



UGR | **Universidad
de Granada**



CSIC

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



ESTACIÓN EXPERIMENTAL DEL ZAIDÍN

Universidad de Granada
Facultad de Ciencias
Consejo Superior de Investigaciones Científicas

TESIS DOCTORAL

**METAGENOMA Y DIVERSIDAD MICROBIANA DE LA RIZOSFERA DE
QUERCÍNEAS DEL ESPACIO PROTEGIDO DE SIERRA NEVADA: EFECTO DEL
GRADIENTE ALTITUDINAL Y DE UN INCENDIO FORESTAL.**

Memoria presentada por D. Antonio José Fernández González, Licenciado en
Biología, para optar al grado de Doctor

Fdo.: **Antonio José Fernández González**

VºBº del director de la Tesis Doctoral

Fdo.: **Dº Manuel Fernández López**
Doctor en Ciencias, sección Biológicas
Investigador Científico del CSIC

Universidad de Granada
Junio de 2014.

Editor: Editorial de la Universidad de Granada
Autor: Antonio José Fernández González
D.L.: GR 2057-2014
ISBN: 978-84-9083-241-7



ugr

Universidad
de Granada



CSIC

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



ESTACIÓN EXPERIMENTAL DEL ZAIDÍN

El doctorando Antonio José Fernández González y el director de la tesis D. Manuel Fernández López garantizamos, al firmar esta Tesis Doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección del director de la Tesis y, hasta donde nuestro conocimiento alcanza, en la realización del trabajo se han respetado los derechos de otros autores al ser citados cuando se han utilizado sus resultados o publicaciones.

Granada, 24 de junio de 2014.

Director de la Tesis

Fdo.: Manuel Fernández López

Doctorando

Fdo.: Antonio José Fernández González

Este trabajo de Tesis Doctoral ha sido realizado en el grupo de investigación de Ecología Genética de la Rizosfera, perteneciente al Departamento de Microbiología y Sistemas Simbióticos de la Estación Experimental del Zaidín (CSIC-Granada).

Durante la realización de la presente Memoria de Tesis Doctoral el doctorando ha colaborado en las siguientes publicaciones y presentaciones en congresos:

Publicaciones:

1. **Quesada del Bosque, M.E., Navajas-Pérez, R., Panero, J.L., Fernández-González, A. and Garrido-Ramos, M.A.** (2011). Sex-chromosome evolution in *Rumex hastatulus* (Polygonaceae), a satellite-DNA evolutionary analysis approach. *Genome* **54**: 253-260.
2. **Fernández-González, A.J., Cobo-Díaz, J.F., Villadas, P.J., Robles, A.B., Toro, N. y Fernández-López, M.** (2011). Análisis de la diversidad procariótica asociada a quercíneas (*Quercus ilex* sp. ballota y *Q. pyrenaica*) para la identificación de bioindicadores asociados a la evolución post-incendio y al cambio climático en Sierra Nevada. *Proyectos de investigación en parques nacionales 2007-2010*. L. Ramírez y B. Asensio (eds.) ISBN 978-84-8014-805-4, pp.: 159-174. Organismo Autónomo Parques Nacionales, Madrid.
3. **Torres-Cortés, G., Millán, V., Fernández-González, A.J., Aguirre-Garrido, J.F., Ramírez-Saad, H.C., Fernández-López, M., Toro, N. and Martínez-Abarca, F.** (2012). Bacterial community in the rhizosphere of a cactus species during dry and rainy seasons assessed by deep sequencing. *Plant and Soil* **357**: 275-288.
4. **Curiel Yuste, J., Barba, J., Fernández-González, A.J., Fernández-López, M., Mattana, S., Martínez-Vilalta, J., Nolis, P. and Lloret, F.** (2012). Changes in soil bacterial community triggered by drought-induced gap succession preceded changes in soil C stocks and quality. *Ecology and Evolution* **2**: 3016-3031.
5. **Cobo-Díaz, J.F., Martínez-Hidalgo, P., Fernández-González, A.J., Martínez-Molina, E., Toro, N., Velázquez, E. y Fernández-López, M.** (2014). The endemism *Genista versicolor*, from Sierra Nevada National Park in Spain, is nodulated by new putative *Bradyrhizobium* species and a novel symbiovar (sierranivadense). *Systematic and Applied Microbiology* **37**(3): 177-185.
6. **Curiel Yuste, J., Fernández-González, A.J., Fernández-López, M., Ogaya, R., Peñuelas, J., Sardans, J. and Lloret, F.** (2014). Strong functional stability of soil microbial

communities adapted to semiarid Mediterranean conditions and subjected to long-term shifts in baseline precipitation. *Soil Biology and Biochemistry* **69**: 223-233.

7. **Curiel Yuste, J., Fernández-González, A.J., Fernández-López, M., Ogaya, R., Peñuelas, J. and Lloret, F.** (2014). Functional diversification within bacterial lineages promotes wide functional overlapping between taxonomic groups in a Mediterranean soil forest. *FEMS Microbiology Ecology* DOI: 10.1111/1574-6941.12373.
8. **Calleja-Cervantes, M.E., Menéndez, S., Fernández-González, A.J., Irigoyen, I., Cibrián-Sabalza, J.F., Toro, N., Aparicio-Tejo, P.M. and Fernández-López, M.** Organic amendments in vineyards: Effects in mineral nutrition, soil quality, microbial community and gaseous emissions after 12 years of application. **Aceptado** en *European Journal of Soil Science*.
9. **Cobo-Díaz, J.F., Fernández-González, A.J., Villadas, P.J., Robles, A.B., Toro, N. and Fernández-López, M.** Assessment of the potential microbial nitrogen pathways in the rhizosphere of a Mediterranean forest after a wildfire. **Sometido** en *Soil Biology and Biochemistry*.

Participación en Congresos:

1. **Manuel Fernández-López, Pablo J. Villadas, Antonio J. Fernández-González and Nicolás Toro.** Analysis of the prokaryotic Diversity of *Quercus* sp. Rhizosphere. Póster del congreso BAGECO 10. Uppsala, Suecia, 15-19 de junio de 2009.
2. **Quesada del Bosque M.E., Navajas-Pérez R., Fernández-González A., Garrido-Ramos M.A.** Evolución del ADN satélite en la planta dioica *Rumex hastatulus* (Polygonaceae). Póster del XXXVII Congreso SEG (Sociedad Española de Genética). Torremolinos, Málaga, del 29 de septiembre al 2 de octubre de 2009.
3. **Fernández-González A.J., Villadas, P.J., Robles A.B., Toro N. y Fernández-López, M.** Análisis de la diversidad procariótica en la rizosfera de roble (*Quercus pyrenaica*) a diferentes altitudes en Sierra Nevada: Bioindicadores de cambio climático. Póster del XXII Congreso Nacional de Microbiología. Almería, 21-24 de septiembre de 2009.
4. **Fernández-López M, Fernández-González AJ, Cobo-Díaz JF, Villadas PJ & Toro N.** Forest rhizospheric ecosystem: prokaryotic diversity versus diazotrophy. Conferencia invitada de la XIII Reunión Nacional de la SEFIN y II congreso Luso-Español de Fijación de Nitrógeno. Zaragoza, 15-18 de junio 2010.

5. **Fernández-González, A.J., Cobo-Díaz J.F., P.J. Villadas, N. Toro y M. Fernández-López.** Identificación de bioindicadores de procesos ambientales en la rizosfera de quercíneas mediante metagenómica. Ponencia en el III Congreso Nacional de Biodiversidad. Granada, 8-10 de junio de 2011.
6. **Arone, G., A.J. Fernández-González, M. Fernández-López, D. Correa y E.J. Bedmar.** Estudio de la biodiversidad microbiana en chacras maiceras mediante pirosecuenciación. Ponencia en el III Congreso Nacional de Biodiversidad. Granada, 8-10 de junio de 2011.
7. **Menéndez, S., M. Calleja, B. Lasa, F. Cibriain, F. Aguirrezabal, I. Oria, I. Irigoyen, P.M. Aparicio-Tejo, A.J. Fernández-González, M. Fernández-López.** Efecto de la aplicación de fertilizantes orgánicos sobre la biodiversidad bacteriana del suelo. Póster del III Congreso Nacional de Biodiversidad. Granada, 8-10 de junio de 2011.
8. **Torres-Cortés G., V. Millán, A.J. Fernández-González, J.F. Aguirre-Garrido, H.C. Ramírez-Saad, M. Fernández-López, N. Toro y F. Martínez-Abarca.** Las estaciones lluviosa y seca determinan perfiles diferentes en la comunidad bacteriana asociada a cactáceas endémicas del altiplano central de México. Ponencia en el III Congreso Nacional de Biodiversidad. Granada, 8-10 de junio de 2011.
9. **Fernández-López, M., Fernández-González, A.J., Villadas, P.J., Cobo-Díaz, J.F. and Toro, N.** Metagenomic approach to the prokaryotic diversity of *Quercus* rhizosphere: biomarkers of forest evolution. Póster del 4th Congress of European Microbiologists FEMS2011. Ginebra, Suiza, 26-30 de junio de 2011.
10. **Curiel Yuste, Jorge. Barba-Ferrer, Josep. Fernández-González, Antonio José. Fernández-Lopez, Manuel. Pariade, Xavier. Martínez-Vilalta, Jordi, Lloret Maya, Francisco.** Microbial community shifts following climate-change driven ecological succession of a Mediterranean forest. Ponencia en el XII Congreso Europeo de la Federación Ecológica. Ávila, 25-29 de septiembre de 2011.
11. **A. Fernández-González, P. J. Villadas, J. Cobo-Díaz, N. Toro, M. Fernández-López, S. G. Tringe.** Diversity and evolution of the microbial communities at the rhizosphere of *Quercus ilex* subsp. *ballota* after a wildfire. Póster del congreso BAGECO 12. Liubliana, Eslovenia, 9-13 de junio de 2013.

12. **Fernández-López, M. Fernández-González, A.J., Cobo-Díaz, J.F., Villadas, P.J., Tringe, S.G., Toro, N.** Diversidad y evolución de las comunidades microbianas en la rizosfera tras un incendio forestal. Ponencia en la II IBEMPA Conference; III Spanish-Portuguese Congress on Nitrogen Fixation; XIV National Meeting of the Spanish Society of Nitrogen Fixation (SEFIN); XXVI Latin American Meeting on Rhizobiology (ALAR). Sevilla, 2-6 de septiembre 2013.

Para la realización del presente trabajo se dispuso de las siguientes formas de **financiación:**

1. Proyecto: Análisis de la diversidad procariótica asociada a quercíneas (*Quercus ilex* y *Q. pyrenaica*) para la identificación de biomarcadores asociados a la evolución post-incendio y al cambio climático en Sierra Nevada. OAPN 021/2007.

Entidad financiadora: Ministerio de Medio Ambiente, Organismo Autónomo Parques Nacionales.

Duración, desde: diciembre 2007 hasta: diciembre 2010

Investigador responsable: Dr. Manuel Fernández López

2. Proyecto: Identificación de bioindicadores microbianos en la rizosfera de quercíneas (*Quercus ilex* y *Q. pyrenaica*) asociados a cambio climático y evolución post-incendio. P08-CVI-03549.

Entidad financiadora: Junta de Andalucía.

Duración, desde: enero 2009 hasta: diciembre 2013

Investigador responsable: Dr. Manuel Fernández López.

3. Beca Predoctoral para la Formación del Profesorado Universitario (FPU).

Entidad financiadora: Ministerio de Educación, Cultura y Deporte.

Duración, desde: agosto 2009 hasta: enero 2014

Agradecimientos

Dicen que es de bien nacido ser agradecido.

Pues a todos los que lean, al menos, esta página del presente trabajo

decirles...

Gracias

Summary

Bacteria are the most abundant and diverse group of cellular microorganisms that inhabit the soil. Their great functional versatility gives them a key role in various biogeochemical cycles. The study of soil bacteria in the carbon (C) cycle has gained momentum in recent years, as the soil is the main reservoir of this element. Strategies to decrease loss of soil C and nitrogen (N) in gaseous form have huge potential to mitigate climate change, which is driven by the disproportionate accumulation of greenhouse gases such as CO₂, CH₄ and N₂O in our atmosphere. The study of microbial communities allows us to know the quality of our soils and their rates of production and retention of these greenhouse gases. Moreover, in the rhizosphere the microbial communities develop a very broad set of metabolic activities, which helps them interact with plants. The functions of rhizosphere microorganisms are poorly understood and their study has led to breakthroughs in both biotechnology and biomedicine. However, more than 99 % of these microorganisms are not culturable with currently available media which has limited taxonomic and functional studies of these organisms. With high-throughput, current generation sequencing (CGS) technologies, it is possible to explore these microbial communities to a depth previously unimaginable due to economic and time constraints. Thus, these new studies allow us to search for microbial biomarkers of environmental changes experienced by some forest formations, such as forests of oaks in the Sierra Nevada Natural Area, as the result of global warming. We can also seek microbial biomarkers of forest fire effects and know the quality and the recovery rate of burned soils.

The results of this study show no effect on the microbial communities by the altitudinal gradient; therefore climatic differences were not substantially different between the bioclimatic stages evaluated. But an effect of vegetation cover distinguished the different composition and structure of microbial communities observed in mature melojo oak forest, immature or expanding melojo oak forest and high mountain areas without melojo oaks. Melojo oak rhizosphere communities were dominated by some families from the *phylum* Proteobacteria, including Bradyrhizobiaceae, Sphingomonadaceae, Hyphomicrobiaceae and Caulobacteraceae and some families from the *phylum* Actinobacteria including Mycobacteriaceae and Nocardioidaceae. Non-rhizosphere communities were more diverse and dominated by families from diverse *phyla* including the Gemmatimonadaceae family from the *phylum* Gemmatimonadetes, the family Planctomycetaceae from the *phylum* Planctomycetes, families Conexibacteraceae and Geodermatophilaceae from the *phylum* Actinobacteria and the family Gp7 from the *phylum* Acidobacteria.

In this study no substantial differences were observed between the rhizosphere microbial communities of melojo oaks and holm oaks. These communities are species rich, with each sample harboring over 5,000 prokaryotic species, dominated by bacterial families belonging to the *phyla* Proteobacteria, Actinobacteria, Acidobacteria, Bacteroidetes, Verrucomicrobia and Planctomycetes, but one of the most abundant groups of bacteria could not be assigned to any known bacterial *phylum*.

Forest fires are one of the main factors reducing the taxonomic and functional diversity of soil microbial communities. We have evaluated the microbial communities associated with holm oaks 3 and 6 years after a forest fire, which occurred in 2005 in the Sierra Nevada Natural Area. This study has allowed us to understand the evolution of these communities in the recovery process and what are the key bacterial groups in that process. The Actinobacterial genera *Arthrobacter*, from the family Micrococcaceae, and *Blastococcus*, from the family Geodermatophilaceae, were the best adapted to the burned soils and the genus *Bradyrhizobium*, from the *phylum* Proteobacteria, was the most negatively affected by fire.

The study of the pan-genome of the genus *Arthrobacter* lets us know what features allow it to better adapt to the changing conditions of a disturbed by wildfire soil. In this case, the enzyme most abundantly represented among the gene sequences of the population of this genus was a heme peroxidase. These heme peroxidases not only allow *Arthrobacter* to be maintained through degradation of lignocellulosic compounds as alternative carbon sources, but also have a great biotechnological potential in waste management for degradation of recalcitrant compounds such as those rich in lignin.

Índice

Introducción	1
1. El estudio de las comunidades microbianas	3
1.1. Historia de la microbiología	3
1.2. El gen 16S <i>rRNA</i> como herramienta filogenética y taxonómica	6
1.3. Últimos avances metodológicos	10
1.4. El Metagenoma	13
1.5. Microorganismos del suelo	15
1.5.1. Microorganismos de la rizosfera	17
1.5.1.1. Efecto del Cambio Climático en las comunidades microbianas rizosféricas	19
1.5.1.2. Efecto de los incendios en las comunidades microbianas rizosféricas	21
2. Parque Natural y Nacional de Sierra Nevada	22
2.1. Gradientes altitudinales en Sierra Nevada.	24
Serie supra-mesomediterránea filábrica y nevadense malacitano-almijareense y alpujarreño-gadoreense silicícola de la encina (<i>Quercus ilex</i> subsp. <i>ballota</i>):	24
Serie mesomediterránea nevadense y malacitano-almijareense silicícola del roble melojo (<i>Quercus pyrenaica</i>):	25
Serie oromediterránea filábrico-nevadense silicícola del enebro rastrero (<i>Juniperus communis</i> subsp. <i>nana</i>):	26
2.2. Incendios forestales	26
Objetivos	29
Material y Métodos	33
3. Área de estudio	35
3.1. Características físico-químicas de los suelos	37
3.2. Inventario de la vegetación acompañante de las zonas de estudio	37
4. Registro de Humedad y Temperatura	38
5. Cultivos celulares y uso de antibióticos	38
3.1. Medios de cultivo bacteriano	39
3.2. Conservación a largo plazo de las cepas bacterianas	39
3.3. Antibióticos	40
6. Aislamiento de bacterias del género <i>Arthrobacter</i>	40
7. Toma de muestras y extracción de DNA	41
5.1. Recolección de las muestras	41
5.2. Extracción de DNA de muestras de suelo	41
5.3. Extracción de DNA plasmídico por precipitación con sales de magnesio	43
5.4. Cuantificación de la concentración de DNA	43
8. Manipulación y amplificación de DNA	44
6.1. Digestión con enzimas de restricción	45
6.2. Aproximaciones basadas en la PCR	45
6.2.1. Amplificación del gen 16S <i>rRNA</i> para TGGE	46
6.2.2. Construcción de librerías clásicas del gen 16S <i>rRNA</i>	47
6.2.3. Construcción de librerías NGS del gen 16S <i>rRNA</i>	50
6.3. Aproximación Metagenómica	55

9. Electroforesis de Ácidos Nucleicos	55
7.1 Electroforesis en geles de agarosa no desnaturalizantes	55
7.2 Electroforesis en gel desnaturalizante con gradiente de temperatura (TGGE)	56
10. Análisis Bioinformático	57
8.1 RDP's Pyrosequencing Pipeline	57
8.2 MOTHUR	62
8.3 FastTree	64
8.4 FastUnifrac	65
8.5 Procesamiento de datos metagenómicos	65
10.5.1. Limpieza de lecturas metagenómicas	65
10.5.2. Solapamiento y ensamblaje de lecturas Illumina	66
10.5.3. Anotación de secuencias metagenómicas	67
8.6 BLAST	69
8.7 STAMP	70
8.8 Ginkgo	71
8.9 CANOCO	71
8.10 Linux	71
Resultados	75
Capítulo 1: Análisis global de los suelos rizosféricos de quercíneas y de las comunidades microbianas que habitan en ellos	77
1. Análisis global de las características físico-químicas de los suelos muestreados	78
2. Análisis de la vegetación acompañante de los sitios muestreados	79
3. Análisis global de las comunidades microbianas de los suelos muestreados	83
4. Análisis de la Temperatura y Humedad de los sitios muestreados en la zona de Cáñar	89
Capítulo 2: Efectos del gradiente altitudinal en la estructura y composición de las comunidades microbianas asociadas a la rizosfera de roble melojo (<i>Q. pyrenaica</i> Willd.)	93
1. Clasificación taxonómica de las secuencias del gen 16S <i>rRNA</i>	94
2. Estudio de la diversidad α mediante el gen 16S <i>rRNA</i>	107
3. Estudio de la diversidad β mediante el gen 16S <i>rRNA</i>	114
Capítulo 3: Influencia de la especie arbórea en la estructura y composición de las comunidades microbianas asociadas a la rizosfera de quercíneas	125
1. Clasificación taxonómica de las secuencias del gen 16S <i>rRNA</i>	126
2. Estudio de la diversidad α mediante el empleo del gen 16S <i>rRNA</i>	139
3. Estudio de la diversidad β mediante empleo del gen 16S <i>rRNA</i>	146
Capítulo 4: Efecto del incendio en la estructura y composición de las comunidades microbianas asociadas a la rizosfera de encina (<i>Q. ilex</i> subsp. <i>ballota</i>)	153
1. Clasificación taxonómica de las secuencias del gen 16S <i>rRNA</i>	154
2. Estudio de la diversidad α mediante el gen 16S <i>rRNA</i>	165
3. Estudio de la diversidad β mediante empleo del gen 16S <i>rRNA</i>	172
Capítulo 5: Dominancia del género <i>Arthrobacter</i> en la rizosfera de encinas quemadas	181
1. Relevancia del género <i>Arthrobacter</i> en las comunidades estudiadas	183
2. Aislamiento de cepas del género <i>Arthrobacter</i>	187
3. Reconstrucción del pan-genoma de <i>Arthrobacter</i>	190
Discusión	197
1. Consideraciones previas	199

2.	Efectos del gradiente altitudinal en las comunidades microbianas	200
3.	Influencia de la especie arbórea en las comunidades microbianas de la rizosfera de quercíneas	202
4.	Efecto del incendio en las comunidades microbianas	204
5.	El género <i>Arthrobacter</i>	207
6.	Consideraciones finales y generales	210
	Conclusions	213
	Bibliografía	217

Introducción

1. El estudio de las comunidades microbianas

1.1. Historia de la microbiología

Hace 350 años el mundo microscópico era totalmente desconocido para el ser humano. Los primeros microorganismos fueron vistos en el año 1663 por Antonie van Leeuwenhoek, él los denominó animáculos y se trataba de bacterias extraídas de la superficie de sus propios dientes. Pudo ver dichos microbios gracias al desarrollo de lentes de aumento muy potentes que él mismo fabricó. Observó microorganismos de una amplia variedad de muestras medioambientales y los describió y dibujó detalladamente, lo que propició la aceptación de la comunidad científica de estas formas de vida microscópicas (Handelsman, 2004). Desde aquel momento se impulsa la observación de los microorganismos a través de microscopios cada vez más sofisticados. Gracias a observaciones meticulosas de dichos microorganismos, por científicos como Louis Pasteur, se pudieron establecer conexiones entre la actividad de estos seres vivos y fenómenos prácticos como por ejemplo, la producción de cerveza y la conexión con enfermedades en plantas, animales y humanos. Pero hasta que no se desarrollaron técnicas de cultivo a finales del siglo XIX los investigadores pudieron hacer poco más que observar. Gracias al aislamiento de los microorganismos en cultivos puros en 1884, Robert Koch y Friedrich Loeffler demostraron la teoría del germen de la enfermedad con una serie de pruebas (los postulados de Koch) para determinar si un microorganismo concreto era el agente causante de una enfermedad específica (de Kruif, 1926). En aquel momento comenzó la llamada Microbiología Moderna, aunque hoy en día la denominamos Microbiología Clásica, centrada en aislar en cultivo puro microorganismos, principalmente bacterias, para hacer estudios tanto morfológicos como fisiológicos muy focalizados en organismos patógenos. Los microorganismos cultivados han sido una inmensa fuente de conocimiento para los investigadores del siglo XX, lo que llevó a grandes avances en la comprensión de la fisiología y genética microbianas. Este hecho propició que pocos microbiólogos intentaran identificar y caracterizar microorganismos que no pueden cultivarse con los medios de cultivo desarrollados por el momento (Handelsman, 2004). Ya en 1986 Grimes y otros autores hicieron saltar la voz de alarma sobre un hecho frecuentemente observado en los laboratorios de microbiología: la drástica diferencia entre las poblaciones microbianas contadas en placas de cultivo y las células vivas contadas en el microscopio a partir de una misma muestra. Esta diferencia es incluso mayor en muestras de suelo, donde solo del 0,1 al 1 % de las bacterias que habitan en él son capaces de crecer en medios de

Introducción

cultivo estándar (Torsvik *et al.* 1990, 2002; Gans *et al.* 2005). Debido al cuello de botella que generó el conocimiento microbiano obtenido solo de cepas aisladas, algunos investigadores como Norman Pace en 1985 optaron por investigar la composición de las comunidades microbianas de distintos ambientes mediante técnicas basadas en secuenciación y que no precisaran del cultivo previo de los individuos (técnicas independientes de cultivo). Esta aproximación llevó al descubrimiento de linajes desconocidos hasta el momento en hábitats de todo el mundo, incluyendo bacterias y parásitos patógenos en el cuerpo humano (Choffnes *et al.* 2013). Pero para disponer de dichas técnicas fueron necesarios algunos descubrimientos fundamentales en la biología molecular:

- 1- El descubrimiento y aislamiento de una DNA polimerasa termo-estable con la que poder amplificar el DNA *in vitro*. Fue en el año 1968 cuando Thomas D. Brock encontró una DNA polimerasa estable hasta los 80° C en la especie bacteriana *Thermus aquaticus*, del *phylum* Deinococcus-Thermus, encontrada en las aguas termales de un manantial del Parque Nacional de Yellowstone (Chien *et al.* 1976).
- 2- La mejora de los métodos de secuenciación hasta conseguir una secuenciación rápida y precisa, gracias al método de los dideoxinucleótidos (ddNTPs) descrito por Sanger en 1977. Este método se ha mejorado durante las tres décadas en las que ha sido la técnica de secuenciación dominante, hoy día clásica, desde el uso de etiquetas radiactivas a fluorescentes en los ddNTPs, y desde el uso de geles de poliacrilamida a la electroforesis capilar para la separación de los fragmentos polimerizados (Rothberg y Leamon, 2008).
- 3- El desarrollo de la técnica de la reacción en cadena de la polimerasa (PCR) desarrollada por Kary Mullis en el año 1983 la cual permite la amplificación de una única molécula de DNA en miles o millones de copias *in vitro* y en pocas horas. Una técnica más rápida, económica y reproducible que la replicación *in vivo* mediante vectores bacterianos replicativos con alto número de copia (Choffnes *et al.* 2013).
- 4- El hallazgo de genes con las características indispensables para ser utilizados como cronómetros evolutivos (Handelsman, 2004). Estos genes ponen las bases sobre la construcción del árbol de la vida en la Tierra actual (TOL, Tree Of Life) y constituyen una herramienta filogenética de un valor incalculable particularmente en microorganismos que no se pueden aislar con los medios de cultivo disponibles. Se trata de los genes que transcriben los RNA ribosómicos (*rRNAs*) y aunque los primeros trabajos evolutivos que realizaron Pace y sus colaboradores en 1985 fueron con los genes 5S y pequeños fragmentos del 16S *rRNA*, con la mejora de las técnicas de

secuenciación y la PCR pronto se impuso la utilización del gen 16S *rRNA* completo (ver apartado 1.2) por su tamaño y por tener más regiones con distintas tasas de acumulación de mutaciones como la mejor herramienta para describir los microorganismos de una muestra ambiental sin cultivarlos (Woese, 1987, 1990).

Antes del uso estandarizado de los genes que transcriben *rRNAs* (rDNAs) impulsado por Woese y Pace a mediados de los años 90, la clasificación de las distintas cepas microbianas aisladas consistía en una batería de pruebas bioquímicas con las que los individuos se clasificaban de acuerdo a rasgos fenotípicos característicos. Estos resultados se complementaban con una técnica molecular que se llevaba a cabo con la extracción del DNA de las distintas cepas a comparar y mediante hibridación heteróloga del material genético de dos cepas (hibridación DNA-DNA), se consensuó que se trataba de la misma especie cuando la similitud DNA-DNA era al menos del 70 % (Woese 1987; Amann *et al.* 1995). Esta técnica molecular, la regla de oro del 70 % de similitud DNA-DNA, se utilizó en los años 90 para estimar la diversidad microbiana no cultivable (Torsvik *et al.* 1990) y se utilizó en algunos trabajos (Amann *et al.* 1992; Fox *et al.* 1992) como técnica confirmatoria para corroborar los resultados obtenidos mediante el uso de los rDNA como marcadores taxonómicos.

Las técnicas independientes de cultivo más utilizadas en los últimos años en estudios de diversidad microbiana, antes de la llegada de los métodos de secuenciación masiva (ver apartado 1.3), consistían en la amplificación de los genes ribosómicos (rDNA) del DNA extraído de las muestras a estudiar y la separación de los amplicones, bien mediante el uso de vectores que eran introducidos en *Escherichia coli* con la posterior individualización de las colonias seleccionadas (cloning) o bien mediante técnicas independientes del clonaje capaces de separar los amplicones de cada muestra en un conjunto de bandas que resulta en un patrón característico para cada una de ellas, conocido como huella genética o fingerprint de la comunidad microbiana (Amann *et al.* 1995; Hill *et al.* 2000). Una de las técnicas de fingerprinting que más se ha aplicado en estos estudios es la de TGGE (Temperature Gradient Gel Electrophoresis), desarrollada en 1981 por Thatcher y Hodson, que permite separar los amplicones mediante su desnaturalización parcial aplicando un gradiente térmico paralelo a la electroforesis en un gel de poliacrilamida (Riesner *et al.* 1989; Hill *et al.* 2000).

1.2. El gen 16S *rRNA* como herramienta filogenética y taxonómica

Los genes que transcriben los *rRNAs* están presentes en todos los organismos vivos debido a la función vital que desempeñan como parte de los ribosomas y que a su vez les confiere un alto grado de constancia funcional (una mutación que hace que pierdan su función puede ser letal o disminuir altamente la supervivencia del individuo), además tienen regiones que acumulan más mutaciones que otras a tasas de cambio muy diferentes (regiones hipervariables). Dichas características los convierte en buenos candidatos para ser utilizados como relojes moleculares en estudios filogenéticos y también pueden ser utilizados en un amplio rango taxonómico (desde dominio a especie) (Woese, 1987).

