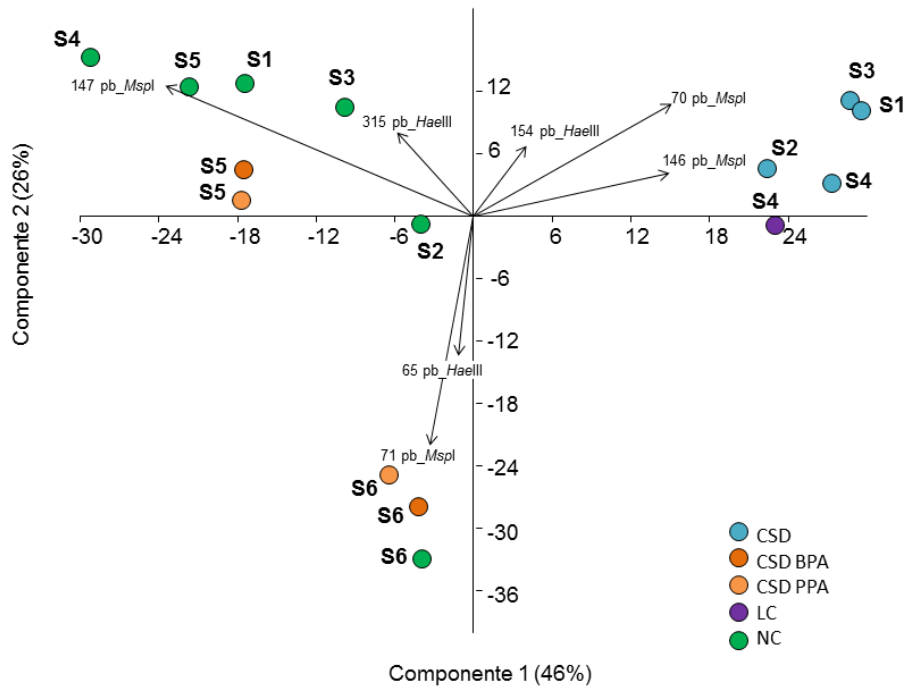


Figura IV. 8. Análisis de componentes principales (PCA) utilizando los tRFs *nifH* que aportaron al menos el 1% del área de fluorescencia total en cada muestra. Solo se muestra el tamaño (pb) de los 7 T-RFs cuya magnitud y dirección fueron relevantes en este análisis. Los colores hacen referencia a los distintos usos (NC, no cultivado) y manejos agrícolas (siembra directa, CSD; con buenas prácticas, CSD BPA, con pobres prácticas CSD PPA y bajo labranza convencional, LC) del suelo.

Restricción in silico de secuencias del gen *nifH*: Asignación taxonómica tentativa de los T-RFs

El método de T-RFLP permite inferir las asignaciones filogenéticas a partir de los tamaños de los T-RFs a través de recursos bioinformáticos, que predicen los tamaños de T-RFs para bacterias conocidas [250]. Dado que en nuestro laboratorio se disponía de los datos de pirosecuenciación del gen *nifH* para los sitios S5 (Pergamino) y S6 (Viale), se recuperaron las secuencias de aquellos OTUs que fueron identificados como abundantes [121]. El resultado de someter estas secuencias a la digestión *in silico* con las endonucleasas *MspI* y *HaeIII* fue comparado con los patrones experimentales de T-RFLP, asumiendo que los T-RFs resultaron de aquellos genomas más abundantes en el ADN metagenómico analizado.

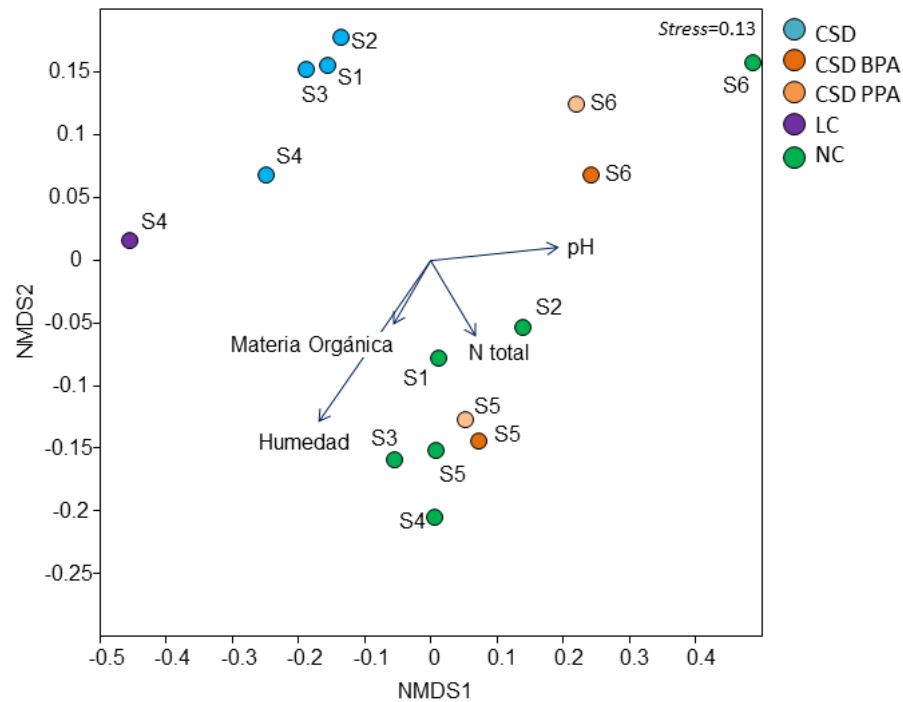


Figura IV. 9. NMDS biplot representando los diferentes suelos en base a la estructura de las comunidades diazotróficas determinada por el análisis de T-RFLP y variables química del suelo. Cada vector apunta a la dirección de aumento para una dada variable y su longitud indica la fuerza de la entre la variable y los valores de ordenación. Los colores hacen referencia a los distintos usos (NC, no cultivado) y manejos agrícolas (siembra directa, CSD; con buenas prácticas, CSD BPA, con pobres prácticas CSD PPA y bajo labranza convencional, LC) del suelo.

A través de este análisis se esperaba identificar coincidencias de tamaño entre las pirosecuencias cortadas *in silico* y los T-RFs empíricos y en consecuencia, realizar una asignación taxonómica de los picos correspondientes a los fragmentos terminales detectados en los electroferogramas. Se adoptó un intervalo de variación de tamaño de ± 3 pb para discriminar identidades entre los T-RFs experimentales y teóricos [251].

El resultado de la asignación taxonómica de los T-RFs en los suelos de Pergamino (S5) y Viale (S6) se muestra en la Figura IV.10. La digestión *in silico* de las secuencias de Pergamino (S5) con las enzimas *MspI* y *HaeIII* reveló un total de seis fragmentos coincidentes en tamaño con T-RFs observados en estos suelos, los cuales presentaron una asignación taxonómica tentativa dentro de los *subclusters nifH* 1A y 1K (Figura V. 10 A). Los *subclusters* 1A y 1K definidos por Zehr *et al.* 2003, se componen de secuencias *nif* de las nitrogenasas convencionales, y los filotipos

representativos de dichos *subclusters* son *Geobacter* y *Anaeromyxobacter* para 1A, y *Rhizobiales*, *Methylosinus/Methylocystis*, *Burkholderia*, para 1K [52]. En el *subcluster* 1A se encontraron los siguientes tres T-RFs: - 147 pb_*MspI*, uno de los T-RFs más abundantes para todos los suelos de Pergamino y mayormente representado en el suelo no cultivado, - 181 pb_*HaeIII* presente solo en el suelo no cultivado y - 315 pb_*HaeIII* hallado en todos los tratamientos, los cuales podrían corresponder a secuencias *nifH* de *Geobacter* sp. *Geobacter pickeringii* y *Geobacter uraniireducens*, respectivamente. Los restantes tres T-RFs fueron asignados al *subcluster* 1K. El T-RF 154 pb_*MspI* revelado solo en los suelos cultivados de S5 presentó semejanza con secuencias de *Bradyrhizobium* y *Methylosinus*. Asimismo, el T-RF 148 pb_*HaeIII*, encontrado solo en BPA, podrían corresponder a secuencias *nifH* de *Bradyrhizobium japonicum*, mientras que el T-RF 154 pb_*HaeIII*, abundante en todos los suelos de Pergamino, podría relacionarse con *Bradyrhizobium elkanii*.

Con respecto a los suelos de Viale (Figura V.10 B), se encontraron seis fragmentos resultantes de la digestión virtual coincidente en tamaño con T-RFs empíricos, los cuales presentaron una asignación taxonómica en los *subclusters* 3B, 1K, 1C y 1J. Los cuatro T-RFs relacionados con los *subclusters* 3B, 1J y 1C se encontraron exclusivamente en los suelos S6. Los T-RFs 107 pb_*MspI* y 113 pb_*MspI*, se relacionaron con secuencias del género *Desulfovibrio* pertenecientes al *subcluster* 3B (secuencias *nifH* de bacterias anaeróbicas estrictas [52]), y se hallaron en los suelos PPA y NC, respectivamente. El T-RF 37 pb_*HaeIII*, más abundante en el ambiente no cultivado se relacionó con secuencias de *Clostridium aerotolerans* del *subcluster* 1C, mientras que el T-RF 61pb_*HaeIII*, más abundante en las malas prácticas, se asoció con el *subcluster* 1J, y en particular con el género *Gluconacetobacter*.

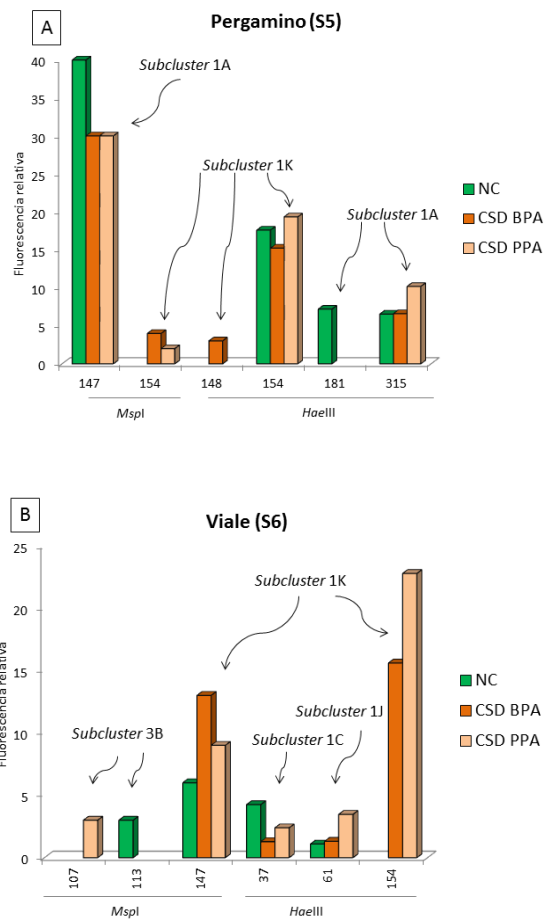


Figura IV.10. Asignación taxonómica tentativa de los T-RFs experimentales en base a la digestión *in silico* de pirosecuencias de *nifH*. (A) Suelos de Pergamino (B) Suelos de Viale. En el eje X de ambos gráficos se encuentran los tamaños en bp de los T-RFs empíricos revelados a partir de la incubación de los amplicones *nifH* con las enzimas *MspI* y *HaeIII*, los cuales presentaron una clasificación taxonómica en *subcluster nifH* [181].

Por otra parte, los T-RFs 147 pb_*MspI*, más abundante en los suelos cultivados, y 154 pb_*HaeIII* presente exclusivamente en los suelos cultivados, se asociaron con secuencias del *subcluster* 1K relacionadas con *Sinorhizobium* sp. y *Bradyrhizobium japonicum*, respectivamente. Estos últimos T-RFs también se encontraron en muestras de otros suelos, incluyendo Pergamino. En el caso del T-RF 154 pb_*HaeIII* la asignación taxonómica 1K con secuencias *nifH* del género *Bradyrhizobium* coincide con la asignación encontrada para el T-RF *HaeIII* del mismo tamaño en Pergamino. Con respecto al T-RF de 147 pb_*MspI*, presente en los tres tratamientos de ambos sitios, la asignación taxonómica fue distinta, por un lado

para Pergamino se asoció con el *subcluster* 1A (*Desulfuromonadales*) y por otra lado en Viale con 1K (*Rhizobiales*).

A partir del análisis exploratorio de asignación taxonómica de T-RFs empíricos, pudimos observar que en Pergamino y Viale los T-RFs asociados al orden *Rhizobiales*, se encontraron particularmente representados en los suelos cultivados. En los ambientes no cultivados las secuencias de *Geobacter* predominan en Pergamino, mientras que en Viale se observa una gran variedad de secuencias. Otros T-RFs relacionados con los órdenes *Desulfovibrionales*, *Clostridiales* y *Rhodospirillales*, podrían constituir grupos taxonómicos asociados a determinada condición de uso y/o manejo de los suelos de Viale.

Con respecto a la diferencia en la asignación taxonómica observada para el T-RF 147 pb_ *MspI*, en Pergamino y Viale, este resultado pone de manifiesto una de las limitaciones del T-FRLP, ya que microorganismos filogenéticamente no relacionados podrían presentar el mismo sitio de reconocimiento y generar T-RFs de igual tamaño.

En el trabajo de Kent *et al.* 2003 sugieren que el uso de múltiples digestiones aumenta la especificidad en la inferencia filogenética de los T-RFs [250]. Otros estudios en donde se ha utilizado el método de T-RFLP para la identificación de taxones en comunidades mixtas, generalmente lo hacen en conjunto con el análisis de secuencias de una biblioteca de clones [210], [221], [252]–[254] de cada sitio o mediante la determinación de T-RFs de cepas cultivadas [255]. Si bien dentro de los métodos independientes de cultivo, la construcción de bibliotecas de clones, ofrece el mayor grado de resolución filogenética, puede resultar una metodología engorrosa cuando se analizan varias muestras en simultáneo, como en el caso de los estudios ecológicos sobre variabilidad temporal o espacial de las comunidades microbianas [250].

Conclusiones

El análisis T-RFLP sobre productos de PCR *nifH* de muestras de suelo resultó en una herramienta útil y efectiva para evaluar la presencia y diversidad de microorganismos fijadores de nitrógeno en suelo provenientes de seis sitios de la Región Pampeana Argentina. Además, este enfoque brinda información acerca de la

estructura microbiana asociada al *pool nifH* y de la potencialidad en la fijación de nitrógeno.

El T-RFLP fue eficiente para revelar cambios de la comunidad diazotrófica por sitio y/o uso agrícola. Para los suelos de los establecimientos productivos La Estrella (S1), La Negrita (S2) y Criadero Klein (S3) y del campo experimental INTA Balcarce, la composición de la comunidad diazotrófica parece estar afectada mayormente por el uso del suelo. En contraste, en los suelos de Pergamino (S5) y Viale (S6) se observa que la ubicación geográfica es el principal conductor de agrupamiento de las comunidades y en segunda instancia el uso del mismo.

Se observó que el impacto provocado en las comunidades diazotróficas por el tratamiento de siembra directa con el manejo BPA o PPA no tiene un efecto diferencial sobre la diversidad de la población fijadora de nitrógeno que pueda ser detectado por este análisis, indicando que el análisis T-RFLP no revela diferencias de manejo agrícola al menos para algunos de suelos analizados. Sin embargo, en el caso del suelo S4 cultivado bajo labranza convencional, el método reveló ciertos fragmentos de restricción terminal específicos de dicho tratamiento, los cuales probablemente colaboraron para que en el análisis de agrupamiento este suelo se separase del resto de los suelos cultivados bajo siembra directa.

También se encontró una distribución diferencial de algunos T-RFs entre usos, manejos y sitios de los suelos, los cuales además presentan una proporción considerable dentro de la comunidad fijadora de N₂ libre. Dichos T-RFs o huellas digitales podrían ser útiles para el desarrollo de marcadores o indicadores biológicos simples y eficientes de calidad de suelo.

Se ha logrado poner en marcha un procedimiento reproducible y útil para generar perfiles de la comunidad diazotrófica en los suelos. Estos perfiles constituyen parámetros biológicos de los suelos que caracterizan a los mismos y por lo tanto proveen herramientas para encarar comparaciones e identificar asociaciones a parámetros físico-químicos de los suelos. El grado de discriminación de diversas secuencias *nifH* podría ser mejorado incorporando al análisis otras enzimas de restricción que aumentan las chances de detectar polimorfismos a lo largo de la región objeto de análisis.

La resolución de la asignación taxonómica de los T-RFs es un tanto limitada, ya que taxones no relacionados entre sí pero, con sitios de reconocimientos que generen fragmentos T-RFs idénticos serán asociados a un mismo filotipo.

Con el reconocimiento de sus limitaciones, el método de T-RFLP es robusto, reproducible y de alto rendimiento que aplicado sistemáticamente a un conjunto de suelos, sitios y tratamientos experimentales puede resultar útil para identificar cambios generales en la composición y diversidad de la comunidad fijadora de N₂ del suelo.

Capítulo V

Caracterización de la comunidad diazotrófica activa en suelo mediante el análisis de ADNc-*nifH*

Introducción

En varias publicaciones se informa sobre investigaciones con enfoque de secuenciación masiva del gen *nifH* en donde describen la ocurrencia de una amplia diversidad de microorganismos del suelo con capacidad potencial para fijar nitrógeno en vida libre, y en las cuales además, el análisis de las secuencias sugiere que una importante proporción de los mismos se corresponden con microorganismos aún sin identificación taxonómica o aún no cultivados *in vitro* [52], [83], [87], [121], [256], [257]. Lo que aparece como interesante en estos estudios es que la comunidad de los microorganismos fijadores es muy abundante y dominante en los suelos [121], [181], [258], [259].

Si bien esta metodología independiente del cultivo celular ha sido muy útil para ahondar el conocimiento sobre la estructura poblacional de diazótrofos en ambientes naturales, la presencia de secuencias *nifH* en el ADN de suelo no implica necesariamente que las mismas correspondan a bacterias activas en la fijación de N₂. La fijación biológica de nitrógeno (FBN) a través de la expresión del sistema de la nitrogenasa es un proceso fuertemente regulado, tanto a nivel transcripcional como hasta niveles de modificación post traduccional de proteínas [260]. La complejidad de la regulación responde en parte dado por la elevada demanda energética durante la fijación de N₂ (16 moléculas de ATP por cada molécula de N₂ fijada) y por la sensibilidad de la nitrogenasa al O₂ [260]–[264]. Ha sido sugerido que la fijación de N₂ en suelos estaría influenciada por varios factores ambientales tales como, la humedad del suelo, pH, tensión de O₂, niveles de carbono orgánico, presencia de compuestos nitrogenados y la disponibilidad de los oligoelementos Mo, Fe y V [258], [265]–[268].

La estimación cuantitativa de contribución o significado ecológico de los diazótrofos de vida libre en suelo es técnicamente difícil, sin recurrir a procedimientos complejos basados en la captación del isótopo estable ¹⁵N₂. En este contexto, la bibliografía muestra variaciones de tasas de fijación con valores comprendidos en un intervalo amplio [81], [269]. Por otro lado, la aplicación del método de reducción de acetileno a la determinación cuantitativa de actividad de nitrogenasa en suelo tiene

ciertas limitaciones [270], sin embargo su uso sigue siendo ampliamente adoptado en estudios ecológicos de suelo [271], [272][273]. Asimismo, también se ha intentado acercar una respuesta basado en el análisis de ácidos nucleicos. En este sentido, Buckley *et al.* 2007 aplicaron un método basado en el uso de sondas de ADN marcadas con el isótopo estable $^{15}\text{N}_2$ ($^{15}\text{N}_2$ -DNA-SIP) para identificar los diazotrofos de vida libre del suelo activos en la fijación de nitrógeno *in situ* [259], [274].

Recientemente, se han publicado estudios que muestran la potencialidades de la metatranscriptómica para evaluar la expresión génica de las comunidades microbianas en ambientes acuáticos [275], en la rizosfera de varios cultivos [276], y en el líquido de drenaje de minas ácidas [277]. En la revisión de Carvalhais *et al.* 2012 se describen las aplicaciones de la metatranscriptómica al ambiente suelo y algunos de los principales desafíos metodológicos relacionados principalmente con la extracción del ARN y la estabilidad del mismo [278].

La identificación de las secuencias *nifH* obtenidas a partir del ARN, nos permite aproximarnos a la definición del perfil de las poblaciones activas, en la expresión de la capacidad de fijación en diferentes ambientes. En el trabajo de Burbano *et al.* 2011 estudiaron la estructura de la comunidad endófito activa en raíces de caña de azúcar a través del análisis secuencias de ADNc-*nifH* y revelaron que la mayor parte de los transcritos presentaban gran homología de secuencias con *nifH* de *Rhizobium rosettiformans* [279]. También, mediante el análisis de secuencias ADNc-*nifH*, en el mar Báltico se encontró gran diversidad de genes *nifH* relacionados con *Alfa-*, *Beta-* y *Gammaproteobacteria* y con bacterias anaeróbicas estrictas [237]. En los suelos de deshielo de valles de la Antártida, también se estudió la estructura de la comunidad diazotrófica activa a través del análisis de ADNc-*nifH*, y los autores sugirieron que los microorganismos activos presentaban homología de secuencias con genes *nifH* de *Cyanobacteria*, *Alfa-*, *Beta-* y *Gammaproteobacteria* [87]. Con respecto a los niveles de expresión de genes *nifH* en suelo, algunas publicaciones muestran niveles relativamente bajos de transcritos del gen *nifH* en suelos agrícolas y forestales [110], [133] y al igual que la diversidad de la comunidad activa, los niveles son afectados por la rotación de cultivos [110], [256].

En el estudio que fuera realizado en nuestro laboratorio en el cual se examinó la estructura de diazótrofos en suelos de la Región Pampeana Argentina mediante el análisis masivo de amplicones *nifH*, se encontró una amplia diversidad bacterias diazotróficas potencialmente fijadores de N₂ y se identificaron también asociaciones entre los perfiles poblacionales y algunas de las características edáficas y/o usos de los suelos [121]. Una pregunta que ha surgido de este estudio es cuales de esos grupos potenciales, contribuye efectivamente al proceso de fijación de N₂ en estos suelos o expresado de otra forma: cuáles son fijadores activos.

En el presente capítulo se encaró la descripción del resultado de la secuenciación masiva de transcritos del gen *nifH* como una estrategia para permitir identificar a los microorganismos diazotróficos probablemente activos en la fijación de N₂ en los suelos de la zona de Pergamino. La región de Pergamino es una de la más importante en producción de cultivos tales como trigo, maíz y soja. Además, en este capítulo se describen los resultados de la cuantificación mediante PCR cuantitativa (qPCR) de la abundancia del gen *nifH* de ciertos grupos taxonómicos específicos que en el relevamiento de transcritos *nifH* aparecieron como predominantes en algunas condiciones de tratamiento agronómico.

Objetivo general

Definir la estructura de la población diazotrófica activa en suelos bajo diferentes manejos agronómicos de la región de Pergamino mediante el análisis de ADNc-*nifH*.

Objetivos específicos

- Evaluar el efecto de diferentes manejos agrícolas y de la profundidad del suelo en la diversidad y composición y de las poblaciones diazotróficas activas mediante la secuenciación masiva de los amplicones ADNc-*nifH*
- Comparar las poblaciones diazotróficas reveladas a partir de la pirosecuenciación de ADNc-*nifH* y ADN-*nifH* extraídos directamente de suelo
- Cuantificar específicamente filotipos de interés en muestras de suelo mediante qPCR del gen *nifH*.

Resultados y discusión

- Algunas consideraciones experimentales

El análisis de los amplicones ADNc-*nifH* se realizó en muestras de suelos de Pergamino bajo siembra directa con dos tipos de manejo agrícola. Se tomaron muestras del horizonte A del suelo, y a su vez sujetas a los manejos: -BPA (**Buenas Prácticas Agrícolas**), rotación intensiva de cultivos, reposición de nutrientes y bajo nivel de uso de agroquímicos y -PPA (**Pobres Prácticas Agrícolas**), monocultivo de soja, baja reposición de nutrientes y elevado uso de agroquímicos. Además, se incluyó en el estudio un suelo testigo de referencia con pasturas naturales, sin registro de actividad agronómica durante al menos los 30 años previos, abreviado NC (**No Cultivado**). Se tomaron muestras en dos profundidades de suelo: 0 a 10 cm y 10 a 20 cm; en el presente capítulo se hará referencia a estas dos profundidades de muestreo como superficial y sub-superficial, respectivamente.

En la Figura V.1 se muestra el flujo de etapas desde la extracción de ARN de suelo hasta la obtención de una base de datos de secuencias de transcriptos *nifH* de alta calidad, denominada aquí base de datos ADNc-*nifH*.

Obtención de ARN de suelo

En general, la manipulación *in vitro* de ARN es intrínsecamente delicada, que sumado a las características particulares y propias de las muestras de suelo, resultó en la necesidad de realizar numerosos ensayos preliminares y de puesta a punto de la técnica de extracción de ARN confiable y reproducible para su adopción como la técnica de uso rutinario en estos estudios.

El objetivo principal fue la obtención de una preparación de ARN que resultara en un sustrato apropiado para las reacciones bioquímicas de síntesis de ADNc y amplificación por PCR. Esto se logró mediante la utilización de un *kit* comercial, con la modificación de algunos pasos durante la extracción, especificados en la sección “Extracción de ARN de muestras de suelo y RT-PCR *nifH*” del Capítulo II Materiales y Métodos. El ARN extraído fue cuantificado mediante espectroscopia UV a 260 nm, mientras que la pureza fue evaluada mediante la relación entre las

absorbancias 260 nm y 280 nm (Tabla V.1). Finalmente, se evaluó la integridad del ARN en una electroforesis en geles de agarosa.

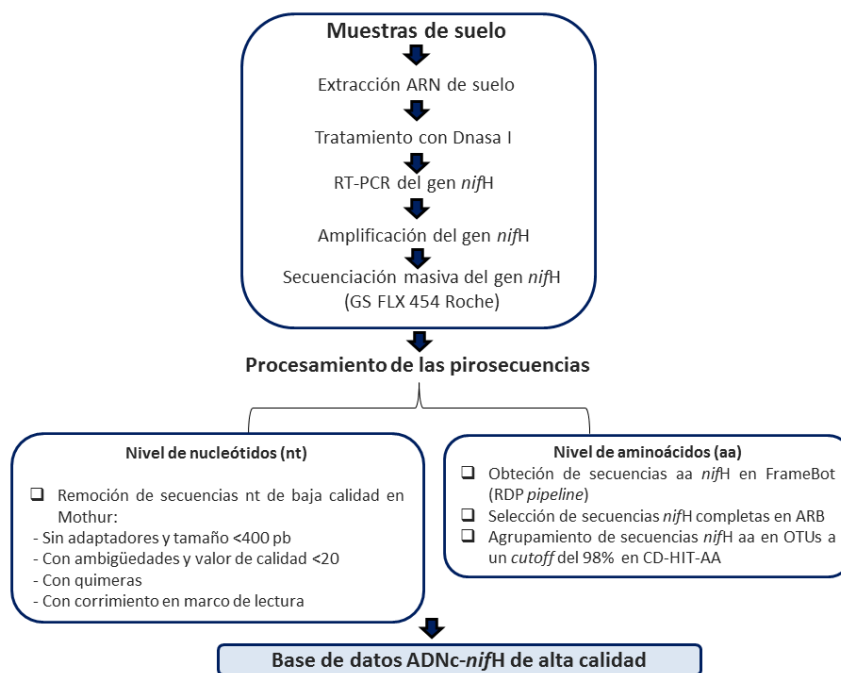


Figura V.1. Esquema del flujo de trabajo implementado para la obtención de la base de datos de secuencias ADNc-*nifH* a partir de ARN de suelo. Se muestran las principales etapas experimentales y del análisis bioinformático para obtener una base de datos ADNc-*nifH* de alta calidad de secuencias.

Los valores del rendimiento de la extracción de ARN fue de entre de 1.2 y 1.75 μg ARN g^{-1} de suelo y la relación 260/280 de la preparación final fue mayor a 1.5 en todos los suelos examinados. La bibliografía muestra variaciones importantes en los rendimientos de ARN extraído de suelo, desde menos de 1 μg hasta 10 μg de ARN g^{-1} de suelo [280]–[286]. Este amplio intervalo de rendimientos en varios protocolos experimentales podría explicarse entre otras causas a la cantidad de microorganismos vivos en la muestra de suelo, la contaminación con ácidos húmicos, o a la pérdida del ARN durante la purificación. Con respecto a los índices de pureza, Sessitsch *et al.* 2002 evaluaron diferentes métodos de extracción de ARN hallando que los valores de $A_{260/280}$ de aquellos métodos que resultaron en ARN de calidad satisfactoria para reacciones bioquímicas posteriores fueron entre 1.5 y 2.2 [280].

Estos valores son coincidentemente con el rango de valores revelados en el presente trabajo.

Tabla V.1 Rendimiento y pureza del ARN tratado con Dnasa I.

Abreviatura	Tratamiento*	Profundidad	Rendimiento ($\mu\text{g ARN g}^{-1}$ suelo)	$A_{260/280}$
NC 10	NC	0-10 cm	1.55	1.83
BPA 10	BPA	0-10 cm	1.30	1.79
PPA 10	PPA	0-10 cm	1.75	1.75
NC 20	NC	10-20 cm	1.25	1.86
BPA 20	BPA	10-20 cm	1.20	1.52
PPA 20	PPA	10-20 cm	1.23	1.60

* BPA, PPA corresponde a los distintos manejos agrícolas, buenas prácticas agrícolas y pobres prácticas agrícolas, respectivamente. NC corresponde al suelo control No Cultivado.

En la Figura V.2 se muestra la imagen de una electroforesis en gel de agarosa con la corrida de alícuotas de preparaciones de ARN. Se observan las bandas correspondientes a los ARNs ribosomales como bandas discretas propias de ARN sin degradación, y además en la parte superior se observan bandas correspondiente al ADN residual. Este ADN residual fue eliminado en una etapa adicional en el tratamiento enzimático con la enzima DnasaI.

El ARN extraído de las seis muestras de suelo de Pergamino presentó una concentración, pureza e integridad aceptable para ser utilizado como templado en la RT-PCR del gen *nifH*. Como se describe en la sección “Amplificación de *nifH* para la pirosecuenciación de la librería *nifH*” del Capítulo II Materiales y Métodos se llevaron a cabo controles para confirmar la eliminación del ADN residual y evitar así amplificaciones inespecíficas a partir de un templado que no correspondiera al ARN.

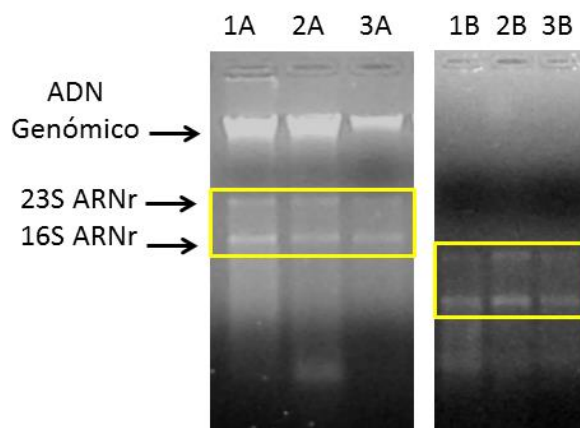


Figura V.2. Análisis electroforético de ARN extraído de los primeros 10 cm de suelo de Pergamino. El ARN se extrajo a partir de 2 gr de suelo utilizando el kit comercial *RNA PowerSoil® Total RNA Isolation* (MOBIO) y se trató con DnasaI para eliminar ADN genómico residual. **(A)** Sin tratamiento con DnasaI, **(B)** Con tratamiento de Dnasa I. **Calles 1-3:** ARN de suelo No Cultivado (NC), ARN de suelo cultivado bajo buenas prácticas agrícolas (BPA) y ARN de suelo cultivado bajo pobres prácticas agrícolas (PPA), respectivamente. Se sembraron 5 μ l de un total de 100 μ l de eluido final de purificación. El recuadro en color amarillo señala a los ARN ribosomales.