Los primeros estudios taxonómicos de muestras ambientales realizados con dichos genes se iniciaron con el transcrito del gen más pequeño, el 5S *rRNA* de 120 nucleótidos aproximadamente, pero debido a ese tamaño su utilización estaba limitada a ecosistemas poco complejos. Con la mejora de la secuenciación Sanger y la PCR se empezaron a utilizar fragmentos amplificados de al menos 1.000 bp (amplicones) de los genes 16S (\approx 1.500 bp) y 23S *rRNA* (\approx 3.000 bp). El tamaño de ambos genes contiene información suficiente para realizar estudios filogenéticos y taxonómicos robustos, incluso en comunidades microbianas muy diversas. Fue más fácil de amplificar completamente el gen 16S lo que hizo que los científicos comenzaran a decantarse por su uso como sonda en trabajos de hibridación con el DNA genómico (DNA-rDNA) de las cepas estudiadas o frente a librerías de DNA extraído de la muestra ambiental, fragmentado y encapsulado en el fago lambda. Posteriormente, se realizaron librerías de dicho gen mediante clonación en *Escherichia coli* que podían secuenciarse, por separado, y compararse con las secuencias conocidas con fines taxonómicos evitando a su vez el tedioso proceso de cribado o screening del método anterior (Woese, 1987; Amann *et al.* 1995).

La primera base de datos que contenía secuencias de genes ribosómicos fue la GenBank creada en 1982 (Bilofsky y Burks, 1988) que era y sigue siendo una base de datos de genes. Sin embargo la primera base de datos específica de genes rDNA fue publicada en 1991 por el grupo de Carl Woese como un proyecto de base de datos ribosomal (en inglés RDP). El RDP comprende un grupo de herramientas en continuo desarrollo para tratar de conocer las comunidades microbianas, centrada principalmente en la identificación de los individuos a través de herramientas taxonómicas que clasifican las secuencias de interés por un método que no necesita conocer ningún rasgo *a priori* de las mismas (Naïve Bayesian method) (Olsen *et al.* 1991; Wang *et al.* 2007). La otra parte fundamental del RDP es su base de datos que está

constituida exclusivamente por secuencias del RNA de la subunidad pequeña (SSU) del ribosoma de procariotas (*Bacteria* y *Archaea*), de plastos y mitocondrias, éste es el transcrito del gen 16S *rRNA*. En sus inicios la base de datos tenía secuencias del *rRNA* SSU de procariotas (16S) y eucariotas (18S) y del *rRNA* LSU (de la subunidad grande del ribosoma) 23S y 28S, respectivamente (Maidak, 1994). Desde el año 2011 se incluyó una base de datos de secuencias del gen 28S *rRNA* (LSU) de hongos y a fecha 3 de marzo de 2014 consta de 2.929.433 secuencias del gen 16S *rRNA* y 95.365 secuencias del gen 28S *rRNA* (<http://rdp.cme.msu.edu/>).

El uso de amplicones del gen 16S *rRNA* ha sido desde entonces un método de rutina para el estudio de las comunidades microbianas de muy diversos ambientes, pero esta metodología tiene algunos puntos débiles que se han tratado de superar con menor o mayor éxito. Uno de estos puntos débiles es el sesgo generado en la amplificación de dicho gen pues los cebadores empleados pueden tener más afinidad a unos taxones que a otros o en algunos microorganismos el DNA adyacente al gen 16S *rRNA* impide la unión de la DNA polimerasa y la amplificación posterior (Amann *et al.* 1995; Hansen *et al.* 1998; Baker *et al.* 2003; Wang *et al.* 2007). Otro punto débil es la formación de artefactos inherentes a la amplificación, como son las mutaciones debidas a errores de la DNA polimerasa, y al origen de la muestra, pues en muestras complejas aparecen quimeras que son más difíciles de detectar cuanto más relacionados filogenéticamente estén los individuos de los que se tomó el molde para amplificar el gen 16S *rRNA* (Shuldiner *et al.* 1989; Wintzingerode *et al.* 1997). Para tratar de solventar el primer punto, Baker y sus colaboradores en 2003 rastrearon las regiones conservadas del gen 16S *rRNA*, que flanquean las regiones hipervariables (Fig. I-1) y obtuvieron cebadores “universales” capaces de amplificar taxones a un rango más amplio del conseguido hasta aquel momento, mejorando sensiblemente la amplificación de individuos del dominio *Archaea*. Respecto al segundo punto, quizás mejor resuelto, la utilización de DNA polimerasas con actividad correctora de errores ha disminuido las mutaciones introducidas por éstas en la PCR. El uso de protocolos de PCR optimizados para reducir el tiempo y la cantidad de ciclos empleados en la reacción (Acinas *et al.* 2005) y el desarrollo y mejora de programas para la detección y eliminación de quimeras (Maidak *et al.* 1999; Ashelford *et al.* 2005, 2006; Nilsson *et al.* 2010; Haas *et al.* 2011; Quince *et al.* 2011; Edgar *et al.* 2011) consiguen reducir el número de secuencias quiméricas hasta valores inferiores al 1 % en los archivos con las secuencias ya procesadas (apartados 8.1 y 8.2 del Material y Métodos).

Para tratar de eliminar las secuencias quiméricas de un modo más eficaz es necesario que las bases de datos contengan secuencias de calidad, pero los controles de calidad iniciales

Introducción

dejaron de ser útiles con la masiva entrada de datos de organismos no cultivados en GenBank y la base de datos RDP. Así otros investigadores crearon una base de datos que tenía en cuenta filtros para quimeras desde un principio, la base de datos del gen 16S *rRNA* Greengenes (DeSantis *et al.* 2006). En 2007 un grupo de científicos alemán recopiló todas las secuencias de genes *rRNA* de todas las bases de datos existentes en aquel momento, filtraron tanto manual como automáticamente secuencias quiméricas y de baja calidad, generando así la primera base de datos de rDNA tanto de SSU (16S/18S) como de LSU (23S/28S) para los tres dominios: *Bacteria*, *Archaea* y *Eukarya*. Esta base de datos, SILVA, sigue en continua expansión y mejora, y en el año 2013 contenía más de 4.200.000 de secuencias (Quast *et al.* 2013). La coexistencia de las 3 bases de datos RDP, Greengenes y SILVA ha fomentado que los desarrolladores de todas ellas sigan mejorándolas, actualmente todas se mantienen actualizadas y realizan filtros de secuencias quiméricas. Durante nuestros estudios se ha utilizado el RDP por ser el que tenía la base de datos más completa al inicio de este trabajo, por albergar una amplia batería de herramientas [clasificación de secuencias, comparación entre distintas librerías y estudios de diversidad α y β entre otras (ver detalladamente en apartado 8.1 de Material y Métodos)] y por tener el método más rápido y fiable, hasta hoy, para clasificar secuencias del gen 16S *rRNA* (Wang *et al.* 2007; Mizrahi-Man *et al.* 2013).

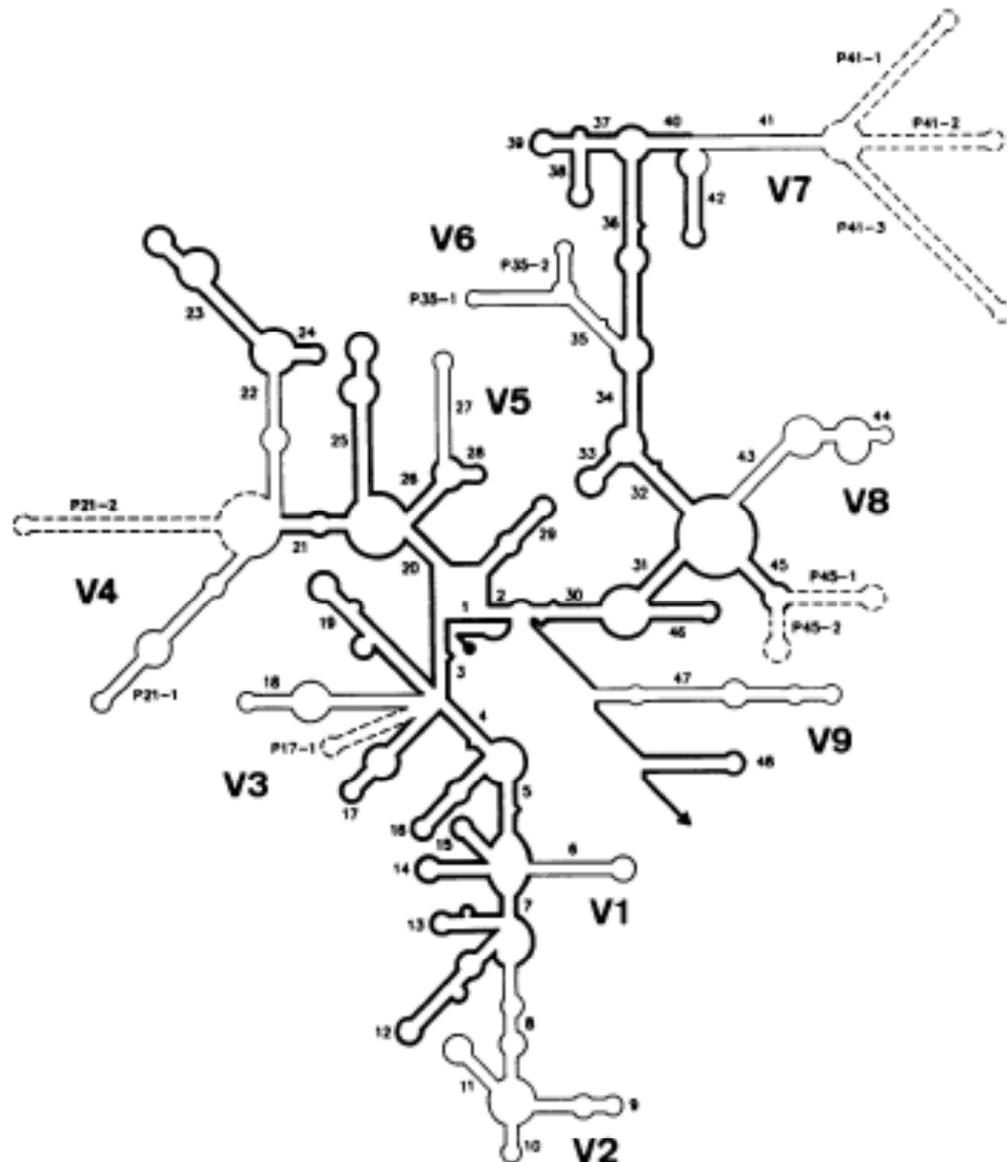


Figura 1. Estructura secundaria del 16S *rRNA*. El extremo 5' está representado con un punto, el 3' con una punta de flecha. Los números pequeños son las hélices comunes en procariotas y eucariotas (sólo números) y las hélices específicas de procariotas (números precedidos de la letra P). En negrita las áreas cuya secuencia está relativamente conservada en todos los procariotas. En línea discontinua aquellas áreas que no aparecen en todos los procariotas. En línea fina se muestran las 9 regiones hipervariables etiquetadas desde la V1 a la V9 (figura tomada del trabajo de Neefs *et al.* 1991).

1.3. Últimos avances metodológicos

Desde la publicación del primer borrador del genoma humano en el año 2001 y la finalización de su secuencia en el año 2003 (International Human Genome Consortium, 2004), los distintos grupos de investigación del Proyecto Genoma Humano (HGP) impulsaron un desarrollo vertiginoso de los métodos de secuenciación para abaratar los costes del proceso. La obtención de dicho genoma se llevó a cabo por el método de Sanger y tuvo un coste aproximado de 300 millones de dólares, así que el Instituto Nacional de Investigación en el Genoma Humano de los Estados Unidos de América (NHGRI) en 2004 fijó el objetivo de conseguir la secuenciación del genoma de un humano por \$1.000. Aunque actualmente el coste de un genoma humano supera en un orden de magnitud la meta propuesta por los investigadores, ronda los 20.000 \$ (McMorrow, 2010), los sistemas de secuenciación masiva o de alto rendimiento siguen evolucionando y actualmente su uso se ha extendido a la inmensa mayoría de grupos de investigación en ciencias de la vida. En microbiología ha supuesto una ventaja en el estudio de las comunidades microbianas a una profundidad antes impensable, debido al coste y tiempo que requería (Tringe y Hugenholtz, 2008; Rincon-Florez *et al.* 2013). Las plataformas de secuenciación masiva más utilizadas actualmente en estos estudios son las plataformas de Roche 454 GS FLX Titanium, 454 GS FLX+ y 454 GS Junior+ (<http://www.454.com>), las plataformas Illumina HiSeq2000 y MiSeq (<http://www.illumina.com>) catalogadas como tecnologías de secuenciación de segunda generación (SGS) para distinguirlas de la tecnología clásica (secuenciación Sanger) de primera generación (FGS) y la plataforma de Pacific Biosciences PacBio RS (<http://www.pacificbiosciences.com/>) que está catalogada como tecnología de secuenciación de tercera generación (TGS) (Schadt *et al.* 2010). Pese al uso de las tecnologías de secuenciación masiva desde su aparición hace casi una década, la primera fue la pirosecuenciación publicada por Margulies *et al.* (2005), hoy día sigue siendo frecuente la denominación de estas tecnologías como secuenciación de próxima generación (NGS, por sus siglas en inglés) (Mardis, 2008; Metzker, 2010; Rincon-Florez *et al.* 2013). Actualmente empiezan a aparecer trabajos que las denominan tecnologías de secuenciación de generación actual (CGS) para distinguirlas de las nuevas tecnologías NGS que empiezan a funcionar o siguen en desarrollo, como es el caso de las tecnologías MinION y GridION de Oxford Nanopore (<https://www.nanoporetech.com/>) (Choffnes *et al.* 2013).

Las tecnologías CGS incluyen una serie de métodos que se pueden agrupar en: preparación de las librerías de DNA, secuenciación y captación de imágenes de alta precisión

durante la misma y finalmente procesamiento de las imágenes para obtener un archivo virtual (SFF o FASTQ) con las secuencias y la calidad de cada nucleótido y el resto de información de la corrida (run) de la muestra. Las diferencias entre estas técnicas y las comentadas anteriormente en cuanto a tamaño, calidad y número de lecturas por run se resumen en la Tabla I-1.

Tabla I-1. Rendimiento de las diferentes plataformas de CGS comparadas con la secuenciación Sanger tradicional. * Tamaño medio de lectura. ** Secuenciación de lecturas pareadas (PE). (Adaptada de Solomon *et al.* 2014).

Plataforma	Tecnología	Tamaño de lectura* (bp)	Tamaño máximo del inserto (bp)	Tiempo de corrida (h - horas; d - días)	Lecturas por corrida	Tasa de error (%)	Aplicación biológica más recomendada
ABI 3730	Sanger	800	> 1.000	2 h	96	0.001	Método clásico de bajo rendimiento
454 GS FLX Titanium (Roche)	Pirosecuenciación (PCR en emulsión)	300-400	800	9 h	10 ⁶	1	Genómica < 7 Mbp Transcriptómica Diversidad 16S rDNA
454 GS FLX+ (Roche)	Pirosecuenciación (PCR en emulsión)	500-600	1.200	23 h	10 ⁶	1	Genómica microbiana Transcriptómica Diversidad 16S rDNA
454 GS Junior+ (Roche)	Pirosecuenciación (PCR en emulsión)	500-600	1.200	9 h	10 ⁵	1	Diversidad 16S rDNA
Illumina HiSeq 2000	Secuenciación por síntesis (PCR en puente)	101-151 2 x 101**	500	9-15 d	3 x 10 ⁹ 6 x 10 ⁹ **	< 1	Genómica Transcriptómica Metagenómica Metatranscriptómica Detección de SNPs
Illumina MiSeq	Secuenciación por síntesis (PCR en puente)	75-300 2 x 300**	500	20-55 h	2.5 x 10 ⁷ 5 x 10 ⁷ **	< 1	Genómica microbiana Transcriptómica Diversidad 16S rDNA
PacBio RS (Pacific Biosciences)	Secuenciación de una molécula en tiempo real (SMRT)	1.100	> 1.000	1.5 h	3.5 x 10 ⁷	8-15	Secuencias de gran tamaño útiles para reensamblado de genomas o transcriptomas

Por el tamaño y calidad de lecturas obtenidas (Tabla I-1) vemos que la pirosecuenciación es la técnica de secuenciación masiva que hasta hoy ofrece datos de calidad al mayor tamaño. Dado que el tamaño de lectura ofrecido por todas las plataformas de Roche es significativamente menor que el obtenido mediante Sanger (Tabla I-1), en estudios de diversidad microbiana mediante amplicones obtenidos por pirosecuenciación (Pyrotags) del gen 16S *rRNA*, ha sido necesaria la elección de parejas de cebadores capaces de amplificar una, dos y finalmente tres regiones hipervariables de dicho gen. Con la tecnología 454 GS FLX de

Introducción

Roche (250 bp), en desuso actualmente, han sido muy utilizadas las regiones hipervariables V1-V2 y V4 (Fig. I-1) para conseguir la clasificación taxonómica más fiable (próxima al 90 %) y más precisa (próxima al 80 %) al nivel más bajo de la base de datos del RDP (género) (Wang *et al.* 2007; Lauber *et al.* 2009). A partir de la tecnología 454 GS FLX Titanium de Roche se han realizado estudios de diversidad con zonas que comprenden varias regiones hipervariables, una de las zonas más utilizadas en dichos estudios es la zona V6-V8 que engloba las regiones V6, V7 y V8 (Fig. I-1). Las secuencias de la zona V6-V8 consiguen una clasificación taxonómica más fiable y precisa que cualquier región hipervariable por sí misma, además están flanqueadas por dos regiones muy conservadas del gen ribosómico, por lo que pueden utilizarse cebadores universales capaces de amplificar el DNA de procariontes con un sesgo menor hacia el dominio *Bacteria* e incluso amplificar el gen 18S de algunos eucariotas (Engelbrektson *et al.* 2010; Lundberg *et al.* 2012). Otra zona utilizada en estudios de diversidad microbiana ha sido la V4-V5 (Torres-Cortés *et al.* 2012; Yuste *et al.* 2012, 2013) que, pese a estar más sesgada al dominio *Bacteria* que la zona V6-V8, consigue clasificar un mayor porcentaje de secuencias a todos los niveles taxonómicos (comunicación personal del Doctor Julien Tremblay; datos de este trabajo). Aunque se han realizado estudios similares con la plataforma MiSeq de Illumina en los que el punto fuerte es la obtención de un número de secuencias (itags) de más de un orden de magnitud superior a los pyrotags obtenidos con las plataformas 454 FLX Titanium y FLX+ de Roche (Bartram *et al.* 2011; Caporaso *et al.* 2011), su punto débil, un tamaño de lectura de 150 bp, ha provocado un retraso en el uso de esta plataforma en estudios de diversidad microbiana. Actualmente la mejora, respecto al tamaño de lectura (Tabla I-1), de la plataforma MiSeq de Illumina comienza a ser una alternativa, por su mayor profundidad, en el estudio de la diversidad de las comunidades microbianas (Lindemann *et al.* 2013).

1.4. El Metagenoma

Handelsman y colaboradores en el año 1998 utilizaron por primera vez el término metagenoma para referirse a la colección de genomas microbianos que clonaron en *E. coli* a partir de DNA extraído de suelo. El DNA extraído de una muestra, sea posteriormente clonado y secuenciado o secuenciado directamente mediante tecnologías CGS, no pertenece exclusivamente a la microbiota de la misma, pues hay que tener en cuenta que también se extrae el DNA de la macrofauna y/o macroflora que coexiste con la comunidad microbiana objeto de estudio, como ocurre por ejemplo al extraer el DNA de la microbiota hospedada en un humano. Un término utilizado habitualmente para distinguir el metagenoma de una muestra del metagenoma exclusivamente microbiano es el microbioma. Pero el término microbioma acuñado por el Nobel en fisiología y medicina Joshua Lederberg en el año 2001 hacía referencia a la comunidad microbiana que hospeda el ser humano a lo largo de su vida y por lo tanto es sinónimo de microbiota, la mayoría lo han utilizado para referirse al metagenoma microbiano humano (Hooper y Gordon, 2001; Mardis, 2008). Otros autores lo han utilizado para referirse al metagenoma microbiano de cualquier ambiente natural (Sleator *et al.* 2008), haciendo una distinción entre microbiota entendida como la comunidad microbiana y microbioma entendido como el conjunto de genomas de la microbiota. Actualmente distintos investigadores están utilizando el término microbioma definido por Lederberg, como sinónimo de microbiota (Lundberg *et al.* 2012; Rincon-Florez *et al.* 2013; Choffnes *et al.* 2013;). La falta de acuerdo en el uso del término microbioma nos hace decantarnos por el uso del término metagenoma tal cual lo definió Handelsman, teniendo en cuenta los métodos de muestreo y extracción del DNA metagenómico estándar empleados en este estudio y la poca proporción de microorganismos eucarióticos presentes en los metagenomas (Piganeau *et al.* 2008; Lin *et al.* 2014). Entendiendo el metagenoma como la información genómica obtenida de la comunidad microbiana, esencialmente de origen procariótico, de un ambiente natural.

Así la metagenómica puede definirse como el análisis genómico, independiente de cultivo, de todos los microorganismos en un nicho en particular (Handelsman *et al.* 1998) que se ha desarrollado con el propósito de descubrir la diversidad estructural y la versatilidad funcional microbianas de ambientes naturales tales como la superficie y cavidades humanas, ambientes marinos y suelo (Sleator *et al.* 2008).

La diversidad microbiana describe la complejidad y variabilidad biológica a distintos niveles de organización de la comunidad, teniendo en cuenta la variabilidad genética entre distintos taxones (especies), el número de taxones (riqueza) y la abundancia relativa (equidad) de cada uno de ellos (Torsvik *et al.* 2002). Para conocer la diversidad microbiana a nivel de especie primero hay que tratar de definir qué es una especie. Actualmente la mayoría de científicos aceptan la clasificación artificial en filotipos u OTUs (unidades taxonómicas operativas) de los microorganismos considerando que son miembros de una misma especie si entre ellos existe una similitud mínima del 97 % al comparar sus respectivas secuencias del gen 16 *rRNA* (Stackebrant y Goebel, 1994; Roselló-Mora y Amann, 2001). Hoy en día, esta técnica es utilizada de forma rutinaria con DNA extraído de las muestras de ambientes tan diversos como el suelo, sin aislar los microorganismos y sin clonar, y directamente amplificado y secuenciado de forma masiva (Torres-Cortés *et al.* 2012; Yuste *et al.* 2012, 2013). Aunque los estudios de diversidad microbiana llevados a cabo sólo con un gen, o parte de éste, tienen sus limitaciones, con los avances en metagenómica de los últimos años empieza a ser usual la comparación con las bases de datos de grandes fragmentos ensamblados de los distintos genomas secuenciados al azar de la comunidad microbiana estudiada, evitando a su vez sesgos de la PCR (Poretsky *et al.* 2014). Este proceso tiene también desventajas comparado con las técnicas basadas en 16S *rRNA*, pues los fragmentos ensamblados (contigs o scaffolds) pueden ser quimeras de individuos de diferentes especies pertenecientes al mismo género o ser una zona del genoma adquirida por transferencia horizontal que dará homología con el genoma del individuo donador, que puede estar filogenéticamente muy alejado del individuo receptor (Wooley y Ye, 2009). Así, algunos investigadores aconsejan el uso de ambas técnicas para un profundo estudio de las comunidades microbianas (Tringe y Hugenholtz, 2008).

1.5. Microorganismos del suelo

El suelo es un ambiente complejo habitado por un amplio rango de organismos eucarióticos, procarióticos y virales. Sabemos que los microorganismos del suelo intervienen en diversos ciclos biogeoquímicos a través de la degradación de la materia orgánica y de su papel en el ciclo del nitrógeno y en la biodisponibilidad de diversos minerales quelados en la matriz inorgánica del suelo, proveyendo así a las plantas con nutrientes esenciales. Aunque poco sabemos todavía sobre la diversidad, las funciones y las interacciones de estos microorganismos (Uroz *et al.* 2013), si sabemos que pueden afectar al crecimiento de las plantas por la liberación de sustancias que estimulan o inhiben su crecimiento y que tienen efecto directo en la arquitectura y fisiología de la raíz (Rincón-Flórez *et al.* 2013). También se ha descrito que la diversidad microbiana, esencialmente procariótica, en el suelo es mucho mayor que la de organismos eucarióticos, pues se considera a los procariotas como el componente principal de la biomasa del mismo y se estima que en 1 gramo de suelo pueden residir del orden de 10^9 células procarióticas de entre 10^4 a 10^6 especies diferentes (Gans *et al.* 2005), siendo uno de los ambientes naturales más diverso y complejo de estudiar a nivel microscópico (Fierer y Lennon, 2011). Se piensa que la elevada heterogeneidad espacial, la estructura polifásica (gases, agua y materia sólida) y las complejas propiedades químicas y biológicas del suelo contribuyen ampliamente a la diversidad microbiana presente en el mismo (Torsvik *et al.* 2002; Daniel, 2005; Roesch *et al.* 2007; Quince *et al.* 2008). La diversidad funcional y taxonómica de las comunidades microbianas del suelo se considera un buen indicador del funcionamiento del ecosistema ya que está fuertemente ligada a factores ambientales tales como las características edáficas (pH, concentración y biodisponibilidad de los nutrientes), las modificaciones climáticas (disminución de lluvias y aumento de temperaturas debidas a la aceleración del calentamiento global) y las interacciones con otros organismos (Uroz *et al.* 2013).

La secuenciación del metagenoma del suelo está aportando un importante valor económico y ambiental a nuestra sociedad. Puesto que, el suelo representa uno de los reservorios genéticos más desconocidos (Lombard *et al.* 2011) su comunidad microbiana se considera una mina de oro repleta de genes y rutas metabólicas que codifican biocatalizadores para procesos de biosíntesis o de biodegradación, incluyendo degradación de contaminantes, síntesis de biocombustible y producción de nuevos fármacos. Además, la secuenciación del metagenoma del suelo está aportando conocimientos de la ecología de microorganismos que

mejoran o perjudican la producción de los cultivos y de aquellos microorganismos que aseguran la calidad y el buen funcionamiento del ecosistema (Vogel *et al.* 2009).

Hasta ahora se conoce que los taxones mayoritarios en las comunidades bacterianas del suelo pertenecen a los *phyla* Proteobacteria, Acidobacteria, Actinobacteria y Firmicutes (Roesch *et al.* 2007; Acosta-Martínez *et al.* 2008; Uroz *et al.* 2010), de los que se tienen bien caracterizados funcionalmente algunos de sus géneros, de los que se tienen cepas aisladas. Sin embargo, de otros *phyla* microbianos dominantes en suelo como Bacteroidetes, Verrucomicrobia, Chloroflexi y Gemmatimonadetes apenas se conocen los rasgos funcionales de los individuos que habitan en el suelo y que pertenecen a dichos taxones. Este hecho es debido principalmente a la dificultad de aislarlos en medios conocidos y a la falta, hasta ahora, de técnicas independientes de cultivo que puedan ayudarnos a esclarecer dichas incógnitas (Uroz *et al.* 2013). También se ha descrito que, de forma general, los microorganismos del suelo pertenecientes a los *phyla* Proteobacteria, Bacteroidetes y Verrucomicrobia son relativamente más abundantes en suelos ricos en nutrientes o en zonas de los mismos en los que las fuentes de carbono son fácilmente accesibles para ellos. Por el contrario, los microorganismos pertenecientes a los *phyla* Acidobacteria, Firmicutes y Chloroflexi son relativamente más abundantes en suelos pobres en nutrientes o con fuentes de carbono recalcitrantes, en su mayoría compuestos aromáticos de difícil degradación. Este patrón oligotrófico se presenta también en microorganismos pertenecientes al dominio Archaea pese a su menor abundancia relativa respecto al dominio Bacteria en la mayoría de suelos naturales (Bergmann *et al.* 2011; Delmont *et al.* 2012; Uroz *et al.* 2013). Pero todos estos estudios se centran en las características funcionales de los microorganismos que habitan el suelo desde un punto de vista general, a nivel de *phylum* y de clase, quedando algunos taxones predominantes en estos ecosistemas, como el *phylum* Actinobacteria, sin una clasificación ecológica clara, pues para algunos autores la abundancia relativa de dicho taxón es independiente de la disponibilidad de los nutrientes (Uroz *et al.* 2013) y para otros disminuye con el aumento de nitrógeno disponible en el suelo (Fierer *et al.* 2012).

1.5.1. Microorganismos de la rizosfera

La rizosfera es la zona del suelo que está bajo la influencia de las secreciones radicales y puede contener hasta 10^{11} células microbianas por gramo de raíz (Egamberdieva *et al.* 2008) y más de 30.000 especies procarióticas (Mendes *et al.* 2011), lo cual la convierte en uno de los ambientes más complejos ya que en ella tienen lugar miles de interacciones con un papel crucial en la salud de la planta. Las plantas secretan a la rizosfera hasta el 40 % de los fotosintatos que producen por medio de sus raíces (Berendsen *et al.* 2012). La mayoría de los suelos son deficientes en carbono y ya que la rizosfera es más rica en compuestos carbonados que el resto del suelo (suelo suelto), la densidad microbiana es entre uno y tres órdenes de magnitud mayor en esta zona debido al intercambio de nutrientes entre la planta, que libera compuestos ricos en carbono, y los microorganismos que rodean sus raíces, que aportan a la planta nitrógeno, fósforo y otros minerales al descomponer la materia orgánica (Singh *et al.* 2004; Smalla *et al.* 2006). La calidad y cantidad de exudados radicales están asociadas a la especie de planta, su estado nutricional, el tipo de suelo y su estadio de desarrollo y el tipo de microorganismos que están interaccionando. Conocer mejor estos elementos puede ayudar a desarrollar técnicas que potencien el crecimiento y mantenimiento de los microorganismos beneficiosos en la rizosfera (Rincón-Flórez *et al.* 2013) para mejorar el rendimiento de los cultivos, recuperar zonas contaminadas o incrementar el éxito en los trabajos de reforestación.

La composición de los microorganismos que colonizan la rizosfera puede ser el resultado de un proceso de selección positiva, negativa o ambas, por parte de la planta (Berg y Smalla, 2009). La colonización de la rizosfera y el establecimiento son importantes para que los microorganismos puedan desarrollar su actividad adecuadamente. Para que se establezca la colonización debe existir un reconocimiento entre la planta y el microorganismo. Dicho reconocimiento se basa en interacciones que implican el intercambio de señales moleculares entre ambos y las más conocidas son el caso de la simbiosis *Rhizobium*-leguminosa que es un ejemplo de interacción positiva y la enfermedad del tumor de cuello producido por especies del género *Agrobacterium* como ejemplo de interacción negativa (Halverson y Stacey, 1986). Tras el reconocimiento se produce la adherencia que en algunos casos se lleva a cabo mediante la formación de biopelículas por parte de los microorganismos para poder adherirse a la raíz (Rodríguez-Navarro *et al.* 2007). En el caso de los endófitos y de algunos patógenos tras la adhesión se produce la invasión del espacio inter o intracelular del tejido radicular. Y finalmente se establece la colonización y el crecimiento de los microorganismos en el nicho

colonizado. Por tanto, la rizosfera puede considerarse como un microambiente único en los ecosistemas terrestres y está integrado en una compleja red alimenticia tanto macro como microbiana. Como se ha dicho anteriormente, un mejor conocimiento de las interacciones planta-microorganismo en la rizosfera podrían ser de gran importancia en proyectos de reforestación de bosques degradados con las especies arbóreas nativas de los mismos. Aunque son bien conocidos los sistemas de interacción planta-microorganismo en algunos sistemas agrícolas, el proceso de enriquecimiento de algunos microorganismos por parte de la planta y los mecanismos de especificidad apenas se conocen en sistemas naturales (Berg y Smalla, 2009). El entendimiento de estos sistemas podría ayudarnos en procesos clave como la descontaminación de espacios naturales mediante la mejora de las técnicas de biorremediación como la bioaumentación, la fitorremediación y la rizorremediación (Kuiper *et al.* 2004).

Las plantas son capaces de moldear su propio microbioma rizosférico (Berendsen *et al.* 2012; Lundberg *et al.* 2012). En algunas especies de plantas se ha demostrado que atraen y hospedan a microorganismos concretos capaces de protegerlas suprimiendo la actividad de organismos patógenos (Mendes *et al.* 2011). Aunque los patógenos pueden afectar de forma severa a la salud de la planta, existen hongos y bacterias alrededor de las raíces o en su interior (endófitos) que compiten con estos patógenos por el espacio y los nutrientes; por tanto ejerciendo un efecto antagonista en ellos (Raaijmakers *et al.* 2009; Nihorimbere *et al.* 2011). A estas bacterias beneficiosas se las conoce generalmente como rizobacterias promotoras del crecimiento vegetal (PGPR).

Las PGPR crecen en el interior, sobre o alrededor de la raíz vegetal y potencian el crecimiento de la planta, incrementan su rendimiento, la protegen frente a patógenos directa o indirectamente y/o reducen el estrés biótico y/o abiótico (como el estrés hídrico) (Vessey, 2003). Algunas de estas bacterias producen fitohormonas que pueden inhibir o promover el crecimiento de las raíces, pueden proteger a la planta ante distintos estreses y mejorar la adquisición de nutrientes a través de la raíz (Berg, 2009; Davies, 2010). Además las PGPR representan una alternativa sostenible a los fertilizantes químicos y los pesticidas (Rincón-Flórez *et al.* 2013).

Para estudiar las comunidades microbianas de la rizosfera se utilizan técnicas moleculares, análisis de la diversidad taxonómica con el gen 16S *rRNA* y funcional con metagenómica, pero debemos tener en cuenta que por ciertas características intrínsecas a las muestras del suelo a menudo supone un desafío. Los problemas más comunes son la presencia

de compuestos orgánicos tales como ácidos húmicos y fúlvicos que inhiben la actividad enzimática en el proceso de amplificación y el rendimiento en el proceso de extracción del DNA (Rincón-Flórez *et al.* 2013). Pese a ello, se ha conseguido demostrar mediante dichas técnicas el efecto de la rizosfera sobre determinados taxones de las comunidades microbianas. Con una selección preferente de algunos microorganismos pertenecientes a las clases β - y γ -Proteobacteria, siendo el género *Burkholderia* (β -Proteobacteria), descrito como solubilizador de nutrientes quelados en la fase mineral del suelo (Uroz *et al.* 2009, 2010), uno de los más abundantes en rizosfera de pastos (Kielak *et al.* 2008) y bosques de quercíneas (Uroz *et al.* 2010). Por otro lado, algunos microorganismos pertenecientes a la clase α -Proteobacteria y la mayoría de bacterias pertenecientes al *phylum* Acidobacteria han mostrado ser relativamente más abundantes en suelos no rizosféricos (Uroz *et al.* 2010).

1.5.1.1. Efecto del Cambio Climático en las comunidades microbianas rizosféricas

Está globalmente aceptado que los microorganismos han tenido un papel clave en la determinación de las concentraciones atmosféricas de los gases de efecto invernadero, dióxido de carbono (CO₂), metano (CH₄) y óxido nitroso (N₂O), en gran parte de la historia de la Tierra. Pero no está tan claro el efecto directo o indirecto que el cambio climático acelerado por la acción humana, que está llevando a una alteración en el régimen de lluvias y un calentamiento global debido al incremento de las concentraciones de estos gases, está ejerciendo sobre las comunidades microbianas heterótrofas terrestres. En la rizosfera la liberación de exudados radicales es el proceso clave por el cual el carbono se transfiere al suelo. Los mecanismos que influyen en el almacenaje del carbono orgánico en el suelo están regulados por la mineralización e inmovilización del mismo mediante la actividad de la comunidad microbiana de la rizosfera. Estos efectos se conocen mejor en las plantas y en los microorganismos simbióticos de las mismas. Así, entender la fisiología y las dinámicas de las comunidades microbianas terrestres es esencial si queremos conocer mejor los mecanismos que intervienen en el flujo de los gases de efecto invernadero teniendo en cuenta que el suelo representa un reservorio masivo de carbono potencialmente volátil y que su mantenimiento en el suelo o su liberación a la atmósfera depende del balance entre la fotosíntesis, la respiración de los organismos descomponedores y la biomineralización (Singh *et al.* 2004, 2010; Trivedi *et al.* 2013). El *phylum* Proteobacteria junto con el *phylum* Actinobacteria son considerados copiotróficos y en términos generales el aumento de individuos pertenecientes a dichos *phyla* en las comunidades microbianas conlleva a un aumento de la descomposición de las fuentes de carbono lábiles y un aumento de CO₂ atmosférico; por el contrario, el *phylum* Acidobacteria

se considera oligotrófico y presenta el patrón opuesto. Pero no todos los grupos taxonómicos de estos *phyla* pueden ser considerados copiotróficos u oligotróficos (Monson *et al.* 2006), además de que algunos de ellos, como las bacterias metanotróficas del *phylum* Proteobacteria que retiran hasta 30 millones de toneladas al año de CH₄ (IPCC, 2007), contribuyen a la reducción atmosférica de otros gases de efecto invernadero, por lo que no podemos utilizar un rango taxonómico tan amplio para predecir los flujos de entrada y salida de los gases de efecto invernadero en el suelo (Fierer *et al.* 2007; Singh *et al.* 2010). Así investigar el papel y las capacidades fisiológicas de cada taxón individualmente nos permitirá predecir si un suelo particular es emisor o sumidero de gases de efecto invernadero. También hay que tener en cuenta los factores abióticos de cada suelo, como son la humedad, la concentración de nutrientes y su biodisponibilidad para predecir si la disminución de las precipitaciones y el aumento de las temperaturas llevará a un aumento (en el caso de humedales) o una disminución (en el caso de bosques templados) de la actividad de las comunidades microbianas que habitan en él (Singh *et al.* 2010).

Debido a la aceleración del cambio climático y a la gestión indebida del suelo, el cual es un recurso no renovable, la presión medioambiental, principalmente el aumento de temperatura y la alteración de las precipitaciones, sobre el mismo es cada vez mayor (Griffiths y Philippot, 2013). Pese a que la supervivencia y el metabolismo de los microorganismos del suelo están muy ligados a la disponibilidad de agua y nutrientes (Daniel, 2005) las comunidades microbianas muestran una fuerte estabilidad funcional debido a dos componentes relacionadas de forma directa con la diversidad de las mismas: resistencia (el grado de insensibilidad de la comunidad ante una perturbación) y resiliencia (la tasa a la que la comunidad regresa a las condiciones previas a la perturbación o se estabiliza en unas nuevas condiciones) las cuales pueden ayudar a contrarrestar el estrés producido por la alteración del Clima (Griffiths y Philippot, 2013; Yuste *et al.* 2013). Así conocer la estructura y las funciones de la comunidad microbiana rizosférica puede mejorar la capacidad de monitorizar las alteraciones del suelo sometido a estrés medioambiental, evaluar su capacidad de recuperación e incluso mejorar la capacidad de detección por medio de bioindicadores microbianos de la alteración de los ecosistemas antes de que el daño sea irreparable (Hartmann *et al.* 2014).

1.5.1.2. Efecto de los incendios en las comunidades microbianas rizosféricas

Se ha descrito el profundo efecto que tienen los incendios en los ecosistemas forestales y particularmente en el suelo rizosférico en el cual un gran número de propiedades físico-químicas se ven alteradas. En la superficie del suelo se llegan a alcanzar los 1.000° C y la eliminación de la vegetación y la hojarasca tras el incendio reduce el albedo y también genera un aumento de la temperatura superficial. Este aumento tan severo de la temperatura del suelo rompe la estructura de la materia inorgánica creando capas hidrofóbicas que disminuyen la infiltración del agua incrementando la erosión del mismo. A su vez la concentración de los nutrientes se altera, con aumentos en fósforo, potasio y magnesio, al tiempo que se forman compuestos tóxicos y recalcitrantes para los microorganismos, y el pH aumenta tras el incendio debido al aumento de cationes básicos depositados en el suelo durante la combustión. Se produce una liberación masiva del carbono almacenado en el mismo contribuyendo además a la acumulación de CO₂ atmosférico. El fuego no solo afecta a la estructura del suelo sino que también afecta a la biomasa y a la estructura de las comunidades microbianas que habitan en él. La biomasa en la mayoría de casos disminuye tras el incendio y puede tardar varios años en recuperarse. La estructura de la comunidad también queda alterada disminuyendo la diversidad (Smith *et al.* 2008; Bárcenas-Moreno *et al.* 2011), por lo que disminuye la resistencia y resiliencia del ecosistema dejándolo temporalmente desprotegido ante otras perturbaciones. Aunque se ha estudiado la alteración de las comunidades microbianas inmediatamente después de un incendio natural o provocado (Bárcenas-Moreno *et al.* 2011), pocos estudios se han llevado a cabo, para determinar la evolución de las comunidades microbianas tras el proceso de recolonización post-incendio en un periodo de tiempo mayor (Fioretto *et al.* 2005; Bárcenas-Moreno *et al.* 2011). Actualmente sólo hay un estudio que haya secuenciado a gran profundidad con técnicas CGS el gen 16S *rRNA* de una comunidad microbiana post-incendio y que haya tenido en cuenta los efectos del incendio a corto plazo, un año y a largo plazo, once años tras el incendio (Xiang *et al.* 2014). El estudio de Xiang y colaboradores se realizó con suelos de un bosque boreal y no se estudió el metagenoma, restringiéndose al estudio de la diversidad taxonómica del mismo. Aun así Xiang demostró que a corto plazo las comunidades microbianas estaban substancialmente alteradas y que tuvieron una asombrosa capacidad de recuperación volviendo a las condiciones pre-incendio once años más tarde. En nuestro trabajo se presenta el primer estudio de la diversidad taxonómica de las comunidades microbianas de un bosque mediterráneo afectado por un incendio a medio plazo, tres años y largo plazo, seis años y se evalúan las capacidades funcionales del género bacteriano dominante en dicha comunidad.

2. Parque Natural y Nacional de Sierra Nevada

La región mediterránea tiene unas características morfológicas, geográficas, históricas y sociales muy particulares que suscitan gran interés científico en el estudio de su clima, pues constituye una zona de transición entre el árido clima del Norte de África y el templado y lluvioso clima de Europa Central. Cada pequeña modificación de este ecotono puede conllevar cambios sustanciales en el clima, haciendo de las montañas mediterráneas una región potencialmente muy vulnerable al cambio climático (Lionello *et al.* 2006). Las predicciones climáticas para estas regiones apuntan a una tendencia de mayores temperaturas, menos precipitación y mayor variabilidad interanual que en montañas europeas no mediterráneas (Giorgi y Lionello, 2008), lo cual conllevará grandes cambios en la vegetación de estas zonas. Estos cambios de vegetación ya se han observado en algunos trabajos, como en el de Ruiz-Labourdette y colaboradores en 2013, donde concluyen que los cambios en las especies arbóreas de las zonas bajas de las montañas mediterráneas se deben principalmente a las limitaciones hídricas que se están dando en estos sitios.

Por lo tanto, de todas las regiones de Europa, la región mediterránea parece la más vulnerable al cambio climático (Schröter *et al.* 2005), siendo las principales amenazas debidas al cambio climático global la disminución de las precipitaciones, el incremento de las temperaturas y el aumento de la frecuencia de los incendios forestales. Estos factores influyen negativamente en los ecosistemas alterando tanto la composición como la riqueza de la comunidad de plantas (Schröter *et al.* 2005), disminuyendo la cantidad y calidad de los recursos del suelo (nutrientes y materia orgánica), afectando a la biomasa y la diversidad microbiana (Guénon *et al.* 2011), y disminuyendo la actividad de mineralización de las comunidades microbianas del suelo (Acea y Carballas, 1996; Sardans y Peñuelas, 2005; Sardans *et al.* 2008). Teniendo en cuenta esto y, como se ha indicado anteriormente, que los ecosistemas de las zonas de montaña son muy sensibles y vulnerables al cambio climático, resulta interesante centrar el estudio de los efectos del cambio climático en un ecosistema mediterráneo de alta montaña, siendo el Parque Nacional de Sierra Nevada el ejemplo más representativo de este tipo de ecosistemas, según el Catálogo de Parques Nacionales del Ministerio de Agricultura, Alimentación y Medio Ambiente (<http://reddeparquesnacionales.mma.es>). Además, hay que destacar que como Reserva de la Biosfera, Sierra Nevada ha sido incluida a nivel mundial entre las 28 reservas de la Biosfera de Montaña que participan en la Red Internacional de Investigación y Seguimiento del Cambio Climático, el proyecto Global Change in Mountain Regions (GLOCHAMORE). Además, en la

segunda fase de este proyecto (GLOCHAMOST), Sierra Nevada sigue siendo una de las 10 reservas de la biosfera seleccionadas a nivel mundial como Observatorio de Cambio Global.

Sierra Nevada se localiza en el Sureste de la Península Ibérica, entre las provincias de Granada y Almería, ubicándose entre las coordenadas geográficas de 36° 55' a 37° 15' de latitud norte, y de 2° 56' a 3° 38' de longitud oeste, aproximadamente. Está delimitada por el valle del río Lecrín al oeste, el valle del río Guadalfeo al sur, por la depresión Bética al norte, por el corredor del río Gérgal al este y por el valle del río Andarax al sureste. En sentido Este-Oeste, tiene una longitud de casi 80 Km y anchura que oscila entre los 15 y 30 Km, con una superficie de 172.238 ha, de las cuales 85.883 constituyen el Parque Nacional (<http://www.juntadeandalucia.es>). En el ámbito internacional, desde 1986 pertenece a la Red Mundial de Reservas de la Biosfera de la UNESCO, zonas ecológicas representativas que constituyen espacios de gran diversidad biológica, con características naturales de gran interés científico (<http://reddeparquesnacionales.mma.es>). Fue declarada Parque Natural en 1989 debido a las singularidades de su flora, fauna, geomorfología y paisaje. Posteriormente, tras la propuesta del Parlamento Autonómico andaluz se declaró Parque Nacional en 1999.

Por su localización biogeográfica, su aislamiento, su rango altitudinal y la diversidad de nichos ecológicos que presenta, este observatorio de cambio global es además un refugio excepcional para la biodiversidad, lo cual hace que se hayan catalogado 2.100 especies vegetales, de las que 66 son endémicas de la región. Dentro de este ecosistema particular, hay que hacer notar que la vertiente Norte de la montaña presenta laderas agrestes y cortadas debido a la mayor erosión, mientras que en la cara Sur son lomas más suaves, aunque igualmente de gran pendiente. Las condiciones que determinan la enorme diversidad climática en Sierra Nevada son su amplitud altitudinal, su latitud y la compleja topografía del terreno. La amplitud altitudinal determina el aumento de la insolación conforme se asciende y provoca oscilaciones térmicas de gran importancia que difieren a lo largo del gradiente altitudinal: por debajo de los 1.500 m de altitud la temperatura media anual oscila entre los 12 y 16 °C; desde los 1.500 m hasta los 2.000 m entre 8 y 12 °C; entre los 2.000 y 2.500 m oscila entre 4 y 8 °C; y a partir de los 3.000 m la temperatura media anual es menor de 0 °C. Dada su posición meridional y el estar en la zona de influencia mediterránea provoca una relativa sequedad entre mayo y octubre, disminuyendo en verano la pluviosidad a los valores mínimos, mientras que en invierno las precipitaciones son casi exclusivamente en forma de nieve a partir de los 2.000 m.

En los ecosistemas de tipo mediterráneo las comunidades vegetales están expuestas a unas condiciones climáticas que pueden considerarse como extremas ya que alternan periodos de heladas invernales y periodos de sequía y alta temperatura estival con lluvias torrenciales ocasionales. Además, las previsiones respecto al clima, por efecto del cambio climático, señalan subidas de temperaturas generales que afectarán especialmente a las máximas de los meses de verano, mientras que la evolución de las precipitaciones es más irregular aunque se prevé un descenso general (Felicísimo *et al.* 2011), lo cual acentúa aún más los estreses propios de los ecosistemas mediterráneos, dando lugar a una posible migración de bosques hacia zonas más elevadas y también un aumento de la frecuencia de incendios forestales.

2.1 Gradientes altitudinales en Sierra Nevada.

En Sierra Nevada pueden observarse todos los pisos bioclimáticos definidos para la región mediterránea, y este trabajo se centra principalmente en las series climatófilas (Valle, 2003) que se detallan a continuación.

Serie supra-mesomediterránea filábrica y nevadense malacitano-almijareense y alpujarreño-gadoreense silicícola de la encina (*Quercus ilex* subsp. *ballota*):

Ocupa una franja altitudinal situada entre 1.400 y 2.000 m. En su estado maduro estaría constituida por un encinar (*Adenocarpus decorticans-Quercetum rotundifoliae* S.), más pobre en especies que el desarrollado sobre sustrato calizo, aunque dada la estructura abierta que presentan la práctica totalidad de los encinares silicícolas, conviven con formaciones de espinales (facies húmedas) y escobonales (*Cytisus scoparii-Adenocarpum decorticantis*) (facies secas), así como con jarales (*Thymus gadorensis- Cistetum laurifolii*), lastonares (*Dactylisotium hispanicae-Festucetum scariosae*) y tomillares subnitrófilos (*Artemisio-Santolinetum rosmarinifoliae helichrysetosum serotini*). La faciación típica de esta serie aparece en el termotipo supramediterráneo con ombrotipo subhúmedo, pero con una fuerte xericidad estival que limita bastante la aparición de caducifolios. La serie en general se halla bastante alterada, aparecen solamente algunos encinares que normalmente están abiertos y poco estratificados. Los factores que han contribuido a esta alteración son el sobrepastoreo, los incendios, la deforestación o prácticas forestales inapropiadas y la agricultura de montaña.

Serie mesomediterránea nevadense y malacitano-almijareense silicícola del roble melojo (*Quercus pyrenaica*):

Su mayor extensión se presenta en barrancos umbríos de Sierra Nevada (Dehesa del Camarate, Maitena, Genil, Monachil, Dílar y Dúrcal) y en distintas zonas de la Alpujarra granadina ocupando las lomas de menor insolación (Cáñar-Soportújar, Barranco del Poqueira, Pitres, Busquístar y Trévez). El rango altitudinal en el que aparecen va desde los 1.100 - 1.200 m hasta 1.800 – 1.900 m, constituyendo el límite altitudinal del bosque. La vegetación potencial es un bosque (*Adenocarpus decorticans* – *Quercetum pyrenaiceae*) donde domina el roble melojo (*Q. pyrenaica* Willd.). En los claros de este bosque aparecen numerosos elementos caducifolios en las facies más húmedas, mientras que en las facies más secas aparecería un escobonal (*Cytisus scoparii-Adenocarpetum decorticans*). Dentro de los factores que determinan la presencia de esta serie, el más importante es la reducción de la xericidad estival, por lo que se presenta en laderas con poca insolación y en fondos de barrancos. Este tipo de formaciones no están bien adaptadas a las tendencias macroclimáticas actuales, estando su supervivencia en peligro debido a diferentes factores, entre los cuales se encuentran el sobrepastoreo, la deforestación y los incendios.

Los bosques autóctonos del Espacio Natural de Sierra Nevada ocupan una superficie aproximada de 15.000 ha. Aunque esta superficie es menor que la que le correspondería de forma natural, debido a que de forma tradicional han sido talados para obtener madera o pastos. El cambio de uso de los suelos durante las últimas décadas ha favorecido la expansión de las áreas forestales y las áreas de matorrales, aunque la tendencia esperada para formaciones claves, como el robledal, es negativa debido a los efectos del cambio climático (Olivares *et al.* 2011). Los robledales de Sierra Nevada o melojares (*Q. pyrenaica*) se localizan en los pisos supra y mesomediterráneos asociados con *Adenocarpus decorticans* (rascaviejas), pudiéndose dar también mezclada únicamente con encina (*Q. ilex* sp. ballota) o con retama (*Retama sphaerocarpa*) (Valle, 2003). En la vertiente Sur de Sierra Nevada, el robledal se extiende por las colinas y no siguiendo el curso de los ríos, gracias a la influencia del aire húmedo mediterráneo (Martínez-Parras y Molero-Mesa, 1982).

El roble melojo está considerado como Especie de Interés Especial según la Ley 8/2003 de flora y fauna andaluza, por lo que es merecedor de una atención especial para su conservación. Sin embargo desde principios del siglo XX se ha documentado un declive de los robledales, que en la década de 1.980 se ha convertido en un problema global afectando a Europa, Asia y América (Mosca *et al.* 2007).

Serie oromediterránea filábrico-nevadense silicícola del enebro rastrero (*Juniperus communis* subsp. *nana*):

Aparece a partir de los 1.800-1.900 m de altitud, llegando hasta los 2.800 m. La comunidad climática es una formación enebral-piornal (*Genisto baeticae – Juniperetum nanae* S.), con especies como: *Juniperus communis* subsp. *nana*, *J. communis* subsp. *hemisphaerica* y *J. sabina* var. *humilis* (este último elemento domina en los afloramientos de rocas ultrabásicas). Como representantes del piornal (formaciones de caméfitos almohadillados) destacamos *Genista versicolor* y *Cytisus galianoi*; esta comunidad ocupa los lugares con suelos más evolucionados (regosoles y cambiosoles dístricos). Es posible observar como la proporción de especies cambia en las distintas facies del piornal: mientras que *C. galianoi* domina en zonas más húmedas y con suelos más evolucionados, *G. versicolor* (anteriormente denominada *G. baetica*) lo hace en condiciones más xéricas y con suelos menos evolucionados.

2.2. Incendios forestales

Los incendios, como eventos ocasionales de origen natural, han influido en la vegetación de la cuenca mediterránea desde antes de la aparición del hombre en este área (Clark, 1996). Además, con el asentamiento del hombre en el área mediterránea, se introdujo otro factor importante, como es el pastoreo, que junto a otros factores (como son las sequías regulares, las elevadas temperaturas o los propios eventos de incendios ocasionales) ha permitido el desarrollo de una vegetación específica y adaptada a estas condiciones (Pausas, 2006) y muy característica de la región mediterránea: la dehesa, donde la acción ganadera juega un papel fundamental en el mantenimiento del ecosistema. El problema es que durante el desarrollo industrial, desde comienzos del siglo XIX, y hasta la actualidad, la estructura socio-económica de gran parte del planeta ha sufrido grandes cambios, siendo el más evidente el éxodo rural en busca de una mejor calidad de vida en núcleos urbanos e industriales. Este éxodo rural, con el consecuente abandono de las prácticas agrícolas, y la industrialización ha provocado cambios en los regímenes históricos de incendios y en la frecuencia e intensidad de los estreses climáticos (Pausas y Fernández-Muñoz, 2011). Además, los modelos de calentamiento global predicen cambios en los patrones espaciales y temporales de precipitaciones para las próximas décadas, con veranos secos en latitudes medias (Solomon *et al.* 2007) e incluyendo cambios en la frecuencia e intensidad de eventos de sequía y olas de calor (IPCC, 2007; Good *et al.* 2008), como ya se ha demostrado en algunos estudios recientes, en los cuales se observan incrementos en la frecuencia de estos eventos (Della-Marta *et al.*

2007; Alexandrian *et al.* 1999), que también provocarán una mayor frecuencia de incendios forestales (Dale *et al.* 2001). Por lo tanto, se puede decir que el abandono rural, el cual implica el uso de las dehesas como zonas de pastoreo, y el cambio climático global son las 2 principales causas del aumento de eventos de incendios forestales. Además, los incendios forestales constituyen una retroalimentación positiva para el cambio climático, ya que contribuyen al calentamiento global de dos formas: por un lado la alta emisión de CO₂ que se produce en un incendio, y por otro lado la eliminación del efecto sumidero de CO₂ que supone la existencia de los bosques y de las comunidades microbianas asociadas a éstos. Los efectos de los incendios de bosques incluyen la aceleración de los ciclos de nutrientes, la mortalidad de árboles, cambios en la dirección de la sucesión ecológica del ecosistema, inducción de la germinación de semillas, pérdida del banco de semillas del suelo, incremento de la heterogeneidad del paisaje, cambios en las capas superficie-suelo y raíz de la planta y tejidos reproductivos, y volatilización de los nutrientes del suelo (Whelan, 1995).

En el caso de nuestro país, la superficie afectada por incendios en España alcanza una media anual de 127.209 hectáreas en el periodo del 2000 al 2009, incluyendo en éstas las formaciones herbáceas, arbustivas y boscosas (<http://www.magrama.gob.es/es/biodiversidad/temas/defensa-contra-incendios-forestales/estadisticas-de-incendios-forestales/>). Dentro de estos incendios está el que ocurrió en el Espacio Protegido de Sierra Nevada en septiembre de 2005. Este incendio afectó a una extensión de 3.416,74 ha (Gómez-Zotano *et al.* 2005), con la presencia de encinas en 412 ha del total de la superficie quemada, claro ejemplo de todo el expuesto anteriormente. Para facilitar la recuperación de estas formaciones vegetales es importante entender los procesos que afectan a la recuperación del suelo y las comunidades microbianas asociadas a éste (Cobo-Díaz, 2013).