A partir de la secuenciación masiva de *ADNc-nifH* se obtuvieron 112.492 lecturas las cuales fueron analizadas para generar una base de datos de secuencias *ADNc-nifH* de alta calidad. Esta base de datos consistió en secuencias con un tamaño comprendido entre 300 y 400 pb, y que además no presentaron quimeras ni corrimiento en el marco de lectura. El curado de la base de datos primarios resultó en 80.787 lecturas y a 28.171 secuencias aminoacídicas únicas.

Debido a que para los organismos procariontes el concepto de especie es difícil de aplicar, generalmente es conveniente utilizar valores de similitud o disimilitud entre secuencias para agruparlas en OTUs (Unidades Taxonómicas Operativas). Las secuencias aminoacídicas fueron agrupadas en base al grado de similitud del 98% y cada uno de los grupos constituyó un OTU. Se adoptó este valor de umbral de semejanza, considerando que generalmente los genes codificadores de proteínas presentan un importante grado de similitud de secuencia a nivel de especie [137].

Este valor también fue usado en el trabajo de Collavino *et al.* 2014 de análisis masivo de secuencias *nifH* de suelos agrícolas [121]. El análisis de agrupamiento resultó en 1661 OTUs, de los cuales 437 OTUs estuvieron representados por al menos 3 secuencias.

Abundancia y distribución de los OTUs

La Figura V.3 muestra la distribución de abundancia, medida en número de secuencias, de los 1661 OTUs resultantes del análisis descrito en la sección precedente. Se encontró una significativa variación en las abundancias de cada uno de los OTUs. Pocos OTUs presentaron alta abundancia, 6 OTUs (0.3%) superaron el número de 1000 secuencias, 11 OTUs (0.7%) presentaron más de 100 y 168 OTUs (10%) al menos 10 secuencias, mientras que 1476 OTUs (89%) estuvieron representados por no más de 3 secuencias. Este perfil de distribución, en el cual la mayoría de los OTUs muestran baja abundancia y solo unos pocos muestran alta representación, también ha sido observado en otros trabajos de pirosecuenciación de muestras ambientales en especial de suelo [121], [150], [175].

Se ha sugerido que la captura de las secuencias de los OTUs “raros” dentro de una comunidad, es aleatoria comparado con los OTUs abundantes, y en consecuencia dicha falta de reproducibilidad puede afectar los análisis estadísticos [150]. Además, la tecnología de pirosecuenciación presenta sus errores propios, con tasas de error de hasta el 3% [162]. Esto resulta en artefactos de secuenciación que en el presente trabajo se trató de eliminar considerando solo aquellos OTUs que en la base de datos están representados por al menos tres secuencias. Criterios semejantes fueron adoptados por otros autores. Por ejemplo, en el análisis que se describe en el artículo de Montecchia *et al.* 2015 solo se consideraron los OTUs con al menos nueve secuencias [150]. Por otro lado, Mirza *et al.* 2014 seleccionaron los OTUs con al menos cinco secuencias [257], mientras que Collavino *et al.* 2014 OTUs con una representación mínima de tres secuencias [121].

Estimación de la diversidad y de la riqueza de la comunidad diazotrófica activa

La diversidad y riqueza de *nifH* en suelo se examinó mediante las curvas de rarefacción y los estimadores SChao1 de riqueza y Shannon-Wiener (H') de

diversidad. Los índices de riqueza proveen información cuantitativa del número de especies presentes en una comunidad, mientras que los índices de diversidad proveen medidas tanto de la abundancia de especies como de su distribución en la comunidad [287]. Las curvas de rarefacción a su vez permiten visualizar el grado de cobertura alcanzado en la secuenciación y comparar la diversidad de OTUs entre las distintas muestras [70].

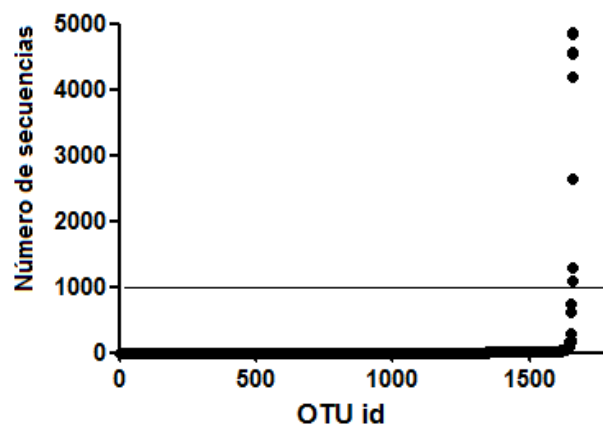


Figura V.3. Abundancia de secuencias para los 1661 OTUs. La abundancia se cuantificó como el número de secuencias por OTU. La línea horizontal separa a los OTUs que presentaron más de 1000 secuencias.

En la Figura V.4 se muestran las curvas de rarefacción del número de secuencias vs. el número de OTUs originados para cada una de las muestras de tratamiento y profundidad. Con un valor de corte de 3.159 secuencias, el valor correspondiente al menor número de secuencias detectado en el caso de la muestra NC 20 cm, la comparación reveló que las muestras de suelo no cultivadas (NC) presentaron el mayor número de OTUs. Mientras que en el resto de las muestras correspondiente a los suelos cultivados en general, se observó dominancia de algunos OTUs. Asimismo, las curvas de rarefacción mostraron que no obstante el esfuerzo de secuenciación ninguna de las muestras alcanzó la fase o valor máximo y constante en el número de OTUs. En particular, en las muestras de los suelos NC para lograr la cobertura de la mayoría de las especies de la comunidad diazotrófica el número de secuencias muestreadas debería ser aún mayor. Para el resto de las muestras, las

curvas de rarefacción se acercan a una asíntota indicando que la secuenciación cubrió una parte importante de la riqueza de secuencias ADNc-*nifH*.

Por otra parte con el estimador de riqueza no paramétrico SChao1 se calculó la cobertura relativa de riqueza (RC_{Chao1}) y como se muestra en la Tabla V.2 alcanzó valores del 67 al 100%. Estos valores indicarían que una fracción sustancial y comparable de riqueza de la comunidad diazotrófica fue evaluada en todas las muestras de suelo.

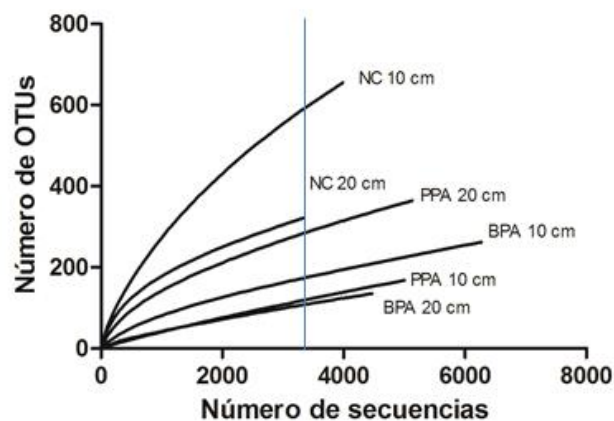


Figura V.4. Estimación de la riqueza de OTUs en las muestras de suelo mediante curvas de rarefacción. La línea vertical indica el punto en la curva correspondiente al menor número de secuencias (3159) de la base de datos de la muestra NC 20 cm. Las curvas de rarefacción se construyeron utilizando el método de re-muestreo sin reemplazo, mediante la ejecución del comando *rarefaction.single* en el entorno Mothur.

Los valores de la diversidad y de la riqueza de la comunidad diazotrófica variaron entre los distintos tratamientos y profundidades de suelo (Tabla V.2). El análisis de varianza simple (*One way ANOVA*) mostró que la diversidad y la riqueza fueron significativamente afectadas por el tratamiento agronómico (Figura V.5). Los suelos no cultivados presentaron valores mayores de H' ($r^2=0.69$; $p=0.0001$) y Chao1 ($r^2=0.64$; $p=0.0002$) comparado con los suelos cultivados. Mientras que no se observaron diferencias significativas a nivel de riqueza y diversidad entre BPA y PPA. Por otra parte, cuando se consideró la variable profundidad no se observaron diferencias significativas ($p>0.63$).

Tabla V.2. Riqueza, diversidad y cobertura relativa de las bibliotecas de amplicones *nifH*. El estimador de riqueza (Chao1) y el índice de diversidad Shannon-Wiener (H') se calcularon para las bibliotecas *nifH* con OTUs definidos al 98% de similitud a nivel de aminoácidos.

	NC		BPA		PPA	
	0-10 cm	10-20 cm	0-10 cm	10-20 cm	0-10 cm	10-20 cm
SeqN	3983	3335	6265	4463	4997	5128
SeqNn	3159	3159	3159	3159	3159	3159
Sobs	231	166	67	33	20	143
H'	3.96	3.32	1.12	0.37	0.11	2.57
Chao1	245	183	67	50	27	175
RC _{Chao1}	94	100	74	91	67	82

SeqN, número de secuencia por muestra, SeqNn, número de secuencia normalizado por muestra, Sobs, número de OTUs detectados a un nivel de distancia filogenética del 2%, RC, cobertura relativa calculada como el número de OTU dividido por la riqueza estimada (SChao1).

El análisis de varianza doble (*Two way ANOVA*) considerando profundidad y tratamiento reveló interacción entre ambos factores. Los valores de H' en los suelos no cultivados y en los suelos agrícolas bajo buenas prácticas agrícolas, fueron significativamente mayores en los primeros 10 cm ($p=0.0003$ y $p<0.0001$, respectivamente), mientras que la mayor diversidad de los suelos con pobres prácticas agrícolas se halló en la sub-superficie ($p<0.0001$) (Figura V.6).

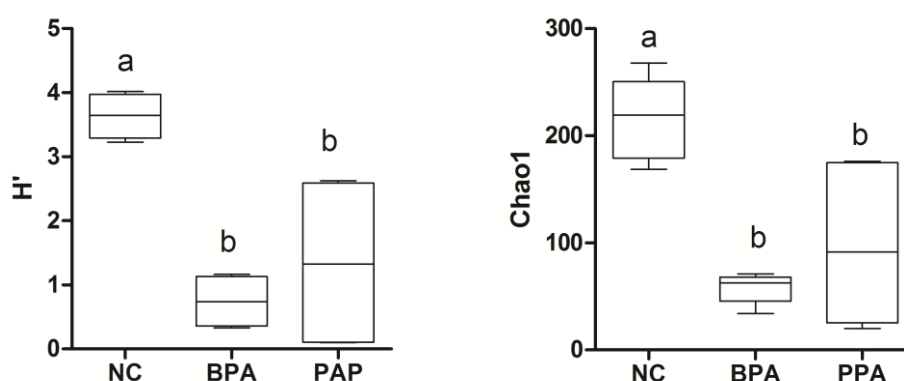


Figura V.5. Efecto del uso y manejo del suelo sobre la diversidad (H') y riqueza (Chao1) de *nifH*. Se aplicó la prueba de Tukey y las medias con distinta letra (a,b) indican diferencias significativas ($p\leq 0.05$). Las abreviaturas NC, BPA y PPA corresponden a no cultivado, buenas prácticas agrícolas y pobres prácticas agrícolas, respectivamente.

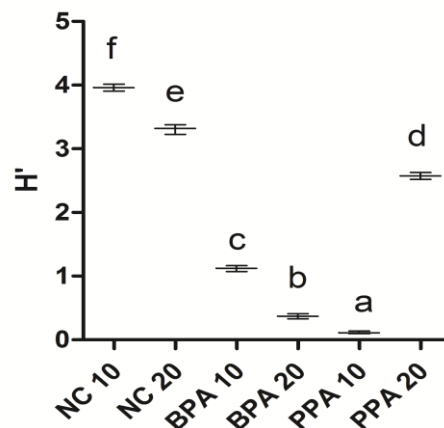


Figura V.6. Efecto del tratamiento y profundidad sobre la diversidad de *nifH*. Se aplicó la prueba de Tukey y las medias con distinta letra (de a-f) indican diferencias significativas ($p \leq 0.05$). La diversidad se da en valores del estimador H' . Las abreviaturas NC, BPA y PPA corresponden a no cultivado, buenas prácticas agrícolas y pobres prácticas agrícolas, respectivamente.

Por lo tanto, podemos inferir que la diversidad y riqueza de los fijadores de N_2 activos en suelo es afectada negativamente por la práctica agrícola comparado con un suelo no perturbado. La profundidad también parece afectar la diversidad diazotrófica en forma dependiente del uso y tratamiento agrícola.

Por otra parte, se compararon los valores de diversidad encontrados en la población activa, con los revelados previamente por la pirosecuenciación de ADN-*nifH* del mismo sitio [121]. Se realizó un análisis de varianza considerando los valores de diversidad del estimador Shannon-Wiener (H'). Como se muestra en la Figura V.7, los valores de H' en los suelos no cultivados no presentan diferencias significativas ($p=0.367$) entre las comunidades potencial y activa. Sin embargo, en los ambientes cultivados la diversidad de la comunidad potencial fue significativamente mayor que la activa (valores de $p < 0,0001$ y $p=0.002$ para BPA y PPA, respectivamente). Esta diferencia surge como consecuencia de que el efecto negativo del uso agronómico en la diversidad diazotrófica es solo observado en la comunidad activa, ya que como fuera previamente reportado los valores de diversidad de la

comunidad potencial no son significativamente diferentes entre los tratamientos analizados [121].

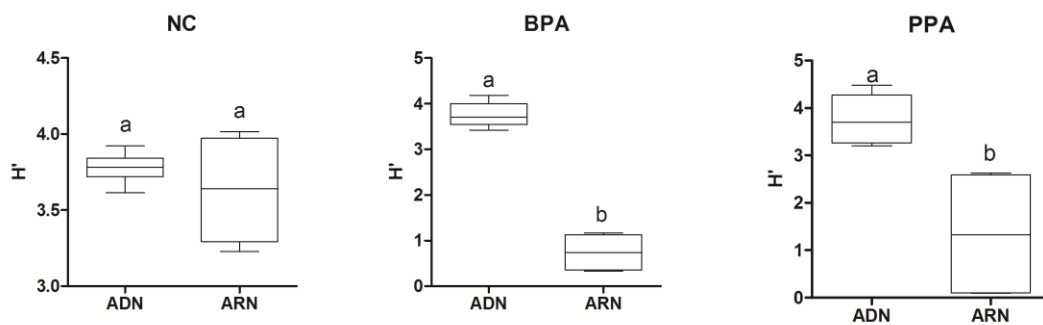


Figura V.7. Análisis comparativo (ANOVA) de la diversidad (H') entre las bases de datos ADN-*nifH* y ARN-*nifH* en los suelos de Pergamino con distinto uso tratamiento agrícola. Se aplicó la prueba de Tukey; las distintas letras (a,b) indican medias con diferencias significativas ($p \leq 0.05$).

Fierer *et al.* 2012 compararon la composición microbiana de dos suelos con distinto uso, uno con sucesión de pastizales y otro con uso agronómico, revelando una menor diversidad microbiana en los suelos con manejo agronómico [173]. Es posible especular que suelos con una mayor diversidad de plantas podrían promover una mayor diversidad bacteriana como es el caso de NC con pastizales naturales en comparación con la rotación de trigo/soja-maíz (BPA) o monocultivo de soja (PPA). Por el contrario, en el trabajo de Figuerola *et al.* 2014 hallaron que la diversidad regional de especies bacterianas en suelos de la Región Pampeana Argentina manejados en siembra directa con rotación de cultivo fue similar a la de los ambientes no cultivados, encontrando una menor diversidad solamente en siembra directa bajo monocultivo de soja [288].

Eilers y colaboradores evaluaron el efecto de la profundidad en la estructura de las comunidades de bacterias y archeas en suelo mediante la pirosecuenciación del gen ARNr 16S observando que la diversidad bacteriana es mayor en los primeros 10 cm disminuyendo un 20% y 40% en los 10-25 cm y 175 cm, respectivamente [191]. Resultados similares fueron obtenidos de otros estudios mediante técnicas de tipo huella digital de ADN, observando una disminución en la diversidad bacteriana en zonas más profundas de suelos [289]–[291]. Asimismo, Baldrian *et al.* 2012 encontró en suelos de bosques de coníferas durante un período de descomposición de materia

orgánica, mayor diversidad microbiana y niveles de transcriptos en el horizonte más superficial de desechos comparado con el horizonte orgánico [292]. En conjunto estos trabajos sugieren que la diversidad de la comunidad bacteriana tiende a disminuir en las mayores profundidades de suelo. En los suelos analizados aquí, también se observó una disminución de la diversidad diazotrófica en los 10-20 cm, excepto para el suelo bajo PPA.

Resumiendo, los resultados del efecto del tratamiento agronómico sobre la diversidad y funciones de las comunidades microbianas son contrastantes. Un conocimiento cada vez más profundo acerca de los efectos de los tratamientos agronómicos sobre la dinámica de las comunidades microbianas, permitiría diseñar estrategias dirigidas a conservar la diversidad de microorganismos y promover la ocurrencia de grupos taxonómicos específicos y claves para los diversos procesos biogeoquímicos que se llevan a cabo en el suelo.

Patrón de diversidad beta de las poblaciones diazotróficas con distinto manejo agrícola y profundidad

La *diversidad beta* (β) se refiere a la diversidad entre hábitats o sitios dentro de un mismo ecosistema o área, es decir, la tasa de recambio entre especies que se produce entre un hábitat y otro adyacente [76]. Para comparar la estructura de las comunidades diazotróficas de las distintas muestras de suelo aplicamos dos protocolos de medida de β diversidad. Uno de ellos consistió en la aplicación de la medida de disimilitud Bray-Curtis usando la matriz de abundancias relativas de los *subclusters* *nifH* [76], [288]. También se aplicó la medida de distancia *weighted* UniFrac entre pares de muestras, la cual utiliza las distancias filogenéticas entre los OTUs y sus abundancias relativas [293].

Los análisis de ordenación PCoA muestran que el patrón de distribución de las comunidades diazotróficas según Bray-Curtis (Panel A de la Figura V.8) resultó similar a la resultante de aplicar el análisis de *weighted* UniFrac (Panel B de la Figura V.8). Ambos criterios de análisis de disimilitud composicional explicaron una alta variabilidad en el componente principal 1 (PC1) con valores del 55.8 % y 47.5 %, respectivamente.

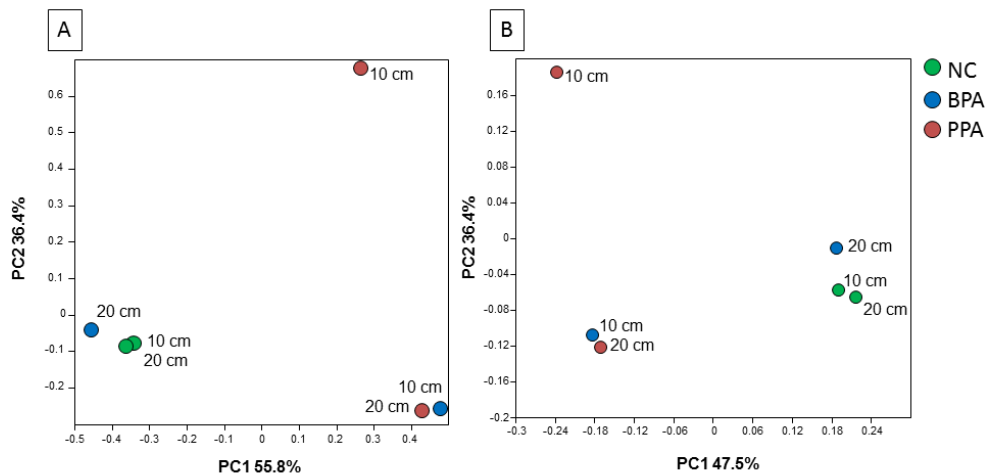


Figura V.8. Análisis de coordenadas principales (PCoA) calculados a partir de la abundancia de los *subclusters* usando Bray Curtis (A), y la abundancia de los OTUs y sus relaciones filogenéticas usando *weighted* UniFrac (B). Los círculos coloreados en azul y bordo corresponden a los distintos manejos agrícolas y en verde al suelo no cultivado.

Se encontró que las coordenadas PC1 y PC2 agrupan las comunidades de las dos profundidades de los suelos no cultivados (NC). Estos suelos NC además se agruparon con la profundidad de 20 cm de BPA, mientras que PPA 20 cm y BPA 10 cm formaron otro grupo. Este análisis muestra que 10 cm de PPA se separa claramente del resto de las muestras (Figura V.8). Este patrón de agrupamiento resultó similar a lo observado aplicando un análisis de agrupamiento o con la representación gráfica de los valores cuantitativos (*heatmap*) utilizando la matriz generada en UniFrac (Figura V.9 A y B, respectivamente).

Por otro lado, la cercanía en la gráfica de la Figura V.8 indica que la comunidad de la sub-superficie (10-20cm) de los suelos mostró un grado de similitud composicional mayor entre sí que con respecto a las comunidades superficiales. Este resultado es apoyado por el análisis de Venn (Figura V.10) donde se observó que en la profundidad de 10-20 cm se compartieron un total de 11 OTUs entre los distintos tratamientos, comparado con los 3 OTUs compartidos en los 0-10 cm.

La significancia de los agrupamientos observados entre las muestras se analizó estadísticamente aplicando la prueba *weighted* UniFrac (Tabla Suplementaria V.1). Comparando los tratamientos, se observó que la comunidad del ambiente no cultivado

fue significativamente diferente a la comunidad del manejo bajo monocultivo ($p \leq 0.003$). Para el caso de las muestras BPA, las relaciones fueron más complejas, la muestra de los 10 cm fue significativamente distinta a todas las muestras, excepto con PPA 20 cm ($p=0.268$), por el contrario la muestra BPA 20 cm fue significativamente distinta solo con la muestra de PPA 10 cm. En relación a la profundidad, se encontró que las comunidades de ambas profundidades de los ambientes NC y BPA no fueron significativamente distintas entre sí en cada tratamiento ($p=0.95$ y $p=0.082$, respectivamente), contrariamente en PPA las comunidades de las dos profundidades fueron diferentes ($p=0.003$).

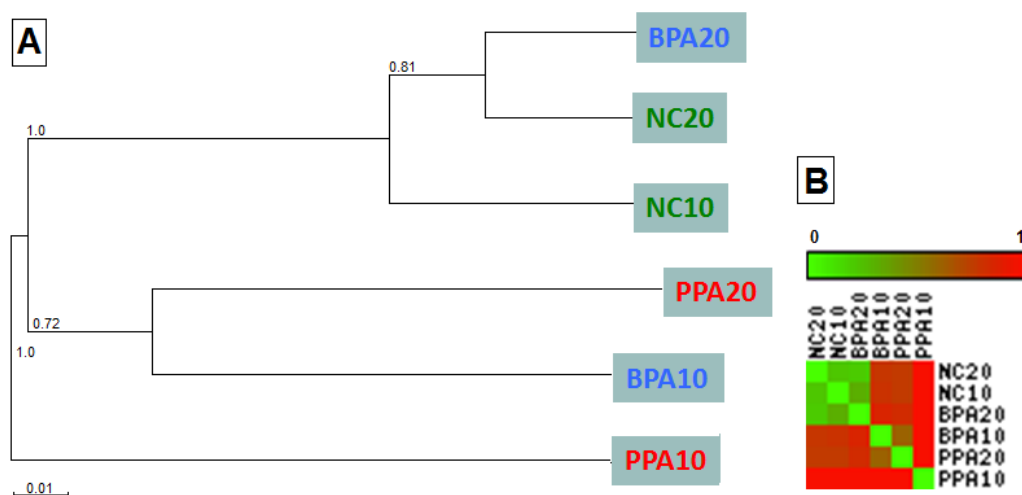


Figura V.9. Análisis de agrupamiento entre las comunidades diazotróficas de los distintos suelos.

El grado de similitud de las comunidades diazotróficas de los distintos tratamientos-profundidades se visualizan en el dendrograma (A) resultante del análisis de agrupamiento UPGMA realizado con el programa UniFrac (*weighted* UniFrac) y en el *heatmap* (B), el cual muestra los valores de disimilitud mediante la paleta de colores desde el rojo (máxima disimilitud) al verde (mínima disimilitud). La matriz utilizada para ambos análisis se construyó en el entorno UniFrac. Las abreviaturas NC, BPA y PPA corresponden a no cultivado, buenas prácticas agrícolas y pobres prácticas agrícolas, respectivamente.

Dada la disimilitud observada entre los ambientes NC y PPA, se puede concluir que el monocultivo de soja (PPA) parece afectar en mayor medida la composición de la comunidad diazotrófica en comparación con el tratamiento que comprende a la rotación de cultivos (BPA). Por otra parte, la profundidad afecta la estructura diazotrófica en los suelos cultivados, siendo particularmente significativa

en el manejo PPA, mientras que en las dos profundidades de los suelos no cultivados las comunidades fueron similares entre sí. Asimismo, considerando que las muestras de los primeros 10 cm son significativamente diferentes entre los tratamientos, mientras que en los 20 cm fueron similares, excepto en PPA y NC, se puede inferir que el manejo del suelo afecta mayormente la comunidad diazotrófica en los primeros 10 cm del suelo. Posiblemente la superficie 0-10 cm experimente una mayor exposición a las condiciones climáticas, al disturbio agrícola, y a los exudados de las raíces de las plantas, comparado con la capa 10-20 cm. En esta profundidad la humedad tiende a ser es más estable, es decir no está tan sujeta a los fenómenos de humedecimiento y secado. Por otro lado, en la sub-superficie el efecto del carbono en nutrientes también será menor (Comunicación personal-F. Behrends Kraemer). Esta “homogeneidad físico-química” a mayor profundidad del suelo podría ser una de los factores responsables que entre los distintos tratamientos se compartan más grupos diazotróficos comparados con la superficie 0-10 cm.

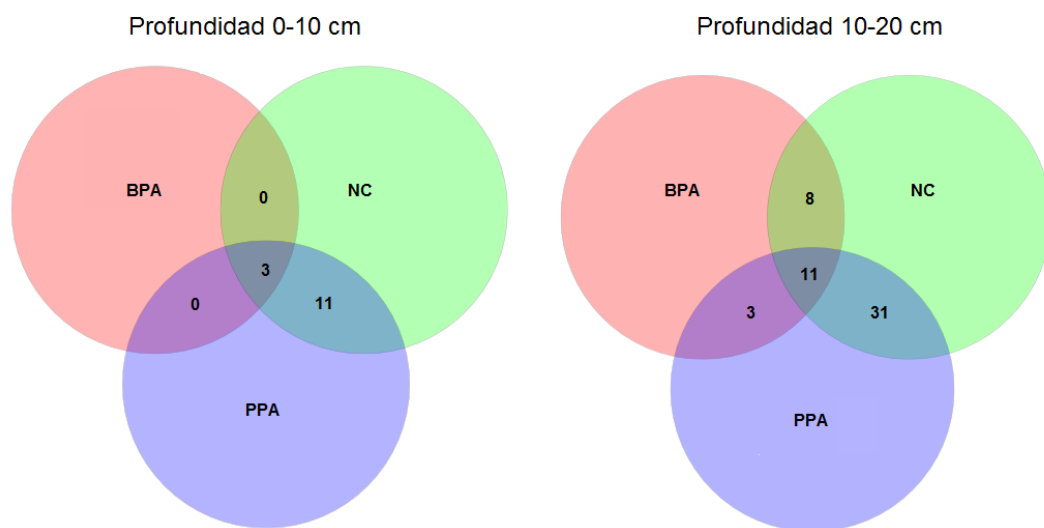


Figura V.10. Diagrama de Venn de los OTUs compartidos entre las muestras de 10 y 20 cm de profundidad. Los colores rosa y celeste corresponden a los manejos BPA y PPA, respectivamente. El color verde corresponde al suelo referencia no disturbado con pasturas naturales, NC. Los números indican el número de OTUs compartidos.

Composición de las comunidades diazotróficas activas en suelo

a. Asignación taxonómica de las secuencias *nifH*

Para la asignación taxonómica de los diferentes OTUs se adoptó la clasificación de las secuencias *nifH* propuesta por Zehr *et al.* 2003 consistente en un agrupamiento en *clusters* y *subclusters* (grupos y sub grupos *nifH*) [52]. A continuación se describen brevemente las características relevantes de los componentes de los *clusters*:

- El *cluster* I se compone de secuencias *nifH* de nitrogenasas convencionales (Fe-Mo) principalmente pertenecientes a miembros de los fila *Proteobacteria*, *Cyanobacteria*, ciertos *Firmicutes* (*Clostridiales* y *Paenibacillus*) y *Actinobacteria* (*Frankia*).
- Las secuencias de las nitrogenasas alternativas (Fe-Fe y Fe-V) se encuentran en el *Cluster* II y lo componen también secuencias *nifH* del grupo de las archeas metanogénicas.
- El *cluster* III comprende secuencias *nifH* de bacterias anaeróbicas estrictas y de bacterias reductoras de sulfato.
- El *cluster* IV se compone de secuencias parálogas de *nifH*.

Se encontró que en la base de datos ADNc-*nifH* predominaron (45%) secuencias de *Deltaproteobacterias* del *subcluster* 1A relacionadas con *Anaeromyxobacter* y *Desulfuromonadales*. Otros taxa dominantes fueron - *Alfaproteobacteria* del *subcluster* 1K representado por el orden *Rhizobiales* (29%), - *Cyanobacteria* del *subcluster* 1B (17%) relacionado con *Nostocales* y - *Deltaproteobacteria* y *Verrucomicrobia* del *subcluster* 3B relacionadas con los órdenes *Desulfovibrionales* y *Verrucomicrobiales* (8%). El resto de los *subclusters* presentaron menos del 1% de las secuencias. La Tabla V.3 resume la distribución de los filotipos *nifH* de toda la base de datos ADNc-*nifH*. Algunos de los *subclusters*, como 1J (*Rhodospirillales*, *Rhizobiales* y *Burkholderiales*) y 1M (*Methylococcales* y *Chromatiales*) que fueran reportados en otros trabajos como diazótrofos abundantes en suelo [52], [83], [121], se encontraron sub-representados en nuestras muestras. Asimismo, tampoco fueron detectadas secuencias relacionadas con *nifH* de Archeas (*Cluster* II) y secuencias parálogas (*Cluster* IV).

Tabla V.3. Filotipos *nifH* de los suelos de Pergamino.

Cluster	Subcluster	Cantidad de secuencias (%)	Cantidad de OTU (%)	Representación en las muestras (%)*	Grupos	Orden**
I	1	0.1	0.5	33	<i>Epsilon</i>	<i>Campylobacterales</i>
	1A	44.8	48.5	83	<i>Delta</i>	<i>Anaeromyxobacter</i> y <i>Desulfuromonadales</i>
	1B	16.7	2.3	50	<i>Cyanobacteria</i>	<i>Nostocales</i>
	1C	0.6	1.1	50	<i>Firmicutes</i>	<i>Clostridiales</i>
	1E	0.2	0.7	33	<i>Firmicutes</i>	<i>Paenibacillus</i>
	1G	0.1	0.2	50	<i>Gamma</i>	<i>Pseudomonadales</i> y <i>Enterobacteriales</i>
	1J	0.4	1.1	83	<i>Alfa</i>	<i>Rhizobiales</i> y <i>Rhodospirillales</i>
	1K	29.0	24.7	100	<i>Alfa</i>	<i>Rhizobiales</i>
	1O	0.1	0.2	17	<i>Gamma</i>	<i>Chromatiales</i>
	1P	0.2	1.1	67	<i>Beta</i> <i>Delta</i> y	<i>Rhodocyclales</i>
III	3B	8.0	19.5	67	<i>Verrucomicrobia</i>	<i>Desulfovibrionales</i> y <i>Verrucomicrobiales</i>

* Presencia del *subcluster* en las bibliotecas analizadas, 100% significa la presencia del *subcluster* en todas las muestras analizadas. ** Ordenes cercanos a las secuencias predominantes del *subcluster*.

b. Distribución de los subclusters *nifH* en las muestras de suelo por tratamiento-profundidad

La Figura V.11 muestra la distribución de los *subclusters* en relación a la profundidad del suelo de 0-10 cm y 10-20 cm. Se encontró una alta proporción de los *subclusters* 1A y 1K en ambas profundidades. En particular, el *subcluster* 1A que comprende las deltaproteobacterias anaeróbicas facultativas, mostró un importante aumento del 27% al 63% en la sub-superficie. Contrariamente, la proporción del *subcluster* 1K, representado principalmente por el orden *Rhizobiales*, descendió en la sub-superficie del 37 a 21%. Por otro lado, los *subclusters* 1B (cianobacterias) y 3B (anaerobios estrictos) presentaron una proporción del 33% en los primeros 10 cm (1B) y del 15% en los 20 cm, respectivamente. Además, se identificaron otros *subclusters* que en su conjunto representaron <3% y <1% en la superficie y sub-superficie, respectivamente.