Objetivos

Las bacterias son indicadores tempranos de cambios en el ambiente debido a su papel en los distintos ciclos biogeoquímicos de los elementos y a su sensibilidad a las alteraciones. Así son esenciales para el cierre de los ciclos del Carbono y del Nitrógeno en el suelo, al tiempo que también resultan fundamentales para favorecer la incorporación de estos elementos a las plantas. Por otra parte, en los próximos años las formaciones vegetales de la región mediterránea se verán sometidas a unas variaciones climáticas que incluyen aumento de temperatura y disminución de la precipitación, según el IPCC. Estos cambios además favorecerán siniestros, como son los incendios forestales, que dificultarán todavía más su supervivencia. Por tanto el objetivo principal del presente trabajo es analizar el papel de las bacterias como posibles bioindicadores de los cambios ambientales que sufren las formaciones boscosas autóctonas del Parque Nacional de Sierra Nevada. La elección de esta zona se debe a que representa el ecosistema de alta montaña mediterránea más meridional de Europa, estando en un equilibrio muy frágil lo que lo hace especialmente sensible a los cambios ambientales. Como elemento vegetal se han elegido el roble melojo (*Quercus pyrenaica* Willd.) ya que está en su límite meridional y casi es un relictos, y la encina (*Q. ilex* subsp. *ballota*) ya que podría ocupar los espacios perdidos por el robledal. Con este bagaje, los objetivos parciales que nos propusimos son:

1. Estudio de la diversidad microbiana en la rizosfera de un robledal en gradiente altitudinal.
2. Análisis de las comunidades microbianas que habitan en la rizosfera de quercíneas (roble melojo y encina) en condiciones naturales.
3. Identificación de bioindicadores de la recuperación de las comunidades microbianas rizosféricas de un encinar tras un incendio forestal.
4. Obtención e identificación del género bacteriano mejor adaptado al estrés post-incendio en la rizosfera de un encinar quemado y reconstrucción de su genoma a partir del metagenoma de la comunidad microbiana a la que pertenece.

Material y Métodos

1. Área de estudio

El área de estudio se localiza en el Parque Nacional de Sierra Nevada, al sureste de España, en la provincia de Granada; en la cual se han seleccionado 3 zonas de estudio. La primera de estas zonas se ubica en la ribera del río Lanjarón, en la cara Sur de Sierra Nevada. Esta zona se eligió porque sufrió un incendio que afectó a 3.416,74 hectáreas en septiembre de 2005, de las cuales 412 ha contenían encinas en su composición (*Quercus ilex* subsp. *ballota*). Dentro de esta zona tenemos 3 sitios de muestreo, dos de ellos en la extensión afectada por el incendio, uno con pies de encinas que rebrotaron después del incendio (LJQ), el otro es un sitio en el que no había encinas antes del incendio y después del mismo crecieron gran cantidad de leguminosas arbustivas y germinaron nuevas encinas a partir de las bellotas del suelo (LJR para suelo sin encinas y LJRE para suelos rizosféricos de encinas). El tercer sitio está en un encinar que no fue afectado por el incendio (LJN), situado ligeramente al norte de la extensión que sí fue afectada.

La segunda de estas zonas se ubica en la ladera del término municipal de Cáñar, también en la ladera Sur de la sierra, donde existen diferentes estadios de un bosque de roble melojo o rebollo (*Quercus pyrenaica* Willd.). Dentro de esta zona tenemos 4 sitios de muestreo siguiendo un gradiente altitudinal. En el sitio más bajo (CNB) el robledal está antropizado y se puede considerar un ecotono del bosque de robles que a la larga podría ser reemplazando por un encinar. El siguiente sitio, siguiendo el gradiente altitudinal, es aquel donde el bosque de robles está bien desarrollado y no presenta alteraciones, invasión de encinas u otros aspectos negativos para el bosque (CNA). Después de éste aparece una franja donde existen robles aislados o formando pequeños grupos, pero sin formar un bosque (CNX). Por último, a igual altitud está el sitio donde no aparecen robles, dominado por un piornal (*Genista versicolor*, antes llamada *G. baetica*), donde se podría expandir el robledal en una situación de cambio climático (CNR).

La tercera y última zona de estudio se encuentra en la Dehesa del Camarate, ubicada en la cara Norte de la sierra, donde se tomaron muestras de dos sitios diferentes, uno formado por un encinar (CTE) y otro por un robledal (CTR).

Material y Métodos

Tabla M-1. Puntos de muestreo georreferenciados. Altitud medida en metros sobre el nivel del mar.

Zona	Sitio	Localización	Altitud	Descripción del sitio	Orientación de la ladera
Ribera del río Lanjarón	LJN	N 36°58'11", W 3°27'37"	1790 msnm	Encinar no quemado	Sur
	LJQ	N 36°57'26", W 3°27'48"	1566 msnm	Encinar quemado, rebrote de encinas.	
	LJR	N 36°57'40", W 3°27'42"	1610 msnm	Matorral y herbáceas quemadas	
	LJRE			Matorral y herbáceas quemadas, germinación de nuevas encinas	
Ladera de Cáñar	CNB	N 36°56'58", W 3°25'03"	1482 msnm	Robledal invadido por encinas	Sur
	CNA	N 36°57'44", W 3°25'09"	1823 msnm	Robledal sin invasión de encinas	
	CNX	N 36°57'56", W 3°25'45"	1945 msnm	Robles aislados, no forman bosque	
	CNR	N 36°57'38", W 3°26'15"	1938 msnm	Suelo no asociado a quercíneas, ubicado en un Piornal	
Dehesa del Camarate	CTE	N 37°11'32", W 3°15'11"	1520 msnm	Encinar no quemado	Norte
	CTR	N 37°10'53", W 3°15'39"	1717 msnm	Robledal	

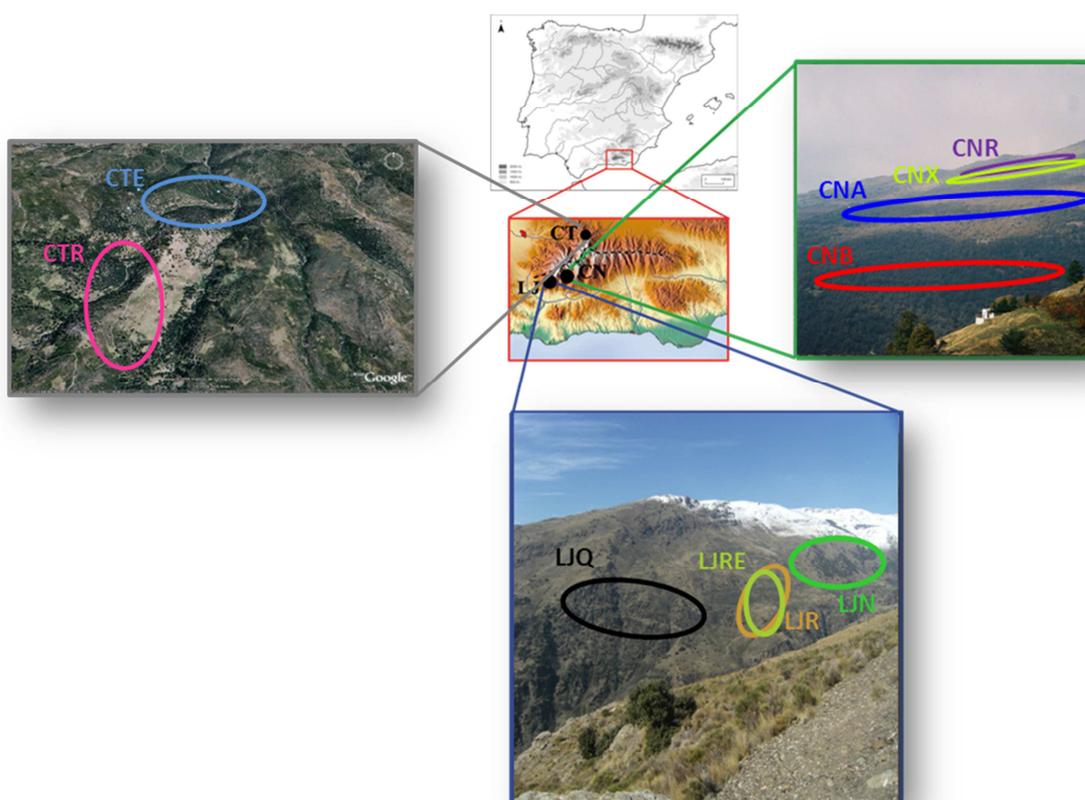


Figura M-1. Detalle de la localización de las zonas de muestreo.

Los acrónimos reflejan en sus dos primeras letras la zona geográfica (Lanjarón, LJ; Cáñar, CN y Dehesa del Camarate, CT), seguidas de una tercera letra que refleja el tratamiento. De esta manera quedaría:

- LJN: Lanjarón, rizosfera de encinas No quemadas
- LJQ: Lanjarón, rizosfera de encinas Quemadas
- LJR: Lanjarón zona quemada de herbáceas, suelo suelto
- LJRE: Lanjarón zona quemada de herbáceas con nuevas Encinas germinadas desde bellota
- CNB: Cáñar, rizosfera de robledal en zona de Baja altitud
- CNA: Cáñar, rizosfera de robledal en la zona Alta
- CNX: Cáñar zona de expansión del robledal, muestra rizosférica
- CNR: Cáñar zona de expansión del robledal, suelo suelto
- CTE: Dehesa del Camarate, rizosfera de Encinas
- CTR: Dehesa del Camarate, rizosfera de Robles

1.1. Características físico-químicas de los suelos

Las características físico-químicas de los suelos muestreados se determinaron, a partir de 2 kg de suelo de cada sitio de muestreo, en el Laboratorio Agroalimentario de Atarfe, Granada, perteneciente a la Consejería de Agricultura y Pesca de la Junta de Andalucía, donde se analizó la textura de los suelos, su conductividad, el contenido en carbono (C), nitrógeno (N), y distintos iones como fósforo, potasio, sodio, magnesio, calcio e hierro, siguiendo métodos usuales.

1.2. Inventario de la vegetación acompañante de las zonas de estudio

Se caracterizó la flora y vegetación en cada una de las 3 parcelas de cada sitio de muestreo (Tabla M-1). Dentro de cada parcela se seleccionaron 3 puntos de muestreo que coinciden con los puntos donde se tomaron los suelos (ver apartado 5.1). El inventario de la flora se acotó a la zona situada en la base de los pies de los robles o encinas, en el caso de

muestras rizosféricas, con el fin de detectar una posible influencia de éstas en la actividad microbiana. Para cada árbol se anotaron las especies presentes y se herborizaron los ejemplares siguiendo el procedimiento habitual para las muestras botánicas: recolección, prensado con almohadillas para secado, mantenimiento en prensa hasta su completa desecación y preparación de los pliegos de cada especie. Las especies vegetales se determinaron siguiendo las claves de determinación de Flora de Andalucía Oriental (Blanca *et al.* 2009).

De cada parcela se aportan los datos de composición florística (abundancia específica por sitio). La abundancia de cada especie se calculó como el coeficiente entre el número de veces que aparece una especie (Sp) en los nueve puntos de muestreo, en relación al total de veces que aparecen todas las especies por sitio ($\sum Sp$): $Abundancia\ Sp = \frac{Sp}{\sum Sp} * 100$

Se compararon los sitios en relación a la abundancia de especies dominantes y se analizaron mediante ANOVA de una vía (factor: sitio) y la comparación de medias (post hoc) se realizó mediante el test de Tukey con un intervalo de confianza del 95 %. En la mayoría de casos se utilizó la prueba no paramétrica de Kruskal-Wallis, al no cumplir los supuestos del ANOVA (normalidad y homocedasticidad). La separación media se realizó mediante el test de Nemenyi (Zar, 1999).

2. Registro de Humedad y Temperatura

Para registrar la humedad relativa y temperatura del aire de la zona de Cáñar se utilizó un HOBO® Data Loggers modelo H8 Pro para cada sitio muestreado, colocados a una altura de 1,60 metros y con orientación 45° Noreste. Para registrar la humedad relativa y temperatura del suelo de estos mismos sitios se utilizó un HOBO® Micro Station con una sonda S-TMB-M006 para registrar temperatura y una sonda S-SMC-M005 para registrar la humedad relativa en cada uno de ellos. Las sondas de suelo se colocaron a una profundidad de 20 centímetros y se les añadió agua inmediatamente después de colocarlas para asentar el terreno.

3. Cultivos celulares y uso de antibióticos

3.1 Medios de cultivo bacteriano

Medio LB

Como medio de cultivo de la cepa DH5 α de *Escherichia coli* se utilizó el medio de Luria-Bertani (LB) (Sambrook *et al.* 1989):

- NaCl 5 g/l
- Triptona (DIFCO) 10 g/l
- Extracto de levadura 5 g/l
- Agua desionizada hasta 1 litro

Para solidificar el medio se añadió agar (PANREAC) al 1,6 %. El medio se esterilizó en autoclave a 120° C durante 20 minutos.

Medio HH'

Se utilizó una modificación del medio selectivo de Hagedorn-Holt (HH) (Hagedorn *et al.* 1975) para aislar cepas de *Arthrobacter*:

- Caldo de Soja Tríplica (TSB) 4 g/l
- Extracto de levadura 2 g/l
- NaCl 30 g/l
- Agua desionizada hasta 1 litro

Para solidificar el medio se añadió agar (PANREAC) al 1,5 %. El medio se esterilizó en autoclave a 120° C durante 20 minutos.

3.2 Conservación a largo plazo de las cepas bacterianas

Las cepas que se consideraron adecuadas (tras realizar la correspondiente secuenciación del gen 16S *rRNA*, ver apartado 6.2.2) fueron crecidas en 3 ml de su correspondiente medio líquido (LB para *E. coli* y HH' sin NaCl para *Arthrobacter*) y a su correspondiente temperatura (37° C para *E. coli* y 30° C para *Arthrobacter*) hasta alcanzar una densidad óptica de 0,8 a una absorbancia de 600 nm de longitud de onda (medida realizada en el espectrofotómetro Novaspec II de Pharmacia Biotech). Una vez obtenida la densidad óptica requerida, se añadieron 750 μ l de este cultivo a un criotubo con 250 μ l de glicerol esterilizado

Material y Métodos

previamente en autoclave a 120° C durante 20 minutos. De esta forma la concentración final de glicerol es del 20 % v/v. Esta operación se realizó por duplicado para tener un tubo “de reserva”. Todo el proceso se realizó en una campana de flujo laminar y con todos los materiales empleados convenientemente esterilizados para evitar contaminaciones. Los dos criotubos fueron debidamente etiquetados, tanto en el tubo como en la tapa, y almacenados a -80° C.

3.3 Antibióticos

La adición de antibióticos a los medios de cultivo se hizo a partir de soluciones 100 veces concentradas preparadas en agua desionizada (para la ampicilina, Ap) o en agua-etanol al 50 % (para la cicloheximida, ClhX). Posteriormente, se esterilizaron por filtración utilizando unidades Minisart® NML (Sartorius) de 0,2 µm de tamaño de poro.

- Ampicilina (Ap) Efecto bacteriostático sobre *E. coli* Preparada a 20 g/l
- Cicloheximida (ClhX) Efecto fungicida Preparada a 10 g/l

4. Aislamiento de bacterias del género *Arthrobacter*

El aislamiento de cepas del género *Arthrobacter* se realizó a partir de una muestra del suelo rizosférico quemado, en concreto LJQ1. Un paso previo al uso del medio HH' fue un pre-enriquecimiento del suelo con agua de Peptona, modificado de Dorothy Jones (2006), del siguiente modo:

- Triptona 5 g/l
- Suelo LJQ1 1 g
- Agua desionizada hasta 30 ml

Para conseguir el pre-enriquecimiento, el medio se dejó en agitación suave en un balancín durante 1 hora a temperatura ambiente. Posteriormente, se dejó el tubo en reposo en posición vertical durante 15 minutos para conseguir la decantación de las partículas más groseras de suelo.

Por último, se sembraron en placas Petri con medio HH' 100 µl de medio de pre-enriquecimiento y 100 µl de cada dilución seriada de éste, hasta la dilución 10^{-7} .

Las placas se incubaron durante 1 semana a 30° C.

5. Toma de muestras y extracción de DNA

5.1 Recolección de las muestras

Para el muestreo se seleccionaron 3 parcelas para cada uno de los sitios de estudio. Estas parcelas fueron escogidas al azar, respetando una línea altitudinal y tomando un transecto de una longitud mínima de 700 m hasta una máxima de 2,5 km aproximadamente. De cada parcela fueron muestreados 3 árboles, con tamaño y estructura de árbol maduro y con al menos 5 metros de separación entre ellos, excepto en LJR y CNR, donde las muestras fueron tomadas en suelo suelto, es decir no asociado a la rizosfera de quercíneas. La toma de muestras fue realizada en cada parcela de la cara Sur de Sierra Nevada los días 26 de abril de 2008 y el 29 de abril de 2011, mientras que en la Dehesa del Camarate, cara Norte, el muestreo se realizó una semana después.

Los puntos de muestreo fueron marcados con estacas de madera o marcando el árbol muestreado, además todas las parcelas fueron registradas con GPS (Global Positioning System) (ver Tabla M-1). Para la obtención de las muestras, en primer lugar se eliminó la vegetación herbácea en un radio de unos 10 cm alrededor del punto de toma de muestra y se descartó el suelo de los primeros 5 cm de profundidad. La muestra consistió en 500 gramos de suelo cogido a una profundidad de 5-25 cm siguiendo una raíz principal del árbol hasta donde se apreciaron raíces laterales jóvenes no suberificadas. Las muestras se mantuvieron refrigeradas durante el transporte en una nevera portátil. Una vez en el laboratorio los suelos se pasaron por un tamiz de 2 mm de malla y se hicieron alícuotas de 0,25 gramos de suelo para extraer el DNA.

5.2 Extracción de DNA de muestras de suelo

La extracción de DNA total para cada suelo se realizó mediante un método directo, usando el “PowerSoil DNA Isolation Kit” de MoBio, a las 24 horas de la toma de las muestras, siguiendo las instrucciones del fabricante:

- Adicionar 0,25 g de muestra de suelo a un “Power Bead Tube” y agitar suavemente la mezcla en un vórtex. El tampón que contiene ayuda a dispersar las partículas del suelo, disuelve los ácidos húmicos y protege los ácidos nucleicos de su degradación.
- Añadir 60 µl de solución C1, la cual contiene SDS y otros agentes químicos que facilitan la lisis celular, y se agita brevemente en un vórtex.
- Cerrar bien los Power Bead Tubes y agitar durante 10 minutos a la máxima velocidad. En este paso se completa la homogenización y lisis celular por la acción combinada de los agentes químicos y la agitación mecánica con las bolitas de cristal que contiene el tubo.
- Centrifugar a 10.000 X g durante 30 segundos a temperatura ambiente.
- Transferir el sobrenadante (400-500 µl) a un tubo eppendorf limpio. El sobrenadante sigue siendo de tono oscuro y puede contener algunas partículas de suelo.
- Adicionar 250 µl de solución C2 y, tras agitarlo durante 5 segundos en el vórtex, se incuba a 4° C durante 5 minutos. La solución C2 contiene un agente que precipita las sustancias inorgánicas y las orgánicas que no sean DNA, como los ácidos húmicos, restos celulares y proteínas, que reducirían la pureza del DNA extraído y lo inhabilitarían para su uso posterior.
- Centrifugar los tubos a 10.000 X g durante 1 minuto a temperatura ambiente.
- Transferir unos 600 µl del sobrenadante a un tubo limpio, evitando arrastrar parte del pellet.
- Adicionar 200 µl de solución C3, agitar en vórtex suavemente e incubar a 4° C durante 5 minutos. La solución C3 contiene un agente precipitante adicional que contribuirá a la pureza del DNA extraído.
- Centrifugar a 10.000 X g durante un minuto a temperatura ambiente.
- Transferir unos 750 µl de sobrenadante a un tubo eppendorf limpio.
- Adicionar 1,2 ml de solución C4 y agitar en vórtex durante 5 segundos. C4 es una solución salina de alta concentración que permitirá la unión cuasi-específica del DNA a las columnas de sílica que se utilizan en pasos posteriores, eliminando así aquellos contaminantes que no se unan a la columna a altas concentraciones de sal.
- Cargar aproximadamente 675 µl dentro de un Spin Filter (columnas de sílica).
- Centrifugar a 10.000 X g durante un minuto a temperatura ambiente.
- Tirar el eluido que ha pasado a través de la columna y repetir el proceso 3 veces en la misma columna, hasta agotar el sobrenadante al que anteriormente añadimos la solución C4. En este momento el DNA se encuentra unido a la columna de sílica.
- Adicionar 500 µl de solución C5 a la columna de sílica y centrifugar a 10.000 X g durante 30 segundos a temperatura ambiente. La solución C5 contiene etanol y se utiliza para limpiar y lavar la columna y el DNA unido a ella, eliminando las sales, ácidos húmicos y otros contaminantes residuales, dejando el DNA unido a la columna.
- Tirar el eluido que pasa a través del tubo.
- Centrifugar a 10.000 X g durante 1 minuto a temperatura ambiente para eliminar el etanol residual que pueda interferir en las aplicaciones posteriores del DNA.

- Colocar el Spin Filter en un tubo eppendorf limpio. Añadir 100 µl de solución C6 (tampón de elución Tris 10 mM) en el centro de la membrana.
- Centrifugar a 10.000 X g durante 30 segundos a temperatura ambiente para recuperar en el tubo el DNA que estaba unido al sílica.
- Tirar el Spin Filter.
- Guardar el DNA extraído a -20° C.

5.3 Extracción de DNA plasmídico por precipitación con sales de magnesio

Esta técnica está basada en el método de Studier (1991), que permite el aislamiento rápido de plásmidos de *E. coli*. Recogemos 1,5 ml de cultivo bacteriano por centrifugación y, tras retirar el sobrenadante, se resuspenden las células en 100 µl de agua desionizada. La lisis tiene lugar por la adición de 100 µl de una solución compuesta por NaOH 0,1 M, EDTA 10 mM y 2 % SDS. Se homogeneiza por agitación la suspensión celular y la solución y los tubos que la contienen se hierven en agua durante 2 min. El DNA cromosómico se precipita por la adición de 50 µl de MgCl₂ 1 M. Tras homogeneizar completamente (aparición de precipitado blanco), se incuba en hielo de 2-5 min. Se centrifuga 1 minuto para sedimentar el precipitado en el fondo del tubo y añadimos 50 µl de acetato potásico 5 M para eliminar proteínas. El tubo invertido se mezcla con vórtex evitando que se levante el precipitado blanco del fondo y se incuba 5 minutos en hielo. Se centrifuga durante 5 minutos y recogemos el sobrenadante a un tubo nuevo donde precipitaremos el DNA plasmídico con 0,6 ml de etanol 100 % guardado a -20° C. Se deja 5 minutos a temperatura ambiente y se centrifuga 5 min. Se lava el sedimento con 200 µl de etanol al 70 % y dejamos secar al vacío. Se resuspende en el volumen adecuado dependiendo de la concentración deseada o necesaria para las diferentes aplicaciones: lo más habitual para un plásmido de alto número de copias es resuspender en 25 µl de agua desionizada con RNAsa a una concentración de 10 µg/ml.

5.4 Cuantificación de la concentración de DNA

El cálculo de la concentración de ácidos nucleicos se estimó siguiendo el método espectrofotométrico descrito en Sambrook *et al.* (1989). Se midió la absorbancia de las muestras en solución acuosa a longitudes de onda 260 y 280 nm utilizando cubetas de cuarzo. La concentración se determinó en función de los valores estándar: 1 U a D.O. de 260 nm se corresponde con 50 µg/ml de DNA de cadena doble, 40 µg/ml de DNA de cadena sencilla y RNA, y 33 µg/ml de oligonucleótidos. La relación D.O. 260 nm / D.O. 280 nm se usó como estimador de la pureza de la preparación, considerándose valores por debajo de 1,8 como indicadores de contaminación por proteínas o fenol.

Generalmente, se ha usado un espectrofotómetro (NanoDrop® ND-1000) que permite trabajar con volúmenes de muestra de 1 µl a 2 µl; facilitando el proceso de análisis.

En el Joint Genome Institute (JGI, Walnut Creek, California, USA) del Departamento de Energía (DOE) de los Estados Unidos de América se calculó la concentración de ácidos nucleicos siguiendo el método fluorimétrico descrito en Schweitzer *et al.* (2003) llevado a cabo, en este caso, con un Qubit® 2.0 Fluorometer (Invitrogen). Esta tecnología se basa en medir la fluorescencia emitida por un marcador fluorescente (fluoróforo) y compararla con una curva patrón para dar un dato de concentración de la muestra, sea DNA, ARN o proteína. El fluoróforo emite muy poca fluorescencia cuando no está unido a su diana, en nuestro caso DNA, pero cuando se une, la intensidad de fluorescencia emitida es varios órdenes de magnitud mayor. Qubit® permite trabajar con volúmenes de muestra de 1 µl a 20 µl y es capaz de medir la concentración en un rango que va desde 10 pg/µl a 1 µg/µl, esta información se detalla en la página web (<http://www.lifetechnologies.com/es/en/home/brands/product-brand/qubit/qubit-fluorometer-vs-nanodrop-nd-1000.html>).

6. Manipulación y amplificación de DNA

6.1 Digestión con enzimas de restricción

La digestión completa del DNA se realizó siguiendo las indicaciones de temperatura y tampón óptimos recomendadas por el proveedor (New England Biolabs). Lo más común es utilizar 1-5 U de enzima/ μg DNA a digerir (entre 0,1 y 0,2 μl de enzima por 10-15 μl de volumen final de la digestión)

Para la digestión con la enzima *NotI*, el DNA se incubó a 37° C durante al menos 1 hora, en la siguiente solución de digestión:

• DNA molde	3 μl
• Tampón 3	1,5 μl
• Albúmina de Suero Bovino (BSA) [1 mg/ml]	1,5 μl
• <i>NotI</i>	0,2 μl
• Agua desionizada	8.8 μl

6.2 Aproximaciones basadas en la PCR

La reacción en cadena de la polimerasa (Polymerase Chain Reaction, PCR) consiste en la amplificación en sucesivos ciclos de un material de partida gracias al uso de una pareja de cebadores. En nuestro caso la hemos utilizado para amplificar insertos que después necesitamos clonar, para la comprobación rutinaria de la presencia de un determinado inserto o para la amplificación del gen ribosomal 16S *rRNA* y su posterior clonación en vectores como pGEM-T Easy o resolución mediante distintas técnicas como la TGGE o la Pirosecuenciación.

Consiste en 3 pasos fundamentales repartidos en 3 ciclos: desnaturalización, anillamiento y extensión. La desnaturalización suele realizarse a 94° C durante 30 segundos aunque el tiempo depende del tipo de molde; al comienzo del programa de PCR suele incluirse un primer ciclo de desnaturalización de 4 minutos para disponer de todo el DNA como cadena simple. La temperatura de anillamiento depende de la composición de los cebadores pero suele rondar los 55° C y el tiempo estándar es 30 segundos. La extensión no es más que la temperatura óptima para la actividad de la enzima (generalmente 72-74° C); el tiempo de este ciclo suele estar en función del tamaño del amplificado y del tipo de enzima usado (por ejemplo, *Taq* polimerasa amplifica un fragmento de ADN de 1 Kb en 30 s). Además el programa

Material y Métodos

de PCR suele terminarse con un ciclo adicional de extensión de 5-10 minutos para completar los fragmentos.

En todos los casos para la amplificación del gen 16S *rRNA* se utilizó el MasterTaq Kit® (5 PRIME), el cual consta de, *Taq* polimerasa 5 U/ μ l, 10x Taq Buffer con sales de Magnesio, solución de Magnesio 25 mM y 5x Taq Master PCR Enhancer (inhibidor de contaminantes de la muestra, como Ácidos Húmicos).

La amplificación se realizó en un termociclador Eppendorf Mastercycler® y el programa empleado varió dependiendo del objetivo. Las reacciones se llevaron a cabo generalmente en un volumen final de 25 μ l conteniendo:

Reactivos	Volumen utilizado	Concentración Final
• 5x TaqMaster PCR Enhancer	5 μ l	1X
• 10x MasterTaq Buffer con 15 mM Mg ²⁺	2,5 μ l	1X (1,5 mM Mg ²⁺)
• 25 mM Magnesium Solution	1 μ l	2,5 mM Mg ²⁺ (final)
• 10 mM dNTP Mix (2,5 mM cada uno)	1 μ l	0,1 mM de cada uno
• Forward primer (10 μ M)	1 μ l	0,4 μ M
• Reverse primer (10 μ M)	1 μ l	0,4 μ M
• DNA molde	2-20 ng	-
• <i>Taq</i> DNA Polymerase (5 U/ μ l)	0,15 μ l	0,75 U*
• Agua milliQ	12,35 μ l	

* En el caso de la polimerasa, no se muestra la concentración final sino la cantidad final añadida.

6.2.1 Amplificación del gen 16S *rRNA* para TGGE

Para construir los perfiles genéticos mediante TGGE (ver apartado 7.2) se amplificó el gen 16S *rRNA* mediante PCR con cebadores universales y cebadores específicos de la clase α -Proteobacteria y de los *phyla* Actinobacteria, Firmicutes y Bacteroidetes (Tabla M-2) siguiendo el protocolo descrito por van Dillewijn *et al.* (2002). El programa utilizado en el primer PCR fue:

- Desnaturalización inicial de 5 minutos a 94° C
- 28 ciclos de:
 - Desnaturalización de 1 minuto a 94° C
 - Anillamiento de 1 minuto a (Tabla M2)
 - Extensión de 2 minutos a 72° C
- Extensión final de 10 minutos a 72° C

El PCR anidado (nested) se llevó a cabo con 1 µl del primer PCR y con los cebadores universales F984GC (5'-CGCCCGGGCGCGCCCGGGCGGGGCGGGGGCACGGGGGGAACGCGAAGAACCTTAC-3') el cual genera una cola GC para evitar la desnaturalización completa del amplicón y R1378 (Tabla M-2) (Heuer *et al.* 1997). El programa utilizado en el nested PCR fue:

- Desnaturalización inicial de 5 minutos a 94° C
- 28 ciclos de:
 - Desnaturalización de 1 minuto a 94° C
 - Anillamiento de 1 minuto a 53° C
 - Extensión de 2 minutos a 72° C
- Extensión final de 10 minutos a 72° C

Tabla M-2. Cebadores utilizados para construir los perfiles genéticos mediante TGGE.
* El cebador universal R1378 se utilizó junto al cebador F243 para amplificar el phylum *Actinobacteria*.