El resultado del análisis que se presenta en la Figura V.12 corresponde a la distribución de los *subclusters* entre los diferentes manejos agrícolas y profundidades del suelo. Al igual que lo observado al analizar la diversidad de secuencias en las muestras, nuevamente encontramos menor diversidad de grupos en los suelos cultivados, ya sean en condiciones de buenas o pobres prácticas agrícolas. Bajo

rotación intensiva de cultivos (BPA) predominó en la superficie el *subcluster* 1K (98%) y en la sub-superficie el 1A (99%). Las cianobacterias (1B) fueron predominantes en la superficie del suelo bajo monocultivo (98%), mientras que los *subclusters* 1K (52%), 3B (45%) y 1A (3%) se encontraron principalmente en la sub-superficie. En contraste, en el suelo sin cultivar, las comunidades diazotróficas en ambas profundidades compartieron 9 *subclusters* y presentaron una proporción comparable de los mismos.

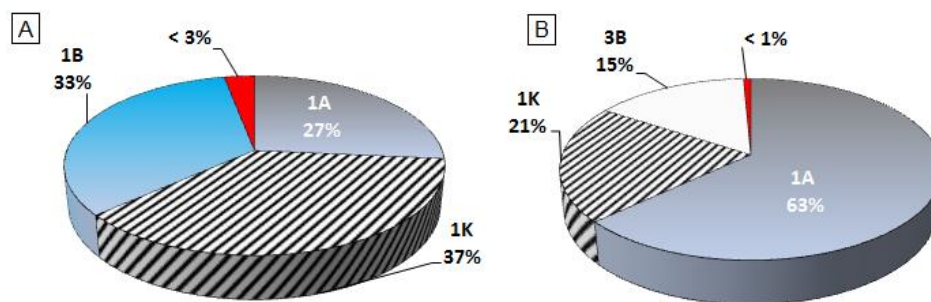


Figura V.11. Distribución de filotipos *nifH* en la muestra superficial (A) y sub-superficial (B) de suelos de Pergamino. Los % se refieren a la abundancia relativa de los diversos *subclusters* en toda la base de datos ADNc- *nifH*. En A la porción color roja que representa < 3% correspondió a los siguientes *subclusters*: 1, 1C, 1E, 1G, 1J, 1O, 1P, 3B y en panel B con <1% a: 1, 1B, 1C, 1E, 1G, 1J, 1P.

Estos resultados indicarían que en los suelos no cultivados la estructura de la comunidad diazotrófica es comparable en ambas profundidades. Por el contrario, en los suelos cultivados se observa, además del efecto por el manejo, un cambio notable de la composición de la comunidad diazotrófica dado por la profundidad.

En un trabajo reciente de Wang *et al.* 2016 se analizó la variación temporal de la abundancia y estructura de la comunidad total y diazotrófica en dos profundidades de suelo 0-20 cm y 20-40 cm, bajo cuatro regímenes de fertilización durante dos años. Concluyeron que la profundidad es el factor más importante que afecta el tamaño y la estructura de las comunidades total y diazotrófica. A su vez observaron que el efecto de la fertilización sobre las comunidades diazotróficas en ambas profundidades fue significativa, aunque estacionalmente fluctuante [294].

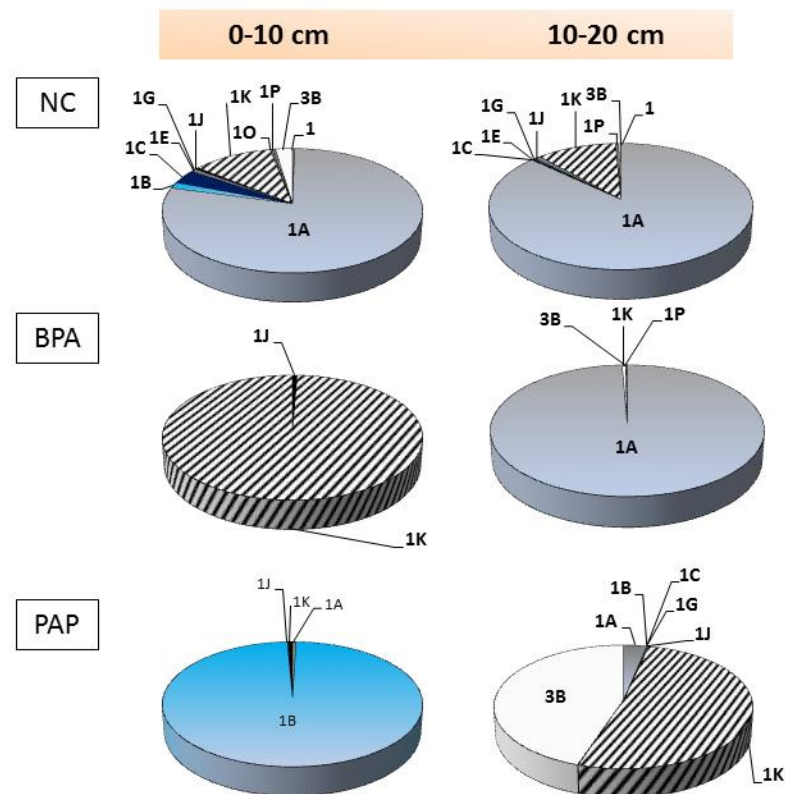


Figura V.12. Distribución de la abundancia de los subclusters *nifH* en la comunidad diazotrófica activa en los distintos tratamientos y profundidades de suelo de Pergamino. Las abreviaturas NC, BPA y PPA corresponden a PPA corresponden a no cultivado, buenas prácticas agrícolas y pobres prácticas agrícolas, respectivamente.

c. Clasificación taxonómica de los OTUs más abundantes

La Figura V.13 muestra la distribución de la abundancia de los 16 OTUs predominantes (presentaron al menos 100 secuencias) en los distintos suelos analizados. Estos 16 OTUs representan una proporción importante (53%) de la base de datos ADNc-*nifH*. Se observa que ciertos OTUs fueron asociados a algunas condiciones de tratamiento de suelo y/o profundidad de muestreo. Algunos OTUs se observaron solo en las capas superficiales de ciertos tratamientos, por ejemplo, los OTUs 12, 376 y 380 en el suelo NC y el OTUs 755 en el suelo PPA. Por el contrario, el OTU 2 se encontró solo en la sub-superficie de los suelos cultivados. Asimismo se observaron OTUs ubicuos, como el OTU 3 presente en todos los suelos analizados y el OTU 8, solo ausente en BPA 10. En la muestra de suelo NC se detectó la mayor cantidad de OTUs abundantes, especialmente en los primeros 10 cm del suelo.

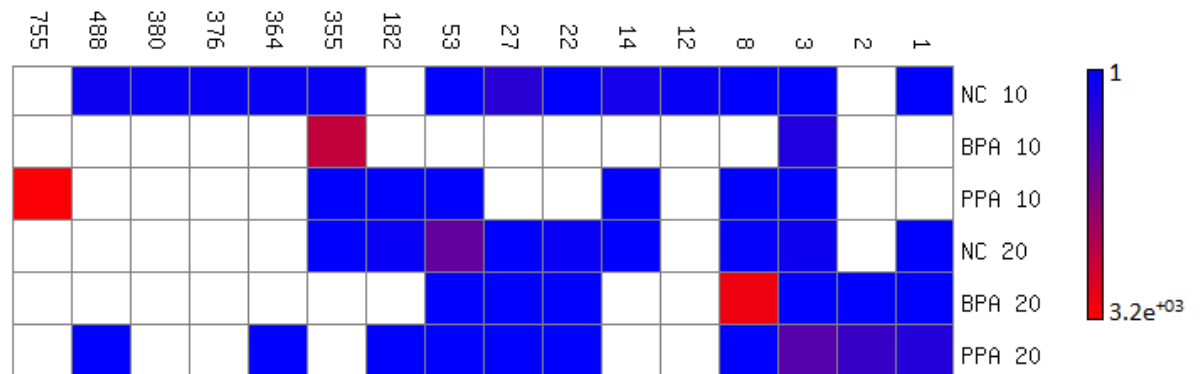


Figura V.13. Distribución de abundancia de los OTUs predominantes en los suelos de Pergamino.

La abundancia de los OTUs se muestra en un *heatmap* mediante la paleta de colores de azul (min) a rojo (máx). Las abreviaturas NC, BPA y PPA corresponden a PPA corresponden a no cultivado, buenas prácticas agrícolas y pobres prácticas agrícolas, respectivamente. Los números 10 y 20 se refieren a las profundidades de muestreo de suelo 0-10 cm y 10-20 cm, respectivamente.

Con el propósito de analizar las relaciones filogenéticas de los 16 OTUs más abundantes, se construyeron árboles de filogenia utilizando la base de datos de referencia de secuencias *nifH* de Zehr *et al.* [181] incorporando además las secuencias de ADN-*nifH* de los suelos de Pergamino [121]. En la Figura V.14 se muestra la relación filogenética de los OTUs dominantes dentro de la base de datos ADNC-*nifH* y su proporción en cada una de las muestras de suelo. A continuación se describen las relaciones encontradas en cada uno de los *subclusters nifH* observados.

- **Subcluster 1A**

Como se muestra en la Figura V.14, 11 OTUs de los 16 OTUs más abundantes fueron clasificados dentro del *subcluster* 1A, 6 de ellos se relacionaron con secuencias de *Geobacter uranireducens*, *Geobacter bemidjiensis* y *Geobacter lovely* y los restantes 5 con secuencias de *Anaeromyxobacter sp.* Fw109-5. En general, se observó que estos OTUs fueron detectados mayoritariamente en los suelos no cultivados, algunos en la superficie (OTUs: 12, 14, 27, 364, 376, 380 y 488) y otros en la sub-superficie (OTUs: 22, 53 y 182) con excepción del OTU 8 el cual predominó en la sub-superficie de BPA. El análisis de componente principales utilizando la matriz de abundancia de los *subclusters* (Figura Suplementaria V.1), mostró que la presencia de

las secuencias 1A fue determinante en la agrupación de las comunidades diazotróficas de ambos suelos NC y BPA de los 20 cm.

El OTU 8 -uno de los más abundantes en la base de datos ADNc-*nifH* (16.3%)- se relacionó con secuencias *nifH* de suelo previamente descritas por Buckley *et al.* 2007, las mismas se posicionaron dentro de un grupo profundamente divergente del *Cluster I* al cual los autores denominaron “grupo de microorganismos no cultivados” [259]. Cabe mencionar que en dicho estudio se utilizó una metodología de análisis basada en principios totalmente diferentes a los usados en el presente trabajo. Utilizaron el isótopo pesado de nitrógeno (^{15}N) como trazador del proceso de fijación de nitrógeno molecular y luego obtuvieron secuencias *nifH* a partir de la amplificación de ADN con la marca ^{15}N . Los autores concluyeron que este “grupo de microorganismos no cultivados” podría constituir una población significativa y activa de fijadores de N_2 de vida libre en suelo. Por otro lado, el OTU 8 también se encontró muy relacionado a clones *nifH* revelados en suelos estudiados por Hsu *et al.* 2009. En dicho trabajo también se analizó la comunidad diazotrófica funcional mediante el isótopo ^{15}N y hallaron como grupos predominantes los *subcluster 1A* y *1K* [258].

Las secuencias *nifH* clasificadas como 1A relacionadas al orden *Anaeromyxobacter* y *Desulfuromonadales* representaron el 45% del total de secuencias en la base de datos ADNc-*nifH*. Coincidentemente, en el censo que realizaron Gaby y Buckley *et al.* 2011 a partir de varias bases de datos encontraron al *subcluster 1A* como uno de los principales y mayoritarios en suelo [83]. En el trabajo de Collavino *et al.* 2014 también encontraron a este filotipo ampliamente distribuido en varios suelos de la región pampeana argentina [121]

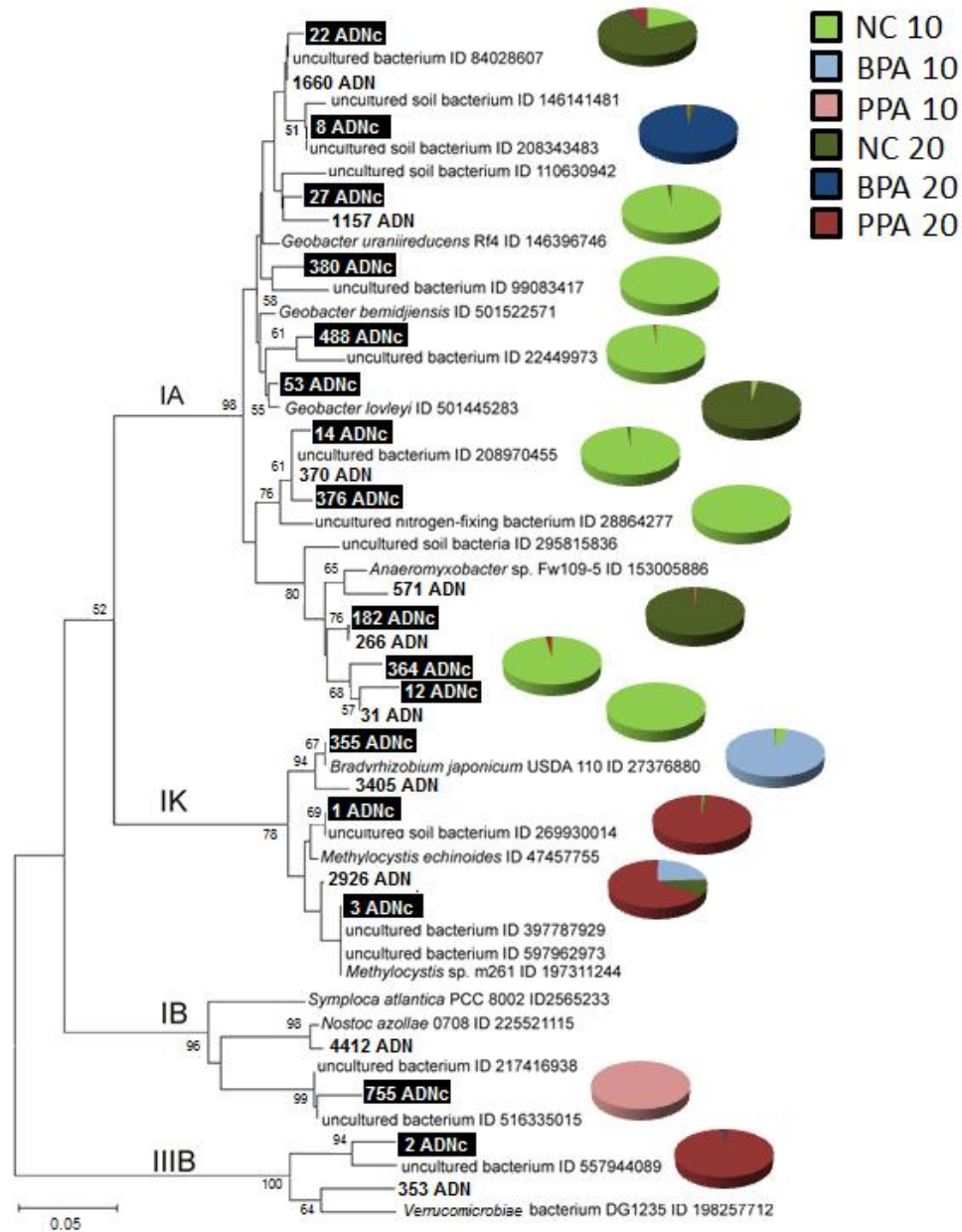


Figura V.14. Relación filogenética de las secuencias aminoacídicas NifH representativas de los OTUs predominantes en los suelos de Pergamino. Estos OTUs abundantes (cajas negras) presentaron al menos 100 secuencias *nifH*. El número que precede la abreviación ADN corresponden a OTUs *nifH* obtenidos a partir del estudio de la comunidad diazotrófica potencial en los suelos de Pergamino [121]. Los *clusters* canónicos de *nifH* se indican sobre las ramas principales. Se muestran los porcentajes de confianza superiores a 50% de 1000 *bootstrap*. Cada OTU se asocia a un gráfico de torta que muestra la proporción del mismo entre las distintas muestras de suelo. El árbol de filogenia se construyó en ARB con secuencias aminoacídicas parciales de NifH usando el método *Neighbour-Joining*.

Asimismo, otros autores examinando secuencias *nifH* de la población diazotrófica asociada a raíces del pasto perenne *Molinia caerulea* en un sitio de Suiza hallaron que más del 50% de las secuencias *nifH* correspondieron al *subcluster* 1A [295]. La predominancia de este grupo potencial fue revelado también en otros ecosistemas tales como en un bosque de coníferas en el norte de Nuevo México [236], en un campo de maíz en el Condado de Clinton, Nueva York [258] y en una sección de deshielo del glaciar Damma en Suiza [296].

Hasta hace poco, el único homólogo de *nifH* de un microorganismo cultivable dentro del *subcluster* 1A pertenecía a *Geobacter metallireducens*, por lo tanto, resultaba difícil determinar la amplitud filogenética del *subcluster*. Los esfuerzos más recientes en secuenciamiento completo de genomas, han revelado que los genes *nifH* de géneros tales como *Pelobacter*, *Desulfuromonas* y *Anaeromyxobacter* afines con miembros de la familia *Geobacteraceae* integran el *subcluster* 1A y se agrupan claramente con las secuencias recuperadas en este estudio. En conjunto estos datos evidencian que las poblaciones clasificadas por su secuencia *nifH* dentro del *subcluster* 1A, las cuales están poco descritas y caracterizadas desde el punto de vista de su fisiología y relación filogenética, aparecen como componentes importantes y ubicuos de los fijadores de N₂ libres en suelos.

La familia *Geobacteraceae*, ha sido caracterizada como un grupo de bacterias anaeróbicas facultativas, y como importantes agentes de la biorremediación, oxidación de compuestos orgánicos y de ciertos metales acoplado a la reducción disimilativa de sulfato de hierro (III), debido a su avanzada capacidad de establecer relaciones sintróficas y a la transferencia extracelular de electrones [297]. Se ha propuesto que su capacidad para fijar N₂ les proporciona una ventaja adaptativa en el subsuelo con nutrientes agotados [298]. Además, la capacidad de los representantes de este clado para utilizar el oxígeno como aceptor alternativo de electrones le da ventajas competitivas en condiciones óxicas/anóxicas fluctuantes del suelo [299]. Estas capacidades funcionales adaptativas de *Geobacter* pueden explicar la importante presencia de este componente de la comunidad fijadora de nitrógeno en suelos, que en algunos casos de suelos argentinos, las propiedades físico-químicas de éstos promoverían el establecimiento y funcionalidad de *Geobacter*.

- ***Subcluster 1K***

Los 3 OTUs 1K abundantes presentaron similitud con secuencias *nifH* del orden *Rhizobiales* (Figura V.14). Los OTUs 1 y 3, representando respectivamente el 2.6% y 8.9% del total de secuencias de la base de datos ADNc-*nifH*, se relacionaron con secuencias *nifH* de las bacterias metanotrofas *Methylocystis/Methylosinus* así como también con secuencias de microorganismos no cultivables del suelo. Estos OTUs se encontraron mayoritariamente en la sub-superficie de PPA. El OTU 355, con una representación del 13.5% en la base de datos ADNc-*nifH* y encontrado principalmente en la superficie de BPA, presentó alta homología con *Bradyrhizobium japonicum* (93% de identidad) y con el OTU 3405 (87% de identidad) hallado previamente en el análisis de ADN-*nifH* de suelos de Pergamino [121]

El *subcluster* 1K comprende grupos filogenéticamente muy diversos (*Alfa*-, *Beta*- y *Gammaproteobacterias*) y se lo encuentra ampliamente distribuido en suelos [121], [233], [300]. En las muestras de suelo analizadas aquí, el *subcluster* 1K se halló representado principalmente por dos grupos bacterianos dentro del orden *Rhizobiales*: los metanotrofos y los rizobios. Las bacterias metanotrofas usan el metano como fuente de carbono y energía, son aeróbicas y se han reportado como diazotrofos abundantes en suelo [121], [258], [296]. La abundancia de este grupo se ha asociado a su ventaja adaptativa de crecimiento en suelos con bajos contenidos de materia orgánica y dependientes de los productos de fermentación de los exudados de raíces o retención de los rastrojos [258], [296]. Con respecto a los rizobios, estos son quimioheterótrofos y aeróbicos capaces de fijar nitrógeno en simbiosis con leguminosas y en algunos casos en vida libre en microaerofilia [271], [301]. En un estudio acerca de la caracterización funcional de la comunidad endófito de raíces de arroz, se encontró que las secuencias de algunos transcritos *nifH* se agruparon con secuencias *nifH* de *Bradyrhizobium* y *Rhizobium spp* [135]. Asimismo, en las raíces del cultivo de caña de azúcar de varios continentes también fue predominante la expresión de genes *nifH* estrechamente relacionados con *Rhizobium rosettiformans* [279]. Por otra parte, Wang *et al.* 2013 estudiando el perfil ecológico de la comunidad diazotrófica mediante el análisis masivo del gen *nifH* en cuatro suelos de diferentes regiones, encontraron un importante número de secuencias *nifH* relacionadas con *Bradyrhizobium* y *Rhizobium*

[302]. Lo que resulta interesante es que en dichos suelos no se encontraban plantas leguminosas hospederas de dichos microorganismos potenciales fijadores de nitrógeno. Los autores proponen que los rizobios presentan un amplio rango de estilos de vida libre o que, alternativamente, microorganismos no cultivables pudieron haber recibido los genes *nifH* mediante transferencia horizontal. Estos resultados, al igual que los observados en nuestras muestras de suelo, demuestran que los rizobios o microorganismos no cultivables portando genes *nifH* de rizobios, representan una parte predominante de la población potencial y activa fijadora de N₂ en vida libre en el suelo.

- ***Subcluster 1B***

El único OTU abundante dentro del *subcluster* 1B, el OTU 755, agrupó con secuencias no cultivables, una de ellas detectada en un análisis de suelos ácidos del sur de China [303] y presentó una similitud baja (81%) con la especie cultivable más cercana, *Nostoc azollae*. Este OTU solo se detectó en los 10 cm del suelo PPA y con el 16.4% presentó la mayor abundancia relativa de secuencias *nifH* (Figura V.13). El análisis de PCA utilizando la matriz de abundancia de los *subclusters* (Figura Suplementaria V.1), mostró que las secuencias 1B se relacionan con PPA 10 cm. Otros 9 OTUs con clasificación taxonómica 1B presentaron una baja representación con <0.8% del total de secuencias en la base de datos ADNc-*nifH* y se hallaron principalmente presentes en NC 10 cm.

Las cianobacterias son fotosintetizadoras oxigénicas, aeróbicas y algunas son capaces de fijar N₂ en vida libre, en asociaciones y en simbiosis con plantas y microorganismos [304]. Se ha sugerido a las cianobacterias como participantes en el proceso de mejoramiento de agregación de suelos, y de jugar un rol colaborativo en la supresión de patógenos y proceso de biorremediación [186], [305].

En el trabajo de Teng y colaboradores *et al.* 2009 analizaron el efecto del agregado de estiércol orgánico sobre la estructura de la comunidad de diazotrofos en suelos ácidos [303], encontrando que en la mayoría de las bibliotecas *nifH* analizadas, las secuencias se relacionaban con secuencias *nifH* de cianobacterias. Además, este grupo de bacterias ha sido reportado como uno de los potenciales grupos diazotróficos

relevantes en el ambiente acuático [52], en la corteza de suelos áridos [306], en campos de cultivo de arroz en Uruguay [307] y en suelos de deshielo de los Valles secos de la Antártida y de los Alpes Suizos [296], [308]. Recientemente, en las matas microbianas de Guerrero Negro, un ambiente que experimenta frecuentes periodos de desecación e inundación y disrupción física, se reveló que el grupo de las cianobacterias es uno de los principales grupos funcionales involucrados en la fijación de N₂ [309]. Estos datos indican la amplia distribución de las cianobacterias fijadoras de nitrógeno en distintos ambientes y regiones de la tierra, y uno de los componentes activos en el funcionamiento de varios ecosistemas incluyendo los suelos.

- ***Subcluster 3B***

En relación con el *subcluster* 3B, se encontró un único grupo de secuencias abundantes, el OTU 2, el cual agrupó con secuencias *nifH* de bacterias no cultivables halladas en madera en descomposición de los árboles *Fagus sylvatica* y *Picea abies* en Alemania [231]. También se relacionó lejanamente (82% de similitud) con secuencias *nifH* del orden *Verrucomicrobiales*. El OTU 2 representó el 3.8% del total de secuencias en la base de datos *nifH*-ADNc y se halló exclusivamente en la sub-superficie de los suelos cultivados con el 99% en PPA y el 1% BPA.

El *subcluster* 3B contiene la mayor diversidad de linajes *nifH* y a su vez es el grupo con menor diversidad muestreada, presenta también grandes bifurcaciones y ramas largas en el árbol de filogenia [181]. Además, el mayor porcentaje de secuencias *nifH* que lo componen provienen del suelo [83]. Este grupo comprende entre otros a los miembros del filo *Verrucomicrobia*, los cuales son abundantes en diversos hábitats, incluyendo suelos, aguas, sedimentos y ambientes extremos como aguas termales y con pH de 2-2.5 [310]–[312]. En particular, en suelo este filo representa el 1.2 al 11% del total de las bacterias del suelo [313], [314]. Han sido revelados representantes del orden *Verrucomicrobiales* preferencialmente en la profundidad de 10 a 50 cm de algunos suelos [191], [315]. Se lo ha propuesto como un filo con estrategias de vida oligotróficas [120], lo cual podría permitirle beneficiarse de los escasos recursos presentes en los estratos más profundos del suelo, sumado a la ventaja adaptativa de realizar fijación de nitrógeno en condiciones anóxicas.

Resumiendo, en esta sección se ha encarado la asociación filogenética de los 16 OTUs que se encontraron en mayor proporción en la base de datos *nifH*-ADNc, y su distribución en los diferentes suelos y profundidades. Este análisis demostró que las secuencias *nifH* predominantes se asociaron a los *subclusters nifH* 1A (*Deltaproteobacteria* y bacterias no cultivables previamente descritas en suelo), 1K (*Alfaproteobacteria*), 1B (*Cyanobacteria*) y 3B (*Verrucomicrobia*). En el caso de los dos últimos grupos, las secuencias halladas presentaron homología muy baja con las secuencias de especies más cercanas cultivables. Esto indicaría que nuestro análisis reveló tanto filotipos asimilables a especies cultivadas, como filotipos nuevos que presentaron alta similitud solo con secuencias *nifH* metagenómicas o no cultivables detectadas en otros ambientes naturales incluyendo el suelo. Asimismo estos *subclusters* presentaron una distribución diferencial en las muestras de suelo bajo distintas condiciones de tratamiento-profundidad (Figura V.14). Los OTUs abundantes comprendidos en la clasificación taxonómica 1A se hallaron principalmente en ambas profundidades de los suelos no cultivados mientras que los OTUs 1K predominaron en los suelos cultivados. Finalmente los OTUs con clasificación 1B y 3B se hallaron mayormente representados en la superficie y sub-superficie, respectivamente, de los suelos bajo monocultivo de soja. En un primer análisis exploratorios estos OTUs *nifH* abundantes de la comunidad diazotrófica activa podrían seleccionarse para estudiar su abundancia, tanto a nivel potencial como por RT-qPCR y comenzar a validar si podrían seleccionarse como candidatos para su adopción como indicadores de calidad de suelo.

La carencia de taxones relacionados con secuencias *nifH* de los géneros bacterianos *Azotobacter* y *Azospirillum* en la base de datos ADNc-*nifH*, resulta sorprendente, ya que son grupos bacterianos ampliamente distribuidos en suelo [316], [317] y es reconocida su capacidad metabólica de fijación de N₂. Al igual que en cualquier metodología basada en la amplificación por PCR, no puede descartarse que los potenciales desvíos/sesgos de los cebadores en la PCR podrían explicar la falta de representación de secuencias de estos géneros bacterianos. Sin embargo, la falta de representación de *Azotobacter* también ha sido reportado en la comunidad diazotrófica potencial de algunos suelos tropicales examinados tanto por métodos

dependientes como independientes de cultivo [318]. Dicho género tampoco ha sido revelado en la comunidad diazotrófica potencial de varios suelos de la región pampeana argentina, mientras que *Azospirillum* representó casi el 10% en la base de secuencias ADN-*nifH* [121] indicando que este grupo no necesariamente es uno de los filotipos activos en la fijación de N₂ en estos suelos de Pergamino. *Azotobacter* es un fijador aeróbico que en el proceso de fijación de nitrógeno requiere de un alto consumo de carbono, se puede especular que en los suelos examinados y en el momento que se tomó la muestra, no se daban las condiciones óptimas para que *Azotobacter* generara transcritos de *nifH*.

Comparación entre la estructura de diazótrofos revelada por análisis de ADN y ADNc del gen nifH

En este capítulo se ha descrito la caracterización del perfil de la comunidad activa en fijación de nitrógeno mediante la secuenciación profunda de ADNc-*nifH*. Por otro lado, nuestro grupo también ha acumulado datos resultantes del análisis de pirosecuenciación del gen *nifH* en el ADN extraído de suelos argentinos [121]. La disponibilidad de estos datos fueron estimulantes para encarar un análisis comparativo con los datos obtenidos en el análisis de las secuencias *nifH* generadas a partir de ARN de suelo. Las secuencias *nifH* que se obtuvieron a partir de ADN de suelos de Pergamino correspondieron a una profundidad de muestreo de 0-20 cm del nivel del suelo. Por lo tanto, para poder comparar el efecto del manejo agrícola en la composición de la comunidad potencialmente presente y activa; fue necesario eliminar el factor profundidad de los datos de secuenciación de ADNc-*nifH*. De esta forma, se combinaron en un único conjunto a las secuencias ADNc-*nifH* correspondientes a las profundidades 0-10 y 10-20 cm para cada uno de los tratamientos NC, BPA y PPA. A continuación se detalla el análisis comparativo de la distribución de los diferentes *subclusters* y OTUs observados en ambas bases de datos.

En la Figura V.15 se muestra la distribución de los principales *subclusters* (> 1%) presentes en las bases de datos *nifH* de ADN y ARN de suelos de Pergamino. Se observa la predominancia de ciertos filotipos de *nifH* en ambas bases de datos, tal es el caso de los *subclusters* 1K (*Alfa* y *Betaproteobacterias*) y 1A

(*Deltaproteobacterias*). Otros *subclusters*, como *Cianobacteria* (1B) y las bacterias anaeróbicas estrictas del subcluster 3B, estuvieron poco representados en la comunidad diazotrófica potencial (el 1.4% y 3%, respectivamente), pero su representación se incrementó considerablemente en la comunidad activa (17% y 8%, respectivamente). Por el contrario, los *subcluster* 1J (*Azospirillum* y *Rhodospirillales*) y 1P (*Rhodocyclales*) fueron grupos predominantes entre las poblaciones potenciales (16% y 4%, respectivamente) mientras que en ARN su representación fue mínima (<0.5%).

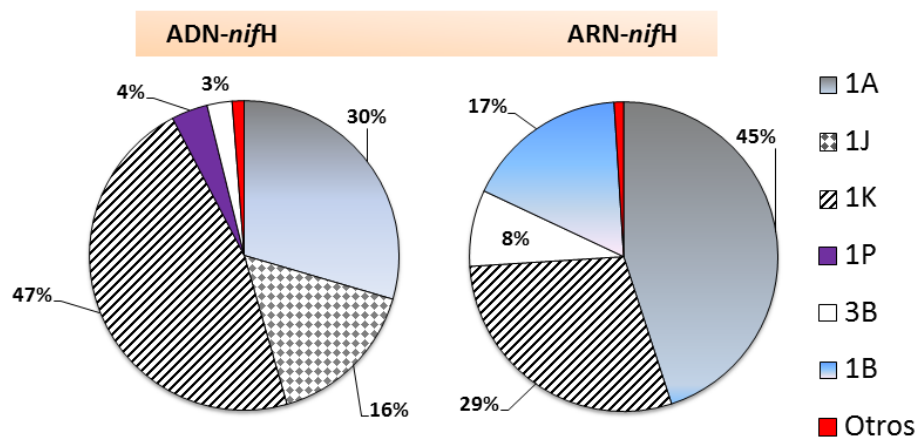


Figura V.15. Comparación de la composición de las comunidades diazotroficas potencial (ADN-*nifH*) y activa (ARN-*nifH*) en suelos de Pergamino. Las lonjas en rojo en las gráficas de torta para ambas bases de datos *nifH* corresponden a los *subclusters* cuya representación fue < al 1% del total de secuencias *nifH*.

En la Figura V.16 se muestra la distribución de los *subclusters* en los distintos tratamientos agronómicos en las dos bases de datos (ADN-*nifH* y ARN-*nifH*). Se hallaron ciertas similitudes entre ambas comunidades. Tanto en la comunidad potencial como en la activa el *subcluster* 1A se encontró en mayor proporción en los suelos sin cultivar (59% y 82%, respectivamente). A su vez, en las dos bases de datos se encontró que la distribución de dicho *subcluster* fue más relevante en el tratamiento BPA (17% en ADN-*nifH* y 41.5% en ARN-*nifH*) comparado con PPA (13% en ADN-*nifH* y 2% en ARN-*nifH*). Por el contrario, 1K se encontró en mayor proporción en los suelos cultivados y en particular en BPA (49% en ADN-*nifH* y 58% en ARN-*nifH*). Asimismo, el *subcluster* 3B también se encontró en mayor proporción en los

suelos cultivados, en BPA (4%) de la comunidad potencial y en PPA (23%) de la comunidad activa. El *subcluster* 1B, detectado en una representación relativamente baja en la mayoría de los suelos, se observó en mayor proporción en los suelos PPA de ambas comunidades.

Concluyendo, se observan importantes puntos de contacto en la composición de las comunidades diazotróficas potencial y activa, la preponderancia de filotipos como 1A y 1K, así como la distribución de los mismos en los diferentes tratamientos, sugiere que los principales grupos encontrados como potenciales son también activos en la fijación de N₂.

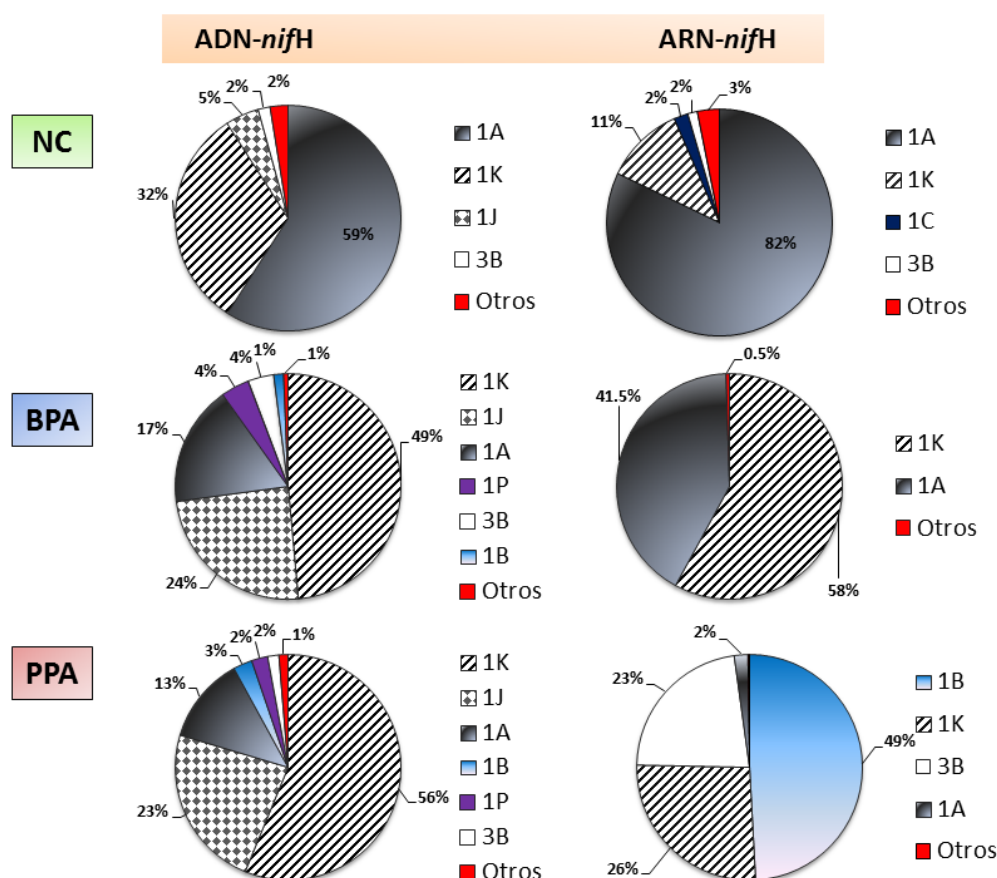


Figura V.16. Composición de la comunidad de diazótrofos revelada por el análisis de ADN y ARN del gen *nifH* en los suelos de Pergamino con distinto tratamiento agronómico. Las porción en rojo en las gráficas de torta para ambas bases de datos *nifH* corresponden a los *subclusters* cuya representación comprendió el rango de 0.2% a 3% del total de secuencias *nifH* en ADN-*nifH* y ARN-*nifH*.

Relación filogenética y distribución de los OTUs nifH dominantes en ADN-nifH y ARN-nifH de suelo.

Con el objeto de profundizar en las relaciones filogenéticas existentes entre los OTUs presentes en las poblaciones diazotróficas potenciales y activas y comparar sus distribuciones en las distintas muestras de suelo, se originó una única base de datos con las secuencias aminoacídicas ADNc-*nifH* y ADN-*nifH*, las mismas se agruparon en OTUs usando el programa CD-Hit y se clasificaron mediante la construcción de árboles de filogenia utilizando la base de datos de referencia de secuencias *nifH* de Zehr *et al.* 2003 [181].

El agrupamiento de las secuencias resultó en un total de 937 OTUs. Los 22 OTUs más abundantes (≥ 200 secuencias) representaron el 60% del total de secuencias *nifH*. En la Figura V.17 se muestran las relaciones entre estos OTUs, clasificados dentro de los *subclusters nifH* 1K, 1A, 1J, 1B y 3B, y su distribución en los distintos tratamientos analizados. Se puede observar que la mayoría de los OTUs abundantes a nivel de ARN también se encontraron representados a nivel de ADN, aunque con distinta proporción relativa.

De los 22 OTUS más abundantes 10 resultaron asociados taxonómicamente al *subcluster* 1K. El sub grupo de secuencias representado por los OTUs 3, 19, 98, 238, 586 y 605 se agrupó principalmente con secuencias *nifH* del género *Bradyrhizobium*. Se observó que estos OTUs fueron en general más abundantes a nivel de ADN, excepto por el OTU 586 que presenta una alta abundancia en las secuencias ADNc en las buenas prácticas. Otros OTUs 1K se relacionaron con secuencias *nifH* de bacterias no cultivables del suelo, entre ellos los OTUs 1, 522 y 657. El OTU 1 se halló presente en todos los suelos examinados tanto a nivel potencial como activo, por el contrario el OTU 657 solo estuvo presente a nivel de ADN (Figura V.17).

Con respecto a los OTUs 588 y 635, asignados al *subcluster* 1J, presentaron homología con dos especies, *Methylobacterium nodulans* y *Gluconacetobacter diazotrophicus*. El OTU 588 se encontró en todos los suelos, tanto a nivel de ADN como ARN, sin embargo es interesante resaltar que en ambos casos fue particularmente abundante en los suelos cultivados bajo rotación de cultivos (BPA).

Ha sido reportado que la presencia de *M. nodulans* es afectada positivamente por la presencia de residuos de maíz [258], coincidentemente el maíz es uno de los cultivos de rotación que presentan estos suelos BPA. Se podría inferir que no solo la presencia de este OTU sino también su actividad fijadora de N₂, se ve favorecida en aquellos suelos que presenten rastrojos del cultivo de maíz. Por otro lado, el OTU 635 solo se reveló a nivel de la comunidad diazotrófica potencial.

Dentro del *subcluster* 1A se revelaron 7 OTUs relacionados principalmente al orden *Desulphuromonadales* y con bacterias no cultivables del suelo. Tanto a nivel de ADN como ARN se observó una mayor predominancia de los OTUs 1A en los suelos no cultivados y bajo rotación intensiva de cultivos (trigo/soja-maíz). En particular, los OTUs 17 y 31 estuvieron más representados en los suelos NC a nivel de ADN mientras que los OTUs 94, y 120 lo hicieron a nivel de ARN.

El único OTU abundante 3B correspondió al OTU 2, se relacionó con secuencias *nifH* de bacterias no cultivables y más lejanamente con el orden *Verrucomicrobiales*. Este OTU se encontró solo en *nifH*-ARN de los suelos cultivados y con mayor proporción en PPA. Dentro del *subcluster* 1B también se reveló un único OTU abundante, el OTU 1303, que presentó mayor homología con secuencias *nifH* de bacterias no cultivables y más lejanamente con el orden *Nostocales*. La mayor proporción de este OTU se registró en las malas prácticas de ambas comunidades diazotróficas.

Este análisis puso en evidencia que los grupos activos (ARN-*nifH*) más abundantes no fueron los más frecuentes en la comunidad diazotrófica potencial (ADN-*nifH*). Por ejemplo, el OTU 586 presentó alrededor de 100 secuencias *nifH* en ADN, mientras que en ARN un total de 2000 secuencias. Esto indicaría que miembros poco abundantes de la comunidad diazotrófica potencial, podrían ser ocasionalmente importantes en la fijación de N₂. Este análisis comparativo también reveló que la mayoría de los grupos activos (ARN-*nifH*) se encontraron representados en la comunidad diazotrófica potencial (ADN-*nifH*). La ausencia del OTU 2 (3B) a nivel ADN pudo deberse a una *sub* representación de dichas secuencias en la base de datos

ADN-*nifH*, y a cuestiones experimentales tales como los desvíos asociados a la reacción por PCR y especificidad de los cebadores.

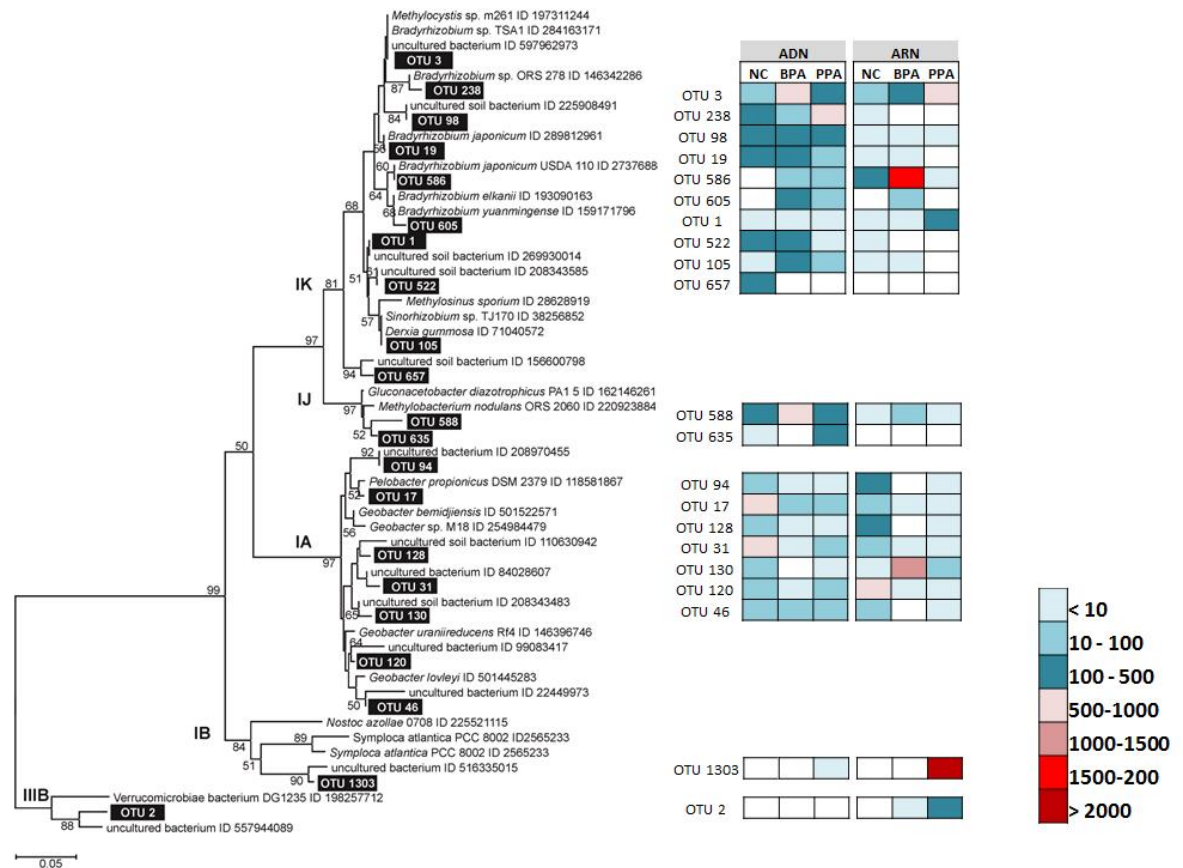


Figura V.17. Relación filogenética de las secuencias aminoacídicas NifH representativas de los OTUs abundantes presentes en ADN y/o ARN de los suelos de Pergamino. Estos OTUs abundantes (cajas negras) presentaron al menos 200 secuencias *nifH*. En el margen derecho de la figura se graficó el *heatmap* con la distribución relativa de cada OTU entre las distintas muestras de suelo. La gama de colores indica la abundancia en número de secuencias *nifH* en dicho OTU y la ausencia de secuencias se señala con el color blanco. El árbol de filogenia se construyó en el entorno ARB con secuencias aminoacídicas parciales de NifH usando el método *Neighbour-Joining*.

Cuantificación en suelo de filotipos de interés mediante qPCR del gen nifH

Se evaluó la abundancia de grupos de interés en el ADN extraído de las muestras de suelo utilizadas para el análisis de transcritos *nifH*, mediante el uso de qPCR y cebadores específicos de dos taxones filogenéticamente relacionados con *Geobacter* y cianobacterias del género *Nostoc*. Se seleccionaron el OTU 755 con clasificación taxonómica *subcluster* 1B y parentesco lejano con secuencias *nifH* del género *Nostoc*, y los 6 OTUs 1A (22, 8, 27, 380,488 y 53) relacionados con *Geobacter*. En la sección “qPCR de grupos específicos de *nifH*” del Capítulo II de Materiales y Métodos se ha descrito el diseño de los cebadores y las condiciones de la qPCR. En la Tabla V.4 se listan los cebadores diseñados en este estudio.

La especificidad de los cebadores se evaluó *in silico* mediante la comparación con genes *nifH* de la base de datos pública NCBI y experimentalmente analizando su especificidad en reacciones de amplificación usando ADN aislado de *Geobacter sp* y *Nostoc sp*. PCC 7120 y además la ausencia de reacción cruzada con ADN de bacterias diazotróficas no relacionadas filogenéticamente con *Nostoc* y *Geobacter*. Como se muestra en la Figura Suplementaria V.2, la amplificación de un fragmento del tamaño esperado solo se observó en los casos de los templados correspondientes a ADN de *Geobacter sp*. y *Nostoc sp*. PCC 7120 (controles positivos).

Tabla V.4. Cebadores utilizados en la PCR cuantitativa qPCR.

Cebador	Blanco	<i>Subcluster</i> <i>nifH</i>	Secuencia (5'-3')	Tamaño de amplificación (pb)
CyanoFw	OTU 755	1B	TGG TAT CAT CAC CGC CAT CAA CTT C	116
CyanoRev	OTU 755	1B	GAA TTG GCA TAG CGA AAC CAC CG	
Geo Fw	OTUs: 8,22, 27,53,380,488	1A	GCC AAR GCG CAG AAT ACG T	314*

* Tamaño esperado de amplificación con el cebador reverso *nifH2* [141]

Para determinar la abundancia relativa de las bacterias diazotróficas de los grupos *nifH* específicos de los *subclusters* 1B y 1A, los datos primarios resultantes de qPCR fueron normalizados con la abundancia del gen ribosomal ARNr 16S. Los niveles de ARNr 16S fueron comparables entre los tratamientos y profundidades de suelo evaluados.

No se detectaron en la muestra PPA 20 cm, niveles apreciables de *nifH* para el grupo 1B relacionado con *Nostoc*. El análisis de ANOVA simple considerando la variable tratamiento mostró que la muestra PPA 10 cm presentó una abundancia de niveles *nifH* significativamente mayor ($r^2=0.85$, $p=0.0002$) comparado con el resto de los tratamientos. Considerando la variable profundidad, hubo diferencias significativas de abundancia ($r^2=0.6$, $p=0.015$), y los mayores niveles de *nifH* en todos los tratamientos se revelaron en los primeros 10 cm. A partir del análisis de ANOVA de dos componentes, se observó que la interacción entre ambas variables (tratamiento y profundidad) fue significativa ($r^2=0.89$, $p<0.0001$), hallándose los mayores niveles en la superficie de los suelos manejados con pobres prácticas (Figura V.18 A).

Con respecto al grupo de diazótrofos 1A relacionado con *Geobacter*, no se detectaron niveles de *nifH* en las muestras PPA. En el análisis de ANOVA simple, se encontraron diferencias significativas de niveles *nifH*, entre NC y BPA ($r^2=0.29$, $p=0.03$), y cuando se consideró la profundidad, la mayor abundancia de *nifH* se registró para los 10-20 cm ($r^2=0.31$, $p=0.026$). La interacción entre profundidad y tratamiento fue significativa ($r^2=0.9$, $p=0.0001$) y BPA 20 cm presentó los mayores valores de abundancia *nifH* (Figura V.18 B).

A partir de los resultados obtenidos mediante el análisis de qPCR de taxones *nifH* específicos se observó que el OTU 1B (cianobacterias) revelado en el análisis de transcritos *nifH* solo en la superficie del suelo PPA, se halló a nivel potencial (ADN) en todos los tratamientos y profundidades de suelo, excepto en PPA 20 cm. La abundancia de este grupo fue mayor en los primeros 10 cm del suelo, observándose el mayor nivel en el manejo bajo monocultivo de soja (PPA). Estos datos sugieren que el desarrollo y adaptación de las cianobacterias fotosintéticas y aeróbicas, puede ser

favorecido en los primeros cm de suelo que presenta mayor exposición solar y disponibilidad de O₂.

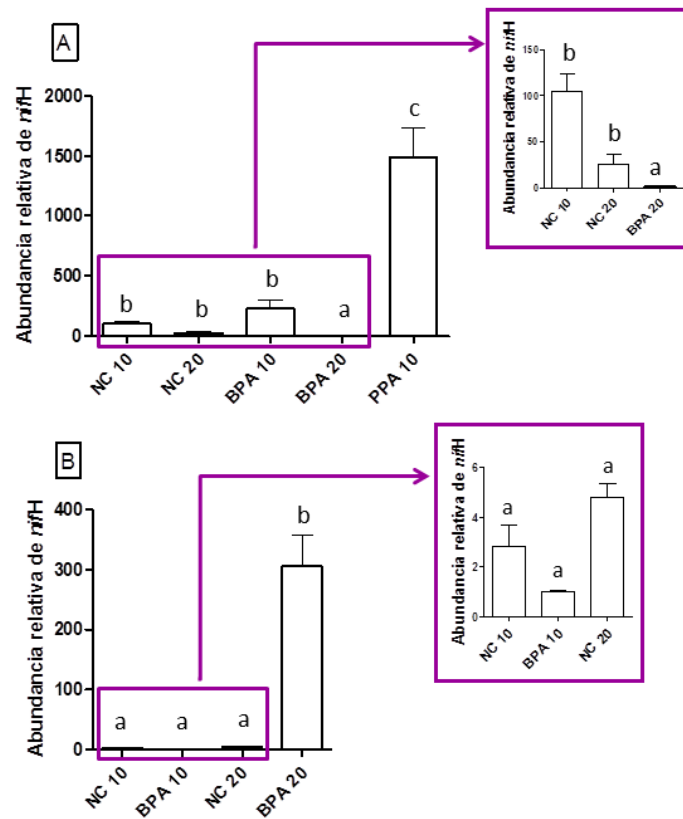


Figura V.18. Efecto del tratamiento y profundidad del suelo sobre la abundancia relativa de *nifH* de taxones específicos de *nifH*. En A se muestran la cuantificación del OTU *nifH* relacionado con *Nostoc* y en B para los OTUs relacionados con *Geobacter*. La abundancia de *nifH*, como unidades relativas normalizadas en base a la abundancia del gen ARNr 16S. Se aplicó la prueba de Tukey, las medias con distinta letra indican diferencias significativas ($p \leq 0.05$). Las abreviaturas NC, BPA y PPA corresponden a no cultivado, buenas prácticas agrícolas y pobres prácticas agrícolas, respectivamente.

Para evaluar si la mayor abundancia encontrada en el suelo PPA podría correlacionarse también con alguna característica físico-química particular de estos suelos, se realizó un análisis de componentes principales exploratorio con los valores de abundancia de los *subclusters nifH* y variables físico-químicas de suelo. Como se

muestra en la Figura Suplementaria V.3 el *subcluster* 1B se asoció con el parámetro físico del suelo *Dmean* (diámetro de la media de poros), la cual se corresponde con la distribución de poros. En el trabajo de tesis doctoral de F. Behrends Kraemer [319] se encontró que el suelo PPA de Pergamino presentó los mayores valores de *Dmean*, y se sugirió que estos valores podrían asociarse con poros de menor estabilidad. Con nuestros datos se podría inferir que la predominancia de este grupo diazotrófico 1B en la superficie de los suelos PPA se relaciona con poros de menor estabilidad. Ya ha sido sugerida la relación de cianobacterias con propiedades físicas del suelo. Por ejemplo, Yeager *et al.* 2004, analizando costras del suelo de la Meseta de Colorado y del desierto de Chihuahua (EEUU), concluyen e identifican una relación entre la inestabilidad del suelo, el estrés hídrico y la predominancia de *Nostoc* y *Scytonema* [129].

El ensayo de qPCR también mostró que el grupo diazotrófico relacionado con *Geobacter* se encuentra en niveles considerablemente mayores en la profundidad 10-20 cm de BPA. Si bien la mayor abundancia de *nifH* de este grupo específico relacionado a *Geobacter* se encontró asociado a los 20 cm de las BPA, en la superficie de NC y BPA también se detectaron niveles apreciables de *nifH*. Los miembros de *Geobacter* son capaces de utilizar la respiración anaeróbica facultativa, y esta propiedad metabólica posiblemente les otorgue cierta ventaja adaptativa permitiendo su distribución en diferentes profundidades del suelo con distintos grados de oxigenación. En el *biplot* de la Figura Suplementaria V.3 el *subcluster* 1A se asoció a las variables - carbono orgánico (OC del inglés *organic carbon*) y - capacidad de intercambio catiónico (CEC del inglés *cation exchange capacity*). Ha sido demostrado por Collavino *et al.* 2014 que la diversidad de la comunidad diazotrófica potencial en varios suelos argentinos es afectada significativamente por los niveles de carbono orgánico del suelo [121]. Podría especularse que los filotipos incluidos en el *subcluster* 1A, al disponer de un metabolismo heterotrófico sean altamente dependientes del uso de los compuestos orgánicos presentes en el suelo. La capacidad de intercambio catiónico resulta un buen indicador de la composición y del comportamiento del suelo. Es sabido que la adsorción de cationes aumenta cuando aumentan los contenidos de partículas finas, con el aumento de minerales de arcilla

2:1 [320] y con el contenido de materia orgánica del suelo [319]. Estos datos nos permiten pensar que la variable CEC puede relacionarse con *Geobacter*, teniendo en cuenta su capacidad para utilizar los óxidos de hierro insolubles como aceptores de electrones para la oxidación de la materia orgánica [233]. El posible rol protagónico de este filotipo 1A relacionado al orden *Desulfuromonadales* en la fijación libre de N₂ en la superficie y sub-superficie resulta sumamente interesante desde un punto de vista agronómico, ya que hemos encontrado que se trata un grupo de diazótrofos abundante y cuyos transcriptos *nifH* han sido detectados en suelo.

En un trabajo reciente de Wang *et al.* 2016 se analizó la variación temporal de la abundancia y estructura de la comunidad bacteriana total y diazotrófica en dos profundidades de suelo 0-20 cm y 20-40 cm, bajo cuatro regímenes de fertilización durante dos años. Observaron que la abundancia de los genes *nifH* y ARNr 16S disminuyó con el aumento de la profundidad, independientemente del régimen de fertilización [294]. Reardon *et al.* 2014 revelaron que la abundancia de bacterias diazotróficas en suelo disminuye exponencialmente a lo largo de las dos profundidades analizadas, 0-2 y 5-10 cm tanto en la rotación de cultivos trigo-barbecho como trigo-arveja. A su vez, no observaron efectos de la fertilización nitrogenada sobre la abundancia de las bacterias diazotróficas en la capa superficial, contrariamente la aplicación de fertilizantes está asociada a una reducción de la abundancia en la profundidad de 5-10 cm [321].

En líneas generales, los resultados de investigaciones sobre trabajos de abundancia de *nifH* tienden a mostrar que los niveles disminuyen en la mayor profundidad de muestreo y en algunos casos que también resultan afectados por la fertilización. En comparación con nuestros resultados, donde se analizó una subpoblación de la comunidad diazotrófica total, encontramos que algunos grupos presentan mayor abundancia en la sub-superficie, mientras que otros en la capa superficial. Los datos que hemos mostrado aquí se refieren al análisis del suelo de sólo una región, nos resulta obvio que para generalizar y adoptar como indicadores biológicos asociados a tipo y uso de los suelos, es necesario extender este análisis a otros suelos que comprendan distintos ambientes y características fisicoquímicas de los suelos. Aún hacen falta más estudios a escala regional/global para poder relacionar

la abundancia de filotipos *nifH* con distintos niveles de profundidad y manejo agrícola, siendo que hay muchas otras variables como el pH, el sitio, el momento de muestreo y rotación de cultivos que también afectan los niveles de *nifH* [110], [121], [294], [322], [323].

Conclusiones

Nos propusimos caracterizar la comunidad diazotrófica activa en suelos considerando dos variables, la profundidad y tratamientos agrícolas. La hipótesis de trabajo consistió primariamente en asumir que los diazótrofos de suelo expresan la nitrogenasa y que es demostrable a través del análisis del ARN extraído directamente del suelo; y en segundo lugar que el tipo de tratamiento agrícola y profundidad puede influir en el perfil de diazótrofos activos. La extracción satisfactoria de ARN del suelo en cuanto a la calidad y cantidad, y a su vez libre de ADN genómico, permitió obtener un *pool* de moléculas de ARN para su posterior análisis. Las etapas adicionales incorporadas durante la purificación y los controles experimentales demostraron que logramos un ARN libre de contaminación de ADN residual. La síntesis de ADNc *nifH* y posterior pirosecuenciación de amplicones *nifH* resultó en una base de datos ADNc-*nifH*, la cual una vez curada se utilizó para varios análisis bioinformáticos y estadísticos. El estudio se limitó a suelos de la región de Pergamino, muestreados solo una vez en el tiempo en un determinado momento, Junio del 2012. Las conclusiones que sacamos de este trabajo tienen que atender en cuanto a generalizaciones, a esta limitación de espacio y tiempo. Resulta importante destacar que la metodología aplicada es apropiada para evaluar la riqueza, diversidad y estructura de la comunidad diazotrófica activa en suelo.

Los diferentes diazótrofos constituyen un grupo que encierra un alto grado de diversidad, algunos de esos grupos se acomodan y distribuyen en las distintos sub-ambientes del suelo en cuanto al tratamiento y profundidad del mismo. Se encontró que la riqueza y diversidad de la comunidad diazotrófica es significativamente mayor en el suelo no cultivado (NC) comparado con los suelos cultivados. La profundidad también afectó significativamente la diversidad diazotrófica en forma dependiente del tratamiento agrícola. Cuando se comparó la diversidad de las poblaciones

diazotróficas, potencial y activa, no se hallaron diferencias significativas para los suelos no cultivados, contrariamente en los suelos cultivados bajo ambos manejos BPA y PPA, la población potencial (ADN-*nifH*) resultó ser más diversa. Estos resultados en su conjunto podrían estar indicando que la comunidad diazotrófica activa es más sensible a factores relacionados con el uso del suelo, y que aquellos grupos que logran transcribir *nif* en los suelos cultivados, podrían presentar mayor capacidad para competir y/o sobrevivir en las condiciones resultantes de aplicar la siembra directa. Los grupos taxonómicos *Cyanobacterias* y *Geobacter* encontrados en nuestro trabajo parecen ser ejemplos ilustrativos de este comportamiento.

La composición de la comunidad diazotrófica de los suelos no cultivados fue similar en ambas profundidades de suelo analizada (0-10 cm y 10-20 cm). Por el contrario, en los suelos cultivados se observó, además del efecto por el manejo, un cambio notable de la composición entre las profundidades. Dado que encontramos que la mayor disimilitud composicional se da entre los ambientes NC y PPA, podemos concluir que el monocultivo de soja (PPA) presentaría un efecto más notable sobre la composición diazotrófica en comparación con el manejo que incorpora la rotación de cultivos (BPA).

Las bacterias del suelo que llevan a cabo la FBN vida libre se distribuyen en diferentes grupos taxonómicos [52]. En la comunidad diazotrófica activa de Pergamino los OTUs *nifH* predominantes se asocian con los géneros *Geobacter* (1A), *Bradyrhizobium* (1K), *Methylocystis* (1K) *Nostocales* (1B) y *Verrucomicrobiales* (3B). Además, otros OTUs *nifH* también abundantes, se asocian filogenéticamente con secuencias *nifH* de bacterias no cultivables del suelo. Esto indica que además de revelar fijadores conocidos, también se revelan microorganismos aún no cultivados que representarían contribuciones significativas al balance de nitrógeno en los suelos, y probablemente microorganismos nuevos. La asociación de rizobios con los suelos cultivados, las cianobacterias con la superficie de los suelos PPA, y taxones relacionados con *Geobacter* en NC y BPA; cada uno de ellos con sus características fisiológicas propias y diferentes entre sí, indican que las bacterias diazotróficas encontrarían nichos ecológicamente diversos para activar la FBN, lo cual es otro indicador de la heterogeneidad del ambiente suelo.

La predominancia de los grupos de *Geobacter* y *rizobios* tanto en la población potencial como activa apunta a identificar rasgos de similitud entre ambas comunidades. Por otro lado, también resultó diferente la proporción de los órdenes *Cyanobacteria* y *Verrucomicrobia*, los cuales están representados pobremente en la comunidad potencial. Con qPCR y oligonucleótidos cebadores específicos para los dos grupos encontrados predominantes en el análisis de ARN, encaramos la evaluación cuantitativa de copias de *nifH* de dichos taxones en muestras de ADN de suelo. El análisis del taxón 1B relacionado lejanamente al orden *Nostocales* se encontró a nivel de ADN en mayor abundancia en la superficie del suelo de los tres tratamientos, pero la presencia de transcriptos *nifH* se limitó solo a la práctica de monocultivo de soja. Encontramos interesante la prevalencia de éste filotipo en suelos, ya que como ha sido sugerido, las cianobacterias desde el punto de vista agronómico no solo mejoran la estructura del suelo y la fertilidad, sino que también podrían favorecer el crecimiento de ciertos cultivos [324].

Una conclusión importante es que este análisis nos permite en futuros estudios dirigir de manera más precisas nuestra atención a algunos componentes de la comunidad de suelo, por ejemplo, cuál es el perfil de los mismos en suelos con diferentes propiedades fisicoquímicas, cómo varían en el curso del tiempo y distintas épocas del cultivo. Más aún, intentar el aislamiento en laboratorio para llevar a cabo una caracterización fisiológica, bioquímica y taxonómica más profunda. Sin dudas los resultados aquí presentados más los futuros, representarían una contribución original e importante a la biología del suelo.