Cebador	Secuencia 5' - 3'	Taxones	Temperatura anillamiento (° C)	Referencia
9bfm	GAGTTTGATYHTGGCTCAG	Bacteria	52	Mühling <i>et al.</i> 2008
1512UR	ACGGHTACCTTGTTACGACTT	Bacteria y Archaea	52	Weisburg <i>et al.</i> 1991
F203α	CCGCATACGCCCTACGGGGGAAAGATTTAT	α-Proteobacteria	56	Gomes <i>et al.</i> 2001
L	CTACGG(AG)TACCTTGTTACGAC	α-Proteobacteria	56	Gomes <i>et al.</i> 2001
F243	GGATGAGCCCGCGGCTA	Actinobacteria	63	Heuer <i>et al.</i> 1997
R1378*	CGGTGTGTACAAGGCCCGGAACG	Bacteria y Archaea	53, 63	Heuer <i>et al.</i> 1997
Firm350f	GGCAGCAGTRGGGAATCTTC	Firmicutes	57	Mühling <i>et al.</i> 2008
Firm814r	ACACYTAGYACTCATCGTTT	Firmicutes	57	Mühling <i>et al.</i> 2008
CFB555f	CCGGAWTYATTGGGTTAAAGGG	Bacteroidetes	61	Mühling <i>et al.</i> 2008
CFB968r	GGTAAGGTTCTCGCGTA	Bacteroidetes	61	Mühling <i>et al.</i> 2008

6.2.2 Construcción de librerías clásicas del gen 16S rRNA

Para construir librerías del gen 16S rRNA fue necesario amplificar dicho gen, mediante PCR, con los cebadores 9bfm específico para el dominio Bacteria y 1512UR universal para los dominios Bacteria y Archaea (Tabla M-2). El programa de PCR utilizado fue:

En el caso de las muestras extraídas de la rizosfera de robles:

- Desnaturalización inicial de 4 minutos a 94° C
- 35 ciclos de:
 - Desnaturalización de 1 minuto a 94° C
 - Anillamiento de 1 minuto a 52° C
 - Extensión de 1,5 minutos a 72° C
- Extensión final de 10 minutos a 72° C

Material y Métodos

En el caso de las muestras extraídas de la rizosfera de encinas se modificó el programa de 35 a 25 ciclos para reducir la aparición de secuencias quiméricas, ya que éstas aumentan exponencialmente pasados 24 ó 25 ciclos (Acinas *et al.* 2005). Las quimeras son consideradas como un artefacto de la PCR, por el cual se produce una extensión incompleta del amplicón y es utilizado en alguno de los siguientes ciclos como cebador, pudiendo unirse a una región homóloga del mismo gen pero de un organismo diferente. Al finalizar la extensión se origina un amplicón con orígenes diferentes en parte de su secuencia. En subsiguientes ciclos esta quimera es amplificada, junto con los amplicones que provienen de un único organismo (Smith *et al.* 2010, Thompson *et al.* 2002, Meyerhans *et al.* 1990, Judo *et al.* 1998, Odelberg, 1995. Ver también el manual de Chimera Uchime del software de MOTHUR en http://drive5.com/usearch/manual/chimera_formation.html). Posteriormente se mezclaron los productos de PCR de los tres árboles de cada parcela, minimizando el sesgo que pueda producir tomar una única muestra de un único árbol y de este modo, tener una muestra representativa de cada parcela. Se limpiaron 60 µl del producto de PCR con las columnas Illustra MicroSpin™ S-300 HR de GE Healthcare, siguiendo las instrucciones del fabricante:

- Resuspender la resina de la columna mediante vórtex.
- Girar la tapa un cuarto de vuelta y quitar el cierre inferior de la columna.
- Poner la columna en un tubo colector y centrifugar a 735 X g durante 1 minuto.
- Poner la columna en un tubo eppendorf DNase-free de 1,5 ml.
- Añadir de 25 a 100 µl de la muestra en el centro de la parte superior de la resina con cuidado.
- Centrifugar a 735 X g durante 2 minutos.
- Almacenar el eluido (la muestra purificada) a -20° C.

Para ligar el DNA purificado en el vector pGEM®-T Easy de Promega se incubó un tubo eppendorf de 1,5 ml a 12° C durante una noche con 12 µl de DNA purificado, 4 µl de agua MilliQ, 2 µl de buffer, 1 µl de vector y 1 µl de ligasa. El producto de esta ligación fue usado para su transformación dentro de la cepa DH5α de *E. coli* siguiendo los siguientes pasos: (Villadas *et al.* 2007).

- Dializar el plásmido durante 15 minutos colocando el producto de la ligación (plásmido) en una membrana de diálisis con un tamaño de poro de 0,025 µm (filtro de nitrocelulosa VSWP de MILLIPORE) que se encuentra flotando en agua destilada estéril dentro de una placa Petri estéril.
- Mientras se dializa el plásmidos, sacar los tubos eppendorf con 50 µl de células electrocompetentes *E. coli* DH5α de -80° C y ponerlos en hielo, junto a las cubetas de electroporación y los tubos eppendorf (estériles) que se van a utilizar.

- Poner 10 μl de plásmido dializado en el tubo eppendorf que contiene las células electrocompetentes (hacer este y los siguientes pasos en campana de flujo laminar para evitar contaminaciones) y mezclar por pipeteo.
- Añadir los 60 μl de la mezcla de células y plásmidos a una cubeta de electroporación.
- Poner la cubeta en el electroporador y darle 2 pulsos a 1.800 V.
- Añadir rápidamente 1.000 μl de medio de cultivo líquido LB en la cubeta, resuspender suavemente por pipeteo y poner la resuspensión de células y plásmidos en un tubo eppendorf.
- Incubar a 37° C durante 1 hora.
- Sembrar en placa diluciones seriadas de la cepa transformada en medio sólido LB con ampicilina y X-Gal e incubar a 37° C durante 24 horas.

Se seleccionaron al menos 30 colonias positivas por muestra crecidas en las placas de LB con ampicilina (200 $\mu\text{g}/\text{ml}$) y X-Gal (40 $\mu\text{g}/\text{ml}$), siendo colonias positivas aquellas que al tener un inserto interrumpiendo el gen *lacZ* del plásmido pGEM-T no generan color azul al crecer en presencia del compuesto químico X-Gal. Cada una de estas colonias se picó y se resuspendió en 100 μl de agua MilliQ. Para lisarlas se incubaron a 96° C durante 10 minutos. El tamaño de inserto de cada colonia lisada fue comprobado mediante una PCR de colonia usando los cebadores T7 (5'-GTAATACGACTCACTATAGG-3') y SP6 (5'-AGCTATTTAGGTGACTATA-3') utilizando las siguientes condiciones de PCR:

- | | | |
|------------------------------------|--------------------|-------------------|
| • DNA molde | 1 μl | |
| • Agua MilliQ | 18,8 μl | |
| • Tampón | 2,5 μl | |
| • Primer T7 (12,5 μM) | | 1 μl |
| • Primer SP6 (12,5 μM) | 1 μl | |
| • dNTPs | 0,5 μl | |
| • <i>Taq</i> DNA polimerasa | | 0,2 μl |

Con el siguiente programa de PCR:

- Desnaturalización inicial de 5 minutos a 94° C
- 30 ciclos de:
 - Desnaturalización de 30 segundos a 94° C
 - Anillamiento de 30 segundos a 56° C
 - Extensión de 2 minutos a 72° C
- Extensión final de 5 minutos a 72° C

Tras la comprobación mediante geles de agarosa (ver apartado 7.1) del resultado de la reacción de PCR, se extrajo el DNA plasmídico de aquellas colonias que presentaron inserto mediante precipitación con sales de Magnesio. Los plásmidos fueron digeridos con la enzima de restricción *NotI* para confirmar el tamaño e integridad del inserto ya purificado. Los insertos de estos plásmidos se amplificaron con los cebadores T7 y SP6 y el producto de esta PCR se mandó a secuenciar después de limpiar en columnas MicroSpin Illustra™ S-300 HR de GE

Healthcare, siguiendo las instrucciones del fabricante. La secuenciación de los clones se realizó mediante tecnología Sanger con un equipo ABI Prism 3130XL Genetic Analyzer, siguiendo las instrucciones del fabricante.

6.2.3 Construcción de librerías NGS del gen 16S rRNA

Para generar estas librerías se han amplificado dos zonas del gen 16S rRNA.

Zona V4-V5 que contiene las regiones hipervariables V4 y V5 de dicho gen (Fig. I-1). Para amplificar dicha zona se utilizaron los cebadores específicos U519F (5'-CAGCMGCCGCGGTAATWC-3') y U926R (5'-CCGTCAATTCCTTTRAGTTT-3') universales para los dominios Bacteria y Archaea (Baker *et al.* 2003). Estos cebadores fueron modificados por la adición de los adaptadores A (5'-CGTATCGCCTCCCTCGGCCATCAG-3') para el cebador Forward y B (5'-CTATGCGCCTTGCCAGCCCGCTCAG-3') para el cebador Reverse, necesarios para llevar a cabo la pirosecuenciación de los amplicones. Dado que, la secuenciación fue bidireccional, es decir, que se utilizaron cebadores complementarios tanto al adaptador A como al adaptador B para llevar a cabo la secuenciación por síntesis, fue necesario el uso de MIDs (Multiplex Identifiers), también llamados Barcodes o códigos de barra, en ambos cebadores. Los MIDs deben ser cortos, entre 5 y 10 nucleótidos, y con una secuencia poco frecuente en la naturaleza, para no interferir en la PCR, además deben tener suficientes diferencias entre ellos para ser discernibles pese a errores introducidos por la polimerasa; por todo ello, se seleccionaron los MIDs a partir de las 48 parejas analizadas por Parameswaran *et al.* (2007), descartando los dos primeros nucleótidos (en posición 5') de los MIDs elegidos con el fin de obtener un cebador útil lo más corto posible y de alta fidelidad. Se utilizó el mismo MID en ambos cebadores para cada muestra, quedando la secuencia definitiva de la pareja de cebadores del siguiente modo:

Cebador Forward (5'-cgtatcgctccctcggccatcag-<XXXXXXXX>-CAGCMGCCGCGGTAATWC-3')

Cebador Reverse (5'-ctatgCGCCTTGCCAGCCCGCTCAG-<XXXXXXXX>-CCGTCAATTCCTTTRAGTTT-3'),

siendo los nucleótidos en minúscula los correspondientes a los adaptadores A y B respectivamente y los MIDs quedan en la zona intermedia representados con el símbolo <XXXXXXXX> por tener una longitud de 8 nucleótidos. Las secuencias de los MIDs utilizados en cada muestra se detallan en la Tabla M-3.

El programa de PCR utilizado fue:

- Desnaturalización inicial de 4 minutos a 94° C
- 25 ciclos de:
 - Desnaturalización de 15 segundos a 94° C
 - Anillamiento de 45 segundos a 55° C
 - Extensión de 1 minuto a 72° C
- Extensión final de 8 minutos a 72° C

Posteriormente se mezclaron los productos de PCR de los tres árboles de cada parcela y también se mezclaron los productos de PCR de las tres parcelas de cada muestra, obteniendo una muestra de cada uno de los sitios de estudio. Se limpiaron 600 µl de PCR de cada muestra con las columnas Amicon® Ultra YM-100 de Millipore, siguiendo las instrucciones del fabricante:

- Añadir producto de PCR a cada filtro.
- Centrifugar 20 minutos a 500 X g.
- Eliminar el eluido y volver a cargar los 300 µl restantes.
- Centrifugar 20 minutos a 500 X g.
- Eliminar el eluido y estimar el volumen retenido.
- Completar hasta 200 µl con agua MilliQ.
- Agitar en vórtex.
- Completar hasta 500 µl con agua MilliQ.
- Centrifugar 30 minutos a 500 X g.
- Eliminar eluido, estimar el volumen retenido y completar hasta 500 µl.
- Realizar los dos pasos anteriores 3 veces.
- Estimar el volumen retenido.
- Centrifugar 5 minutos a 1.000 X g con el filtro invertido.

Posteriormente, se cargaron 0,5 µl de cada muestra en gel de agarosa para confirmar el tamaño e integridad de los amplicones. El siguiente paso fue hacer una dilución 1/10 de 2 µl de cada muestra. Con la ayuda del programa Quantity One se ajustaron los volúmenes necesarios de cada muestra para conseguir una cantidad equimolecular de cada una de ellas. Tras mezclar las 11 muestras se volvieron a concentrar y limpiar con las columnas YM-100. Una dilución 1/10 de la mezcla se cuantificó en gel de agarosa y Nanodrop 1000 y se envió la mitad de la mezcla (MIX) a la empresa LifeSequencing SL (Valencia, España) para su pirosecuenciación con la tecnología 454 GS FLX Titanium y los reactivos para secuenciación bidireccional en un cuarto de placa, con un rendimiento esperado de entre 125.000 y 200.000 lecturas.

Material y Métodos

Tabla M-3. MIDs utilizados en cada muestra.

Muestra	Descripción	MID
CNA	Cáñar zona alta del robledal	CCGTTACG
CNB	Cáñar zona de baja altitud del robledal	CCTGGCAT
CNX	Cáñar zona de expansión del robledal, muestra rizosférica	TTCAAGTC
CNR	Cáñar zona de expansión del robledal, suelo suelto	CCTCCTGA
CTR	Dehesa del Camarate, rizosfera de robles	TTAGGTAC
LJN	Lanjarón encinas no quemadas	AAGAACGT
LJQ	Lanjarón encinas quemadas	GGCAATGC
LJR	Lanjarón zona quemada de herbáceas, suelo suelto	CCATTGCA
CTE	Dehesa del Camarate, rizosfera de encinas	GGCTTCAG

Zona V6-V8 que contiene las regiones hipervariables V6, V7 y V8 de dicho gen (Fig. I-1). Para amplificar dicha zona se utilizaron los cebadores específicos 926F (5'-AAACTYAAAKGAATTGRCGG-3') y 1392R (5'-ACGGGCGGTGTGTRC-3') universales para los dominios Bacteria y Archaea (Engelbrektsen *et al.* 2010). Estos cebadores fueron modificados por la adición de los adaptadores A (5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3') para el cebador Forward y B (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3') para el cebador Reverse, necesarios para llevar a cabo la pirosecuenciación de los amplicones. Dado que, la secuenciación fue unidireccional, es decir, que se utilizaron cebadores complementarios solo para uno de los adaptadores, en este caso para el adaptador B, para llevar a cabo la secuenciación por síntesis, fue necesario el uso de MIDs solo en el cebador reverso. Los MIDs utilizados se obtuvieron de Sogin *et al.* (2006) y Parameswaran *et al.* (2007), descartando los cinco primeros nucleótidos (en posición 5') de los MIDs elegidos con el fin de obtener un cebador útil lo más corto posible. Quedando la secuencia definitiva de la pareja de cebadores del siguiente modo:

Cebador Forward (5'-cctatcccctgtgtgccttggcagtctcag-AAACTYAAAKGAATTGRCGG-3')
 Cebador Reverse (5'-ccatctcatccctgctgtctccgactcag-<XXXXX>-ACGGGCGGTGTGTRC-3'),
 siendo los nucleótidos en minúscula los correspondientes a los adaptadores A y B respectivamente y los MIDs quedan en la zona intermedia representados con el símbolo <XXXXX> por tener una longitud de 5 nucleótidos. Las secuencias de los MIDs utilizados en cada muestra se detallan en la Tabla M-4.

En este caso las condiciones de PCR fueron:

- DNA molde 10 ng
- Agua MilliQ hasta 20 μ l
- Tampón 5X GC PCR 4 μ l
- GC Melt Solution (5M) 2 μ l
- Primer 926F (25 nM) 1 μ l
- Primer 1392R (25 nM) 1 μ l
- dNTPs 0,4 μ l
- Advantage GC 2 Polimerase Mix 0,4 μ l

El programa de PCR utilizado fue:

- Desnaturalización inicial de 3 minutos a 95° C
- 25 ciclos de:
 - Desnaturalización de 30 segundos a 95° C
 - Anillamiento de 45 segundos a 50° C
 - Extensión de 1 minuto a 68° C
- Extensión final de 10 minutos a 68° C

En este caso se mezclaron los productos de PCR de los tres árboles de cada parcela, para conseguir una muestra homogénea de cada una de éstas. De modo que, se obtuvo una muestra de cada una de las tres parcelas de cada sitio estudiado. El producto de PCR de cada muestra se purificó con el kit Agencourt® AMPure® XP de Beckman Coulter, siguiendo las instrucciones del fabricante:

- Añadir 90 μ l de Agencourt® AMPure® XP a cada muestra de 50 μ l.
- Mezclar los reactivos con el producto de PCR pipeteando 10 veces.
- Incubar 5 minutos a temperatura ambiente.
- Colocar la microplaca sobre la placa magnética Agencourt® SPRIPlate 96 Super Magnet Plate.
- Esperar 2 minutos o hasta que la mezcla se aclare.
- Aspirar y descartar el sobrenadante sin arrastrar perlas magnéticas.
- Añadir 200 μ l de Etanol 70 %.
- Incubar 30 segundos a temperatura ambiente.
- Aspirar y descartar el Etanol.
- Realizar los 3 pasos anteriores 2 veces.
- Incubar 3 minutos a temperatura ambiente.
- Retirar la microplaca de la placa magnética Agencourt® SPRIPlate 96 Super Magnet Plate.
- Añadir 40 μ l de tampón TE y mezclar 10 veces mediante pipeteo.
- Colocar la microplaca sobre la placa magnética Agencourt® SPRIPlate 96 Super Magnet Plate.
- Esperar 1 minuto para que las perlas magnéticas se separen.
- Transferir el sobrenadante a una nueva microplaca sin arrastrar perlas magnéticas.

Material y Métodos

Posteriormente, se cuantificaron las muestras con un Qubit® 2.0 Fluorometer y se diluyeron con agua a 10 ng/μl. Se hicieron dos mezclas equimoleculares, la primera contiene treinta muestras, procedentes todas ellas del muestreo de 2008 y la segunda contiene otras treinta, tres de las cuales proceden del muestreo de 2008 y veintisiete del muestreo de 2011 (ver Tabla M-4). Cada MIX fue pirosecuenciado en media placa con los reactivos de secuenciación unidireccional en un equipo 454 GS FLX Titanium en el centro de investigación Joint Genome Institute (JGI, Walnut Creek, California, USA), con un rendimiento esperado de entre 400.000 y 500.000 lecturas por MIX.

Tabla M-4. MIDs utilizados en cada muestra para cada MIX pirosecuenciado. Las muestras en **negrita** tuvieron que ser resecuenciadas.

MIX 1	MID	MIX 2	MID
CNA1_08	CTACT	LJR1_08	CTACT
CNA2_08	CTCGC	LJR2_08	CTCGC
CNA3_08	ATGCT	LJR3_08	TCTGC
CNB1_08	TACGT	CNA1_11	TACGT
CNB2_08	TAGAG	CNA2_11	TAGAG
CNB3_08	TCATC	CNA3_11	TCATC
CNR1	TCTAT	CNB1_11	TCTAT
CNR2	TGCTG	CNB2_11	TGCTG
CNR3	TGTAG	CNB3_11	TGTAG
CNXP1	ACGTC	CTR1_11	ACGTC
CNXP2	ACACG	CTR2_11	ACACG
CNXP3	AGTGT	CTR3_11	AGTGT
CNXV1	AGCAC	CNX1_11	AGCAC
CNXV2	ATGCT	CNX2_11	ATGCT
CNXV3	ATATG	CNX3_11	ATATG
CNXO1	CATGC	LJN1_11	CATGC
CNXO2	CACAG	LJN2_11	CACAG
CNXO3	CGATG	LJN3_11	CGATG
CTR1_08	TCTCG	LJQ1_11	TCTCG
CTR2_08	TCTGC	LJQ2_11	TCTGC
CTR3_08	TCGCT	LJQ3_11	TCGCT
LJN1_08	TCGTG	CTE1_11	TCGTG
LJN2_08	TCGAC	CTE2_11	TCGAC
LJN3_08	TCAGT	CTE3_11	TCAGT
LJQ1_08	CATGC	LJR1_11	CGATG
LJQ2_08	CACAG	LJR2_11	TCTCG
LJQ3_08	TGTGT	LJR3_11	TGTGT
CTE1_08	TGACT	LJRE1	TGACT
CTE2_08	ATATG	LJRE2	TCGCT
CTE3_08	TATCT	LJRE3	TATCT

6.3 Aproximación Metagenómica

Gracias a las nuevas técnicas de secuenciación masiva podemos obtener una enorme cantidad de información taxonómica y funcional de forma directa de las comunidades microbianas a bajo coste. Para llevar a cabo este tipo de estudio se pirosecuenciaron 5 µg de DNA, a una concentración de 300 ng/µl, de cada una de las siguientes muestras, LJN, LJQ, CNA y CNB en un equipo 454 GS FLX Titanium de Roche, realizado por la empresa LifeSequencing SL, a la cual se le pidieron unas 250.000 lecturas de cada muestra con un tamaño de 400 bp. También se utilizaron otros 5 µg de DNA, a una concentración de 300 ng/µl, de cada una de las cuatro muestras anteriores y de la muestra CNX para realizar una secuenciación masiva mediante la tecnología HiSeq2000 PE 101 de Illumina®, realizado en el Beijing Genomics Institute (Tai Po, Hong Kong, China). En este caso, el DNA de cada muestra se fragmentó de forma mecánica hasta un tamaño aproximado de 170 bp y se secuenciaron 101 bp de cada secuencia por ambas hebras para obtener 2 lecturas pareadas (Paired-end) de cada secuencia. Se le pidieron 2x180 millones de lecturas pareadas (1 línea del secuenciador HiSeq2000) de cada muestra.

7. Electroforesis de Ácidos Nucleicos

La electroforesis es una técnica separativa basada en la migración diferencial en sentido y velocidad de partículas cargadas en el seno de un campo eléctrico, que fue desarrollada por Tiselius en 1937. Consiste en la separación de macromoléculas en función de su tamaño y carga. Los ácidos nucleicos, con carga negativa, migran hacia el polo positivo.

7.1 Electroforesis en geles de agarosa no desnaturalizantes

El análisis de productos de PCR, de DNA total o plasmídico o de fragmentos de restricción se llevó a cabo mediante electroforesis horizontal en geles de agarosa (agarosa SeaKemR LE, Cambrex/Iberlabo) al 0,8 % en tampón TAE 1x: [TrisHCl 40 mM, EDTA Na₂ 2 mM, ácido acético glacial 0,1142 % (v/v)]. El voltaje habitual de trabajo fue de 120 V aunque determinados procedimientos requirieron velocidades inferiores, todos ellos llevados a cabo

en cubetas de la empresa BioRad. Como tampón de carga se utilizó una solución 6 X de 0,25 % (p/v) de azul de bromofenol y glicerol 45 % (v/v).

El revelado de los geles de agarosa se realizó por inmersión de los mismos en una solución de bromuro de etidio 1 mg/ml en agua durante 15-20 minutos. Este intercalante permite visualizar DNA o RNA en un transiluminador que emite luz UV de 365 nm de longitud de onda. Las fotografías de los geles se realizaron con Gelprinter y Gelstation de TDI para grabar y procesar las imágenes.

7.2 Electroforesis en gel desnaturalizante con gradiente de temperatura (TGGE)

Para la resolución de los amplificadores del gen ribosomal 16S *rRNA* en función de su secuencia, se empleó un sistema de electroforesis basado en geles desnaturalizantes con gradiente de temperatura (TGGE Maxi, BIOMETRA), que fue utilizado según las especificaciones del fabricante. Se utilizaron geles con un 6 % de acrilamida/bisacrilamida, 1 % TAE, 8 M Urea, 20 % Formamida desionizada y 2 % Glicerol. La polimerización fue iniciada con 100 µl de TEMED y 200 µl de Persulfato Amónico al 10 %. Posteriormente, se cargaron entre 2-8 µl de producto de PCR al cual se le añadió 1/5 de tampón de carga (5 X TAE, 0,1 % Tritón X-100 y 0,01 de azul de bromofenol). La electroforesis de los fragmentos del gen 16S *rRNA* se realizó durante 16 h a 130 V, bajo un gradiente térmico de 44,5° C a 56° C

La presencia de las diferentes bandas del gel fue evidenciada mediante tinción con nitrato de plata (Heuer *et al.* 1997). Todos los pasos se realizaron en agitación: dos lavados de 3 minutos en solución de fijación, incubación durante 15 minutos en solución de tinción, cuatro lavados (aproximadamente 4 segundos cada vez) con agua desionizada, inmersión en solución de revelado hasta que se observen todas las bandas, lavado con agua desionizada e inmersión en la solución de parada durante 10 minutos. Finalmente, se dejó reposar en solución de conservación durante 7 minutos.

Tabla M-5. Composición de las soluciones empleadas en la tinción con plata.

Solución	Composición
Fijación	10 % Etanol, 0,5 % Ácido Acético
Tinción	0,2 % Nitrato de plata
Revelado	1,5 % NaOH, 0,01 Borohidruro Sódico, 0,15 % Formaldehído
Paro	0,75 % p/v Carbonato sódico
Conservación	25 % w/w etanol, 10 % v/v glicerol

8. Análisis Bioinformático

8.1 RDP's Pyrosequencing Pipeline

La secuenciación del gen 16S *rRNA* proveniente de muestras medioambientales es uno de los métodos más utilizados para determinar la composición de las comunidades de bacterias y arqueas. Las nuevas técnicas de secuenciación masiva, como la pirosecuenciación, son cada vez más utilizadas gracias a que producen gran cantidad de secuencias y evitan sesgos derivados del proceso de clonación, permitiendo un análisis más profundo de dichas comunidades. Por ello, hemos utilizado este tipo de aproximación y para procesar las secuencias obtenidas mediante pirosecuenciación del gen 16S *rRNA* de nuestras muestras, utilizamos el paquete de herramientas web RDP (Ribosomal Database Project, <http://rdp.cme.msu.edu/>). El software del servidor RDP nos permite procesar de forma bastante simplificada muestras con gran número de secuencias y además, permite obtener los resultados en formatos muy utilizados por los paquetes estadísticos más comunes en ecología, como son SPADE, EstimateS, MOTHRUR y R (Cole *et al.* 2009). Las herramientas básicas que hemos utilizado para analizar nuestras muestras son:

- **Pipeline Initial Process:** Gracias a esta herramienta se procesaron los archivos fasta y quality, actualmente se introduce directamente el archivo sff, de cada ronda de secuenciación. Es necesario un archivo de texto, en el que en cada línea se escribe el código de barras (MID) perteneciente a una muestra y seguido de un tabulador se escribe el nombre de dicha muestra, este archivo

tiene tantas líneas como muestras se mezclaron para la misma ronda de secuenciación. Es necesario introducir el Forward Primer (Cebador Específico que amplifica desde la región 5' del gen 16S *rRNA*). El Reverse Primer es opcional, pero si se indica sólo devuelve secuencias que tienen ambos cebadores. En el caso de pirosecuenciación bidireccional con amplicones de 407 bp, pusimos ambos cebadores, puesto que más del 90 % de las lecturas son secuenciadas completamente. Además fue necesario poner el cebador de la región 3' como Forward Primer y el de la región 5' como Reverse Primer para obtener todas aquellas lecturas que fueron secuenciadas desde la región 3' del gen. Posteriormente se unen los dos archivos fasta de cada muestra con el comando *cat*. En el caso de pirosecuenciación unidireccional, sólo pusimos el cebador de la región 3' como Forward Primer, ya que la secuenciación fue en sentido 3'-5' respecto al gen. Al obtener fragmentos de 585 bp, la cantidad de fragmentos que se secuencian completamente es mucho menor y por ello sólo pusimos el Forward Primer. **Filtros:** Para ambos cebadores se cambia este valor de 0 a 2, ya que ambos son degenerados y encontramos muchas variantes de los mismos en las secuencias de partida. Número de N (ambigüedades) lo dejamos en 0. En el tamaño mínimo de las secuencias se deja en 150 nt. No marcamos la opción Keep primers (mantener los cebadores). Antes de alinear las secuencias se procesaron con las herramientas del paquete informático **MOTHUR** para eliminar las posibles quimeras (ver más adelante).