Capítulo VI

Conclusiones generales y .perspectivas

La actividad del complejo agroindustrial de nuestro país tiene el potencial suficiente para abastecer el mercado interno y además exportar bienes y servicios de clase mundial. La producción primaria de origen agropecuario depende de un nivel de productividad creciente para permanecer competitiva en el futuro inmediato. Una de las limitantes más importantes de dicha producción es sostener la fertilidad de los suelos, siendo la disponibilidad de N una de las determinantes de la fertilidad. La consideración del relativo alto costo económico del uso de fertilizantes químicos y su impacto ecológico, pone a la FBN como una alternativa con un rol clave en los sistemas productivos y sustentables en el suelo. El estudio de las comunidades diazotróficas del suelo aplicando métodos independientes del cultivo en laboratorio, permite además de obtener una visión amplia de la diversidad de esa comunidad bacteriana proveer información útil para intentar el desarrollo de políticas ambientales que preserven y aseguren la explotación sustentable de los ambientes naturales y agrícolas.

En este trabajo de tesis se obtuvo información relevante acerca de la diversidad y la estructura de la comunidad diazotrófica en suelos con distintos manejos agronómicos, localizados en la principal región agrícola productiva de la Argentina, la denominada Región Pampeana. El estudio de la comunidad fijadora de N_2 libre comprendió diferentes enfoques. En uno de ellos se aplicó un criterio de análisis metagenómico, en el cual se encaró la secuenciación masiva al azar del ADN total del suelo. Estos datos nos fueron útiles para intentar un examen *in silico* del metagenoma del suelo, haciendo hincapié en la identificación de componentes asociados al ciclo del nitrógeno. Los resultados de este análisis nos mostraron que el suelo presenta capacidades metabólicas ubicuas en el ambiente en una escala global, y que ciertas categorías funcionales jerárquicas, como el metabolismo de carbohidratos, el metabolismo de proteína, y los mecanismos de virulencia y defensa, se hallan ampliamente representados en suelo comparado con otros ambientes (ej. acuáticos). Con respecto al metabolismo de N, hallamos que las respectivas representaciones de las comunidades de FBN y de nitrificación son bajas. Esta observación contrasta con los resultados obtenidos al relevar las comunidades fijadoras de nitrógeno a través de la secuenciación profunda de amplicones *nifH*. Dado que está aceptado que los

procesos mencionados son fundamentales en la incorporación y transformación del N del suelo, es razonable inferir que esa baja representación podría adjudicarse a la limitada cobertura de secuenciación de los metagenomas. Esos resultados son coincidentes con los de otros autores, los cuales fueron publicados recientemente durante o después de encarados nuestros estudios [170]. Estas características del nivel de cobertura nos limitaron en el alcance y la capacidad de esta estrategia para revelar un número apreciable de secuencias asignadas a dichos procesos y consecuentemente identificar *in silico* a componentes del ciclo de nitrógeno. Sin embargo, resultó apropiado para el estudio de las comunidades involucradas en la reducción asimilativa de nitrato y en la desnitrificación, lo cual reflejaría su predominancia entre las comunidades del ciclo de N en los suelos analizados de la región pampeana. Se encontraron diversos filotipos asociados a las diferentes etapas del ciclo del N. Algunos de estos corresponden a filo ampliamente distribuidos en suelo como *Proteobacteria*, *Firmicutes*, *Acidobacteria* y *Actinobacteria*, mientras que otros como *Deinococcus thermus* y *Aquificae* comprenden bacterias que suelen habitar ambientes con condiciones de temperatura más extremas. Además, encontramos al dominio *Archaea* y un grupo de bacterias no cultivables representados en nuestros datos. Estas observaciones reflejan el amplio rango de microorganismos que potencialmente podrían estar involucrados en los diferentes procesos metabólicos del ciclo del N en suelo.

La disponibilidad de herramientas con un grado de baja complejidad experimental es deseable para evaluar las distintas comunidades microbianas presentes en los suelos. Encontramos aplicando la técnica de tipo huella digital T-RFLP, que es posible revelar diferencias en la estructura diazotrófica potencial (ADN) entre suelos agrícolas cultivados y no cultivados de seis localidades geográficas. Los resultados obtenidos indican que los dos factores evaluados -manejo agronómico y ambiente- presentaron un efecto notable en los perfiles T-RFLP de la comunidad diazotrófica. En algunos suelos la composición de la comunidad diazotrófica parece estar afectada mayormente por el uso agrícola del suelo (S1 a S4), mientras que en otros (S5 y S6) el principal factor efector fue la localización geográfica. La técnica no alcanza a revelar cambios en los perfiles T-RFLP de suelos bajo siembra directa con

las dos diferentes prácticas analizadas, consistentes esencialmente en el monocultivo sin reposición de nutrientes (PPA) *versus* la rotación intensiva con fertilización balanceada (BPA). Los resultados indican que son más simples de identificar las diferencias entre efectos por la actividad agrícola (suelos cultivados vs. suelos no cultivados) que las diferencias atribuibles a las distintas calidades de uso y manejo del suelo en BPA y PPA. Es interesante concluir que la generación de perfiles T-RFLP, propios y característicos de un tipo de suelo y del uso agrícola, ofrece un camino para evaluar en forma comparativa los mismos, y la dinámica de las comunidades diazotróficas en el curso del tiempo. Este tipo de análisis permitiría integrar un indicador biológico a los ya disponible parámetros fisicoquímicos usados para caracterizar a los suelos. Por ejemplo, la distribución diferencial de algunos componentes diazotróficos revelados como distintos T-RFs entre usos, manejos y sitios de los suelos, los cuales además presentaron una proporción considerable dentro de la comunidad diazotrófica, podrían brindar información útil para el desarrollo de indicadores biológicos simples y eficientes de calidad de suelo y/o utilizarse para el monitoreo de cambios en la estructura poblacional diazotrófica en una escala de tiempo o tratamientos experimentales.

Basado en la pirosecuenciación de secuencias *nifH*, obtenidas a partir de la amplificación de transcritos *nif*, encaramos la tarea de identificar a los grupos taxonómicos candidatos activos en la fijación de N₂ en vida libre. Se analizaron suelos de Pergamino con distinto manejo agronómico y profundidad. La diversidad diazotrófica es significativamente mayor en el suelo no cultivado (NC) comparado con los suelos cultivados y, a su vez fue afectada por la profundidad en forma dependiente del tratamiento agrícola. El efecto de la profundidad fue principalmente notable en la composición diazotrófica de los suelos cultivados, ya que en los suelos no cultivados la composición fue semejante en ambas profundidades. Además, la mayor disimilitud composicional se reveló entre los suelos no cultivados y los suelos bajo monocultivo. Estos datos indican que tanto el uso como la profundidad del suelo, determinan en cierta medida cuales diazótrofos dentro de la amplia diversidad, son activados para realizar fijación de nitrógeno en vida libre y también ponen de manifiesto que el monocultivo de soja parece tener un efecto importante sobre la

estructura de la comunidad diazotrófica, comparado con el tratamiento que comprende a la rotación de cultivos. Las características de manejo en las prácticas pobres, o algunas de las mismas, forzarían al amplio espectro de diazótrofos presentes en el suelo, a una redistribución diferencial en las capas de distinta profundidad de los respectivos suelos respecto a la de los ambientes no perturbados.

Encontramos interesante en este examen demostrar la predominancia de los filotipos *nifH* relacionados con los órdenes *Desulfuromonadales*, *Nostocales* y *Rhizobiales* en la comunidad diazotrófica activa. Por otro lado, reforzamos las evidencias de presencia de bacterias no cultivables y aún no caracterizadas, como miembros activos en la fijación en vida libre de N₂. Estas observaciones son una demostración de la potencialidad del estudio a gran escala de un gen funcional de interés, mediante pirosecuenciación. Se revelaron distintos perfiles de distribución de grupos predominantes entre los distintos suelos examinados. Así, las secuencias del orden *Rhizobiales* se asocian con los suelos cultivados, mientras que las cianobacterias se encuentran principalmente en la capa superficial de los suelos con monocultivo, y por el contrario, los taxones relacionados con *Geobacter* se asocian a los suelos no cultivados o con rotación de cultivos. Estas bacterias diazotróficas caracterizadas por presentar diferentes fisiologías encontrarían en el contexto de la heterogeneidad estructural del suelo, micro-nichos ecológicamente diversos en los cuales cada grupo encuentra condiciones apropiadas para activar la fijación biológica de nitrógeno. Coincidentemente, *Geobacter* y los rizobios fueron encontrados como grupos predominantes en los mismos manejos de suelo en la comunidad diazotrófica potencial (ADN) indicando que estos grupos son importantes componentes del proceso a nivel potencial y activo. En contraste, los órdenes *Cyanobacteria* y *Verrucomicrobia* pobremente representados en la comunidad potencial son grupos predominantes de la comunidad activa de los suelos.

Fue reportado por nuestro grupo de trabajo, que la diversidad de la comunidad diazotrófica potencial en los suelos de Pergamino es similar entre tratamientos. Interesantemente, en los suelos no perturbados hallamos que la diversidad diazotrófica fue similar entre la comunidad potencial y activa, mientras que para los suelos cultivados se encontraron valores significativamente mayores en la comunidad

potencial. Los resultados tanto del análisis del ADN extraído del suelo como los aquí presentados sobre el análisis del ARN, indican una alta redundancia funcional en fijación biológica de nitrógeno en los suelos no disturbados. En relación a los suelos agrícolas, la menor diversidad funcional puede ser producto de las perturbaciones propias del uso, menor diversidad de plantas presentes, agregado de fertilizantes, inoculación con fijadores simbióticos, aplicación de agroquímicos etc., las cuales parecen afectar la actividad de varios de los grupos diazotróficos presentes en los suelos cultivados.

La cuantificación por qPCR de grupos específicos relacionados con secuencias de *Geobacter* y *Nostoc*, mostró que los niveles de copias de *nifH*-ADN fueron significativamente más abundantes en aquellos suelos donde también se reveló el mayor número de transcritos *nifH*. Existen investigaciones previas, basadas en distintas metodologías de estudio, que demuestran la presencia y el protagonismo de secuencias *nifH* relacionadas con *Geobacter* en el proceso de fijación de N₂ en suelos agrícolas [83], [121], [233], [296]. Se ha propuesto que sus capacidades metabólicas le permiten adaptarse a condiciones cambiantes, de disponibilidad de nutrientes o disposición de O₂, lo que resultaría en su prevalencia con respecto a otros grupos diazotróficos. Asimismo, hallamos secuencias relacionadas con *Nostoc* mayormente asociadas a la superficie y predominantes en el monocultivo de soja. Hemos encontrado que el mayor nivel de copias de este filotipo se relacionó principalmente con parámetros físicos relacionados con menor estabilidad de suelo. En trabajos previos se ha encontrado que la mayor abundancia de *Nostoc* en líquenes de suelos áridos se encuentra principalmente relacionada con factores como inestabilidad del suelo y estrés hídrico [129].

Los datos primarios obtenidos con la pirosecuenciación del gen *nifH*, fueron apropiados para estimar la diversidad, describir la composición de la comunidad activa, y compararlos con los datos obtenidos previamente sobre la estructura de la comunidad diazotrófica potencial. Esta información que encontramos auspiciosa invita a extender los estudios con preguntas más precisas sobre algunos grupos, ya que demostramos que algunos diazótrofos son interesantes, en sus perfiles distintivos de presencia y actividad en las diferentes condiciones analizadas. Entendemos que

después de la evaluación de un número mayor de diferentes suelos y condiciones de uso, podrían resultar validadas esas características distintivas para su ulterior adopción como indicadores biológicos de tipo y manejo de suelo. Este trabajo nos permite avanzar y contribuir con la demostración y documentación de los efectos del ambiente sobre la diversidad de comunidades diazotroficas e identificar asociaciones con algunas características edáficas y tipo de manejo agrícola. Esto abre nuevas avenidas para investigaciones sobre la biología del suelo, en particular con procesos biogeoquímicos claves para el estado nutricional y fertilidad de los suelos como lo representa la fijación biológica de nitrógeno. Resulta esencial aumentar el número de muestras analizadas, y extender estos análisis a otros tipos de suelos y regiones para así disponer de una cobertura más amplia de la ocurrencia de estos filotipos *nifH* de particular interés.

Combinando los resultados descriptos en el trabajo anterior de nuestro grupo de investigación [121] y los aquí presentados, podemos concluir que las características edáficas y el uso agronómico afectan significativamente la estructura de la comunidad diazotrófica y que la distribución de algunos grupos filogenéticos se encuentra asociada a características particulares del suelo, abriendo así la posibilidad de su aplicación como marcadores de calidad. La realización de nuevas investigaciones para validar estas observaciones es un objetivo de los próximos nuevos proyectos de investigación.

Material Suplementario

Capítulo III

Subsystem Spreadsheet								
Organism	Variant	HutH	HutU	HutI	GluF	HutG	NfoD	ForI
<i>Bacteroides thetaiotaomicron</i>	1	Q8A4B3	Q8A4A9	Q8A4B1	Q8A4B0			
<i>Desulfotela psychrophila</i>	1	g51246205	g51246204	g51246203	g51246202			
<i>Halobacterium sp.</i>	2	Q9HQD5	Q9HQD8	Q9HQD6		Q9HQD7		
<i>Deinococcus radiodurans</i>	2	Q9RZ06	Q9RZ07	Q9RZ05		Q9RZ04		
<i>Bacillus subtilis</i>	2	P10944	P25503	P42084		P42088		
<i>Caulobacter crescentus</i>	3	P58062	Q9A9M1	P58079			Q9A9M0	Q9A9L9
<i>Pseudomonas putida</i>	3	Q88CZ7	Q88CZ6	Q88CZ9			Q88D00	Q88CZ3
<i>Xanthomonas campestris</i>	3	Q8PAA7	P58988	Q8PAA6			Q8PAA8	Q8PAA5
<i>Listeria monocytogenes</i>	-1							

Figura Suplementaria III.1. Hoja tipo de cálculo representando al subsistema del proceso metabólico degradación de histidina. El subsistema se halla poblado por 9 organismos, donde cada fila representa un organismo y cada columna un rol funcional del subsistema “degradación de histidina”. Los genes que llevan a cabo los roles funcionales específicos en el respectivo organismo “pueblan” la celda respectiva. El sombreado gris en las celdas indica cercanía de los genes en el cromosoma. Existen tres variedades de la degradación de histidina presente en este subsistema poblado, todas las variantes utilizan las mismas tres enzimas en los primeros tres pasos de la conversión de histidina. La variante 1 (sombreada en celeste) está presente en *Bacteroides thetaiotaomicron* y *Desulfotela psychrophila* y aquí la glutamato formiminotransferasa (EC 2,1,2,5, GluF) realiza la conversión de formino glutamato a glutamato. La variante 2 (sombreado en amarillo) está presente en *Halobacterium sp.*, *Deinococcus radiodurans* y *Bacillus subtilis*, aquí la conversión se lleva a cabo por la formiminoglutamasa (EC 3,5,3,8, HutG). La variante 3 (sombreada en verde) está presente en *Caulobacter crescentus*, *Pseudomonas putida*, y *Xanthomonas campestris*. Aquí, el formimino glutamato se convierte a glutamato por la vía N-Formyl-L-Glutamato por dos enzimas: Formiminoglutámico iminohidrolasa (EC 3,5,3,13, ForI) y N-Formylglutamato deformylasa (EC 3,5,1,68, NfoD).

Tabla Suplementaria III.1. Metagenomas de suelos estudiados en otras regiones del mundo

Plataforma de secuenciación 454 GS-FLX Titanium Roche

1. Estación Experimental de Embrapa Soja (Londrina, Brasil)

Tratamiento/Características	N° de lecturas	Largo de lecturas (promedio)
Labranza convencional con rotación de cultivos	1,050,712	304
Labranza convencional con sucesión de cultivos	1,080,923	307
Siembra directa con rotación de cultivos	913,435	306
Siembra directa con sucesión de cultivos 4 réplicas X tratamiento (16 muestras de suelo)	1,034,153	311

Referencia Souza et al., 2013 (*Soil metagenomics reveals differences under conventional and no-tillage with crop rotation or succession*)

2. Cedar Creek Ecosystem Science Reserve y Kellog Biological Station (Estados Unidos)

Tratamiento/Características	N° de lecturas	Largo de lecturas (promedio)
3 niveles de Nitrógeno (N) X 2 sitios X 3 réplicas por sitio y nivel de N (18 muestras de suelo)	1,350,000	384

Referencia Fierer et al., 2012 (*Comparative metagenomic, phylogenetic and physiological analysis of soil microbial communities across nitrogen gradients*)

3. Rothamsted Research (Park Grass) Site (Reino Unido)

Tratamiento/Características	N° de lecturas	Largo de lecturas (promedio)
Distinta profundidad de suelo, escala temporal y protocolo de extracción (13 muestras de suelo)	12,575,129	358 ± 31,8 bp

Referencia Delmont et al., 2012 (*Structure fluctuation and magnitude of a natural grassland soil metagenome*)

Tabla Suplementaria III.2. Distribución relativa (en porcentaje del total de lecturas anotadas) de los 28 subsistemas metabólicos jerárquicos en los metagenomas de suelos argentinos.

Subsistema Jerárquico	Promedio (%)
<i>Clustering-based subsystems</i>	15,48 ± 0,07
<i>Carbohydrates</i>	11 ± 0,11
<i>Amino Acids and Derivatives</i>	8,93 ± 0,11
<i>Miscellaneous</i>	8,39 ± 0,09
<i>Protein Metabolism</i>	7,66 ± 0,11
<i>Cofactors Vitamins Prosthetic Groups Pigments</i>	6,68 ± 0,05
<i>DNA Metabolism</i>	3,88 ± 0,08
<i>RNA Metabolism</i>	3,81 ± 0,04
<i>Respiration</i>	3,77 ± 0,10
<i>Cell Wall and Capsule</i>	3,39 ± 0,09
<i>Fatty Acids Lipids and Isoprenoids</i>	3,16 ± 0,05
<i>Nucleosides and Nucleotides</i>	3,11 ± 0,07
<i>Membrane Transport</i>	2,98 ± 0,10
<i>Virulence Disease and Defense</i>	2,77 ± 0,11
<i>Stress Response</i>	2,69 ± 0,04
<i>Metabolism of Aromatic Compounds</i>	2,05 ± 0,09
<i>Regulation and Cell signaling</i>	1,57 ± 0,06
<i>Phages Prophages Transposable elements Plasmids</i>	1,38 ± 0,06
<i>Sulfur Metabolism</i>	1,37 ± 0,03
<i>Cell Division and Cell Cycle</i>	1,35 ± 0,03
<i>Nitrogen Metabolism</i>	1,09 ± 0,04

<i>Phosphorus Metabolism</i>	0,89 ± 0,03
<i>Motility and Chemotaxis</i>	0,85 ± 0,05
<i>Secondary Metabolism</i>	0,55 ± 0,04
<i>Iron acquisition and metabolism</i>	0,49 ± 0,05
<i>Potassium metabolism</i>	0,34 ± 0,03
<i>Dormancy and Sporulation</i>	0,20 ± 0,01
<i>Photosynthesis</i>	0,15 ± 0,02

Tabla Suplementaria III.3. Referencia de los subsistemas en el PcoA

Subsistema Jerárquico	Número de referencia
<i>Amino Acids and Derivatives</i>	1
<i>Carbohydrates</i>	2
<i>Cell Division and Cell Cycle</i>	3
<i>Cell Wall and Capsule</i>	4
<i>Clustering-based subsystems</i>	5
<i>Cofactors, Vitamins, Prosthetic Groups, Pigments</i>	6
<i>DNA Metabolism</i>	7
<i>Dormancy and Sporulation</i>	8
<i>Fatty Acids, Lipids, and Isoprenoids</i>	9
<i>Iron acquisition and metabolism</i>	10
<i>Membrane Transport</i>	11
<i>Metabolism of Aromatic Compounds</i>	12
<i>Miscellaneous</i>	13
<i>Motility and Chemotaxis</i>	14
<i>Nitrogen Metabolism</i>	15
<i>Nucleosides and Nucleotides</i>	16
<i>Phages, Prophages, Transposable elements, Plasmids</i>	17
<i>Phosphorus Metabolism</i>	18
<i>Photosynthesis</i>	19
<i>Potassium metabolism</i>	20
<i>Protein Metabolism</i>	21
<i>RNA Metabolism</i>	22
<i>Regulation and Cell signaling</i>	23
<i>Respiration</i>	24
<i>Secondary Metabolism</i>	25
<i>Stress Response</i>	26
<i>Sulfur Metabolism</i>	27
<i>Virulence, Disease and Defense</i>	28

Capítulo IV

Tabla Suplementaria IV.1. Rendimiento y pureza de ADN de suelo. Las mediciones se realizaron con el espectrofotómetro Nanodrop.

Abreviatura	Rendimiento ($\mu\text{g ADN-1 suelo}$)	$A_{260/280}$	$A_{260/230}$
S1 CSD	1.9	2.1	1.6
S1 NC	1.6	1.8	1.0
S2 CSD	2.0	2.2	1.8
S2 NC	1.8	1.9	0.9
S3 CSD	1.5	1.9	1.5
S3 NC	1.6	2.0	1.2
S4 CSD	1.3	2.1	1.0
S4 LC	1.7	1.8	0.9
S4 NC	1.6	1.9	1.5
S5 CSD BPA	1.9	1.8	1.2
S5 CSD PPA	2.0	1.5	1.6
S5 NC	1.9	1.6	1.3
S6 CSD BPA	1.8	1.9	1.5
S6 CSD PPA	1.5	1.7	0.9
S6 NC	1.7	1.9	1.8

Tabla Suplementaria IV.2. Reproducibilidad de la etapa de PCR. En a y b se muestran los valores de la altura de los picos obtenidos a partir de la digestión con *MspI* de dos amplicones *nifH*.

Muestra	Tamaño (pb)	Altura	
		a	b
S1 CSD	69	118	
	71	7800	8820
	139		54
	140		82
	146	3200	3248
	154	350	387
	160	502	654
	161	650	745
S2 NC	69	1541	1055
	72	305	270
	109	117	151
	111	108	120
	115	621	527
	140	102	161
	147	2004	3083
	156	1267	2786
S3 NC	69	640	680
	160	590	720
	162	590	720
S3 NC	71	1658	1869
	114	197	140

	145	360	882
	147	3179	4188
	156	227	250
	162	664	1064
S4 LC	69	81	
	71	2600	2800
	108	161	
	115	224	149
	144	247	350
	147	2879	3000
	154	249	
	157	455	344
	159	349	351
	161	335	379
	162	132	128
S4 NC	71	96	397
	147	8463	8426
	154	97	
	160	286	338
	162	930	1016
	177	464	404

Tabla Suplementaria IV.3. Reproducibilidad de la etapa de PCR. En a y b se muestran los valores de la altura de los picos obtenidos a partir de la digestión con *HaeIII* de dos amplicones *nifH*.

Muestra	Tamaño (pb)	Altura	
		a	b
S1 CSD	45	153	120
	61	949	798
	62	729	624
	64	574	479
	146	217	170
	148	89	75
	154	6814	5472
	178	1007	767
	179	1654	1271
	181	502	359
	182	671	494
	312	111	87
	314	105	84
	316	400	326
	S2 NC	62	185
66		68	
69		131	
70		64	

	71	51	
	154	4821	3898
	178	337	321
	179	511	403
	197	1393	1289
S3 NC	62	304	230
	64	234	187
	70	167	
	147	126	
	152	328	271
	154	227	352
	169	625	681
	179	570	359
	198	170	78
	313	270	196
	315	2004	1396
	356	201	111
S4 LC	62	1008	902
	63	137	
	153	1912	1701
	155	102	
	171	69	
	177	915	727
	180	1560	1247
	314	842	556
S4 NC	62	249	
	154	2013	1780
	169	76	
	178	1025	612
	180	886	468
	197	759	664
	315	2900	2450
	356	564	415

Tabla Suplementaria IV.4. Reproducibilidad de la etapa de digestión con *MspI* o *HaeIII*. En a y b se muestran los valores de la altura de los picos correspondientes a los duplicados del producto de digestión.

<i>MspI</i>		Altura		<i>HaeIII</i>		Altura	
Muestra	Tamaño (pb)	a	b	Muestra	Tamaño (pb)	a	b
S3 BPA	70	472	492	S3 BPA	62	1163	1118
	71	3912	4027		64	4010	4069
	147	1395	1439		70	234	209
	154	2453	2472		154	6236	6366
	159	188	188		175	594	552
	161	432	435		180	2684	2400
S4 NC	70	120	144	199	187	166	
	147	8629	9072	313	168	147	
	154	101	114	315	1355	1302	
	161	730	726	355	370	324	
	177	150	180	S4 NC	62	730	642
					178	2771	2889
			180		1625	1605	
			197		999	995	
			315	9450	9183		
			356	744	765		

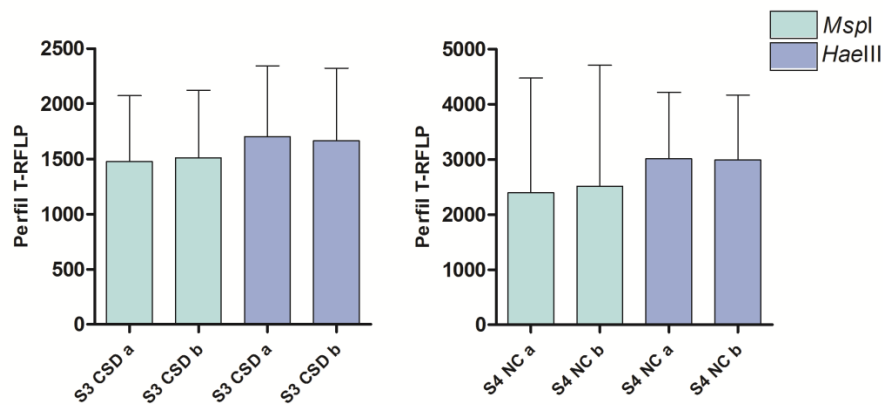


Figura Suplementaria IV.1. Reproducibilidad de la digestión enzimática. Se muestran en distinto color los perfiles T-RFLP de las muestras de suelos S3 CSD y S4 2 NC que se obtuvieron a partir de incubación de los amplicones *nifH* con las endonucleasas *MspI* y *HaeIII* generando los duplicados de digestión a y b. En el eje Y se representa el valor de fluorescencia correspondiente a la integración de todos los picos del electroferograma.

Tabla Suplementaria IV.5. Perfil T-RFLP de las 15 muestras de suelo analizadas. Se muestra la proporción relativa (%) del área de fluorescencia de todos los T-RFs generados a partir de la incubación de los amplicones *nifH* con las enzimas *MspI* y *HaeIII*.

T-RFs	S1 CSD	S1 NC	S2 CSD	S2 NC	S3 CSD	S3 NC	S4 CSD	S4 LC	S4 NC	S5 CSD BPA	S5 CSD PPA	S5 NC	S6 CSD BPA	S6 CSD PPA	S6 NC
69_Mspl				13											
70_Mspl	36	5	27		30	13	6	16	1			4			
71_Mspl										8	6		26	34	23
104_Mspl															
107_Mspl														3	
109_Mspl															2
113_Mspl															3
114_Mspl						1				2	1	1	6	2	
115_Mspl				3				2							
139_Mspl													2		
140_Mspl															
144_Mspl								2							
145_Mspl															7
146_Mspl	9		7		18		43	25							
147_Mspl		37		14		25			43	30	30	40	13	9	6
150_Mspl															
154_Mspl	2				1			2		4	2				
156_Mspl			6	7		3									
157_Mspl								3							
158_Mspl															7
159_Mspl								1							
160_Mspl			3	5					1						2
161_Mspl	4	7	6	4	2	6	2	1	4	5	12	4	2	2	
177_Mspl						1	0	0	2						
338_Mspl		2	2												
37_HaeIII													1	2	4
40_HaeIII		7											7	3	
41_HaeIII															8
42_HaeIII			15				2								
56_HaeIII															2
61_HaeIII													1	3	1
62_HaeIII	4	2		9	2	4	3	8	1						
63_HaeIII										3	4	7			
64_HaeIII	2				8										
65_HaeIII													15	10	23
70_HaeIII															
71_HaeIII														2	2

78_HaeIII							1							
94_HaeIII	2									2	6	6		
146_HaeIII														
147_HaeIII			1											
148_HaeIII										3				
151_HaeIII								3						
152_HaeIII						6								5
153_HaeIII								18						
154_HaeIII	34	27	19	16	32	9	16		10	15	19	18	16	23
169_HaeIII						2								
174_HaeIII														
175_HaeIII														2
178_HaeIII		4	2	2		7	2	5	5	12	5			1
180_HaeIII	2	6			1			10	4		1		6	3
181_HaeIII			2				11					7		
182_HaeIII	1									3				
197_HaeIII		2	5	18			1		3					
198_HaeIII												4		
199_HaeIII														
313_HaeIII					1	1		5						
315_HaeIII	2	3	1		3	17	7		27	7	10	7		
350_HaeIII										2				
354_HaeIII												2		
355_HaeIII				4	1					4	6		2	3

Tabla Suplementaria IV.6. Propiedades químicas de los suelos.

Abreviatura	pH	Humedad	Materia Orgánica	N total
S1 CSD	6.06	22.00	2.60	0.19
S1 NC	5.40	26.10	6.20	0.35
S2 CSD	5.60	23.70	2.70	0.18
S2 NC	6.21	26.70	5.80	0.31
S3 CSD	5.38	21.30	2.90	0.21
S3 NC	5.90	21.80	3.60	0.23
S4 CSD	5.19	26.25	5.28	0.24
S4 LC	5.15	30.54	5.58	0.22
S4 NC	5.63	35.0	5.60	0.34
S5 CSD BPA	5.60	19.5	1.77	0.16
S5 CSD PPA	6.00	20.0	1.91	0.14
S5 NC	5.80	26.7	2.99	0.25
S6 CSD BPA	6.50	19.8	3.58	0.30
S6 CSD PPA	5.90	9.9	2.65	0.20
S6 NC	6.10	10.6	3.57	0.27

Capítulo V

Tabla Suplementa V.1. Significancia del agrupamiento de las comunidades diazotróficas según la prueba de significancia *Weighted normalized UniFrac*.

Muestra 1	Muestra 2	p valor
BPA 10	BPA 20	0.082
BPA 10	NC 10	0.003
BPA 10	NC 20	0.007
BPA 10	PPA 10	0.0
BPA 10	PPA 20	0.268
BPA 20	NC 10	1.0
BPA 20	NC 20	1.0
BPA 20	PPA 10	0.0
BPA 20	PPA 20	0.053
NC 10	NC 20	0.953
NC 10	PPA 10	0.003
NC 10	PPA 20	0.0
NC 20	PPA 10	0.0
NC 20	PPA 20	0.0
PPA 10	PPA 20	0.0

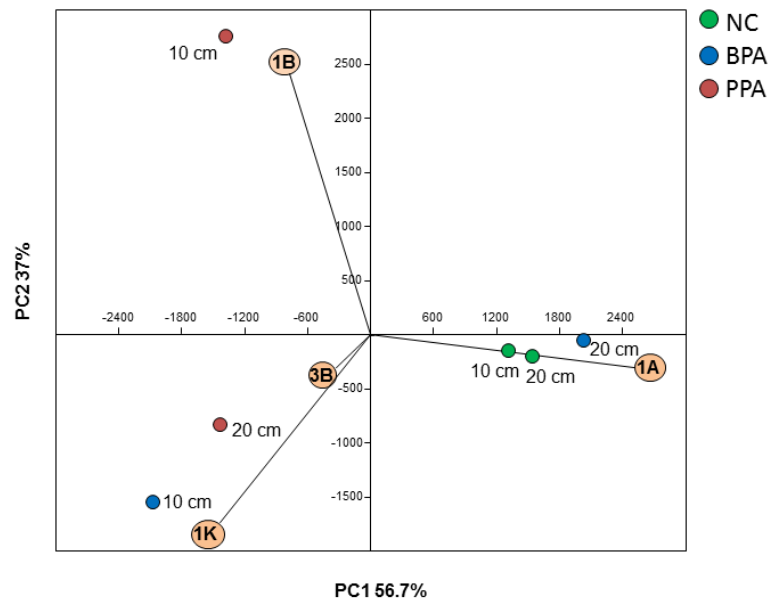


Figura Suplementaria V.1. Representación *biplot* del análisis de componentes principales basado en la abundancia de los *subclusters nifH*. Solo se muestran los *subclusters* con una representación mayor al 1% en la base de datos ADNc-*nifH*. Los círculos coloreados en azul y bordo corresponden a los distintos manNCjos agrícolas y en verde al suelo control.

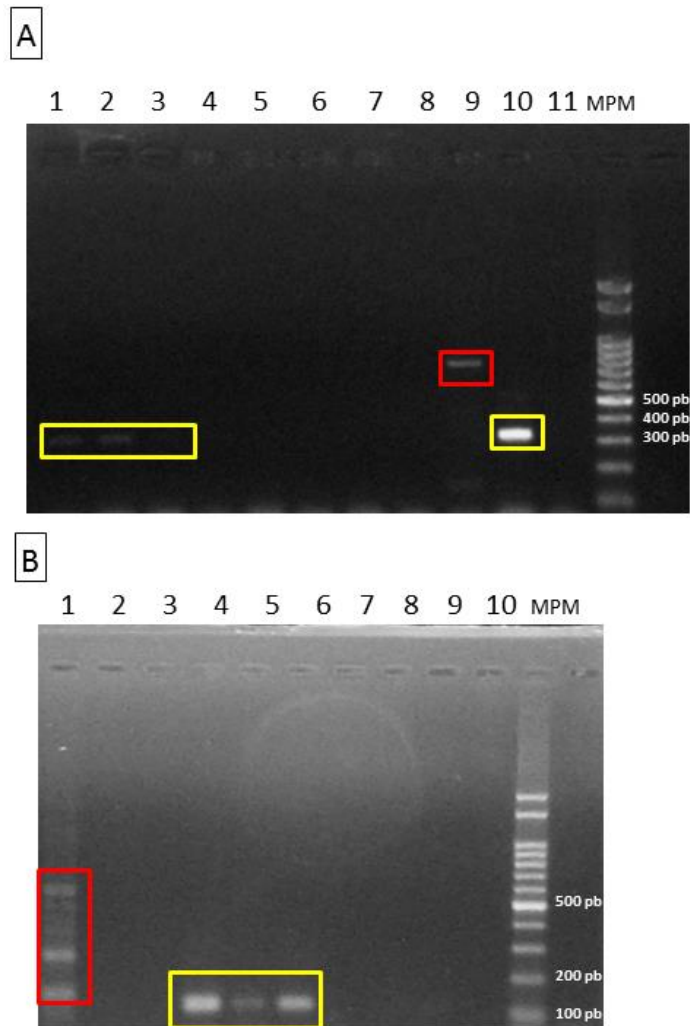


Figura Suplementaria V.2. Productos resultantes de la amplificación por PCR del gen *nifH* de filotipos de interés a partir de ADN metagenómico de suelo y ADN de bacterias diazotróficas. (A) PCR con cebadores para OTUs relacionados a *Geobacter*. Se remarca en amarillo el amplicón esperado de 314 pb, y en rojo un amplicón inespecífico. Calles: 1 NC 10 cm; 2 NC 20 cm; 3 BPA 10 cm; 4 *Rhizobium sp.*; 5 *Herbaspirillum sp.*; 6 *Burkholderia sp.*; 7 *Nostoc sp.*; 8 *Enterobacter sp.*; 9 *Azotobacter sp.*; 10 *Geobacter sp.*; 11 Control negativo. **(B)** PCR con cebadores para OTUs relacionado a *Nostoc*. Se remarca en amarillo el amplicón esperado de 116 pb, y en rojo amplicones inespecíficos. Calles: 1 *Rhizobium sp.*; 2 *Azotobacter sp.*; 3 *Burkholderia sp.*; 4 *Nostoc sp.*; 5 NC 10 cm; 6 PAP 10 cm; 7 *Enterobacter sp.*; 8 *Geobacter sp.*; 9 *Herbaspirillum sp.*; 10 Control negativo. MPM corresponde a marcador de peso molecular de 100pb (PB-L Productos Bio-Logicos®).

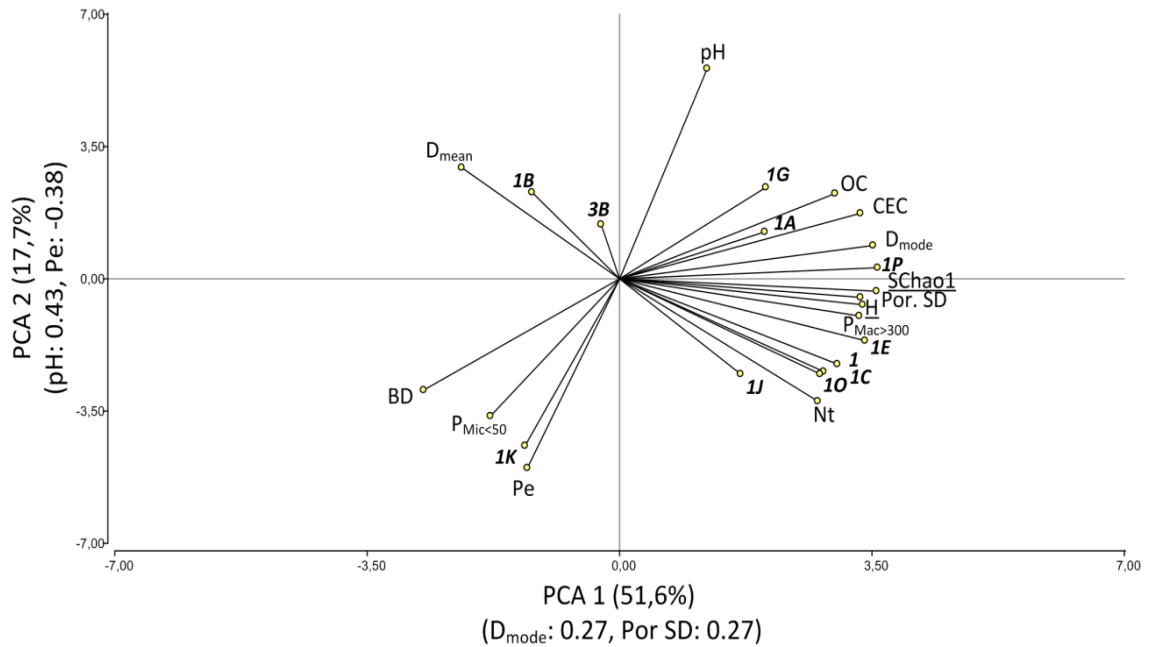


Figura Suplementaria V.3. Representación *biplot* del análisis de componentes principal basado en la abundancia de *subclusters nifH*, variables fisicoquímicas, y los estimadores de diversidad (H) y riqueza (SChao1). Las abreviaturas en inglés de las variables fisicoquímicas D_{mean}, BD, P_{Mic<50}, Nt, P_{Mac>300}, D_{mode}, CEC y OC se refieren respectivamente a: Diámetro de la media de poros, densidad aparente, Porosidad macroporos menor a 50 μm , Porosidad macroporos mayor a 300 μm , Diámetro de la moda de poros, Capacidad de intercambio catiónico y carbono orgánico total. Se usó la matriz de correlaciones de las variables analizadas para el análisis PCA. Los datos físicos del suelo para este análisis provienen del manuscrito en proceso de elaboración con el Dr. F. Behrends Kraemer, y los mismos también se encuentran en su trabajo de tesis doctoral [319]

Referencias Bibliográficas

- [1] P. Robe, R. Nalin, C. Capellano, T. M. Vogel, and P. Simonet, "Extraction of DNA from soil," *Eur. J. Soil Biol.*, vol. 39, no. 4, pp. 183–190, Oct. 2003.
- [2] B. Nogales, "La microbiología del suelo en la era de la biología molecular : descubriendo la punta del iceberg," *Ecosistemas*, vol. 14, no. 2, pp. 41–51, 2005.
- [3] E. A. Paul, *Soil Microbiology, Ecology and Biochemistry*. Academic Press, 2006.
- [4] L. Ranjard, F. Poly, J. Combrisson, A. Richaume, and S. Nazaret, "A single procedure to recover DNA from the surface or inside aggregates and in various size fractions of soil suitable for PCR-based assays of bacterial communities," *Eur. J. Soil Biol.*, vol. 34, no. 2, pp. 89–97, Apr. 1998.
- [5] J.-S. Kim and D. E. Crowley, "Size fractionation and microbial community structure of soil aggregates," *J. Agric. Chem. Environ.*, vol. 02, no. 04, pp. 75–80, Nov. 2013.
- [6] L. R. Bakken and V. Lindahl, "Recovery of bacterial cells from soil," in *Nucleic acids in the environment*, Springer B., 1995, pp. 9–27.
- [7] P. M. Haygarth and K. Ritz, "The future of soils and land use in the UK: Soil systems for the provision of land-based ecosystem services," *Land use policy*, vol. 26, no. 1, pp. 187–197, Dec. 2009.
- [8] R. Daniel, "The metagenomics of soil," *Nat. Rev. Microbiol.*, vol. 3, no. 6, pp. 470–8, Jun. 2005.
- [9] V. Torsvik and L. Øvreås, "Microbial diversity and function in soil: from genes to ecosystems," *Curr. Opin. Microbiol.*, vol. 5, no. 3, pp. 240–5, Jun. 2002.
- [10] T. L. Kieft, E. Soroker, and M. K. Firestone, "Microbial biomass response to a rapid increase in water potential when dry soil is wetted," *Soil Biol. Biochem.*, vol. 19, no. 2, pp. 119–126, Jan. 1987.
- [11] J. K. Jansson, J. D. Neufeld, M. A. Moran, and J. A. Gilbert, "Omics for understanding microbial functional dynamics.," *Environ. Microbiol.*, vol. 14, no. 1, pp. 1–3, Jan. 2012.
- [12] B. A. Needelman, "What Are Soils?," *Nat. Educ. Knowl.*, vol. 4, no. 3, p. 2, 2013.
- [13] C. Ettema, "Spatial soil ecology," *Trends Ecol. Evol.*, vol. 17, no. 4, pp. 177–183, Apr. 2002.
- [14] A. Konopka, "What is microbial community ecology?," *ISME J.*, vol. 3, no. 11, pp. 1223–1230, Nov. 2009.
- [15] M. Levine and H. W. Schoenlein, "A compilation of culture media for the cultivation of microorganisms," in *A Compilation of Culture Media for the Cultivation of Microorganisms*, 1930.
- [16] H. Al-Awadhi, N. Dashti, M. Khanafer, D. Al-Mailem, N. Ali, and S. Radwan, "Bias problems in culture-independent analysis of environmental bacterial communities: a representative study on hydrocarbonoclastic bacteria," *Springerplus*, vol. 2, no. 1, p. 369, Jan. 2013.
- [17] I. M. Head, J. R. Saunders, and R. W. Pickup, "Microbial Evolution, Diversity, and

- Ecology: A Decade of Ribosomal RNA Analysis of Uncultivated Microorganisms,” *Microb. Ecol.*, vol. 35, no. 1, pp. 1–21, Jan. 1998.
- [18] A. Ogram, G. Sayler, and T. Barkay, “The extraction and purification of microbial DNA from sediments,” *J. Microbiol. Methods*, vol. 7, no. 2, pp. 57–66, 1987.
- [19] W. E. Holben, J. K. Jansson, B. K. Chelm, and J. M. Tiedje, “DNA Probe Method for the Detection of Specific Microorganisms in the Soil Bacterial Community,” *Appl. Environ. Microbiol.*, vol. 54, no. 3, pp. 703–711, Mar. 1988.
- [20] R. Steffan and J. Goksøyr, “Recovery of DNA from soils and sediments,” *Appl. Environ. Microbiol.*, vol. 54, no. 12, pp. 2908–2915, 1988.
- [21] L. Leff and J. Dana, “Comparison of methods of DNA extraction from stream sediments,” *Appl. Environ. Microbiol.*, vol. 61, no. 3, pp. 1141–1143, 1995.
- [22] E. M. Gabor, E. J. de Vries, and D. B. Janssen, “Efficient recovery of environmental DNA for expression cloning by indirect extraction methods,” *FEMS Microbiol. Ecol.*, vol. 44, no. 2, pp. 153–163, May 2003.
- [23] J. Zhou, M. a Bruns, and J. M. Tiedje, “DNA recovery from soils of diverse composition,” *Appl. Environ. Microbiol.*, vol. 62, no. 2, pp. 316–22, Feb. 1996.
- [24] T. O. Delmont, P. Robe, S. Cecillon, I. M. Clark, F. Constancias, P. Simonet, P. R. Hirsch, and T. M. Vogel, “Accessing the soil metagenome for studies of microbial diversity,” *Appl. Environ. Microbiol.*, vol. 77, no. 4, pp. 1315–1324, 2011.
- [25] J. Theron and T. E. Cloete, “Molecular Techniques for Determining Microbial Diversity and Community Structure in Natural Environments,” *Crit. Rev. Microbiol.*, vol. 26, no. 1, pp. 37–57, Jan. 2000.
- [26] J. Borneman, P. W. Skroch, K. M. O’Sullivan, J. A. Palus, N. G. Rumjanek, J. L. Jansen, J. Nienhuis, and E. W. Triplett, “Molecular microbial diversity of an agricultural soil in Wisconsin,” *Appl. Environ. Microbiol.*, vol. 62, no. 6, pp. 1935–1943, Jun. 1996.
- [27] E. Smit, P. Leeflang, S. Gommans, J. van den Broek, S. van Mil, and K. Wernars, “Diversity and seasonal fluctuations of the dominant members of the bacterial soil community in a wheat field as determined by cultivation and molecular methods,” *Appl. Environ. Microbiol.*, vol. 67, no. 5, pp. 2284–2291, May 2001.
- [28] M. R. Liles, B. F. Manske, S. B. Bintrim, J. Handelsman, and R. M. Goodman, “A census of rRNA genes and linked genomic sequences within a soil metagenomic library,” *Appl. Environ. Microbiol.*, vol. 69, no. 5, pp. 2684–2691, May 2003.
- [29] H. Urakawa, K. Kita-Tsukamoto, and K. Ohwada, “Microbial diversity in marine sediments from Sagami Bay and Tokyo Bay, Japan, as determined by 16S rRNA gene analysis,” *Microbiology*, vol. 145 (, no. 1, pp. 3305–3315, Nov. 1999.
- [30] J. P. Bowman and R. D. McCuaig, “Biodiversity, community structural shifts, and biogeography of prokaryotes within Antarctic continental shelf sediment,” *Appl. Environ. Microbiol.*, vol. 69, no. 5, pp. 2463–2483, May 2003.
- [31] S. Tringe and P. Hugenholtz, “A renaissance for the pioneering 16S rRNA gene,” *Curr. Opin. Microbiol.*, vol. 11, no. 5, pp. 442–446, 2008.

- [32] M. F. Polz and C. M. Cavanaugh, "Bias in template-to-product ratios in multitemplate PCR.," *Appl. Environ. Microbiol.*, vol. 64, no. 10, pp. 3724–30, Oct. 1998.
- [33] G. Muyzer, E. C. de Waal, and A. G. Uitterlinden, "Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA," *Appl. Environ. Microbiol.*, vol. 59, no. 3, pp. 695–700, Mar. 1993.
- [34] A. Felske, H. Rheims, A. Wolterink, E. Stackebrandt, and A. D. Akkermans, "Ribosome analysis reveals prominent activity of an uncultured member of the class Actinobacteria in grassland soils," *Microbiology*, vol. 143, no. 9, pp. 2983–2989, Sep. 1997.
- [35] H. Heuer, M. Krsek, P. Baker, K. Smalla, and E. Wellington, "Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients," *Appl. Environ. Microbiol.*, vol. 63, no. 8, pp. 3233–3241, Aug. 1997.
- [36] D. H. Lee, Y. G. Zo, and S. J. Kim, "Nonradioactive method to study genetic profiles of natural bacterial communities by PCR-single-strand-conformation polymorphism," *Appl. Environ. Microbiol.*, vol. 62, no. 9, pp. 3112–3120, Sep. 1996.
- [37] F. Schwieger and C. C. Tebbe, "A new approach to utilize PCR-single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis," *Appl. Environ. Microbiol.*, vol. 64, no. 12, pp. 4870–4876, Dec. 1998.
- [38] W. Liu, T. L. Marsh, and H. Cheng, "Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA," *Appl. Environ. Microbiol.*, vol. 63, no. 11, pp. 4516–4522, 1997.
- [39] P. Wikström, "Biomonitoring complex microbial communities using random amplified polymorphic DNA and principal component analysis," *FEMS Microbiol. Ecol.*, vol. 28, no. 2, pp. 131–139, Feb. 1999.
- [40] L. Espinosa-Asuar, A. E. Escalante, L. I. Falcón, G. Bonilla-Rosso, S. Ramírez-Barahona, L. E. Eguiarte, and V. Souza, "Comparación de tres métodos moleculares para el análisis de procariontes ambientales en el mar del canal de Yucatán, México," *Hidrobiológica*, vol. 24, no. 3, pp. 257–270, 2014.
- [41] A. M. Osborn, E. R. B. Moore, and K. N. Timmis, "An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics," *Env. Microbiol.*, vol. 2, no. 1, pp. 39–50, 2000.
- [42] L. Tedersoo, R. H. Nilsson, K. Abarenkov, T. Jairus, A. Sadam, I. Saar, M. Bahram, E. Bechem, G. Chuyong, and U. Kõljalg, "454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases," *New Phytol.*, vol. 188, no. 1, pp. 291–301, Oct. 2010.
- [43] J. Dunbar and S. Barns, "Empirical and theoretical bacterial diversity in four Arizona soils," *Appl. ...*, 2002.
- [44] A. A. Andongma, L. Wan, Y. C. Dong, P. Li, N. Desneux, J. A. White, and C. Y. Niu, "Pyrosequencing reveals a shift in symbiotic bacteria populations across life stages of

- Bactrocera dorsalis,” *Sci. Rep.*, vol. 5, p. 9470, Jan. 2015.
- [45] S. A. Jackson, J. Kennedy, J. P. Morrissey, F. O’Gara, and A. D. W. Dobson, “Pyrosequencing reveals diverse and distinct sponge-specific microbial communities in sponges from a single geographical location in Irish waters,” *Microb. Ecol.*, vol. 64, no. 1, pp. 105–116, Jul. 2012.
- [46] T. Zhang, M.-F. Shao, and L. Ye, “454 pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants,” *ISME J.*, vol. 6, no. 6, pp. 1137–1147, Jun. 2012.
- [47] E. R. Mardis, “The impact of next-generation sequencing technology on genetics.,” *Trends Genet.*, vol. 24, no. 3, pp. 133–41, Mar. 2008.
- [48] J. Shendure and H. Ji, “Next-generation DNA sequencing.,” *Nat. Biotechnol.*, vol. 26, no. 10, pp. 1135–1145, Oct. 2008.
- [49] L. F. W. Roesch, R. R. Fulthorpe, A. Riva, G. Casella, A. K. M. Hadwin, A. D. Kent, S. H. Daroub, F. A. O. Camargo, W. G. Farmerie, and E. W. Triplett, “Pyrosequencing enumerates and contrasts soil microbial diversity.,” *ISME J.*, vol. 1, no. 4, pp. 283–90, Aug. 2007.
- [50] B. J. Campbell, S. W. Polson, T. E. Hanson, M. C. Mack, and E. A. G. Schuur, “The effect of nutrient deposition on bacterial communities in Arctic tundra soil,” *Environ. Microbiol.*, vol. 12, no. 7, pp. 1842–1854, 2010.
- [51] J. H. Rotthauwe, K. P. Witzel, and W. Liesack, “The ammonia monooxygenase structural gene amoA as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations.,” *Appl. Environ. Microbiol.*, vol. 63, no. 12, pp. 4704–12, Dec. 1997.
- [52] J. P. Zehr, B. D. Jenkins, S. M. Short, and G. F. Steward, “Minireview Nitrogenase gene diversity and microbial community structure : a cross-system comparison,” *Environ. Microbiol.*, vol. 5, no. 7, pp. 539–554, 2003.
- [53] J. H. J. Leveau, “The magic and menace of metagenomics: prospects for the study of plant growth-promoting rhizobacteria,” *Eur. J. Plant Pathol.*, vol. 119, no. 3, pp. 279–300, Sep. 2007.
- [54] N. G. Yok and G. L. Rosen, “Combining gene prediction methods to improve metagenomic gene annotation.,” *BMC Bioinformatics*, vol. 12, no. 1, p. 20, Jan. 2011.
- [55] M. Rho, H. Tang, and Y. Ye, “FragGeneScan: predicting genes in short and error-prone reads,” *Nucleic Acids Res.*, vol. 38, no. 20, p. e191, Nov. 2010.
- [56] T. M. Vogel, P. Simonet, J. K. Jansson, P. R. Hirsch, J. M. Tiedje, J. D. van Elsas, M. J. Bailey, R. Nalin, and L. Philippot, “TerraGenome: a consortium for the sequencing of a soil metagenome,” *Nat. Rev. Microbiol.*, vol. 7, no. 4, pp. 252–252, Apr. 2009.
- [57] J. Jansson, “Towards terra terra: Terabase sequencing of terrestrial metagenomics,” *Microbe*, July 2011, 2011, Sep. 2012.
- [58] L. G. Wall, “Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches,” F. J. de Bruijn, Ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2011.

- [59] N. Rascovan, B. Carbonetto, S. Revale, M. D. Reinert, R. Alvarez, A. M. Godeas, R. Colombo, M. Aguilar, M. V. Novas, L. Iannone, A. M. Zelada, A. Pardo, G. Schrauf, A. Mentaberry, and M. P. Vazquez, "The PAMPA datasets: a metagenomic survey of microbial communities in Argentinean pampean soils," *Microbiome*, vol. 1, no. 1, p. 21, 2013.
- [60] J. Handelsman, "Metagenomics: application of genomics to uncultured microorganisms," *Microbiol. Mol. Biol. Rev.*, vol. 68, no. 4, pp. 669–685, Dec. 2004.
- [61] G. W. Tyson, J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield, "Community structure and metabolism through reconstruction of microbial genomes from the environment.," *Nature*, vol. 428, no. 6978, pp. 37–43, Mar. 2004.
- [62] J. C. Venter, K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Neelson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith, "Environmental genome shotgun sequencing of the Sargasso Sea," *Science*, vol. 304, no. 5667, pp. 66–74, Apr. 2004.
- [63] T. Thomas, J. Gilbert, and F. Meyer, "Metagenomics - a guide from sampling to data analysis.," *Microb. Inform. Exp.*, vol. 2, no. 1, p. 3, Jan. 2012.
- [64] S. G. Tringe, C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz, and E. M. Rubin, "Comparative metagenomics of microbial communities," *Science*, vol. 308, no. 5721, pp. 554–557, Apr. 2005.
- [65] N. Fierer, C. L. Lauber, K. S. Ramirez, J. Zaneveld, M. a Bradford, and R. Knight, "Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients.," *ISME J.*, vol. 6, no. 5, pp. 1007–17, May 2012.
- [66] J. Qin, R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J.-M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H. B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Doré, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, P. Bork, S. D. Ehrlich, and J. Wang, "A human gut microbial gene catalogue established by metagenomic sequencing.," *Nature*, vol. 464, no. 7285, pp. 59–65, Mar. 2010.
- [67] P. Lorenz and J. Eck, "Metagenomics and industrial applications.," *Nat. Rev. Microbiol.*, vol. 3, no. 6, pp. 510–6, Jun. 2005.
- [68] H. K. Allen, L. a Moe, J. Rodbumrer, A. Gaarder, and J. Handelsman, "Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil.," *ISME J.*, vol. 3, no. 2, pp. 243–251, 2009.
- [69] A. Escalante Lozada, G. Gosset Lagarda, A. Martínez Jiménez, and F. Bolívar Zapata, "Diversidad bacteriana del suelo: Métodos de estudio no dependiente del cultivo microbiano e implicaciones biotecnológicas," *Agrociencia*, vol. 38, no. 6, pp. 583–592, 2004.

- [70] A. E. Escalante, "Ecología molecular en el estudio de comunidades bacterianas," in *Ecología Molecular*, L. E. E. X. Aguirre, V. Souza, Ed. México DF: Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), 2008, pp. 393–424.
- [71] Y. Huang, B. Niu, Y. Gao, L. Fu, and W. Li, "CD-HIT Suite: a web server for clustering and comparing biological sequences," *Bioinformatics*, vol. 26, no. 5, pp. 680–682, Mar. 2010.
- [72] P. D. Schloss, S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, and C. F. Weber, "Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities.," *Appl. Environ. Microbiol.*, vol. 75, no. 23, pp. 7537–41, Dec. 2009.
- [73] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool.," *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–10, Oct. 1990.
- [74] Q. Wang, G. M. Garrity, J. M. Tiedje, and J. R. Cole, "Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy.," *Appl. Environ. Microbiol.*, vol. 73, no. 16, pp. 5261–7, Aug. 2007.
- [75] W. Ludwig, O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüssmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. H. Schleifer, "ARB: a software environment for sequence data," *Nucleic Acids Res.*, vol. 32, no. 4, pp. 1363–1371, Jan. 2004.
- [76] A. E. Magurran, *Measuring Biological Diversity*, John Wiley. 2005.
- [77] R. Whittaker, "Evolution and measurement of species diversity," *Taxon*, 1972.
- [78] M. B. Carbonetto, "Diversidad de las comunidades microbianas de los suelos pampeanos. Enfoques ecológicos y metagenómicos," Universidad de Buenos Aires, 2014.
- [79] R. Kindt and R. Coe, "Tree diversity analysis: a manual and software for common statistical methods for ecological and biodiversity studies," 2005.
- [80] M. G. Klotz and L. Y. Stein, "Nitrifier genomics and evolution of the nitrogen cycle," *FEMS Microbiol. Lett.*, vol. 278, no. 2, pp. 146–56, Jan. 2008.
- [81] P. M. Vitousek, D. N. L. Menge, S. C. Reed, and C. C. Cleveland, "Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems," *Philos. Trans. R. Soc. London B, Biol. Sci.*, vol. 368, no. 1621, pp. 1–9, Jul. 2013.
- [82] J. N. Galloway, J. D. Aber, J. W. Erisman, S. P. Seitzinger, R. W. Howarth, E. B. Cowling, and B. J. Cosby, "The nitrogen cascade," *Bioscience*, vol. 53, no. 4, pp. 341–356, Apr. 2003.
- [83] J. C. Gaby and D. H. Buckley, "A global census of nitrogenase diversity," *Environ. Microbiol.*, vol. 13, no. 7, pp. 1790–9, Jul. 2011.

- [84] C. Rösch, A. Mergel, H. Bothe, and C. Ro, "Biodiversity of Denitrifying and Dinitrogen-Fixing Bacteria in an Acid Forest Soil," *Appl. Environ. Microbiol.*, vol. 68, no. 8, pp. 3818–3829, 2002.
- [85] J. P. Zehr, M. T. Mellon, and S. Zani, "New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of Nitrogenase (nifH) genes," *Appl. Environ. Microbiol.*, vol. 64, no. 9, pp. 3444–50, Sep. 1998.
- [86] H. Farnelid, A. F. Andersson, S. Bertilsson, W. A. Al-Soud, L. H. Hansen, S. Sørensen, G. F. Steward, Å. Hagström, and L. Riemann, "Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria," *PLoS One*, vol. 6, no. 4, p. e19223, Jan. 2011.
- [87] T. D. Niederberger, J. A. Sohm, J. Tirindelli, T. Gunderson, D. G. Capone, E. J. Carpenter, and S. C. Cary, "Diverse and highly active diazotrophic assemblages inhabit ephemerally wetted soils of the Antarctic Dry Valleys," *FEMS Microbiol. Ecol.*, vol. 82, no. 2, pp. 376–390, 2012.
- [88] H. Bürgmann, S. Meier, M. Bunge, F. Widmer, and J. Zeyer, "Effects of model root exudates on structure and activity of a soil diazotroph community," *Environ. Microbiol.*, vol. 7, no. 11, pp. 1711–1724, Nov. 2005.
- [89] I. R. Kennedy and N. Islam, "The current and potential contribution of asymbiotic nitrogen fixation to nitrogen requirements on farms: a review," *Anim. Prod. Sci.*, vol. 41, no. 3, pp. 447–457, May 2001.
- [90] J. P. Schimel and J. Bennett, "NITROGEN MINERALIZATION: CHALLENGES OF A CHANGING PARADIGM," *Ecology*, vol. 85, no. 3, pp. 591–602, Mar. 2004.
- [91] J. Zhang, P. Wu, B. Hao, and Z. Yu, "Heterotrophic nitrification and aerobic denitrification by the bacterium *Pseudomonas stutzeri* YZN-001," *Bioresour. Technol.*, vol. 102, no. 21, pp. 9866–9869, Nov. 2011.
- [92] H. S. Joo, M. Hirai, and M. Shoda, "Characteristics of ammonium removal by heterotrophic nitrification-aerobic denitrification by *Alcaligenes faecalis* No. 4," *J. Biosci. Bioeng.*, vol. 100, no. 2, pp. 184–191, Aug. 2005.
- [93] F. J. Verhagen, H. Duyts, and H. J. Laanbroek, "Competition for Ammonium between Nitrifying and Heterotrophic Bacteria in Continuously Percolated Soil Columns," *Appl. Environ. Microbiol.*, vol. 58, no. 10, pp. 3303–3311, Oct. 1992.
- [94] L. Lu, W. Han, J. Zhang, Y. Wu, B. Wang, X. Lin, J. Zhu, Z. Cai, and Z. Jia, "Nitrification of archaeal ammonia oxidizers in acid soils is supported by hydrolysis of urea," *ISME J.*, vol. 6, no. 10, pp. 1978–1984, 2012.
- [95] T. Jin, T. Zhang, and Q. Yan, "Characterization and quantification of ammonia-oxidizing archaea (AOA) and bacteria (AOB) in a nitrogen-removing reactor using T-RFLP and qPCR," *Appl. Microbiol. Biotechnol.*, vol. 87, no. 3, pp. 1167–1176, Jul. 2010.
- [96] P. Junier, O. S. Kim, V. Molina, P. Limburg, T. Junier, J. F. Imhoff, and K. P. Witzel, "Comparative in silico analysis of PCR primers suited for diagnostics and cloning of ammonia monooxygenase genes from ammonia-oxidizing bacteria," *FEMS Microbiol. Ecol.*, vol. 64, no. 1, pp. 141–152, Apr. 2008.