- **Aligner:** Las secuencias procesadas con la herramienta anterior se alinearon mediante esta aplicación, la cual utiliza el programa Infernal aligner (Nawrocki y Eddy, 2007) para llevar a cabo dicho alineamiento, teniendo en cuenta la estructura secundaria del gen 16S *rRNA*. Actualmente, es capaz de alinear hasta 1 millón de secuencias por proceso. En el caso del modelo Bacteria de la estructura secundaria del gen 16S *rRNA*, utilizado en todos los análisis, las secuencias se alinean teniendo como referencia la secuencia del gen 16S *rRNA* completo de la cepa de *E. coli* J01695. Todas las muestras fueron alineadas tanto por separado, una a una, como juntas.
- **Complete Linkage Clustering:** Con esta herramienta agrupamos las secuencias en Unidades Taxonómicas Operativas (OTUs por su abreviatura en inglés) a la distancia especificada, en nuestro caso al 3 % de distancia, ya que es el valor aceptado para considerar dichos OTUs como especies cuando se compara el

gen 16S *rRNA* (Bond *et al.* 1995, Borneman *et al.* 1997, Hugenholtz *et al.* 1998, Everett *et al.* 1999, McCaig *et al.* 1999, Sait *et al.* 2002). El método de agrupamiento utilizado es el Furthest Neighbor (vecino más alejado), por el cual todas las secuencias dentro de un OTU se difieren como máximo del resto de secuencias de este mismo OTU en un 3 %. Es decir, la similitud entre todas las secuencias de un OTU determinado es igual o superior al 97 %. Esta herramienta computa la matriz de distancias para todas las secuencias (matriz que el programa no nos devuelve) y a partir de ella, realiza el agrupamiento (clustering). Al analizar con este programa varias muestras al mismo tiempo, obtuvimos un archivo con el número de secuencias de cada muestra analizada que pertenecen a un mismo OTU, información necesaria para conocer la diversidad β . Actualmente, es capaz de agrupar hasta 150.000 secuencias por proceso.

- **Índices de Shannon y Chao1:** El archivo de clustering se introduce en esta aplicación para obtener los índices de diversidad α de cada muestra. Además del índice de diversidad de Shannon y el índice de riqueza estimada de Chao1, nos devuelve el índice de equidad de Pielou (Evenness).
 - **Índice de Chao1 (riqueza estimada):** Es un estimador del número de especies en una comunidad basado en el número de especies raras en la muestra (Chao, 1984; Chao y Lee, 1992; Smith y van Belle, 1984). Se calcula aplicando la siguiente fórmula:
$$\text{Chao1} = S + \frac{2a}{2b}$$
 Donde S es el número de especies en una muestra, a es el número de especies que están representadas solamente por un único individuo en esa muestra (número de “singletons”) y b es el número de especies representadas por exactamente dos individuos en la muestra (número de “doubletons”, Colwell, 1997; Colwell y Coddington, 1994).
 - **Índice de Shannon (diversidad estimada):** También conocido como índice de Shannon-Wiener o Shannon-Weaver. Este índice de abundancia proporcional relaciona la cantidad de especies presentes con el número de individuos de cada una. Expresa, por tanto, el grado promedio de incertidumbre en la predicción de la especie de un individuo escogido al azar desde una comunidad (Magurran, 1988). Su

fórmula es: $H' = -\sum p_i \ln(p_i)$ Donde p_i es la abundancia relativa del OTU i .

- **Índice de Pielou (equidad):** mide la proporción de la diversidad observada con relación a la máxima diversidad esperada. Su valor oscila entre 0 y 1, siendo 1 el valor correspondiente a situaciones donde todas las especies son igualmente abundantes (Magurran, 1988). Se calcula aplicando la fórmula: $J' = \frac{H'}{\ln S}$ Donde H' es el índice de Shannon y S el número de OTUs. El logaritmo neperiano de S es la forma de calcular el máximo valor de H' ($H' \max$) que se puede dar en una población de S número de OTUs. En nuestro caso J' se representa como E (evenness). Este índice es muy práctico cuando se quiere comparar la diversidad entre dos muestras que no tienen el mismo número de OTUs, ya que el índice de Shannon aumenta con el aumento del número de OTUs. Por lo tanto, ayuda a estandarizar la diversidad y que ésta no dependa del número de OTUs obtenidos.
- **Índice de Jaccard y Sørensen:** El archivo de clustering se introduce en esta aplicación para obtener los índices de diversidad β de todas las muestras. Se utilizó el índice de Jaccard por ser el más usado habitualmente en este tipo de análisis además de que ambos mostraron resultados equivalentes.
 - **Índice de Jaccard:** estima la similitud entre dos muestras distintas, basado en la probabilidad que cada individuo de cada muestra tiene de pertenecer a una especie (OTU) compartida por ambas muestras. Se calcula la similitud basada en abundancia con la siguiente fórmula: $Jaccard\ Index = \frac{UV}{U+V-UV}$ Donde U es el total de sumar las abundancias relativas de las especies compartidas de la primera muestra y V es el total de sumar las abundancias relativas de las especies compartidas de la segunda muestra (Chao *et al.* 2006). Con los valores de similitud de dicho índice para todas las muestras, comparadas dos a dos, esta herramienta genera una matriz de similitudes y un dendrograma basado en Agrupamiento Jerárquico Aglomerativo (Agglomerative Hierarchical Clustering).
- **Rarefacción:** Introdujimos el archivo de clustering en esta aplicación para obtener las curvas de esfuerzo de cada muestra.

- Curva de esfuerzo:** muestra el cambio en el valor esperado de riqueza de especies de acuerdo al tamaño de las muestras. Calcula el número esperado de especies de cada muestra si todas las muestras fueran reducidas a un tamaño estándar. Es decir, si la muestra fuera considerada de n individuos, donde $n < N$, ¿cuántas especies se habrían registrado? Se calcula con la siguiente fórmula:

$$E(Sn) = \sum_{i=1}^N \left[1 - \left(\frac{N-N_i}{n} \right)^n \right] \quad \text{Donde } E(Sn) \text{ es el número de especies}$$

- (OTUs) esperado, N es el número total de individuos en la muestra, N_i el número total de individuos de la especie i , y n submuestra de tamaño estándar de N . Se representa con un gráfico, con el número de secuencias en el eje de abscisas (X) y el número esperado de OTUs en el eje de ordenadas (Y) (Fernández *et al.* 2001). Proporciona la información de si el esfuerzo de muestreo realizado ha sido suficiente para tener una muestra representativa de la comunidad de origen o no. También informa de si un mayor esfuerzo de muestreo, aumentaría o no, significativamente el registro de especies anteriormente no registradas (Kenneth *et al.* 1975).
- Representative Sequence:** El archivo de clustering se introdujo en esta aplicación para obtener una secuencia representativa de cada OTU. Esta herramienta elige la secuencia con la mínima suma de cuadrados de distancias respecto a todas las demás de su mismo OTU.
 - Cluster to R Formatter:** Introdujimos el archivo de clustering en esta aplicación para obtener una matriz con una muestra por fila y un OTU por columna. Esta matriz o tabla se traspuso en una hoja de cálculo (MS Excel 2010) de modo que se trabajó en ella con un OTU por fila y una muestra por columna. Las distintas muestras fueron comparadas 2 a 2 aplicando la fórmula de **RDP Lib Compare**, para conocer qué OTUs eran significativamente más abundantes en una muestra respecto a la otra.
 - RDP Classifier:** Con esta herramienta clasificamos las secuencias filtradas por **Pipeline Initial Process** o las secuencias representativas de cada OTU. Esta aplicación utiliza un árbol filogenético pre-computado con asignación taxonómica y mediante el algoritmo naïve Bayesian clasificó nuestras secuencias, dándole a cada una de ellas un valor de bootstrapping (entre 0 % y 100 %), para cada nivel taxonómico: Dominio, *phylum*, clase, orden, familia y

género. Podemos descargar el programa `rdp_classifier` y utilizarlo de forma local, útil cuando el servidor está saturado o falla la conexión a Internet.

- **RDP Lib Compare:** Esta herramienta utiliza el RDP Classifier para clasificar dos muestras y estima la probabilidad de observar la diferencia en un determinado taxón utilizando la siguiente prueba estadística:

$$p(y|x) = \binom{N_2}{N_1}^y \frac{(x+y)!}{x!y! \left(1 + \frac{N_2}{N_1}\right)^{(x+y+1)}} \quad \text{Donde, } x \text{ es el número de secuencias}$$

observadas en la muestra "1" en un determinado taxón; y es el número de secuencias observadas en la muestra "2" en este mismo taxón; N_1 es el número de secuencias de la muestra "1"; N_2 es el número de secuencias de la muestra "2".

- **MOTHUR: Phylip Distance Matrix:** Introdujimos todos los archivos de alineamiento, procesados con **Aligner**, en esta aplicación para obtener una matriz de distancias en formato Phylip Lower Triangular. Este algoritmo anota en una matriz simétrica las distancias no corregidas de cada pareja de secuencias previamente alineadas. Actualmente, es capaz de analizar hasta 50.000 secuencias por envío. Este tipo de matrices ocupan varias decenas de Gigas cuando provienen de, al menos 10 muestras con 10.000 secuencias cada una. Con esta matriz se llevó a cabo un análisis estadístico llamado *AMOVA* dentro del paquete bioinformático **MOTHUR**.
- **Good's coverage:** Se calcula con el archivo de clustering. Es un índice que estima la proporción de OTUs observados respecto al total de OTUs esperados en la población. No está incluido en el paquete de herramientas web RDP, pero es fácil de obtener en una hoja de cálculo, como las de Microsoft Excel, mediante la siguiente fórmula: $G = 1 - \frac{n}{N}$ Donde n es el número de singltons (OTUs formados por una única secuencia) y N es el número de secuencias de la muestra (Good, 1953).

8.2 MOTHUR

MOTHUR (<http://www.mothur.org/>) es un paquete informático que combina gran cantidad de programas especializados en el análisis de datos moleculares útiles en ecología microbiana. Actualmente permite procesar las secuencias brutas, tal cual nos llegan del secuenciador, hasta la obtención de datos como la diversidad α y β (Schloss *et al.* 2009). Al igual que RDP, esta herramienta surgió por la necesidad de procesar miles de secuencias provenientes de las nuevas metodologías de secuenciación masiva; principalmente para estudios de diversidad y por tanto, enfocada en el análisis de secuencias del gen 16S *rRNA*. Pero, es una aplicación mucho más flexible que RDP a la hora de analizar otros genes de interés. El autor ofrece una guía *on-line* (http://www.mothur.org/wiki/Schloss_SOP) para utilizar los programas más frecuentes para cumplir nuestros objetivos de análisis (Schloss *et al.* 2011).

Ya que en este trabajo las secuencias se procesaron mediante RDP, se detallan solo las aplicaciones utilizadas en MOTHUR que complementan de forma sencilla las limitaciones del software de RDP.

- **Chimera.uchime:** Después de filtrar las secuencias por **Pipeline Initial Process**, en RDP, buscamos las secuencias que potencialmente son quimeras mediante esta herramienta. A partir del archivo fasta procesado por RDP y con un alineamiento de referencia, con secuencias completas del gen 16S *rRNA* revisadas previamente y confirmadas como secuencias no quiméricas, podemos analizar la probabilidad de que fracciones de una misma secuencia pertenezcan a individuos alejados filogenéticamente. La base de datos que se ha usado proviene de SILVA (<http://www.arb-silva.de>) y contiene secuencias del gen 16S *rRNA* pertenecientes al dominio Bacteria, el enlace a dicho alineamiento está en el propio paquete MOTHUR. La forma de ejecutar esta aplicación es :
 - `chimera.uchime(fasta=secuencias.fasta, reference=silva.gold.align)`
 - `remove.seqs(accnos=nombre_quimeras.accnos, fasta=secuencias.fasta)`
 - Devuelve un archivo secuencias.pick.fasta sin las quimeras.
- **Sub.sample:** Con este comando pudimos normalizar todas nuestras muestras, tomando siempre como valor de normalización el número de secuencias de la muestra más pequeña. Es necesario normalizar nuestras muestras para poder

hacer estudios comparativos entre los resultados obtenidos de diversidad α y β . Si nuestra muestra más pequeña tuviera 8.250 secuencias, la forma de ejecutar el comando sería:

- `sub.sample(fasta=secuencias.pick.fasta, size=8250)`
- Devuelve un archivo `secuencias.pick.subsample.fasta` con 8.250 secuencias.
- **Libshuff:** Es un test para describir si 2 o más comunidades tienen la misma estructura utilizando el estadístico de Cramer-von Mises. Para más de dos muestras se aplica el integral de Libshuff y una corrección de Bonferroni a los p-valores. El p-valor sale siempre significativo cuando se trabaja con miles de secuencias, por lo que sólo fue utilizado para tratar las secuencias provenientes de genotecas clásicas. Para utilizar este comando necesitamos una matriz de distancias con todas las secuencias a comparar y un archivo de grupos que detalla a qué muestra pertenece cada secuencia utilizada en la matriz.
 - `make.groups(fasta=muestra1.fasta-muestra2.fasta, groups=A-B)`
 - `dist.seqs(fasta=secuencias.pick.subsample.fasta)`
 - `libshuff(phylip=secuencias.pick.subsample.phylip.dist, group=secuencias.groups)`
- **Amova:** Es un test estadístico análogo al ANOVA. Es un análisis molecular de la varianza. Este método prueba la hipótesis de que la diversidad genética entre dos poblaciones no es significativamente diferente de la diversidad genética resultante de una población mezcla de ambas poblaciones. Este comando también necesita la matriz de distancias con todas las secuencias a comparar y un archivo de grupos, que en este caso se llama archivo diseño. Ambos archivos se pueden obtener directamente de la aplicación **MOTHUR: Phylip Distance Matrix** de RDP:
 - `amova(phylip=secuencias.pick.subsample.phylip.dist, design=secuencias.groups)`

8.3 FastTree

El software FastTree infiere árboles filogenéticos muy próximos a los producidos por Máxima Verosimilitud a partir de alineamientos de secuencias nucleotídicas o aminoacídicas. Actualmente es la herramienta, para este propósito, más rápida y capaz de producir árboles filogenéticos con alineamientos de hasta 1 millón de secuencias (Price *et al.* 2010).

Para obtener el árbol resultante al mejor rendimiento, utilizamos el comando FastTreeMP capaz de utilizar más de 1 CPU. Los parámetros utilizados fueron los recomendados por los autores en su página web (<http://www.microbesonline.org/fasttree/>):

- `FastTreeMP -nt -gtr -spr 4 -mlacc 2 -slownni alnsecuencias.fasta > alnsecuencias.tree`

8.4 FastUnifrac

Esta herramienta se utiliza para la comparación de comunidades microbianas a partir de su información filogenética (<http://unifrac.colorado.edu/>). Con esta aplicación pudimos complementar los estudios de diversidad β de las comunidades microbianas realizados por distancia genética en RDP y MOTHUR. Para realizar el análisis necesitamos un árbol filogenético de todas nuestras secuencias (realizado con FastTree), un archivo de grupos (en el que identificamos cada secuencia a su muestra de origen, obtenido por MOTHUR o RDP) y por último un archivo de categorías en el que se detallan los metadatos (alto, bajo; quemado, no quemado) de cada una de las muestras utilizadas en el análisis (Hamady *et al.* 2010).

8.5 Procesamiento de datos metagenómicos

Tras la extracción y secuenciación masiva del DNA genómico de las comunidades microbianas que habitan en los suelos (metagenomas según Handelsman *et al.* 1998) de las muestras previamente indicadas (ver en el apartado 6.3) se llevaron a cabo los siguientes pasos para procesar y analizar los datos adquiridos.

8.5.1. Limpieza de lecturas metagenómicas

A partir de las secuencias brutas (Raw) obtenidas tras la secuenciación directa del DNA se procedió al curado de las mismas de este modo:

- Para las secuencias obtenidas mediante el sistema 454:
 - Se eliminaron las secuencias repetidas con la herramienta informática 454 Replicate Filter web tool (Gómez-Álvarez *et al.* 2009), con un corte al 90 % de identidad de secuencia, un valor de 0 en la diferencia de longitud y un valor de 10 en el número de nucleótidos del inicio de la secuencia a analizar.
 - Se realizó un filtrado de las secuencias de baja complejidad (ricas en repeticiones de mono/dinucleótidos) y baja calidad (valor de Q promedio inferior a 20) con la herramienta informática SeqTrim pipeline (Falgueras *et al.* 2010), usando los parámetros por defecto.
- Para las secuencias obtenidas mediante el sistema Illumina:
 - Utilizando la herramienta Combine separate Read1 and Read2 del servidor web GALAXY/JGI (<https://galaxy.jgi-psf.org/>) se fusionaron los dos archivos FastQ, Read1 perteneciente a la región 5' de las lecturas y Read2 perteneciente a la región 3' de las mismas, devolviendo un archivo FastQ con las lecturas pareadas ordenadas de forma contigua.
 - Utilizando la herramienta Import Illumina Data del servidor web GALAXY/JGI se descomprimió el archivo FastQ, se eliminaron todas aquellas secuencias con más de un nucleótido ambiguo "N", se eliminaron las regiones 3' de las secuencias que tenían una calidad inferior a Q³, observando dicho valor de tres en tres nucleótidos en dirección 3' → 5' (este filtro es laxo, ya que el programa FLASH utiliza su propio filtro de calidad más estricto). Se eliminaron aquellas secuencias que tras los pasos anteriores tienen un tamaño menor de 50 nucleótidos. Por último, eliminó todas aquellas secuencias de baja complejidad, siendo éstas, las lecturas que tienen más del 90 % de su secuencia constituida por repeticiones mono/dinucleotídicas

8.5.2. Solapamiento y ensamblaje de lecturas Illumina

Puesto que, se solicitó a la empresa BGI que las librerías tuvieran un tamaño promedio de 170 nucleótidos y la tecnología de secuenciación utilizada solo permite leer los 100 primeros nucleótidos, se realizó la lectura de la misma secuencia desde ambos extremos. Posteriormente se solaparon “*in silico*” las secuencias apareadas.

- Las secuencias ya curadas fueron reconstruidas mediante el solapamiento de cada secuencia (leída en dirección 5' → 3') con su pareja (leída en dirección 3' → 5') utilizando el programa FLASH v 1.0.3, albergado también en el servidor web GALAXY/JGI, utilizando los parámetros por defecto.
- Se realizaron varios ensamblajes de las secuencias, tanto las solapadas como las no solapadas, utilizando el programa SOAPdenovo (Li *et al.* 2010), empleando para cada ensamblaje tamaños de k-mer de 71, 75, 79, 83, 87 y 91. Posteriormente se fusionaron los seis ensamblajes resultantes eliminando los contigs redundantes, manteniendo siempre el de mayor tamaño. Para llevar a cabo este procedimiento y conocer qué tamaños de k-mer eran los más apropiados se utilizó la herramienta mgaSOAPdenovo del servidor web GALAXY/JGI.
- **Contigs:** secuencias contiguas no solapantes de un genoma (Edwards *et al.* 1990; Zerbino *et al.* 2008). En un metagenoma desconocemos la contigüidad de estas secuencias al tratarse de fragmentos secuenciados al azar de, en nuestro caso, miles de genomas de diferentes individuos. Así en metagenómica, se considera como contig la secuencia de mayor tamaño, de las obtenidas por secuenciación masiva “reads”, que es reconstruida tras la identificación de reads solapantes, posterior alineamiento y fusión de las mismas (Namiki *et al.* 2012).

8.5.3. Anotación de secuencias metagenómicas

Para la anotación taxonómica de las secuencias ya curadas, previas al solapamiento y ensamblado, procedentes de la secuenciación Illumina se utilizaron varios scripts escritos en lenguaje de programación PERL que procedieron de la siguiente forma:

Material y Métodos

- **Clustering:** Las secuencias se fragmentaron en k-mers de 20 nucleótidos y estos 20-mers se compararon entre sí y se agruparon todos los que eran 100 % idénticos. Se seleccionó solo un 20-mer representativo de cada grupo para eliminar redundancia y reducir sensiblemente el tiempo y el esfuerzo de análisis.
- **BLASTn:** Los 20-mers representativos se enfrentaron a la base de datos nucleotídica de SILVA completa (16S/18S, SSU y 23S/28S, LSU de los tres dominios *Bacteria*, *Archaea* y *Eukarya*) y se eliminaron todos aquellos 20-mers que no mostraron homología.
- **Clasificación:** Se recuperaron aquellas secuencias de las que procedían los 20-mers con homología tras el BLASTn y se clasificaron con la herramienta rdp_classifier versión 2.3, pero en vez de utilizar el árbol filogenético pre-computado con la base de datos RDP, se utilizó un árbol filogenético pre-computado en el JGI con la base de datos SILVA.

Para la anotación taxonómica y funcional tanto de las secuencias obtenidas por pirosecuenciación como de los contigs obtenidos tras ensamblar las secuencias obtenidas por Illumina, se procedió de la misma forma siguiendo los siguientes pasos:

- **MG-RAST:** Se enfrentaron las secuencias y los contigs de los diferentes metagenomas a diferentes bases de datos (M5NR database que integra las bases de datos IMG, TrEMBL, PATRIC, SwissProt, GenBank, Refseq, SEED, eggNOG y KEGG para proteínas y Greengenes, Silva y RDP para rRNA) implementadas en el servidor web MG-RAST (Meyer *et al.* 2008), con un corte máximo de e-valor de 10^{-5} , una identidad mínima del 60 % y una longitud mínima de alineamiento de 50 bp.
- **IMG:** Se enfrentaron las secuencias y los contigs de los diferentes metagenomas a diferentes bases de datos (PSSMs database de COGs del NCBI y Pfam-A database del instituto Sanger para proteínas y el programa SPARTAN para la identificación y anotación de los diferentes tRNA y rRNA) integradas en el servidor web IMG/MER (Markowitz *et al.* 2011), utilizando los parámetros por defecto.
- **BLAST(n/x):** Se enfrentaron las secuencias y los contigs de los diferentes metagenomas a las bases de datos NT (compuesta por secuencias nucleotídicas) y NR (compuesta por secuencias aminoácidas no redundantes de proteínas) del NCBI (ver apartado 8.6).

Para el reclutamiento genómico de los contigs del metagenoma de la muestra LJQ secuenciado mediante Illumina pertenecientes al pan-genoma del género *Arthrobacter* frente a las cepas de referencia *A. globiformis* NBRC12137 y *A. sp.* FB24 se realizaron los siguientes pasos en el programa **MUMmer** (<http://mummer.sourceforge.net/>):

- *Nucmer --coords -p archivo_salida Reference.fasta Query.fasta*
- *Show-coords -rcl archivo_salida.delta > archivo_salida.coords*
- *Mummerplot --filter --coverage --layout -p nombre_salida -R Reference.fasta -Q Query.fasta --medium --png -t título_gráfico archive_salida.delta*

8.6 BLAST

BLAST (Basic Local Alignment Search Tool) herramienta de búsqueda basada en alineamiento local, es la herramienta más utilizada para calcular similitud entre secuencias homólogas. Mediante la búsqueda de similitudes entre secuencias se pueden inferir funciones de nuevos genes, predecir si pertenecen a familias génicas descritas y explorar relaciones evolutivas. Actualmente con la secuenciación de genomas completos, podemos utilizar esta herramienta también para predecir la localización de regiones codificantes de proteínas en el DNA genómico secuenciado (Altschul, 1990).

La posibilidad de utilizar BLAST de forma local, es decir, instalándolo en nuestro propio ordenador y descargando en éste las bases de datos del NCBI, nos permitió analizar más de 1 millón de secuencias en menos de dos meses. BLAST utiliza enormes bases de datos previamente indexadas por lo que no necesita gran cantidad de memoria RAM, pero el rendimiento mejora de forma casi lineal cuando aumentamos el número de CPUs utilizadas.

Tanto en la búsqueda de homólogos para las secuencias del gen 16S *rRNA* mediante BLASTn (para confirmar su identificación taxonómica con respecto a la obtenida por RDP y tratar de indentificar las especies de las que proceden) cómo para relacionar posibles ORFs (Open Reading Frames o marcos abiertos de lectura) con funciones ya descritas, mediante BLASTx, ajustamos el e-valor a un máximo de 10^{-10} . Este es un parámetro que describe el número de hits (aciertos) que uno puede esperar que aparezcan por azar cuando se realiza la búsqueda en una base de datos de un tamaño concreto. No pudimos reducir más el valor máximo de E-value ya que este parámetro tiene en cuenta también el tamaño del alineamiento entre la secuencia problema y su homóloga en la base de datos y nosotros hemos trabajado con secuencias cortas.

Sólo el formato cero '0' de BLAST local nos permite conocer la descripción de las secuencias de la base de datos con las que tienen homología cada una de nuestras secuencias problema (query). Pero, este formato es difícil de utilizar para obtener tablas de resultados,

para ello, es necesario cambiar el parámetro *outfmt* (formato de salida de BLAST) a 6 ó 7. Ambos formatos son tabulares, pero el 7 además, nos devuelve la cabecera con la información a la que corresponde cada una de las columnas. Actualmente BLAST no muestra la descripción de las secuencias de la base de datos en formato tabular, solo muestra el GI (Gene Identifier) y el Acc (Accession Number), por lo que fue necesario diseñar un Script que extrajera las descripciones de cada uno de ellos, a partir de su GI. Este script crea un listado con todos los GIs de los hits resultantes de BLAST(n/x) y mediante el comando de BLAST *blastdbcmd* extrae de la base de datos indexada la descripción correspondiente a cada GI (gracias a la opción *-outfmt %t'*). Finalmente el script crea una columna más en el resultado tabular de BLAST (6 ó 7) y añade dicha descripción.

La versión de dicho programa utilizada en nuestros trabajos fue BLAST+ 2.2.27 (Camacho *et al.* 2009), obtenida desde <ftp://ftp.ncbi.nih.gov/blast/executables/blast+/>.

8.7 STAMP

STAMP (Statistical Analysis of Metagenomic Profiles, <http://kiwi.cs.dal.ca/Software/STAMP>) es un paquete estadístico para analizar perfiles metagenómicos, tanto perfiles taxonómicos como perfiles funcionales. Este programa está prediseñado para funcionar en Microsoft Windows, pero puede instalarse también en Linux y Macintosh. Para llevar a cabo el estudio estadístico de los perfiles taxonómicos encontrados en nuestras muestras se utilizaron los parámetros aconsejados por Parks y Beiko (2010):

- Comparación de dos muestras sin réplicas evaluada mediante el test exacto de Fisher con un intervalo de confianza del 95 % evaluado por el método de diferencia de proporciones de Newcombe-Wilson, teniendo en cuenta ambas colas. La corrección de los p-valores obtenidos al hacer test múltiples fue la aplicada por Storey o la de Benjamini-Hoechberg cuando la primera no era aplicable (en aquellos casos en los que los p-valores no estaban distribuidos uniformemente).
- Comparación de múltiples grupos (múltiples muestras con réplicas) evaluada mediante ANOVA con un intervalo de confianza del 95 % evaluado por el método de Eta al cuadrado, para corregir el efecto del tamaño de muestra. Los grupos con diferencias significativas fueron *a posteriori* evaluados dos a dos por el test de Tukey-Kramer con un intervalo de confianza del 95 %. La corrección de los p-valores

obtenidos al hacer test múltiples fue la aplicada por Storey o la de Benjamini-Hochberg cuando la primera no era aplicable.

8.8 Ginkgo

Ginkgo (<http://biodiver.bio.ub.es/veganaweb/main/?section=../ginkgo/content.jsp>) es una herramienta de análisis multivariante (De Cáceres *et al.* 2007), solo para Microsoft Windows, con la que podemos confirmar dendrogramas de Agrupamiento Jerárquico Aglomerativo de distintos tipos (Simple, Completo, UPGMA, weighted) o hacerlos cuando la herramienta **Complete Linkage Clustering** del paquete RDP no lo permite por limitación en el número de secuencias analizadas. También se ha utilizado esta herramienta para realizar Análisis de Correspondencias (CA), resultando en un biplot en el que se sobreponen un Análisis de Coordenadas Principales (PCoA) junto con los vectores de todos los taxones que generan la ordenación obtenida, pudiendo filtrar aquellos taxones que más contribuyen a la ordenación de las muestras evaluadas en el PCoA.

8.9 CANOCO

CANOCO™ for Windows v4.5 es un completo programa de análisis estadístico multivariante (Ter Braak, 2002). Con este programa se han realizado análisis canónicos, que consiguen la ordenación de los datos incluyendo regresión. De este modo, permiten relacionar la estructura de la comunidad microbiana (medida como abundancia relativa de cada familia) con parámetros ambientales, como son las características físico-químicas del suelo del que se obtuvo el DNA de dicha comunidad. Para ello, se ha llevado a cabo el Análisis de Correspondencias Canónicas (CCA) que es la forma canónica del CA y es adecuado tanto para datos cuantitativos como para datos cualitativos que presenten una relación unimodal (Legendre, 1998). Antes de realizar estos análisis algunos de los parámetros físico-químicos medidos fueron transformados. Así, los valores de concentración en partes por millón (ppm) de los minerales medidos fueron transformados aplicando el logaritmo neperiano (Ln), como aconseja Clarke (1993). La textura del suelo se obtuvo de la fórmula propuesta por Crowther

$$(2014): \textit{Textura} = \frac{\% \textit{arena}}{\% \textit{arcilla} + \% \textit{limo}}$$

8.10 Linux

Actualmente la mayoría de herramientas bioinformáticas están disponibles de forma remota mediante páginas Web o de forma descargable, local, para los 3 sistemas operativos principales, Microsoft Windows, Apple Macintosh y Linux. Sólo los sistemas operativos basados en UNIX, Macintosh y Linux, son de Código Abierto (Open Source) y por tanto, permiten una mejor adecuación y optimización de dichas herramientas. Para poder utilizar nuestras herramientas bioinformáticas al máximo rendimiento permitido por el Hardware del Ordenador que hemos utilizado, nos decantamos por el sistema operativo basado en Linux, Ubuntu 10.04 LTS. Macintosh, al igual que muchas distribuciones de Linux como, Centos, Debian, etc. son de código abierto pero son de pago. Ubuntu es una distribución de código abierto, acogida a la licencia de software libre GNU GPL y gratuita cada vez más utilizada por la comunidad científica (www.ubuntu.com).

A través de la WIKIPEDIA y de libros generales de informática (Guillén 2007, Coullingbourne 2009) podemos conocer los comandos más frecuentes para utilizar Ubuntu, por lo que aquí se exponen solo aquellos comandos más utilizados para facilitar el análisis de los datos obtenidos.

- **top** y **htop**: son 2 comandos que nos permiten conocer el uso de nuestros procesadores y nuestra RAM, así como, conocer los procesos (programas) que están utilizando dichos recursos.
- **less archivo**: comando que nos permite ver el contenido de un archivo de texto directamente en el terminal. Sólo carga el contenido del archivo que cabe en el tamaño de ventana de nuestro terminal, por lo que podemos abrir un archivo de varios Gigabytes en segundos. Al ser un lector de textos, no nos permite editar el archivo, pero introduciendo el símbolo, barra diagonal '/', dentro del programa podemos buscar cualquier patrón en el texto. De este modo podemos buscar la secuencia de nombre '*Bradyrhizobium*' en un archivo fasta del siguiente modo: `/>Bradyrhizobium` (presionar enter). Para salir del visor de textos *less* presionamos 'q'.
- **tail -n X archivo**: si queremos conocer la información de las últimas X líneas en un archivo de texto.
- **grep -c ">" archivo.fasta**: gracias a este comando podemos conocer el número de secuencias que contiene un archivo en formato fasta. Cada secuencia en el archivo viene encabezada por una línea que comienza por el símbolo mayor que '>' seguida

del nombre de dicha secuencia y opcionalmente lleva información adicional. Este comando detecta el patrón entre comillas en el archivo indicado y con la opción `-c` cuenta las líneas en las que ha aparecido dicho patrón.

- **ssh -Y usuario@ip:** gracias a este comando podemos acceder de forma remota al ordenador (servidor) en el que queremos realizar un determinado análisis. Siempre que el servidor utilice cualquier distribución Linux como sistema operativo. Es necesario conocer el nombre de usuario, la IP del servidor al que accedemos y la contraseña de dicho usuario.
- **VNC: Escritorio Remoto:** para poder visualizar el Escritorio del servidor en nuestro monitor y no solo una Terminal, como en el caso de SSH, podemos utilizar VNC. A través de ssh accedemos al servidor y una vez dentro ejecutamos `vncserver :1 -geometry 1280x768` (de este modo activamos el Escritorio remoto número 1 con una resolución de pantalla de 1280x768 píxeles por pulgada). Volviendo al ordenador en el que estamos trabajando, mediante `exit`, ejecutamos `vncviewer ip:1`. De este modo se abrirá una ventana, con la resolución indicada anteriormente, mostrando el escritorio del servidor.
- **screen:** con este comando podemos crear un terminal virtual que no se cerrará ni detendrá los procesos que se lleven a cabo en él, aunque cerremos el terminal real. De este modo, podemos conectar a un servidor remoto mediante SSH y crear un terminal virtual en dicho servidor. El proceso que ejecutemos en el servidor no se detendrá hasta terminar, incluso si salimos del terminal virtual `Ctrl+A+D` y cerramos la conexión SSH.
- **time:** comando utilizado delante del programa a utilizar para conocer cuánto tiempo necesita dicho programa en realizar el análisis. El tiempo se muestra en segundos y al finalizar el proceso. Para conocer cuánto tarda un BLASTn lo ejecutamos de este modo: `time blastn -query secuencias.fasta -db nt -outfmt 6 -out resultado.txt`.
- **&:** el símbolo Ampersand sirve para liberar el terminal en el que estamos trabajando y dejar el análisis que queremos hacer en segundo plano. De este modo, podemos ejecutar el BLASTn anterior y seguir trabajando en el mismo terminal así: `time blastn -query secuencias.fasta -db nt -outfmt 6 -out resultado.txt &`.
- **Expresiones regulares frecuentes:** cuando queremos buscar o buscar y reemplazar en un archivo de texto los espacios, es suficiente con presionar la barra espaciadora, dentro de la función buscar (`Ctrl+F` en editor de textos Gedit, después de `'/'` en visor de textos less), aunque la expresión regular es `\s`. Pero si queremos buscar y/o reemplazar

un tabulador o un salto de carro (enter), las expresiones regulares que debemos utilizar son `\t` para el tabulador y `\n` para el salto de carro. Para buscar un punto `.` en el visor less debemos poner antes del punto, el símbolo barra inversa `\`, ya que el punto es tomado automáticamente por el buscador del programa como la expresión regular con significado cualquier carácter.

Resultados

Capítulo 1: Análisis global de los suelos rizosféricos de quercíneas y de las comunidades microbianas que habitan en ellos

Como fase previa al análisis en profundidad de las comunidades microbianas que habitan en los suelos, se realizaron distintos estudios de los sitios de muestreo (Tabla M-1). Para ello se llevó a cabo un análisis global de las características físico-químicas de dichos suelos, así como un estudio de la vegetación acompañante de los sitios de muestreo y los perfiles genéticos de las comunidades microbianas mediante TGGE. Además, en la zona de Cañar se registraron la temperatura del aire y la precipitación, así como la temperatura y humedad relativa de los suelos, para determinar cómo afectan estos parámetros del gradiente altitudinal a las comunidades microbianas de los suelos estudiados en la zona.

1 **Análisis global de las características físico-químicas de los suelos muestreados**

A partir de 2 kg de suelo de cada sitio muestreado en la primavera de 2008 (Tabla M-1) se determinaron las características físico-químicas de los mismos (Tabla R1-1). LJR y LJRE son dos tratamientos, no asociado y asociado a rizosfera de encina, respectivamente, que provienen de las mismas parcelas y comparten el mismo tipo de suelo, reflejado en la Tabla R1-1 como LJR. Estos suelos presentan texturas de tipo franco (CTR, CNA, CTE, LJN y LJQ) y de tipo franco-arenoso (CNB, CNX, CNR y LJR), todos ellos están clasificados como suelos de tipo phaeozem háplico con origen silíceo. El pH en los suelos de encinares sanos es ligeramente ácido, siendo el suelo del encinar no quemado de la ladera sur (LJN) el más ácido con un pH de 6,1, seguido del suelo del encinar de la ladera norte (CTE) con un pH de 6,6. Los sitios afectados por el incendio varían en su valor de pH dependiendo de si se trata del encinar (LJQ) o no (LJR), pues el primero es el suelo más alcalino (pH 7,6) y el segundo es ligeramente ácido (pH 6,5). El suelo del robledal de la ladera norte (CTR) es ligeramente ácido y con el mismo valor de pH que el encinar de esa zona (pH de 6,6). Por el contrario, el pH en los suelos del robledal de la ladera sur es neutro (pH de 7,2 en CNB y 7,1 en CNA), el suelo del robledal en expansión (CNX) y el suelo no asociado a rizosfera de roble (CNR) son ligeramente ácidos (6,6 en CNX y 6,2 en CNR) (Tabla R1-1). Los valores de agua útil son mayores en la ladera norte (20,01 % en CTE y 17,14 % en CTR) que en la ladera sur, entre 16,52 % en CNA, más arcilloso, y 10,39 % en CNX, más arenoso (Tabla R1-1). Un perfil parecido se observa en el contenido de materia orgánica de estos suelos, siendo los más ricos los de la ladera norte (9,24 % en CTR y 8,30 % en CTE); en la ladera sur vemos que los suelos de Cáñar tienen un bajo contenido de materia orgánica (de 6,07% en CNA a 3,44 % en CNR) y en Lanjarón vemos diferencias en el contenido de materia orgánica entre el encinar sano (7,61 % en LJN) y los sitios afectados por el incendio (4,54 % en LJN y LJR) (Tabla R1-1). Los valores de nitrógeno total (de 0,206 % en CNB a 0,427 % en CTE) y la proporción C/N (de 11,95 en LJN a 7,66 % en CNR) son característicos de suelos fértiles, excepto en CTR en el que los valores de nitrógeno total son más bajos (0,145 %) y la proporción C/N (36,62 %) es más alta, debido probablemente a la gran cantidad de raíces leñosas extraídas junto con el suelo de este sitio (Tabla R1-1). Los suelos que no pertenecen a bosques tienen una salinidad más baja (de 0,080 mmhos/cm³ en LJR a 0,046 mmhos/cm³ en CNR), mientras que en los suelos de bosque, el encinar quemado (LJQ) presenta el mayor nivel de salinidad con 0,220 mmhos/cm³ (Tabla R1-1). En los robledales el valor más bajo de potasio se observa en el sitio sin robles (CNR) con 90 ppm, el resto de sitios

tienen valores medios de potasio (de 205 ppm en CNX y CNB, a 365 ppm en CNA). Algo parecido se observa en el encinar, en el que el suelo no rizosférico (LJR) tiene el menor valor de potasio con 170 ppm y el resto de sitios tiene valores medios y altos de potasio (de 265 ppm en CTE a 445 ppm en LJN) (Tabla R1-1). La concentración de fósforo asimilable es muy variable entre los distintos suelos muestreados (de 4,0 ppm en CNB a 27,0 ppm en CNA) sin atender a una distribución por localización geográfica ni por especie de planta a la que están asociados (Tabla R1-1).

Tabla R1-1. Características físico-químicas de los suelos muestreados.

Parámetros	CTR	CNB	CNA	CNX	CNR	CTE	LJN	LJQ	LJR
Arcilla (%)	24,00	16,35	22,25	11,75	13,70	21,70	21,00	20,50	12,05
Arena (%)	40,65	63,50	44,25	62,44	54,53	43,60	45,74	49,54	56,08
Limo (%)	35,35	20,15	33,50	25,81	31,77	34,70	33,26	29,96	31,87
Tipo de suelo	Franco	Franco-arenoso	Franco	Franco-arenoso	Franco-arenoso	Franco	Franco	Franco	Franco-arenoso
pH (H2O)	6,6	7,2	7,1	6,6	6,2	6,6	6,1	7,6	6,5
pH (CIK)	5,9	6,4	6,4	5,8	5,4	6,0	5,7	7,0	6,0
Agua útil (%)	17,14	11,10	16,52	10,39	13,25	20,01	17,11	16,43	15,20
Salinidad (mmhos/cm ³)	0,140	0,120	0,160	0,070	0,046	0,110	0,140	0,220	0,080
Materia Orgánica Total (%)	9,24	3,88	6,07	3,82	3,44	8,30	7,61	4,54	4,54
N total (%)	0,145	0,206	0,301	0,279	0,258	0,427	0,366	0,233	0,250
Proporción C/N	36,62	10,82	11,59	7,87	7,66	11,17	11,95	11,19	10,44
Fósforo asimilable (ppm)	8,0	4,0	27,0	20,0	6,0	7,0	8,0	5,2	20,0
Potasio (ppm)	355	205	365	205	90	265	445	330	170
Sodio (ppm)	0,065	0,052	0,610	0,043	0,130	0,050	0,065	0,065	0,040
Magnesio (ppm)	3,050	1,933	2,508	1,283	0,623	2,470	4,363	7,203	1,480
Calcio (%)	15,030	8,748	12,939	6,942	4,350	15,400	12,650	17,972	8,650
Hierro (ppm)	74	28	51	37	29	59	122	42	40

2 Análisis de la vegetación acompañante de los sitios muestreados

Para alcanzar una mayor comprensión en el estudio de las comunidades microbianas asociadas a la rizosfera de quercíneas se ha llevado a cabo el análisis de la vegetación acompañante al roble melojo en cada uno de los sitios en la zona de Cáñar. De igual manera se analizó la vegetación acompañante a las encinas en sitios afectados y no afectados por el

Resultados Capítulo 1

incendio, zona de Lanjarón. Ambas zonas se sitúan en la ladera sur del espacio protegido de Sierra Nevada (Tabla M-1).

El estudio de la vegetación acompañante de los sitios muestreados en la zona de Cáñar (Tabla M-1), muestra diferencias significativas en la abundancia de las especies dominantes para los 4 sitios analizados (Tabla R1-2). De modo que la especie *Festuca elegans* (con el nombre vulgar de Cerrillo) solo aparece en las parcelas del robledal alto (CNA) con un p-valor = 0,0129, dicha especie es la gramínea vivaz (perenne) que caracteriza a los melojares descritos para la faciación típica (Lorite 2002, Valle et al. 2007) situados en el piso supramediterráneo, como ocurre en este sitio. *Festuca indigesta* (con el nombre vulgar de Rompebarrigas) es característica de los pastizales oromediterráneos y se encuentra situado entre los claros del piornal, en nuestro estudio sólo aparece en los sitios más elevados (CNR y CNX), aunque su abundancia fue significativamente mayor en el CNR con un p-valor = 0,0215 (Tabla R1-2). *Festuca scariosa* (Lastón) aparece en todos los sitios, lo que indica la mayor plasticidad de esta especie, aunque fue dominante en el robledal en expansión (CNX) con un p-valor = 0,0205. Por otro lado, el robledal bajo (CNB) se diferencia de los anteriores por el dominio de la leguminosa *Ulex parviflorus* (Aulaga) con un p-valor = 0,0129 (Tabla R1-2), que aparece como acompañante de las etapas de degradación de los melojares situados en cotas inferiores (faciación mesomediterránea). La leguminosa *Adenocarpus decorticans* (Rascaviejas) sólo se encontró en los dos sitios del robledal maduro (CNA y CNB), siendo más abundante en el alto que en el bajo, aunque sin diferencias significativas (Tabla R1-2). Dicha especie caracteriza la serie de vegetación de los melojares. Como puede verse en el dendrograma realizado a partir de la composición florística (presencia de especies) de cada muestra (Fig. R1-1), se distingue claramente (59,5 % de distancia) el robledal maduro (CNA y CNB) de la zona en expansión (CNX) y del suelo suelto (CNR). Los sitios más altos (CNX y CNR) son más parecidos entre sí respecto a su composición florística (una distancia máxima del 29,2 %) que los sitios del robledal maduro, CNA y CNB (agrupados al 35,7 % de distancia) (Fig. R1-1).

Tabla R1-2. Abundancia de especies vegetales acompañantes dominantes en los sitios muestreados en la zona de Cáñar. Letras diferentes en una misma fila indican diferencias estadísticamente significativas evaluadas mediante test no paramétrico de Kruskal-Wallis o ANOVA (test post-hoc de Nemenyi y Tukey respectivamente). \pm Error estándar.

Especie	CNA	CNB	CNR	CNX
<i>Festuca elegans</i>	2,33 \pm 0,3 a	0,0 \pm 0,0 b	0,0 \pm 0,0 b	0,0 \pm 0,0 b
<i>Festuca indigesta</i>	0,0 \pm 0,0 b	0,0 \pm 0,0 b	3,0 \pm 0,0 a	1,33 \pm 0,6 b
<i>Festuca scariosa</i>	0,67 \pm 0,3 b	0,33 \pm 0,3 b	0,67 \pm 0,3 b	2,67 \pm 0,3 a
<i>Ulex parviflorus</i>	0,0 \pm 0,0 b	2,34 \pm 0,3 a	0,0 \pm 0,0 b	0,0 \pm 0,0 b
<i>Adenocarpus decorticans</i>	0,67 \pm 0,3 b	0,33 \pm 0,0 b	0,0 \pm 0,0 b	0,0 \pm 0,0 b

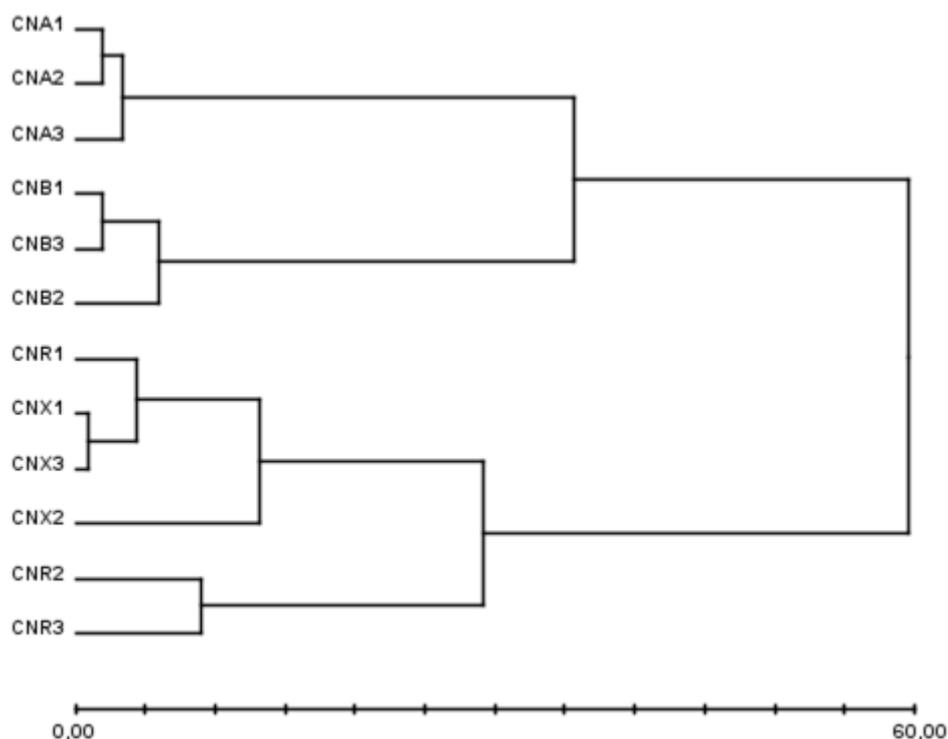


Figura R1-1. Agrupamiento jerárquico aglomerativo usando la matriz de distancias euclídeas al cuadrado y el algoritmo Ward teniendo en cuenta la abundancia de especies vegetales acompañantes del roble melojo de las 12 parcelas muestreadas.

El estudio de la vegetación acompañante de los sitios muestreados en los encinares de la zona de Lanjarón (Tabla M-1), muestra diferencias significativas en la abundancia de las especies dominantes para los 3 sitios analizados (Tabla R1-3). De modo que las especies *Adenocarpus decorticans* y *Vicia lathyroides* (Alverjana), ambas leguminosas, son significativamente más abundantes en el sitio sin encinas (LJR) que en el encinar quemado (LJQ) con p-valores de 0,0303 y 0,0233, respectivamente, quedando el encinar no quemado (LJN) en una abundancia intermedia sin significación estadística respecto a los otros dos sitios (Tabla R1-3). Las gramíneas anuales también son más abundantes en LJR. La primera de ellas, *Bromus sterilis* (Espigajo o Cebada bravía) sólo aparece en LJR y su abundancia es significativa, con un p-valor de 0,0129; la segunda, *Bromus tectorum* (Espiguilla colgante) solo se observó en los sitios afectados por el incendio (LJQ y LJR) aunque su abundancia resultó marginalmente significativa con un p-valor de 0,0611. Al contrario, las dos especies de gramíneas vivaces *Festuca scariosa* y *Dactylis glomerata* (Triguera fina o Dactilo) sólo se observaron en el encinar no quemado (LJN), aunque solo la primera especie muestra diferencias estadísticamente significativas respecto a su abundancia, con un p-valor de 0,0132 y la segunda resultó ser

Resultados Capítulo 1

marginalmente significativa con un p-valor de 0,0691 (Tabla R1-3). Como puede verse en el dendrograma realizado a partir de la composición florística de cada muestra (Fig. R1-2), se distinguen claramente (61,6 % de distancia) los sitios afectados por el incendio (LJQ y LJR) del encinar no quemado (LJN). Todos los sitios quemados están agrupados a una distancia del 32,0 %, quedando las 3 muestras del encinar quemado (LJQ) en el centro del grupo a una distancia del 17,3 % (Fig. R1-2).

Tabla R1-3. Abundancia de especies vegetales acompañantes dominantes en los sitios muestreados en la zona de Lanjarón. Letras diferentes en una misma fila indican diferencias estadísticamente significativas evaluadas mediante test no paramétrico de Kruskal-Wallis o ANOVA (test post-hoc de Nemenyi y Tukey respectivamente). * Diferencias marginalmente significativas. \pm Error estándar.

Especie	LJN	LJQ	LJR
<i>Adenocarpus decorticans</i>	0.67 \pm 0.33 ab	0 b	2.00 \pm 0.67 a
<i>Vicia lathyroides</i>	0.33 \pm 0.13 ab	0 b	2.00 \pm 0.0 a
<i>Festuca scariosa</i>	2.00 \pm 0.67 a	0 b	0 b
<i>Dactylis glomerata</i> *	1.33 \pm 0.67	0	0
<i>Bromus sterilis</i>	0 b	0 b	1.7 \pm 0.3a
<i>Bromus tectorum</i> *	0	1.7 \pm 0.9	1.7 \pm 0.9

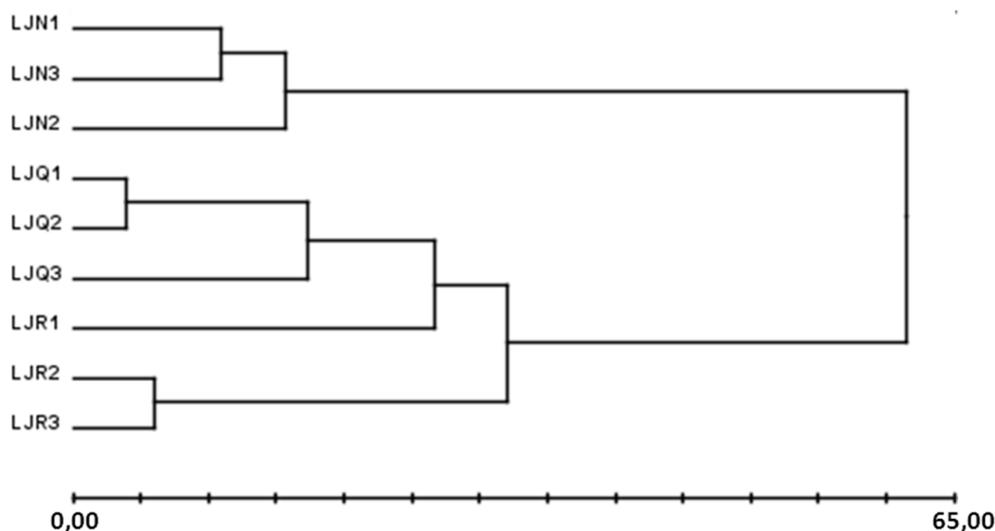


Figura R1-2. Agrupamiento jerárquico aglomerativo usando la matriz de distancias euclídeas al cuadrado y el algoritmo Ward teniendo en cuenta la abundancia de especies vegetales acompañantes de la encina de las 9 parcelas muestreadas.

3 Análisis global de las comunidades microbianas de los suelos muestreados

Se analizaron los perfiles genéticos, obtenidos mediante TGGE de la zona V6-V8 del gen *16S rRNA*, de las comunidades microbianas de los sitios muestreados en primavera de los años 2008 y 2011 (Tabla M-1).

En el análisis de las comunidades microbianas asociadas a roble y la muestra de suelo suelto de la zona de expansión del robledal (CNR), utilizando cebadores universales para el gen *16S rRNA*, se observa mayor relación entre todos los sitios muestreados en 2008 (Fig. R1-3a) que en 2011 (Fig. R1-3b). En el muestreo de 2008 se separan las muestras en dos grupos, el primero con las muestras de suelo suelto (CNR) más una muestra del robledal de Camarate (CTR3); el segundo grupo se subdivide en un grupo con una similitud superior al 88 % en el que aparecen muestras de robledal y fuera de dicho subgrupo se encuentran las muestras de pies de roble de la zona de expansión del robledal (CNX) (Fig. R1-3a). En el análisis de las muestras tomadas en 2011 (Fig. R1-3b) no se observa agrupación de las muestras por el sitio de origen. Analizando los perfiles generados por secuencias del gen *16S rRNA* pertenecientes al *phylum* Actinobacteria de las mismas muestras (Fig. R1-4), se observa que las muestras de CNR y CNX son las más alejadas del resto en el muestreo de 2008 (Fig. R1-4a). En el muestreo de 2011 se observan dos grupos (Fig. R1-4b), el primero con las muestras de expansión (CNX) y el robledal a mayor altitud (CNA) y el segundo con las muestras del robledal a menor altitud (CNB) junto con las muestras del robledal de la ladera norte (CTR). En este análisis también se observa que las comunidades microbianas de los distintos sitios estudiados son más parecidas entre sí en el muestreo de 2008 que en el de 2011 (Fig. R1-4).

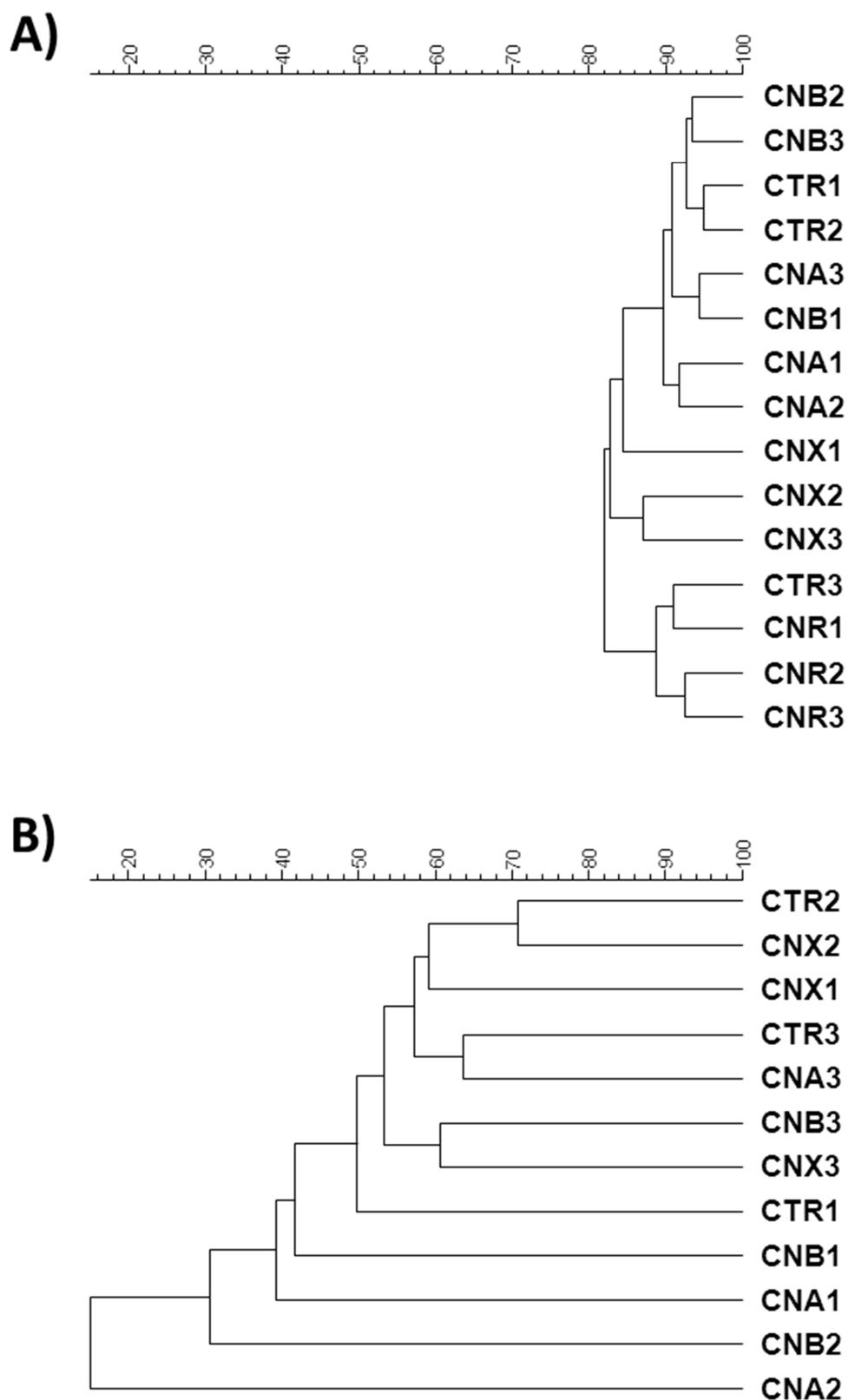


Figura R1-3. Agrupamiento jerárquico aglomerativo usando el coeficiente de correlación de Pearson y el análisis UPGMA de los perfiles de TGGE generados con cebadores universales para el gen 16S *rRNA*. Dendrograma de los sitios asociados a roble y suelo suelto (CNR) muestreados en 2008 (A) y de los sitios asociados a roble en 2011 (B). La escala numérica muestra la similitud entre los sitios de muestreo.