- [97] C. Moreno-Vivián, P. Cabello, M. Martínez-Luque, R. Blasco, and F. Castillo, "Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases.," *J. Bacteriol.*, vol. 181, no. 21, pp. 6573–84, Nov. 1999.
- [98] W. G. Zumft, "Cell biology and molecular basis of denitrification," *Microbiol. Mol. Biol. Rev.*, vol. 61, no. 4, pp. 533–616, Dec. 1997.
- [99] M. Hayatsu, K. Tago, and M. Saito, "Various players in the nitrogen cycle: Diversity and functions of the microorganisms involved in nitrification and denitrification," *Soil Sci. Plant Nutr.*, vol. 54, no. 1, pp. 33–45, Feb. 2008.
- [100] J. Ollivier, S. Töwe, A. Bannert, B. Hai, E. M. Kastl, A. Meyer, M. X. Su, K. Kleineidam, and M. Schloter, "Nitrogen turnover in soil and global change," *FEMS Microbiol. Ecol.*, vol. 78, no. 1, pp. 3–16, 2011.
- [101] A. E. Santoro, K. L. Casciotti, and C. A. Francis, "Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current," *Environ. Microbiol.*, vol. 12, no. 7, pp. 1989–2006, Jul. 2010.
- [102] G. Braker, J. Zhou, L. Wu, a H. Devol, and J. M. Tiedje, "Nitrite reductase genes (nirK and nirS) as functional markers to investigate diversity of denitrifying bacteria in pacific northwest marine sediment communities," *Appl. Environ. Microbiol.*, vol. 66, no. 5, pp. 2096–2104, May 2000.
- [103] D. J. Scala and L. J. Kerkhof, "Horizontal Heterogeneity of Denitrifying Bacterial Communities in Marine Sediments by Terminal Restriction Fragment Length Polymorphism Analysis," *Appl Env. Microbiol*, vol. 66, no. 5, pp. 1980–1986, 2000.
- [104] J. J. Rich and D. D. Myrold, "Community composition and activities of denitrifying bacteria from adjacent agricultural soil, riparian soil, and creek sediment in Oregon, USA," *Soil Biol. Biochem.*, vol. 36, no. 9, pp. 1431–1441, Sep. 2004.
- [105] P. Vitousek, J. Aber, and R. Howarth, "Human alteration of the global nitrogen cycle: causes and consequences," 1997.
- [106] E. F. Viglizzo, F. C. Frank, L. V. Carreño, E. G. Jobbágy, H. Pereyra, J. Clatt, D. Pincén, and M. F. Ricard, "Ecological and environmental footprint of 50 years of agricultural expansion in Argentina," *Glob. Chang. Biol.*, vol. 17, no. 2, pp. 959–973, Feb. 2011.
- [107] R. Derpsch, "Global Overview of Conservation Agriculture Adoption," 2009. [Online]. Available: <http://www.fao.org/>.
- [108] P. R. Hobbs, K. Sayre, and R. Gupta, "The role of conservation agriculture in sustainable agriculture.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 363, no. 1491, pp. 543–55, Feb. 2008.
- [109] H. Nacke, A. Thürmer, A. Wollherr, C. Will, L. Hodac, N. Herold, I. Schöning, M. Schruppf, and R. Daniel, "Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils.," *PLoS One*, vol. 6, no. 2, p. e17000, Jan. 2011.
- [110] C. H. Orr, A. James, C. Leifert, J. M. Cooper, and S. P. Cummings, "Diversity and activity of free-living nitrogen-fixing bacteria and total bacteria in organic and conventionally managed soils," *Appl. Environ. Microbiol.*, vol. 77, no. 3, pp. 911–919,

- Feb. 2011.
- [111] A. Keuter, E. Veldkamp, and M. D. Corre, "Asymbiotic biological nitrogen fixation in a temperate grassland as affected by management practices," *Soil Biol. Biochem.*, vol. 70, pp. 38–46, Mar. 2014.
- [112] D. A. Bossio, K. M. Scow, N. Gunapala, and K. J. Graham, "Determinants of Soil Microbial Communities: Effects of Agricultural Management, Season, and Soil Type on Phospholipid Fatty Acid Profiles," *Microb. Ecol.*, vol. 36, no. 1, pp. 1–12, Jul. 1998.
- [113] S. Colvan, J. Syers, and A. O'Donnell, "Effect of long-term fertiliser use on acid and alkaline phosphomonoesterase and phosphodiesterase activities in managed grassland," *Biol. Fertil. Soils*, vol. 34, no. 4, pp. 258–263.
- [114] Y. Clough, A. Kruess, and T. Tschardt, "Organic versus conventional arable farming systems: Functional grouping helps understand staphylinid response," *Agric. Ecosyst. Environ.*, vol. 118, no. 1–4, pp. 285–290, Jan. 2007.
- [115] L. Donnison, "Management influences on soil microbial communities and their function in botanically diverse haymeadows of northern England and Wales," *Soil Biol. Biochem.*, vol. 32, no. 2, pp. 253–263, Feb. 2000.
- [116] A. K. Patra, L. Abbadie, A. Clays-Josserand, V. Degrange, S. J. Grayston, N. Guillaumaud, P. Loiseau, F. Louault, S. Mahmood, S. Nazaret, L. Philippot, F. Poly, J. I. Prosser, and X. Le Roux, "Effects of management regime and plant species on the enzyme activity and genetic structure of N-fixing, denitrifying and nitrifying bacterial communities in grassland soils," *Environ. Microbiol.*, vol. 8, no. 6, pp. 1005–16, Jun. 2006.
- [117] S. a Wakelin, M. J. Colloff, P. R. Harvey, P. Marschner, A. L. Gregg, and S. L. Rogers, "The effects of stubble retention and nitrogen application on soil microbial community structure and functional gene abundance under irrigated maize," *FEMS Microbiol. Ecol.*, vol. 59, no. 3, pp. 661–70, Mar. 2007.
- [118] S. A. Wakelin, A. L. Gregg, R. J. Simpson, G. D. Li, I. T. Riley, and A. C. McKay, "Pasture management clearly affects soil microbial community structure and N-cycling bacteria," *Pedobiologia (Jena)*, vol. 52, no. 4, pp. 237–251, May 2009.
- [119] E. L. M. Figuerola, L. D. Guerrero, S. M. Rosa, L. Simonetti, M. E. Duval, J. A. Galantini, J. C. Bedano, L. G. Wall, and L. Erijman, "Bacterial indicator of agricultural management for soil under no-till crop production," *PLoS One*, vol. 7, no. 11, p. e51075, Jan. 2012.
- [120] B. Carbonetto, N. Rascovan, R. Álvarez, A. Mentaberry, and M. P. Vázquez, "Structure, composition and metagenomic profile of soil microbiomes associated to agricultural land use and tillage systems in Argentine Pampas," *PLoS One*, vol. 9, no. 6, p. e99949, 2014.
- [121] M. M. Collavino, H. J. Tripp, I. E. Frank, M. L. Vidoz, P. A. Calderoli, M. Donato, J. P. Zehr, and O. M. Aguilar, "nifH pyrosequencing reveals the potential for location-specific soil chemistry to influence N₂-fixing community dynamics," *Environ. Microbiol.*, vol. 16, no. 10, pp. 3211–3223, Apr. 2014.

- [122] M. E. Duval, J. A. Galantini, J. O. Iglesias, S. Canelo, J. M. Martinez, and L. Wall, "Analysis of organic fractions as indicators of soil quality under natural and cultivated systems," *Soil Tillage Res.*, vol. 131, pp. 11–19, Jul. 2013.
- [123] R. P. Colombo, L. Fernández Bidondo, V. A. Silvani, M. B. Carbonetto, N. Rascovan, M. J. Bompadre, M. Pérgola, G. Cuenca, and A. M. Godeas, "Diversity of arbuscular mycorrhizal fungi in soil from the Pampa Ondulada, Argentina, assessed by pyrosequencing and morphological techniques.," *Can. J. Microbiol.*, vol. 60, no. 12, pp. 819–27, Dec. 2014.
- [124] O. Barsky and S. Bearzotti, *El Desarrollo agropecuario pampeano*. Instituto Nacional de Estadísticas y Censos, 1991.
- [125] R. Alvarez and H. S. Steinbach, "A review of the effects of tillage systems on some soil physical properties, water content, nitrate availability and crops yield in the Argentine Pampas," *Soil Tillage Res.*, vol. 104, no. 1, pp. 1–15, Jun. 2009.
- [126] F. Meyer, D. Paarmann, M. D'Souza, R. Olson, E. M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening, and R. A. Edwards, "The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes," *BMC Bioinformatics*, vol. 9, no. 1, p. 386, Jan. 2008.
- [127] S. Powell, D. Szklarczyk, K. Trachana, A. Roth, M. Kuhn, J. Muller, R. Arnold, T. Rattei, I. Letunic, T. Doerks, L. J. Jensen, C. von Mering, and P. Bork, "eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges," *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D284–D289, Jan. 2012.
- [128] J. A. Fish, B. Chai, Q. Wang, Y. Sun, C. T. Brown, J. M. Tiedje, and J. R. Cole, "FunGene: the functional gene pipeline and repository," *Front. Microbiol.*, vol. 4, no. 4, p. 291, Jan. 2013.
- [129] C. M. Yeager, J. L. Kornosky, D. C. Housman, E. E. Grote, J. Belnap, and C. R. Kuske, "Diazotrophic Community Structure and Function in Two Successional Stages of Biological Soil Crusts from the Colorado Plateau and Chihuahuan Desert," *Appl. Environ. Microbiol.*, vol. 70, no. 2, pp. 973–984, 2004.
- [130] F. Widmer, B. T. Shaffer, L. a Porteous, and R. J. Seidler, "Analysis of nifH gene pool complexity in soil and litter at a Douglas fir forest site in the Oregon cascade mountain range," *Appl. Environ. Microbiol.*, vol. 65, no. 2, pp. 374–80, Feb. 1999.
- [131] S. Zani, M. T. Mellon, J. L. Collier, and J. P. Zehr, "Expression of nifH Genes in Natural Microbial Assemblages in Lake George, New York, Detected by Reverse Transcriptase PCR," *Appl. Environ. Microbiol.*, vol. 66, no. 7, pp. 3119–3124, Jul. 2000.
- [132] C. J. Smith, B. S. Danilowicz, A. K. Clear, F. J. Costello, B. Wilson, and W. G. Meijer, "T-Align, a web-based tool for comparison of multiple terminal restriction fragment length polymorphism profiles.," *FEMS Microbiol. Ecol.*, vol. 54, no. 3, pp. 375–380, Nov. 2005.
- [133] H. Izumi, I. C. Anderson, I. J. Alexander, K. Killham, and E. R. B. Moore, "Diversity and expression of nitrogenase genes (nifH) from ectomycorrhizas of Corsican pine (*Pinus nigra*)," *Environ. Microbiol.*, vol. 8, no. 12, pp. 2224–2230, Dec. 2006.

- [134] T. Ueda, Y. Suga, N. Yahiro, and T. Matsuguchi, "Remarkable N₂-fixing bacterial diversity detected in rice roots by molecular evolutionary analysis of nifH gene sequences," *J. Bacteriol.*, vol. 177, no. 5, pp. 1414–1417, Mar. 1995.
- [135] A. Sessitsch, P. Hardoim, J. Döring, A. Weilharter, A. Krause, T. Woyke, B. Mitter, L. Hauberg-Lotte, F. Friedrich, M. Rahalkar, T. Hurek, A. Sarkar, L. Bodrossy, L. van Overbeek, D. Brar, J. D. van Elsas, and B. Reinhold-Hurek, "Functional characteristics of an endophyte community colonizing rice roots as revealed by metagenomic analysis," *Mol. Plant. Microbe. Interact.*, vol. 25, no. 1, pp. 28–36, Jan. 2012.
- [136] R. C. Edgar, B. J. Haas, J. C. Clemente, C. Quince, and R. Knight, "UCHIME improves sensitivity and speed of chimera detection," *Bioinformatics*, vol. 27, no. 16, pp. 2194–2200, Aug. 2011.
- [137] K. T. Konstantinidis and J. M. Tiedje, "Genomic insights that advance the species definition for prokaryotes," *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 7, pp. 2567–2572, Feb. 2005.
- [138] J. P. Zehr, L. L. Crumbliss, M. J. Church, E. O. Omoregie, and B. D. Jenkins, "Nitrogenase genes in PCR and RT-PCR reagents: implications for studies of diversity of functional genes.," *Biotechniques*, vol. 35, no. 5, pp. 996–1002, 1004–5, Nov. 2003.
- [139] J.-W. Park and D. E. Crowley, "Normalization of soil DNA extraction for accurate quantification of target genes by real-time PCR and DGGE.," *Biotechniques*, vol. 38, no. 4, pp. 579–86, Apr. 2005.
- [140] "PAST: Palentiological Statistics software package for education and data analysis." Hammer and Harper, 2012.
- [141] J. P. Zehr and L. a McReynolds, "Use of degenerate oligonucleotides for amplification of the nifH gene from the marine cyanobacterium *Trichodesmium thiebautii*," *Appl. Environ. Microbiol.*, vol. 55, no. 10, pp. 2522–6, Oct. 1989.
- [142] K. Clarke, "Non-parametric multivariate analyses of changes in community structure," *Aust. J. Ecol.*, vol. 18, no. 1, pp. 117–143, 1993.
- [143] R. C. Di Rienzo JA, Casanoves F, Balzarini MG, Gonzalez L, Tablada M, "InfoStat versión 2016." Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina, 2016.
- [144] D. Borcard, P. Legendre, and P. Drapeau, "Partialling out the Spatial Component of Ecological Variation," *Ecology*, vol. 73, no. 3, p. 1045, Jun. 1992.
- [145] J. C. Gower, "Some distance properties of latent root and vector methods used in multivariate analysis," *Biometrika*, vol. 53, no. 3–4, pp. 325–338, Dec. 1966.
- [146] A. Ramette, "Multivariate analyses in microbial ecology.," *FEMS Microbiol. Ecol.*, vol. 62, no. 2, pp. 142–60, Nov. 2007.
- [147] S. W. Culman, H. G. Gauch, C. B. Blackwood, and J. E. Thies, "Analysis of T-RFLP data using analysis of variance and ordination methods: a comparative study," *J. Microbiol. Methods*, vol. 75, no. 1, pp. 55–63, Sep. 2008.
- [148] V. Torsvik, J. Goksøyr, and F. L. Daae, "High diversity in DNA of soil bacteria," *Appl. Environ. Microbiol.*, vol. 56, no. 3, pp. 782–787, Mar. 1990.

- [149] R. E. Ley, J. K. Harris, J. Wilcox, J. R. Spear, S. R. Miller, B. M. Bebout, J. A. Maresca, D. A. Bryant, M. L. Sogin, and N. R. Pace, “Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat.,” *Appl. Environ. Microbiol.*, vol. 72, no. 5, pp. 3685–3695, May 2006.
- [150] M. S. Montecchia, M. Tosi, M. A. Soria, J. A. Vogrig, O. Sydorenko, and O. S. Correa, “Pyrosequencing Reveals Changes in Soil Bacterial Communities after Conversion of Yungas Forests to Agriculture,” *PLoS One*, vol. 10, no. 3, p. e0119426, Jan. 2015.
- [151] K. Chen and L. Pachter, “Bioinformatics for whole-genome shotgun sequencing of microbial communities.,” *PLoS Comput. Biol.*, vol. 1, no. 2, pp. 106–112, Jul. 2005.
- [152] T. O. Delmont, E. Prestat, K. P. Keegan, M. Faubladier, P. Robe, I. M. Clark, E. Pelletier, P. R. Hirsch, F. Meyer, J. a Gilbert, D. Le Paslier, P. Simonet, and T. M. Vogel, “Structure, fluctuation and magnitude of a natural grassland soil metagenome,” *ISME J.*, vol. 6, no. 9, pp. 1677–1687, Sep. 2012.
- [153] N. Fierer, J. W. Leff, B. J. Adams, U. N. Nielsen, S. T. Bates, C. L. Lauber, S. Owens, J. a Gilbert, D. H. Wall, and J. G. Caporaso, “Cross-biome metagenomic analyses of soil microbial communities and their functional attributes,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 52, pp. 21390–21395, 2012.
- [154] V. M. Markowitz, I.-M. A. Chen, K. Chu, E. Szeto, K. Palaniappan, M. Pillay, A. Ratner, J. Huang, I. Pagani, S. Tringe, M. Huntemann, K. Billis, N. Varghese, K. Tennessen, K. Mavromatis, A. Pati, N. N. Ivanova, and N. C. Kyrpides, “IMG/M 4 version of the integrated metagenome comparative analysis system,” *Nucleic Acids Res.*, vol. 42, no. Database issue, pp. D568–D573, Jan. 2014.
- [155] R. Seshadri, S. A. Kravitz, L. Smarr, P. Gilna, and M. Frazier, “CAMERA: a community resource for metagenomics,” *PLoS Biol.*, vol. 5, no. 3, p. e75, Mar. 2007.
- [156] D. H. Huson, A. F. Auch, J. Qi, and S. C. Schuster, “MEGAN analysis of metagenomic data.,” *Genome Res.*, vol. 17, no. 3, pp. 377–386, Mar. 2007.
- [157] J. Zhou, Z. He, Y. Yang, Y. Deng, S. G. Tringe, and L. Alvarez Cohen, “High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis : Open and Closed Formats,” *MBio*, vol. 6, no. 1, pp. 1–17, 2015.
- [158] N. Fierer, C. L. Lauber, K. S. Ramirez, J. Zaneveld, M. A. Bradford, and R. Knight, “Comparative metagenomic , phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients,” pp. 1–11, 2011.
- [159] R. G. Pacchioni, F. M. Carvalho, C. E. Thompson, A. L. F. Faustino, F. Nicolini, T. S. Pereira, R. C. B. Silva, M. E. Cantão, A. Gerber, A. T. R. Vasconcelos, and L. F. Agnez-Lima, “Taxonomic and functional profiles of soil samples from Atlantic forest and Caatinga biomes in northeastern Brazil,” *Microbiologyopen*, vol. 3, no. 3, pp. 299–315, 2014.
- [160] H. Noguchi, J. Park, and T. Takagi, “MetaGene: prokaryotic gene finding from environmental genome shotgun sequences.,” *Nucleic Acids Res.*, vol. 34, no. 19, pp. 5623–5630, Jan. 2006.
- [161] J. Besemer and M. Borodovsky, “GeneMark: web software for gene finding in

- prokaryotes, eukaryotes and viruses,” *Nucleic Acids Res.*, vol. 33, no. Web Server issue, pp. W451–W454, Jul. 2005.
- [162] K. J. Hoff, “The effect of sequencing errors on metagenomic gene prediction,” *BMC Genomics*, vol. 10, no. 1, p. 1, Jan. 2009.
- [163] T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen, “Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB,” *Appl. Environ. Microbiol.*, vol. 72, no. 7, pp. 5069–72, Jul. 2006.
- [164] J. R. Cole, B. Chai, R. J. Farris, Q. Wang, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, A. M. Bandela, E. Cardenas, G. M. Garrity, and J. M. Tiedje, “The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data,” *Nucleic Acids Res.*, vol. 35, no. Database issue, pp. D169–72, Jan. 2007.
- [165] E. Pruesse, C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Glöckner, “SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB,” *Nucleic Acids Res.*, vol. 35, no. 21, pp. 7188–96, Jan. 2007.
- [166] L. K. McNeil, C. Reich, R. K. Aziz, D. Bartels, M. Cohoon, T. Disz, R. A. Edwards, S. Gerdes, K. Hwang, M. Kubal, G. R. Margaryan, F. Meyer, W. Mihalo, G. J. Olsen, R. Olson, A. Osterman, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, D. A. Rodionov, X. Shi, O. Vassieva, V. Vonstein, O. Zagnitko, F. Xia, J. Zinner, R. Overbeek, and R. Stevens, “The National Microbial Pathogen Database Resource (NMPDR): a genomics platform based on subsystem annotation,” *Nucleic Acids Res.*, vol. 35, no. Database issue, pp. D347–53, Jan. 2007.
- [167] T. Key and R. Overbeek, “Subsystems Annotation : What Should be the Goal of the Annotation Effort ?,” pp. 1–12.
- [168] M. Punta, P. C. Coghill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, A. Bateman, and R. D. Finn, “The Pfam protein families database,” *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D290–D301, Jan. 2012.
- [169] R. C. Souza, M. E. Cantão, A. T. R. Vasconcelos, M. A. Nogueira, and M. Hungria, “Soil metagenomics reveals differences under conventional and no-tillage with crop rotation or succession,” *Appl. Soil Ecol.*, vol. 72, pp. 49–61, Oct. 2013.
- [170] K. Yu and T. Zhang, “Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge,” *PLoS One*, vol. 7, no. 5, p. e38183, Jan. 2012.
- [171] S. Uroz, P. Ioannidis, J. Lengelle, A. Cébron, E. Morin, M. Buée, and F. Martin, “Functional assays and metagenomic analyses reveals differences between the microbial communities inhabiting the soil horizons of a Norway spruce plantation,” *PLoS One*, vol. 8, no. 2, p. e55929, Jan. 2013.
- [172] J. A. Gilbert, D. Field, Y. Huang, R. Edwards, W. Li, P. Gilna, and I. Joint, “Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities,” *PLoS One*, vol. 3, no. 8, p. e3042, Jan. 2008.

- [173] N. Fierer, C. L. Lauber, K. S. Ramirez, J. Zaneveld, M. a Bradford, and R. Knight, "Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients," *ISME J.*, vol. 6, no. 5, pp. 1007–1017, 2012.
- [174] T. O. Delmont, A. M. Eren, L. Maccario, E. Prestat, Ö. C. Esen, E. Pelletier, D. Le Paslier, P. Simonet, and T. M. Vogel, "Reconstructing rare soil microbial genomes using in situ enrichments and metagenomics.," *Front. Microbiol.*, vol. 6, p. 358, Jan. 2015.
- [175] M. L. Sogin, H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl, "Microbial diversity in the deep sea and the underexplored 'rare biosphere'.,," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 32, pp. 12115–20, Aug. 2006.
- [176] N. Korotkova, L. Chistoserdova, V. Kuksa, and M. E. Lidstrom, "Glyoxylate regeneration pathway in the methylotroph *Methylobacterium extorquens* AM1.,," *J. Bacteriol.*, vol. 184, no. 6, pp. 1750–8, Mar. 2002.
- [177] S. Kolb, "Aerobic methanol-oxidizing bacteria in soil.,," *FEMS Microbiol. Lett.*, vol. 300, no. 1, pp. 1–10, Nov. 2009.
- [178] E. V Koonin, "Orthologs, paralogs, and evolutionary genomics.,," *Annu. Rev. Genet.*, vol. 39, pp. 309–38, Jan. 2005.
- [179] C. a Pfister, F. Meyer, and D. a Antonopoulos, "Metagenomic profiling of a microbial assemblage associated with the California mussel: a node in networks of carbon and nitrogen cycling.,," *PLoS One*, vol. 5, no. 5, p. e10518, Jan. 2010.
- [180] J. M. Tiedje, "Ecology of denitrification and dissimilatory nitrate reduction to ammonium," in *Biology of Anaerobic Microorganism*, A. J. B. Zehnder, Ed. Wiley, New York, 1988, pp. 179–224.
- [181] J. P. Zehr, B. D. Jenkins, S. M. Short, and G. F. Steward, "Nitrogenase gene diversity and microbial community structure: a cross-system comparison.,," *Environ. Microbiol.*, vol. 5, no. 7, pp. 539–54, Jul. 2003.
- [182] P. M. Vitousek, K. Cassman, C. Cleveland, T. Crews, C. B. Field, N. B. Grimm, R. W. Howarth, R. Marino, L. Martinelli, E. B. Rastetter, and J. I. Sprent, "Towards an ecological understanding of biological nitrogen fixation," *Biogeochemistry*, vol. 57–58, no. 1, pp. 1–45.
- [183] H. P. Horz, J. H. Rotthauwe, T. Lukow, and W. Liesack, "Identification of major subgroups of ammonia-oxidizing bacteria in environmental samples by T-RFLP analysis of amoA PCR products.,," *J. Microbiol. Methods*, vol. 39, no. 3, pp. 197–204, Feb. 2000.
- [184] J. Galantini and R. Rosell, "Long-term fertilization effects on soil organic matter quality and dynamics under different production systems in semiarid Pampean soils," *Soil Tillage Res.*, vol. 87, no. 1, pp. 72–79, May 2006.
- [185] T. P. Cáceres, M. Megharaj, and R. Naidu, "Biodegradation of the pesticide fenamiphos by ten different species of green algae and cyanobacteria.,," *Curr. Microbiol.*, vol. 57, no. 6, pp. 643–6, Dec. 2008.
- [186] L. I. Domracheva, I. G. Shirokikh, and A. I. Fokina, "Anti-Fusarium activity of

- cyanobacteria and actinomycetes in soil and rhizosphere,” *Microbiology*, vol. 79, no. 6, pp. 871–876, Dec. 2010.
- [187] L. Philippot, “Denitrifying genes in bacterial and Archaeal genomes,” *Biochim. Biophys. Acta - Gene Struct. Expr.*, vol. 1577, no. 3, pp. 355–376, Sep. 2002.
- [188] R. W. Ye and S. M. Thomas, “Microbial nitrogen cycles: physiology, genomics and applications,” *Curr. Opin. Microbiol.*, vol. 4, no. 3, pp. 307–312, Jun. 2001.
- [189] D. J. Richardson*, B. C. Berks, D. A. Russell, S. Spiro, and C. J. Taylor, “Functional, biochemical and genetic diversity of prokaryotic nitrate reductases,” *Cell. Mol. Life Sci.*, vol. 58, no. 2, pp. 165–178, Feb. 2001.
- [190] D. Slade and M. Radman, “Oxidative stress resistance in *Deinococcus radiodurans*,” *Microbiol. Mol. Biol. Rev.*, vol. 75, no. 1, pp. 133–91, Mar. 2011.
- [191] K. G. Eilers, S. Debenport, S. Anderson, and N. Fierer, “Digging deeper to find unique microbial communities: The strong effect of depth on the structure of bacterial and archaeal communities in soil,” *Soil Biol. Biochem.*, vol. 50, pp. 58–65, Jul. 2012.
- [192] W. Wei, K. Isobe, T. Nishizawa, L. Zhu, Y. Shiratori, N. Ohte, K. Koba, S. Otsuka, and K. Senoo, “Higher diversity and abundance of denitrifying microorganisms in environments than considered previously,” *ISME J.*, vol. 9, no. 9, pp. 1954–65, Sep. 2015.
- [193] E. Wessén, S. Hallin, and L. Philippot, “Differential responses of bacterial and archaeal groups at high taxonomical ranks to soil management,” *Soil Biol. Biochem.*, vol. 42, no. 10, pp. 1759–1765, Oct. 2010.
- [194] N. L. Ward, J. F. Challacombe, P. H. Janssen, B. Henrissat, P. M. Coutinho, M. Wu, G. Xie, D. H. Haft, M. Sait, J. Badger, R. D. Barabote, B. Bradley, T. S. Brettin, L. M. Brinkac, D. Bruce, T. Creasy, S. C. Daugherty, T. M. Davidsen, R. T. DeBoy, J. C. Detter, R. J. Dodson, A. S. Durkin, A. Ganapathy, M. Gwinn-Giglio, C. S. Han, H. Khouri, H. Kiss, S. P. Kothari, R. Madupu, K. E. Nelson, W. C. Nelson, I. Paulsen, K. Penn, Q. Ren, M. J. Rosovitz, J. D. Selengut, S. Shrivastava, S. A. Sullivan, R. Tapia, L. S. Thompson, K. L. Watkins, Q. Yang, C. Yu, N. Zafar, L. Zhou, and C. R. Kuske, “Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils,” *Appl. Environ. Microbiol.*, vol. 75, no. 7, pp. 2046–56, Apr. 2009.
- [195] T. O. Delmont, P. Robe, I. Clark, P. Simonet, and T. M. Vogel, “Metagenomic comparison of direct and indirect soil DNA extraction approaches,” *J. Microbiol. Methods*, vol. 86, no. 3, pp. 397–400, Sep. 2011.
- [196] N. Fierer and R. B. Jackson, “The diversity and biogeography of soil bacterial communities,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 3, pp. 626–31, Jan. 2006.
- [197] P. C. Baveye, “To sequence or not to sequence the whole-soil metagenome?,” *Nat. Rev. Microbiol.*, vol. 7, no. 10, p. 756; author reply 756–7, Oct. 2009.
- [198] A. Bollmann, G. S. Bullerjahn, and R. M. McKay, “Abundance and diversity of ammonia-oxidizing archaea and bacteria in sediments of trophic end members of the Laurentian Great Lakes, Erie and Superior,” *PLoS One*, vol. 9, no. 5, p. e97068, Jan. 2014.

- [199] M. Pester, T. Rattei, S. Flechl, A. Gröngroft, A. Richter, J. Overmann, B. Reinhold-Hurek, A. Loy, and M. Wagner, "amoA-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of amoA genes from soils of four different geographic regions.," *Environ. Microbiol.*, vol. 14, no. 2, pp. 525–39, Feb. 2012.
- [200] N. J. Ritchie, M. E. Schutter, R. P. Dick, and D. D. Myrold, "Use of Length Heterogeneity PCR and Fatty Acid Methyl Ester Profiles To Characterize Microbial Communities in Soil," *Appl. Environ. Microbiol.*, vol. 66, no. 4, pp. 1668–1675, Apr. 2000.
- [201] E. Avannis-Aghajani, K. Jones, A. Holtzman, T. Aronson, N. Glover, M. Boian, S. Froman, and C. Brunk, "Molecular technique for rapid identification of mycobacteria," *J. Clin. Microbiol.*, vol. 34, no. 1, pp. 98–102, Jan. 1996.
- [202] M. A. J. Hullar, L. A. Kaplan, and D. A. Stahl, "Recurring seasonal dynamics of microbial communities in stream habitats.," *Appl. Environ. Microbiol.*, vol. 72, no. 1, pp. 713–22, Jan. 2006.
- [203] E. Katsivela, E. R. B. Moore, D. Maroukli, C. Strömpl, D. Pieper, and N. Kalogerakis, "Bacterial community dynamics during in-situ bioremediation of petroleum waste sludge in landfarming sites," *Biodegradation*, vol. 16, no. 2, pp. 169–180, Mar. 2005.
- [204] M. Noll, D. Matthies, P. Frenzel, M. Derakshani, and W. Liesack, "Succession of bacterial community structure and diversity in a paddy soil oxygen gradient.," *Environ. Microbiol.*, vol. 7, no. 3, pp. 382–95, Mar. 2005.
- [205] M. Schmidt, A. Priemé, and P. Stougaard, "Bacterial diversity in permanently cold and alkaline ikaite columns from Greenland.," *Extremophiles*, vol. 10, no. 6, pp. 551–62, Dec. 2006.
- [206] J. E. Thies, "Soil Microbial Community Analysis using Terminal Restriction Fragment Length Polymorphisms," *Soil Sci. Soc. Am. J.*, vol. 71, no. 2, p. 579, Mar. 2007.
- [207] O. R. Kotsyurbenko, K.-J. Chin, M. V Glagolev, S. Stubner, M. V Simankova, A. N. Nozhevnikova, and R. Conrad, "Acetoclastic and hydrogenotrophic methane production and methanogenic populations in an acidic West-Siberian peat bog.," *Environ. Microbiol.*, vol. 6, no. 11, pp. 1159–73, Nov. 2004.
- [208] A. I. Leïbo, A. I. Netrusov, and R. Conrad, "[Effect of hydrogen concentration on the hydrogenotrophic methanogenic community structure studied by T-RELP analysis of 16S rRNA gene amplicons].," *Mikrobiologiya*, vol. 75, no. 6, pp. 786–91, Jan. .
- [209] Y. Lu, T. Lueders, M. W. Friedrich, and R. Conrad, "Detecting active methanogenic populations on rice roots using stable isotope probing.," *Environ. Microbiol.*, vol. 7, no. 3, pp. 326–36, Mar. 2005.
- [210] M. M. Moeseneder, C. Winter, J. M. Arrieta, and G. J. Herndl, "Terminal-restriction fragment length polymorphism (T-RFLP) screening of a marine archaeal clone library to determine the different phylotypes.," *J. Microbiol. Methods*, vol. 44, no. 2, pp. 159–72, Mar. 2001.
- [211] S. Weber, "Methanogenic populations involved in the degradation of rice straw in anoxic paddy soil," *FEMS Microbiol. Ecol.*, vol. 38, no. 1, pp. 11–20, Dec. 2001.

- [212] R. B. Franklin and A. L. Mills, "Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field.," *FEMS Microbiol. Ecol.*, vol. 44, no. 3, pp. 335–46, Jun. 2003.
- [213] D. L. Mummey and P. D. Stahl, "Spatial and temporal variability of bacterial 16S rDNA-based T-RFLP patterns derived from soil of two Wyoming grassland ecosystems.," *FEMS Microbiol. Ecol.*, vol. 46, no. 1, pp. 113–20, Oct. 2003.
- [214] T. Lukow¹, P. Dunfield, and W. Liesack, "Use of the T-RFLP technique to assess spatial and temporal changes in the bacterial community structure within an agricultural soil planted with transgenic and non-transgenic potato plants.," *FEMS Microbiol. Ecol.*, vol. 32, no. 3, pp. 241–247, Jun. 2000.
- [215] A. T. Mintie, R. S. Heichen, K. Cromack, D. D. Myrold, and P. J. Bottomley, "Ammonia-oxidizing bacteria along meadow-to-forest transects in the Oregon Cascade Mountains.," *Appl. Environ. Microbiol.*, vol. 69, no. 6, pp. 3129–36, Jun. 2003.
- [216] J. J. Rich, R. S. Heichen, P. J. Bottomley, K. Cromack, and D. D. Myrold, "Community Composition and Functioning of Denitrifying Bacteria from Adjacent Meadow and Forest Soils," *Appl. Environ. Microbiol.*, vol. 69, no. 10, pp. 5974–5982, Oct. 2003.
- [217] C. Rösch and H. Bothe, "Improved Assessment of Denitrifying , N₂ -Fixing , and Total-Community Bacteria by Terminal Restriction Fragment Length Polymorphism Analysis Using Multiple Restriction Enzymes," *Appl. Environ. Microbiol.*, vol. 71, no. 4, pp. 2026–2035, 2005.
- [218] Z. Tan, T. Hurek, and B. Reinhold-Hurek, "Brief report Effect of N-fertilization , plant genotype and environmental conditions on nifH gene pools in roots of rice," *Env. Microbiol.*, vol. 5, no. 10, pp. 1009–1015, 2003.
- [219] S. R. Mohanty, P. L. E. Bodelier, V. Floris, and R. Conrad, "Differential effects of nitrogenous fertilizers on methane-consuming microbes in rice field and forest soils.," *Appl. Environ. Microbiol.*, vol. 72, no. 2, pp. 1346–54, Feb. 2006.
- [220] K. Bruce, "Analysis of mer Gene Subclasses within Bacterial Communities in Soils and Sediments Resolved by Fluorescent-PCR-Restriction Fragment Length Polymorphism Profiling," *Appl. Envir. Microbiol.*, vol. 63, no. 12, pp. 4914–4919, Dec. 1997.
- [221] J. Dunbar, L. O. Ticknor, and C. R. Kuske, "Assessment of Microbial Diversity in Four Southwestern United States Soils by 16S rRNA Gene Terminal Restriction Fragment Analysis," *Appl Env. Microbiol.*, vol. 66, no. 7, pp. 2943–2950, 2000.
- [222] F. Poly, L. J. Monrozier, and R. Bally, "Improvement in the RFLP procedure for studying the diversity of nifH genes in communities of nitrogen fixers in soil," *Res. Microbiol.*, vol. 152, no. 1, pp. 95–103, Jan. 2001.
- [223] C. Singh, R. Soni, S. Jain, S. Roy, and R. Goel, "Diversification of nitrogen fixing bacterial community using nifH gene as a biomarker in different geographical soils of Western Indian Himalayas," *J. Environ. Biol.*, vol. 31, no. 5, pp. 553–556, Sep. 2010.
- [224] V. M. Conn and C. M. M. Franco, "Analysis of the endophytic actinobacterial

- population in the roots of wheat (*Triticum aestivum* L.) by terminal restriction fragment length polymorphism and sequencing of 16S rRNA clones,” *Appl. Environ. Microbiol.*, vol. 70, no. 3, pp. 1787–1794, Mar. 2004.
- [225] Y. Shi, M. Tapa, C. Li, H. Yang, T. Zhang, Y. Gao, J. Sun, J. Zeng, Q. Lin, Z. Cao, K. Outi, Y. Li, and K. Lou, “Diversity and space-time dynamics of endophytic archaea from sugar beet in the north slope of Tianshan Mountain revealed by 454 pyrosequencing and T-RFLP,” *World J. Microbiol. Biotechnol.*, vol. 31, no. 7, pp. 1031–1039, Jul. 2015.
- [226] A. Hiraishi, M. Iwasaki, and H. Shinjo, “Terminal restriction pattern analysis of 16S rRNA genes for the characterization of bacterial communities of activated sludge,” *J. Biosci. Bioeng.*, vol. 90, no. 2, pp. 148–156, Jan. 2000.
- [227] P. E. Saikaly, P. G. Stroot, and D. B. Oerther, “Use of 16S rRNA gene terminal restriction fragment analysis to assess the impact of solids retention time on the bacterial diversity of activated sludge,” *Appl. Environ. Microbiol.*, vol. 71, no. 10, pp. 5814–5822, Oct. 2005.
- [228] U. M. E. Schütte, Z. Abdo, S. J. Bent, C. Shyu, C. J. Williams, J. D. Pierson, and L. J. Forney, “Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities,” *Appl. Microbiol. Biotechnol.*, vol. 80, no. 3, pp. 365–80, Sep. 2008.
- [229] J. C. Gaby and D. H. Buckley, “A comprehensive evaluation of PCR primers to amplify the *nifH* gene of nitrogenase,” *PLoS One*, vol. 7, no. 7, p. e42149, Jan. 2012.
- [230] F. Poly, L. Ranjard, S. Nazaret, F. Gourbière, and L. J. Monrozier, “Comparison of *nifH* gene pools in soils and soil microenvironments with contrasting properties,” *Appl. Environ. Microbiol.*, vol. 67, no. 5, pp. 2255–62, May 2001.
- [231] B. Hoppe, T. Kahl, P. Karasch, T. Wubet, J. Bauhus, F. Buscot, and D. Krüger, “Network analysis reveals ecological links between N-fixing bacteria and wood-decaying fungi,” *PLoS One*, vol. 9, no. 2, p. e88141, Jan. 2014.
- [232] A. Welsh, D. J. Burke, and D. Hahn, “Analysis of nitrogen-fixing members of the epsilon subclass of Proteobacteria in salt marsh sediments,” *Appl. Environ. Microbiol.*, vol. 73, no. 23, pp. 7747–52, Dec. 2007.
- [233] S. T. Berthrong, C. M. Yeager, L. Gallegos-Graves, B. Steven, S. a Eichorst, R. B. Jackson, and C. R. Kuske, “Nitrogen fertilization has a stronger effect on soil nitrogen-fixing bacterial communities than elevated atmospheric CO₂,” *Appl. Environ. Microbiol.*, vol. 80, no. 10, pp. 3103–12, May 2014.
- [234] M. R. R. Coelho, I. E. Marriel, S. N. Jenkins, C. V. Lanyon, L. Seldin, and A. G. O’Donnell, “Molecular detection and quantification of *nifH* gene sequences in the rhizosphere of sorghum (*Sorghum bicolor*) sown with two levels of nitrogen fertilizer,” *Appl. Soil Ecol.*, vol. 42, no. 1, pp. 48–53, May 2009.
- [235] L. F. W. Roesch, F. L. Olivares, L. M. Pereira Passaglia, P. A. Selbach, E. L. S. Sá, and F. a. O. Camargo, “Characterization of diazotrophic bacteria associated with maize: effect of plant genotype, ontogeny and nitrogen-supply,” *World J. Microbiol. Biotechnol.*, vol. 22, no. 9, pp. 967–974, Mar. 2006.

- [236] C. M. Yeager, D. E. Northup, C. C. Grow, M. Barns, C. R. Kuske, and S. M. Barns, "Changes in Nitrogen-Fixing and Ammonia-Oxidizing Bacterial Communities in Soil of a Mixed Conifer Forest after Wildfire Changes in Nitrogen-Fixing and Ammonia-Oxidizing Bacterial Communities in Soil of a Mixed Conifer Forest after Wildfire," *Appl. Environ. Microbiol.*, vol. 71, no. 5, pp. 2713–2722, 2005.
- [237] H. Farnelid, M. Bentzon-Tilia, A. F. Andersson, S. Bertilsson, G. Jost, M. Labrenz, K. Jürgens, and L. Riemann, "Active nitrogen-fixing heterotrophic bacteria at and below the chemocline of the central Baltic Sea.," *ISME J.*, vol. 7, no. 7, pp. 1413–23, Jul. 2013.
- [238] M. Fürnkranz, W. Wanek, A. Richter, G. Abell, F. Rasche, and A. Sessitsch, "Nitrogen fixation by phyllosphere bacteria associated with higher plants and their colonizing epiphytes of a tropical lowland rainforest of Costa Rica.," *ISME J.*, vol. 2, no. 5, pp. 561–70, May 2008.
- [239] T. L. Marsh, "Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products.," *Curr. Opin. Microbiol.*, vol. 2, no. 3, pp. 323–7, Jun. 1999.
- [240] A. L. Flores-Mireles, S. C. Winans, and G. Holguin, "Molecular characterization of diazotrophic and denitrifying bacteria associated with mangrove roots.," *Appl. Environ. Microbiol.*, vol. 73, no. 22, pp. 7308–21, Nov. 2007.
- [241] P. Junier, T. Junier, and K.-P. Witzel, "TRiFLe, a program for in silico terminal restriction fragment length polymorphism analysis with user-defined sequence sets.," *Appl. Environ. Microbiol.*, vol. 74, no. 20, pp. 6452–6, Oct. 2008.
- [242] J. J. Engebretson and C. L. Moyer, "Fidelity of select restriction endonucleases in determining microbial diversity by terminal-restriction fragment length polymorphism.," *Appl. Environ. Microbiol.*, vol. 69, no. 8, pp. 4823–9, Aug. 2003.
- [243] F. Li, M. A. J. Hullar, and J. W. Lampe, "Optimization of terminal restriction fragment polymorphism (TRFLP) analysis of human gut microbiota.," *J. Microbiol. Methods*, vol. 68, no. 2, pp. 303–11, Feb. 2007.
- [244] C.-P. Yu, R. Ahuja, G. Saylor, and K.-H. Chu, "Quantitative molecular assay for fingerprinting microbial communities of wastewater and estrogen-degrading consortia.," *Appl. Environ. Microbiol.*, vol. 71, no. 3, pp. 1433–44, Mar. 2005.
- [245] A. M. Noguez, H. T. Arita, A. E. Escalante, L. J. Forney, F. García-Oliva, and V. Souza, "Microbial macroecology: highly structured prokaryotic soil assemblages in a tropical deciduous forest," *Glob. Ecol. Biogeogr.*, vol. 14, no. 3, pp. 241–248, May 2005.
- [246] H. E. Reed and J. B. H. Martiny, "Microbial composition affects the functioning of estuarine sediments.," *ISME J.*, vol. 7, no. 4, pp. 868–79, Apr. 2013.
- [247] E. da C Jesus, T. L. Marsh, J. M. Tiedje, and F. M. de S Moreira, "Changes in land use alter the structure of bacterial communities in Western Amazon soils.," *ISME J.*, vol. 3, no. 9, pp. 1004–11, Sep. 2009.
- [248] M. S. Girvan, J. Bullimore, J. N. Pretty, A. M. Osborn, and A. S. Ball, "Soil type is the primary determinant of the composition of the total and active bacterial communities

- in arable soils.," *Appl. Environ. Microbiol.*, vol. 69, no. 3, pp. 1800–9, Mar. 2003.
- [249] J.-B. Ramond, F. Tshabuse, C. W. Bopda, D. A. Cowan, and M. I. Tuffin, "Evidence of variability in the structure and recruitment of rhizospheric and endophytic bacterial communities associated with arable sweet sorghum (*Sorghum bicolor* (L) Moench)," *Plant Soil*, vol. 372, no. 1–2, pp. 265–278, May 2013.
- [250] A. D. Kent, D. J. Smith, B. J. Benson, W. Triplett, and E. W. Triplett, "Web-Based Phylogenetic Assignment Tool for Analysis of Terminal Restriction Fragment Length Polymorphism Profiles of Microbial Communities Web-Based Phylogenetic Assignment Tool for Analysis of Terminal Restriction Fragment Length Polymorphism Profiles of," *Appl. Environ. Microbiol.*, vol. 69, no. 11, p. 6798, 2003.
- [251] B. Sercu, L. C. Van De Werfhorst, J. L. S. Murray, and P. A. Holden, "Cultivation-independent analysis of bacteria in IDEXX Quanti-Tray/2000 fecal indicator assays.," *Appl. Environ. Microbiol.*, vol. 77, no. 2, pp. 627–33, Jan. 2011.
- [252] M. W. Friedrich, D. Schmitt-Wagner, T. Lueders, and A. Brune, "Axial Differences in Community Structure of Crenarchaeota and Euryarchaeota in the Highly Compartmentalized Gut of the Soil-Feeding Termite *Cubitermes orthognathus*," *Appl. Environ. Microbiol.*, vol. 67, no. 10, pp. 4880–4890, Oct. 2001.
- [253] F. Inagaki, Y. Sakihama, A. Inoue, C. Kato, and K. Horikoshi, "Molecular phylogenetic analyses of reverse-transcribed bacterial rRNA obtained from deep-sea cold seep sediments," *Environ. Microbiol.*, vol. 4, no. 5, pp. 277–286, May 2002.
- [254] T. Lueders and M. Friedrich, "Archaeal Population Dynamics during Sequential Reduction Processes in Rice Field Soil," *Appl. Environ. Microbiol.*, vol. 66, no. 7, pp. 2732–2742, Jul. 2000.
- [255] W. B. Nilsson and M. S. Strom, "Detection and identification of bacterial pathogens of fish in kidney tissue using terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes.," *Dis. Aquat. Organ.*, vol. 48, no. 3, pp. 175–85, Apr. 2002.
- [256] P. Gonzalez Perez, J. Ye, S. Wang, X. Wang, and D. Huang, "Analysis of the occurrence and activity of diazotrophic communities in organic and conventional horticultural soils," *Appl. Soil Ecol.*, vol. 79, pp. 37–48, 2014.
- [257] B. S. Mirza, C. Potisap, K. Nüsslein, B. J. M. Bohannon, and J. L. M. Rodriguesa, "Response of free-living nitrogen-fixing microorganisms to land use change in the amazon rainforest," *Appl. Environ. Microbiol.*, vol. 80, no. 1, pp. 281–288, 2014.
- [258] S.-F. Hsu and D. H. Buckley, "Evidence for the functional significance of diazotroph community structure in soil.," *ISME J.*, vol. 3, no. 1, pp. 124–36, Jan. 2009.
- [259] D. H. Buckley, V. Huangyutitham, S.-F. Hsu, and T. a Nelson, "Stable isotope probing with $^{15}\text{N}_2$ reveals novel noncultivated diazotrophs in soil.," *Appl. Environ. Microbiol.*, vol. 73, no. 10, pp. 3196–204, May 2007.
- [260] Y. B. Chen, B. Dominic, M. T. Mellon, and J. P. Zehr, "Circadian rhythm of nitrogenase gene expression in the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. strain IMS 101.," *J. Bacteriol.*, vol. 180, no. 14, pp. 3598–605, Jul. 1998.

- [261] J. A. Coyer, A. Cabello-Pasini, H. Swift, and R. S. Alberte, "N₂ fixation in marine heterotrophic bacteria: dynamics of environmental and molecular regulation.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 8, pp. 3575–80, Apr. 1996.
- [262] T. Hoover, "Control of nitrogen fixation genes in *Klebsiella pneumoniae*," in *Prokaryotic nitrogen fixation*, E. W. Triplett, Ed. Norfolk, United Kingdom: Horizon Scientific Press, 2000, pp. 131–147.
- [263] M. J. Merrick and R. A. Edwards, "Nitrogen control in bacteria.," *Microbiol. Rev.*, vol. 59, no. 4, pp. 604–22, Dec. 1995.
- [264] K. Kim, Y. Zhang, and G. P. Roberts, "Correlation of activity regulation and substrate recognition of the ADP-ribosyltransferase that regulates nitrogenase activity in *Rhodospirillum rubrum*," *J. Bacteriol.*, vol. 181, no. 5, pp. 1698–702, Mar. 1999.
- [265] R. Brouzes, J. Lasik, and R. Knowles, "The effect of organic amendment, water content, and oxygen on the incorporation of ¹⁵N₂ by some agricultural and forest soils.," *Can. J. Microbiol.*, vol. 15, no. 8, pp. 899–905, Aug. 1969.
- [266] P. O'Toole and R. Knowles, "Efficiency of acetylene reduction (nitrogen fixation) in soil: Effect of type and concentration of available carbohydrate," *Soil Biol. Biochem.*, vol. 5, no. 6, pp. 789–797, Dec. 1973.
- [267] M. M. Roper and N. A. Smith, "Straw decomposition and nitrogenase activity (C₂H₂ reduction) by free-living microorganisms from soil: Effects of pH and clay content," *Soil Biol. Biochem.*, vol. 23, no. 3, pp. 275–283, Jan. 1991.
- [268] C. Limmer and H. L. Drake, "Non-symbiotic N₂-fixation in acidic and pH-neutral forest soils: Aerobic and anaerobic differentials," *Soil Biol. Biochem.*, vol. 28, no. 2, pp. 177–183, Feb. 1996.
- [269] J. N. Galloway, F. J. Dentener, D. G. Capone, E. W. Boyer, R. W. Howarth, S. P. Seitzinger, G. P. Asner, C. C. Cleveland, P. A. Green, E. A. Holland, D. M. Karl, A. F. Michaels, J. H. Porter, A. R. Townsend, and C. J. Vo, *Nitrogen cycles : past , present , and future*. 2004.
- [270] G. Hardarson and S. K. A. Danso, "Methods for measuring biological nitrogen fixation in grain legumes," *Plant Soil*, vol. 152, no. 1, pp. 19–23, 1993.
- [271] H. H. Zahran, "Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate.," *Microbiol. Mol. Biol. Rev.*, vol. 63, no. 4, pp. 968–89, table of contents, Dec. 1999.
- [272] C. Sicking, M. Bruschi, A. Lindackers, K.-U. Riedel, B. Schubert, N. Isakovic, C. Krall, W. Klipp, T. Drepper, K. Schneider, and B. Masepohl, "Identification of two new genes involved in diazotrophic growth via the alternative Fe-only nitrogenase in the phototrophic purple bacterium *Rhodobacter capsulatus*," *J. Bacteriol.*, vol. 187, no. 1, pp. 92–8, Jan. 2005.
- [273] N. Wu, Y. M. Zhang, and A. Downing, "Comparative study of nitrogenase activity in different types of biological soil crusts in the Gurbantunggut Desert, Northwestern China," *J. Arid Environ.*, vol. 73, no. 9, pp. 828–833, 2009.
- [274] D. H. Buckley, V. Huangyutitham, S.-F. Hsu, and T. A. Nelson, "¹⁵N₂-DNA-stable isotope probing of diazotrophic methanotrophs in soil," *Soil Biol. Biochem.*, vol. 40,

- no. 6, pp. 1272–1283, Jun. 2008.
- [275] Y. Shi, G. W. Tyson, and E. F. DeLong, “Metatranscriptomics reveals unique microbial small RNAs in the ocean’s water column,” *Nature*, vol. 459, no. 7244, pp. 266–9, May 2009.
- [276] T. R. Turner, K. Ramakrishnan, J. Walshaw, D. Heavens, M. Alston, D. Swarbreck, A. Osbourn, A. Grant, and P. S. Poole, “Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants,” *ISME J.*, vol. 7, no. 12, pp. 2248–58, Dec. 2013.
- [277] L.-X. Chen, M. Hu, L.-N. Huang, Z.-S. Hua, J.-L. Kuang, S.-J. Li, and W.-S. Shu, “Comparative metagenomic and metatranscriptomic analyses of microbial communities in acid mine drainage,” *ISME J.*, vol. 9, no. 7, pp. 1579–1592, Dec. 2014.
- [278] L. C. Carvalhais, P. G. Dennis, G. W. Tyson, and P. M. Schenk, “Application of metatranscriptomics to soil environments,” *J. Microbiol. Methods*, vol. 91, no. 2, pp. 246–51, Nov. 2012.
- [279] C. S. Burbano, Y. Liu, K. L. Rösner, V. M. Reis, J. Caballero-Mellado, B. Reinhold-Hurek, and T. Hurek, “Predominant nifH transcript phylotypes related to *Rhizobium rosettiformans* in field-grown sugarcane plants and in Norway spruce,” *Environ. Microbiol. Rep.*, vol. 3, no. 3, pp. 383–9, Jun. 2011.
- [280] A. Sessitsch, S. Gyamfi, N. Stralis-Pavese, A. Weilharter, and U. Pfeifer, “RNA isolation from soil for bacterial community and functional analysis: evaluation of different extraction and soil conservation protocols,” *J. Microbiol. Methods*, vol. 51, no. 2, pp. 171–9, Oct. 2002.
- [281] H. Burgmann, F. Widmer, W. V. Sigler, and J. Zeyer, “mRNA Extraction and Reverse Transcription-PCR Protocol for Detection of nifH Gene Expression by *Azotobacter vinelandii* in Soil,” *Appl. Environ. Microbiol.*, vol. 69, no. 4, pp. 1928–1935, Apr. 2003.
- [282] J. Borneman and E. W. Triplett, “Rapid and direct method for extraction of RNA from soil,” *Soil Biol. Biochem.*, vol. 29, no. 11–12, pp. 1621–1624, Nov. 1997.
- [283] T. Mendum, “The detection of Gram-negative bacterial mRNA from soil by RT-PCR,” *FEMS Microbiol. Lett.*, vol. 164, no. 2, pp. 369–373, Jul. 1998.
- [284] C. Mettel, Y. Kim, P. M. Shrestha, and W. Liesack, “Extraction of mRNA from soil,” *Appl. Environ. Microbiol.*, vol. 76, no. 17, pp. 5995–6000, Sep. 2010.
- [285] Y. Wang, S. Morimoto, N. Ogawa, T. Oomori, and T. Fujii, “An improved method to extract RNA from soil with efficient removal of humic acids,” *J. Appl. Microbiol.*, vol. 107, no. 4, pp. 1168–77, Oct. 2009.
- [286] Y. Wang, J. Shimodaira, T. Miyasaka, S. Morimoto, T. Oomori, N. Ogawa, M. Fukuda, and T. Fujii, “Detection of bphAa gene expression of *Rhodococcus* sp. strain RHA1 in soil using a new method of RNA preparation from soil,” *Biosci. Biotechnol. Biochem.*, vol. 72, no. 3, pp. 694–701, Mar. 2008.
- [287] J. B. Hughes, J. J. Hellmann, T. H. Ricketts, and B. J. M. Bohannan, “Counting the Uncountable : Statistical Approaches to Estimating Microbial Diversity,” *Appl.*

- Environ. Microbiol.*, vol. 67, no. 10, pp. 4399–4406, 2001.
- [288] E. L. M. Figuerola, L. D. Guerrero, D. Türkowsky, L. G. Wall, and L. Erijman, “Crop monoculture rather than agriculture reduces the spatial turnover of soil bacterial communities at a regional scale.,” *Environ. Microbiol.*, vol. 17, no. 3, pp. 678–88, Mar. 2015.
- [289] N. Fierer, J. P. Schimel, and P. a Holden, “Variations in microbial community composition through two soil depth profiles,” *Soil Biol. Biochem.*, vol. 35, no. 1, pp. 167–176, Jan. 2003.
- [290] C. Will, A. Thürmer, A. Wollherr, H. Nacke, N. Herold, M. Schrumpf, J. Gutknecht, T. Wubet, F. Buscot, and R. Daniel, “Horizon-specific bacterial community composition of German grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes.,” *Appl. Environ. Microbiol.*, vol. 76, no. 20, pp. 6751–9, Oct. 2010.
- [291] A. Agnelli, J. Ascher, G. Corti, M. T. Ceccherini, P. Nannipieri, and G. Pietramellara, “Distribution of microbial communities in a forest soil profile investigated by microbial biomass, soil respiration and DGGE of total and extracellular DNA,” *Soil Biol. Biochem.*, vol. 36, no. 5, pp. 859–868, May 2004.
- [292] P. Baldrian, M. Kolařík, M. Stursová, J. Kopecký, V. Valášková, T. Větrovský, L. Zifčáková, J. Snajdr, J. Rídl, C. Vlček, and J. Voříšková, “Active and total microbial communities in forest soil are largely different and highly stratified during decomposition.,” *ISME J.*, vol. 6, no. 2, pp. 248–58, Feb. 2012.
- [293] C. Lozupone, M. E. Lladser, D. Knights, J. Stombaugh, and R. Knight, “UniFrac: an effective distance metric for microbial community comparison.,” *ISME J.*, vol. 5, no. 2, pp. 169–72, Feb. 2011.
- [294] J. Wang, D. Zhang, L. Zhang, J. Li, W. Raza, Q. Huang, and Q. Shen, “Temporal variation of diazotrophic community abundance and structure in surface and subsoil under four fertilization regimes during a wheat growing season,” *Agric. Ecosyst. Environ.*, vol. 216, pp. 116–124, Jan. 2016.
- [295] J. Hamelin, N. Fromin, and S. Tarnawski, “nifH gene diversity in the bacterial community associated with the rhizosphere of *Molinia coerulea*, an oligonitrophilic perennial grass,” *Env. Microbiol.*, vol. 4, no. 8, pp. 477–481, 2002.
- [296] L. Duc, M. Noll, B. E. Meier, H. Bürgmann, and J. Zeyer, “High diversity of diazotrophs in the forefield of a receding alpine glacier,” *Microb. Ecol.*, vol. 57, no. 1, pp. 179–190, Jan. 2009.
- [297] J. E. Butler, N. D. Young, and D. R. Lovley, “Evolution of electron transfer out of the cell: comparative genomics of six *Geobacter* genomes.,” *BMC Genomics*, vol. 11, no. 1, p. 40, Jan. 2010.
- [298] D. E. Holmes, K. P. Nevin, and D. R. Lovley, “Comparison of 16S rRNA, nifD, recA, gyrB, rpoB and fusA genes within the family Geobacteraceae fam. nov.,” *Int. J. Syst. Evol. Microbiol.*, vol. 54, no. Pt 5, pp. 1591–9, Sep. 2004.
- [299] W. C. Lin, M. V. Coppi, and D. R. Lovley, “*Geobacter sulfurreducens* Can Grow with Oxygen as a Terminal Electron Acceptor,” *Appl. Environ. Microbiol.*, vol. 70, no. 4,

- pp. 2525–2528, Apr. 2004.
- [300] J. H. P. Kahindi, P. Woome, T. George, F. M. de Souza Moreira, N. K. Karanja, and K. E. Giller, “Agricultural intensification, soil biodiversity and ecosystem function in the tropics: the role of nitrogen-fixing bacteria,” *Appl. Soil Ecol.*, vol. 6, no. 1, pp. 55–76, Aug. 1997.
- [301] M. Laranjo, A. Alexandre, and S. Oliveira, “Legume growth-promoting rhizobia: an overview on the *Mesorhizobium* genus,” *Microbiol. Res.*, vol. 169, no. 1, pp. 2–17, Jan. 2014.
- [302] Q. Wang, J. F. Quensen, J. A. Fish, T. K. Lee, Y. Sun, J. M. Tiedje, and J. R. Cole, “Ecological patterns of *nifH* genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool,” *MBio*, vol. 4, no. 5, pp. e00592–13, Jan. 2013.
- [303] Q. Teng, B. Sun, X. Fu, S. Li, Z. Cui, and H. Cao, “Analysis of *nifH* gene diversity in red soil amended with manure in Jiangxi, south China,” *J. Microbiol.*, vol. 47, no. 2, pp. 135–141, 2009.
- [304] B. A. Whitton, Ed., *Ecology of Cyanobacteria II: Their Diversity in Space and Time*, vol. 27. Springer Science & Business Media, 2012.
- [305] O. Malam Issa, C. Défarge, Y. Le Bissonais, B. Marin, O. Duval, A. Bruand, L. P. D’Acqui, S. Nordenberg, and M. Annerman, “Effects of the inoculation of cyanobacteria on the microstructure and the structural stability of a tropical soil,” *Plant Soil*, vol. 290, no. 1–2, pp. 209–219, Dec. 2006.
- [306] C. M. Yeager, J. L. Kornosky, R. E. Morgan, E. C. Cain, F. Garcia-Pichel, D. C. Housman, J. Belnap, and C. R. Kuske, “Three distinct clades of cultured heterocystous cyanobacteria constitute the dominant N₂-fixing members of biological soil crusts of the Colorado Plateau, USA,” *FEMS Microbiol. Ecol.*, vol. 60, no. 1, pp. 85–97, Apr. 2007.
- [307] P. Irisarri, S. Gonnet, and J. Monza, “Cyanobacteria in Uruguayan rice fields: diversity, nitrogen fixing ability and tolerance to herbicides and combined nitrogen,” *J. Biotechnol.*, vol. 91, no. 2–3, pp. 95–103, Oct. 2001.
- [308] P. Cavacini, “Soil algae from northern Victoria Land (Antarctica),” *Polar Biosci.*, vol. 14, pp. 45–60, Feb. 2001.
- [309] D. Woebken, L. C. Burow, F. Behnam, X. Mayali, A. Schintlmeister, E. D. Fleming, L. Prufert-bebout, S. W. Singer, T. M. Hoehler, J. Pett-ridge, A. M. Spormann, A. Lo, M. Wagner, P. K. Weber, and B. M. Bebout, “Revisiting N₂ fixation in Guerrero Negro intertidal microbial mats with a functional single-cell approach,” pp. 1–12, 2014.
- [310] A. F. Khadem, A. Pol, M. S. M. Jetten, and H. J. M. Op den Camp, “Nitrogen fixation by the verrucomicrobial methanotroph ‘*Methylacidiphilum fumariolicum*’ SolV,” *Microbiology*, vol. 156, no. Pt 4, pp. 1052–9, Apr. 2010.
- [311] P. F. Dunfield, A. Yuryev, P. Senin, A. V Smirnova, M. B. Stott, S. Hou, B. Ly, J. H. Saw, Z. Zhou, Y. Ren, J. Wang, B. W. Mountain, M. A. Crowe, T. M. Weatherby, P. L. E. Bodelier, W. Liesack, L. Feng, L. Wang, and M. Alam, “Methane oxidation by

- an extremely acidophilic bacterium of the phylum Verrucomicrobia.,” *Nature*, vol. 450, no. 7171, pp. 879–82, Dec. 2007.
- [312] A. Pol, K. Heijmans, H. R. Harhangi, D. Tedesco, M. S. M. Jetten, and H. J. M. Op den Camp, “Methanotrophy below pH 1 by a new Verrucomicrobia species.,” *Nature*, vol. 450, no. 7171, pp. 874–8, Dec. 2007.
- [313] P. Sangwan, X. Chen, P. Hugenholtz, and P. H. Janssen, “Chthoniobacter flavus gen. nov., sp. nov., the first pure-culture representative of subdivision two, Spartobacteria classis nov., of the phylum Verrucomicrobia.,” *Appl. Environ. Microbiol.*, vol. 70, no. 10, pp. 5875–81, Oct. 2004.
- [314] A. Kielak, A. S. Pijl, J. A. Van Veen, and G. A. Kowalchuk, “Differences in vegetation composition and plant species identity lead to only minor changes in soil-borne microbial communities in a former arable field,” *FEMS Microbiol. Ecol.*, vol. 63, no. 3, pp. 372–382, Mar. 2008.
- [315] G. T. Bergmann, S. T. Bates, K. G. Eilers, C. L. Lauber, J. G. Caporaso, W. A. Walters, R. Knight, and N. Fierer, “The under-recognized dominance of Verrucomicrobia in soil bacterial communities,” *Soil Biol. Biochem.*, vol. 43, no. 7, pp. 1450–1455, Jul. 2011.
- [316] J. Liu, M. Peng, and Y. Li, “Phylogenetic diversity of nitrogen-fixing bacteria and the nifH gene from mangrove rhizosphere soil.,” *Can. J. Microbiol.*, vol. 58, no. 4, pp. 531–9, Apr. 2012.
- [317] W. Shu, G. P. Pablo, Y. Jun, and H. Danfeng, “Abundance and diversity of nitrogen-fixing bacteria in rhizosphere and bulk paddy soil under different duration of organic management,” *World J. Microbiol. Biotechnol.*, vol. 28, no. 2, pp. 493–503, Jul. 2011.
- [318] J. A. Izquierdo and K. Nüsslein, “Distribution of extensive nifH gene diversity across physical soil microenvironments.,” *Microb. Ecol.*, vol. 51, no. 4, pp. 441–52, May 2006.
- [319] F. Behrends Kraemer, “Influencia de la granulometría y la mineralogía en el comportamiento hidro-físico y estructural en suelos con distinta intensidad y secuencia de cultivos bajo siembra directa,” Universidad de Buenos Aires, 2015.
- [320] Y. A. Martel, C. R. De Kimpe, and M. R. Laverdière, “Cation-exchange Capacity of Clay-rich Soils in Relation to Organic Matter, Mineral Composition, and Surface Area,” *Soil Sci. Soc. Am. J.*, vol. 42, no. 5, p. 764, May 1978.
- [321] C. Reardon, H. Gollany, and S. Wuest, “Diazotroph community structure and abundance in wheat–fallow and wheat–pea crop rotations,” *Soil Biol. Biochem.*, 2014.
- [322] A. Mergel, K. Kloos, and H. Bothe, “Seasonal fluctuations in the population of denitrifying and N₂-fixing bacteria in an acid soil of a Norway spruce forest,” *Plant Soil*, vol. 230, no. 1, pp. 145–160.
- [323] B. Hai, N. Diallo, and S. Sall, “Quantification of key genes steering the microbial nitrogen cycle in the rhizosphere of sorghum cultivars in tropical agroecosystems,” *Appl. ...*, 2009.
- [324] M. P. Maqubela, P. N. S. Mnkeni, O. M. Issa, M. T. Pardo, and L. P. D’Acqui, “Nostoc cyanobacterial inoculation in South African agricultural soils enhances soil

structure, fertility, and maize growth,” *Plant Soil*, vol. 315, no. 1–2, pp. 79–92, Aug. 2008.