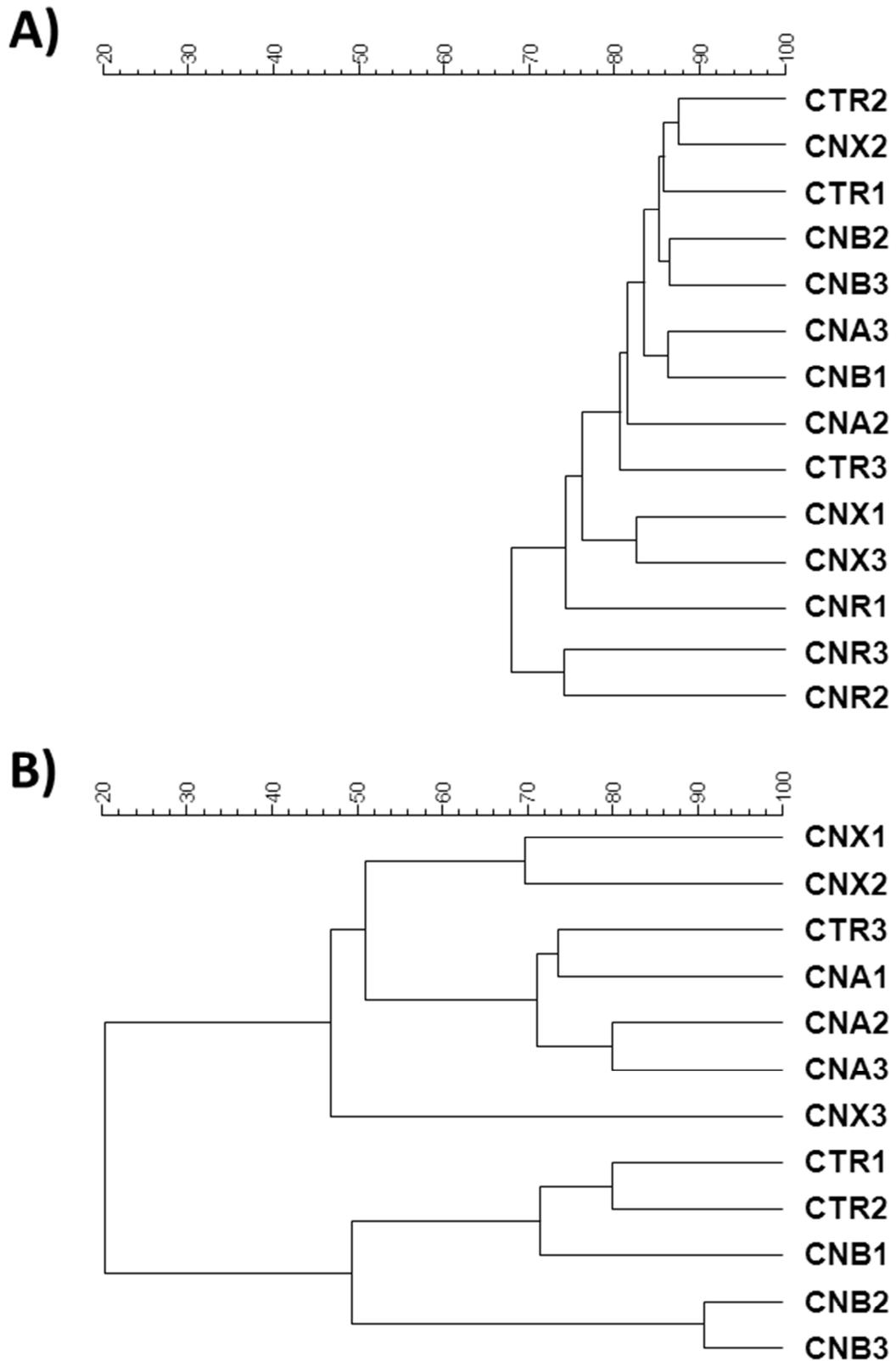


Figura R1-4. Agrupamiento jerárquico aglomerativo usando el coeficiente de correlación de Pearson y el análisis UPGMA de los perfiles de TGGE generados con cebadores específicos del *phylum* Actinobacteria para el gen 16S *rRNA*. Dendrograma de los sitios asociados a roble y suelo suelto (CNR) muestreados en 2008 (A) y de los sitios asociados a roble en 2011 (B). La escala numérica muestra la similitud entre los sitios de muestreo.

En cuanto al análisis de las comunidades microbianas asociadas a encina, y la muestra de suelo suelto de la zona afectada por el incendio (LJR), utilizando cebadores universales para el gen 16S *rRNA*, también se observa mayor relación entre todos los sitios muestreados en 2008 (Fig. R1-5a) que en 2011 (Fig. R1-5b). En el muestreo de 2008 se distinguen dos grupos, uno con las muestras del encinar quemado (LJQ) y otro con todas las demás (Fig. R1-5a). En el segundo grupo podemos distinguir un subgrupo con las muestras del encinar no quemado de la ladera sur (LJN) junto con dos muestras del encinar de la ladera norte (CTE) separado de otro subgrupo formado por 2 de las 3 muestras del suelo suelto afectado por el incendio (LJR) (Fig. R1-5a). En el muestreo de 2011 se observan 2 grupos (Fig. R1-5b). El primer grupo se subdivide en un grupo con las muestras del encinar no quemado de la ladera sur (LJN) y el resto de muestras pertenecen al encinar quemado (LJQ) y suelo suelto afectado por el incendio (LJR). El segundo grupo se compone por las muestras del encinar de la ladera norte (CTE) y las encinas no quemadas germinadas en el sitio afectado por el incendio (LJRE) (Fig. R1-5b). Analizando los perfiles generados por secuencias del gen 16S *rRNA* pertenecientes al *phylum* Actinobacteria de las mismas muestras (Fig. R1-6), observamos que se separan los sitios no quemados (LJN y CTE) de los sitios quemados (LJQ, LJR y LJRE). El patrón obtenido con cebadores de actinobacterias de la rizosfera de las encinas no quemadas germinadas en el sitio afectado por el incendio (LJRE) muestreadas en 2011, tienen un perfil genético más parecido al procedente de la rizosfera de las encinas quemadas (LJQ) y del suelo suelto afectado por el mismo incendio (LJR) (Fig. R1-6b).

Los análisis de los perfiles genéticos generados con la clase α -Proteobacteria y los *phyla* Firmicutes y Bacteroidetes no mostraron agrupación de las muestras por el sitio de origen. Siendo el *phylum* Actinobacteria el único capaz de mostrar diferencias entre las distintas muestras (Fig. R1-4 y R1-6), con un papel más relevante en la distinción entre sitios afectados y no afectados por un incendio (Fig. R1-6), incluso 6 años después del mismo (Fig. R1-6b).

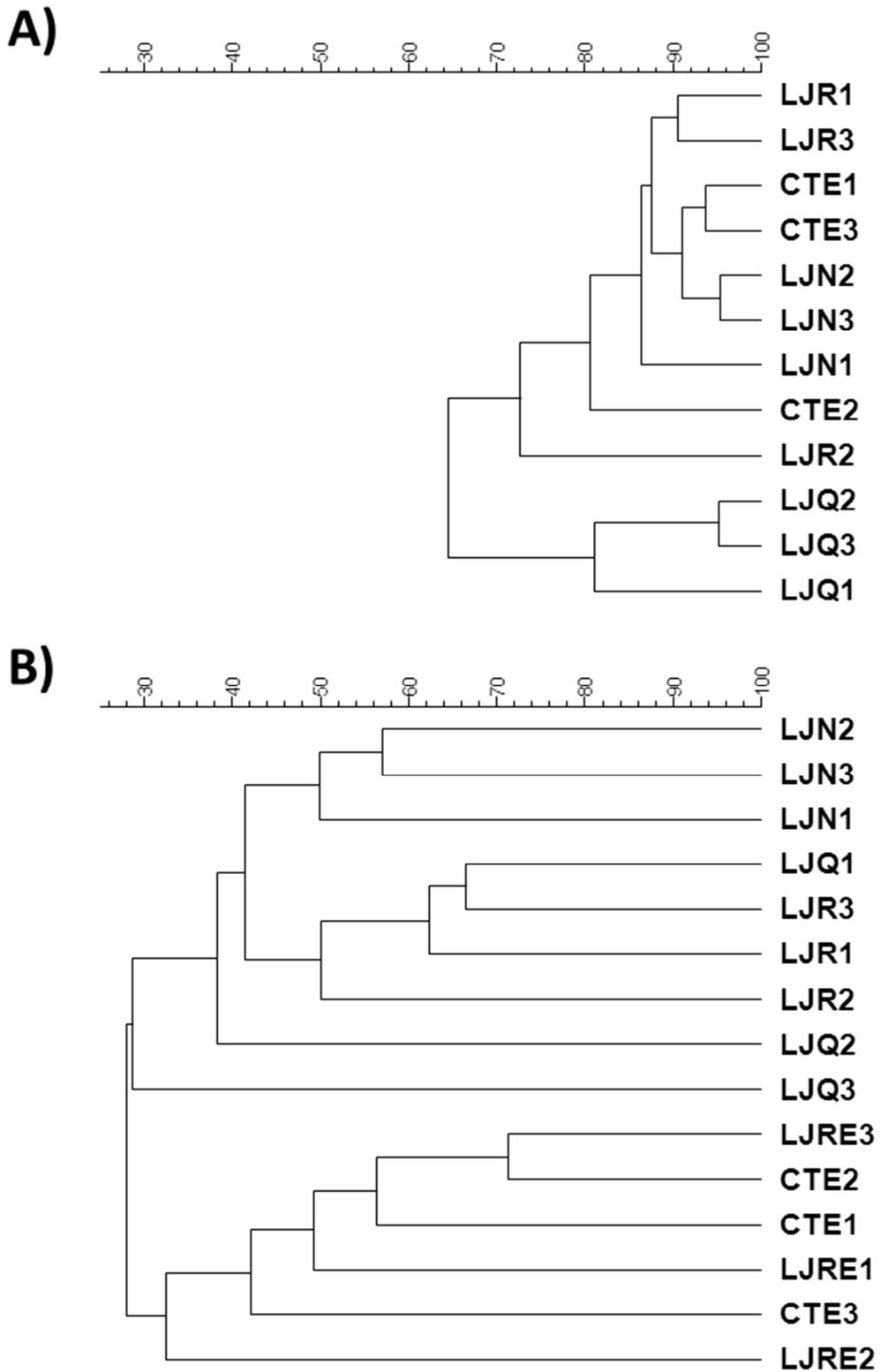


Figura R1-5. Agrupamiento jerárquico aglomerativo usando el coeficiente de correlación de Pearson y el análisis UPGMA de los perfiles de TGGE generados con cebadores universales para el gen 16S *rRNA*. Dendrograma de los sitios asociados a encina y suelo suelto (LJR) muestreados en 2008 (A) y en 2011 (B). La escala numérica representa la similitud entre los sitios de muestreo.

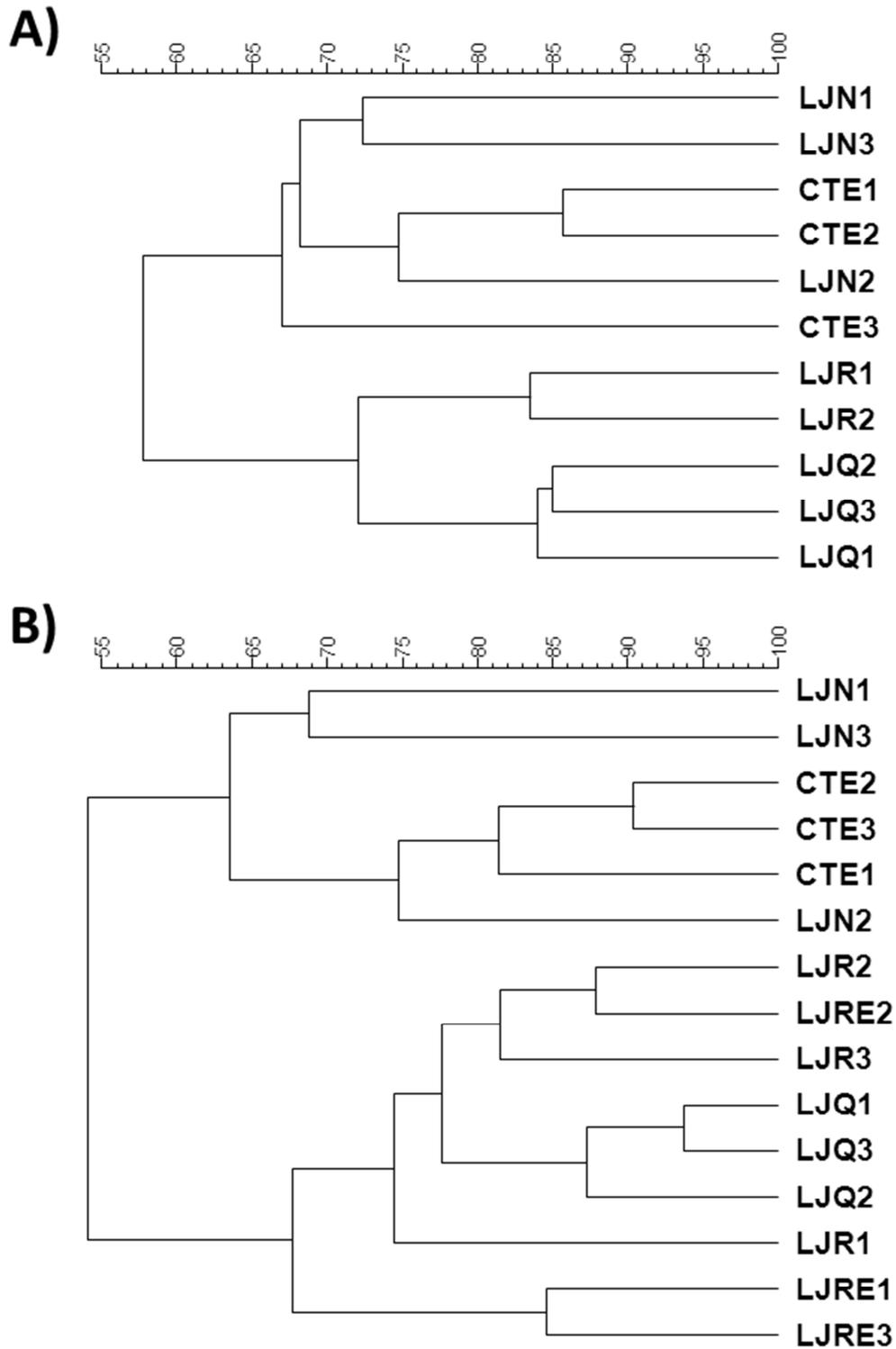


Figura R1-6. Agrupamiento jerárquico aglomerativo usando el coeficiente de correlación de Pearson y el análisis UPGMA de los perfiles de TGGE generados con cebadores específicos del *phylum* Actinobacteria para el gen 16S *rRNA*. Dendrograma de los sitios asociados a encima y suelo suelto (LJR) muestreados en 2008 (A) y en 2011 (B). La escala numérica representa la similitud entre los sitios de muestreo.

4 Análisis de la Temperatura y Humedad de los sitios muestreados en la zona de Cañar

Tras analizar los datos obtenidos del registro diario de temperatura del aire de la zona de Cañar los 6 meses previos a cada muestreo (2008 y 2011) y a una altitud de 1.735 m, observamos que no hay diferencias significativas entre ellos; siendo la temperatura media en el mes de muestreo (abril) de $7,9\text{ }^{\circ}\text{C} \pm 3,9\text{ }^{\circ}\text{C}$ en 2008 y de $10,6\text{ }^{\circ}\text{C} \pm 4,3\text{ }^{\circ}\text{C}$ en 2011 (Tabla R1-4). La temperatura del suelo se registró diariamente a 3 altitudes distintas: el sitio más bajo (CNB) donde hay un robledal bien formado a 1.482 m; sitio de altitud intermedia, que es el más alto (CNA) donde aparece el robledal a 1.823 m; y los dos sitios de mayor altitud, uno tomado como expansión del robledal (CNX) y el otro como suelo suelto (CNR) a 1.945 m, muestreados en la primavera del año 2011 (Tabla R1-4). Los datos obtenidos del registro de la temperatura del suelo, de los 6 meses previos al muestreo, muestran un gradiente altitudinal, registrando los valores más altos en el sitio de muestreo más bajo (CNB) y los valores más bajos en los suelos de muestreo a mayor altitud (CNX y CNR) (Tabla R1-4 y Fig. R1-7); sin embargo, no hay diferencias significativas en la temperatura media de cada mes entre los sitios muestreados. Siendo la temperatura media del mes de abril de $11,6\text{ }^{\circ}\text{C} \pm 1,5\text{ }^{\circ}\text{C}$ en CNB, de $10,8\text{ }^{\circ}\text{C} \pm 1,6\text{ }^{\circ}\text{C}$ en CNA y de $9,1\text{ }^{\circ}\text{C} \pm 2,0\text{ }^{\circ}\text{C}$ en CNX y CNR (Tabla R1-4).

Tabla R1-4. Temperatura media mensual de los 6 meses previos al muestreo. ^a Temperatura registrada en aire. ^s Temperatura registrada en suelo. * Meses del año anterior al de muestreo. \pm Desviación estándar.

	Temperatura media ($^{\circ}\text{C}$) ^a		Temperatura media ($^{\circ}\text{C}$) ^s		
	2008	2011	2011		
	Cañar		CNB	CNA	CNX/R
Noviembre*	-	$6,3 \pm 2,7$	$8,6 \pm 1,4$	$7,8 \pm 1,3$	$6,2 \pm 1,6$
Diciembre*	-	$5,3 \pm 4,4$	$6,7 \pm 1,9$	$5,4 \pm 1,8$	$3,7 \pm 1,8$
Enero	$5,4 \pm 3,0$	$3,3 \pm 3,6$	$5,6 \pm 1,1$	$4,4 \pm 1,2$	$2,9 \pm 1,1$
Febrero	$5,3 \pm 1,5$	$4,3 \pm 3,2$	$5,2 \pm 1,5$	$4,2 \pm 1,6$	$2,4 \pm 1,4$
Marzo	$7,2 \pm 4,3$	$4,3 \pm 3,5$	$6,9 \pm 1,8$	$5,7 \pm 1,7$	$4,0 \pm 1,9$
Abril	$7,9 \pm 3,9$	$10,6 \pm 4,3$	$11,6 \pm 1,5$	$10,8 \pm 1,6$	$9,1 \pm 2,0$

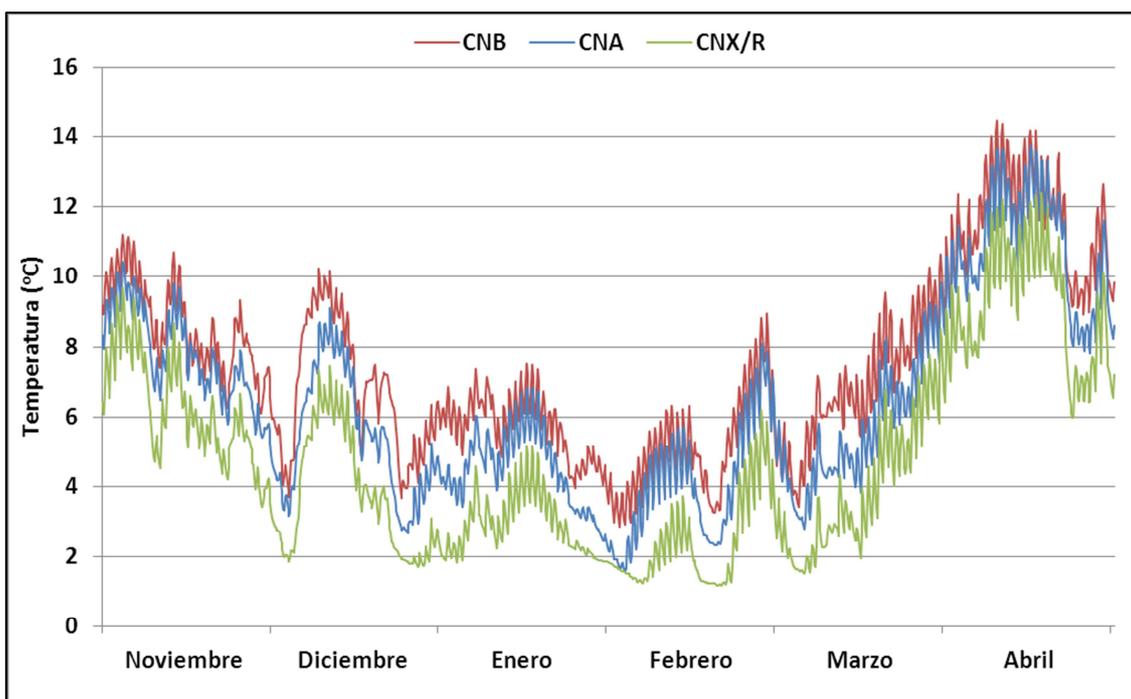


Figura R1-7. Temperatura del suelo a 20 cm de profundidad registrada cada 4 horas en los 6 meses previos al muestreo del año 2011 en los sitios estudiados a 3 altitudes (CNB, CNA y CNX/CNR) en la zona de Cáñar.

Tras analizar los datos obtenidos del registro diario de precipitación en la zona de Cáñar, los 6 meses previos a cada muestreo (2008 y 2011), a una altitud de 1.735 m, observamos diferencias significativas entre ellos; siendo la precipitación acumulada durante los 6 meses previos al muestreo de 538,8 mm en 2008 y de 1.099,6 mm en 2011 (Tabla R1-5). Las diferencias más notables se observaron en los meses de noviembre (165,7 mm en 2007 y 273,7 mm en 2010), diciembre (120,3 mm en 2007 y 547,7 mm en 2010) y marzo (14,0 mm en 2008 y 81,9 mm en 2011) (Tabla R1-5). La humedad relativa del suelo se registró diariamente a 3 altitudes distintas: el sitio más bajo (CNB) a 1.482 m; sitio de altitud intermedia (CNA) a 1.823 m; y los dos sitios de mayor altitud (CNX y CNR) a 1.945 m, muestreados en la primavera del año 2011 (Tabla R1-5). Los datos obtenidos del registro de la humedad relativa del suelo, de los 6 meses previos al muestreo, muestran un gradiente altitudinal, registrando los valores más bajos en el sitio de muestreo más bajo (CNB) y los valores más altos en los suelos de muestreo a mayor altitud (CNX y CNR) y a altitud intermedia (CNA) (Tabla R1-5 y Fig. R1-8); sin embargo, no hay diferencias significativas en la humedad relativa media de cada mes entre los sitios muestreados. Siendo la humedad relativa media del mes de abril de $192,3 \text{ L/m}^3 \pm 51,9 \text{ L/m}^3$ en CNB, de $207,8 \text{ L/m}^3 \pm 25,9 \text{ L/m}^3$ en CNA y de $209,0 \text{ L/m}^3 \pm 46,3 \text{ L/m}^3$ en CNX y CNR (Tabla R1-5).

Tabla R1-5. Precipitación acumulada y Humedad relativa media mensual de los 6 meses previos al muestreo. ^a Precipitación acumulada medida con sonda de aire. ^s Humedad relativa registrada en suelo. * Meses del año anterior al de muestreo. ± Desviación estándar.

	Precipitación acumulada (mm) ^a		Humedad Relativa media (L/m ³) ^s		
	2008	2011	2011		
	Cañar		CNB	CNA	CNX/R
Noviembre*	165,7	273,7	170,8 ± 68,6	220,5 ± 40,1	216,9 ± 36,3
Diciembre*	120,3	547,7	205,3 ± 45,0	253,5 ± 37,7	234,1 ± 41,0
Enero	68,9	49,3	202,4 ± 42,3	246,0 ± 36,8	222,3 ± 36,9
Febrero	56,2	45,8	205,0 ± 41,3	253,4 ± 36,0	229,6 ± 39,6
Marzo	14,0	81,9	205,5 ± 44,2	257,5 ± 35,6	227,0 ± 39,4
Abril	113,6	101,2	192,3 ± 51,6	207,8 ± 25,9	209,0 ± 46,3
Total	538,8	1099,6			

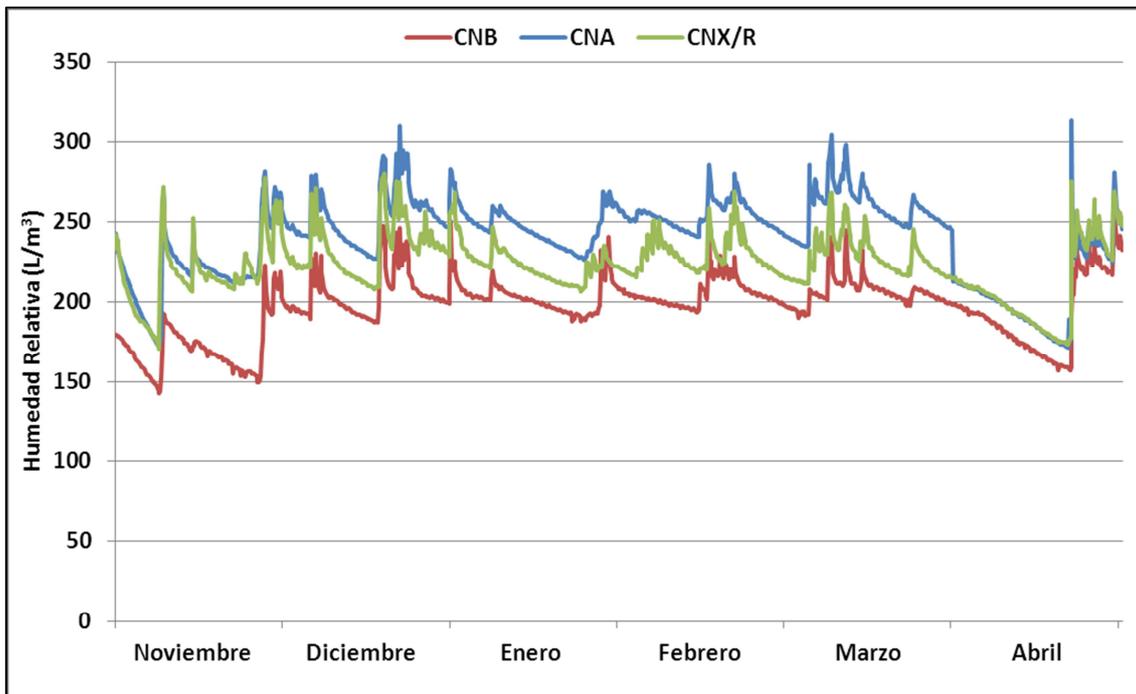


Figura R1-7. Humedad relativa del suelo registrada cada 4 horas en los 6 meses previos al muestreo de 2011 en los sitios estudiados a 3 altitudes (CNB, CNA y CNX y CNR) en la zona de Cañar.

Capítulo 2: Efectos del gradiente altitudinal en la estructura y composición de las comunidades microbianas asociadas a la rizosfera de roble melojo (*Q. pyrenaica* Willd.)

Se determinó la influencia del gradiente altitudinal en la estructura y composición de las comunidades microbianas analizando el DNA medioambiental extraído de la rizosfera de los robles a 3 altitudes distintas: el sitio más bajo (CNB) donde hay un robledal bien formado a 1.482 m sobre el nivel del mar; sitio de altitud intermedia, que es el más alto (CNA) donde aparece el robledal a 1.823 m; y el sitio de mayor altitud, tomado como expansión del robledal (CNX) a 1.945 m, muestreados en la primavera de los años 2008 y 2011. En la zona de mayor altitud se muestreó también el suelo suelto (CNR) en la primavera del año 2008.

1 Clasificación taxonómica de las secuencias del gen 16S *rRNA*

En el estudio clásico de librerías génicas (genotecas), se analizaron solo los sitios bajo (CNB) y alto (CNA) muestreados en el año 2008. De ellos se obtuvieron al menos 75 clones por sitio con un amplicón casi completo del gen 16S *rRNA* (≈ 1460 bp), de dichos clones se consiguieron un total de 122 secuencias de buena calidad y no quiméricas, 61 secuencias en cada sitio. Tras analizar nuestras secuencias frente a la base de datos de Bacteria y Archaea del RDP (16S *rRNA* training set 9) vemos que, a nivel de *phylum*, conseguimos clasificar el 91,8 % de las secuencias en ambos sitios, siendo el *phylum* Proeobacteria el mayoritario en ambos sitios con una abundancia relativa del 36,1 % en CNB y un 27,9 % en CNA. El segundo *phylum* con mayor abundancia en CNA es Actinobacteria con un 19,7 %, siendo los demás *phyla* menores al 10 % para este sitio. En CNB el segundo *phylum* es Bacteroidetes con un 16,4 %, el tercero es Acidobacteria con un 11,5 %, estando el *phylum* Actinobacteria en cuarta posición con un 9,8 % seguido del resto de *phyla* también menores al 10 % (Fig. R2-1). A nivel de familia conseguimos clasificar el 52,5 % de las secuencias en CNB y el 42,6 % en CNA, siendo la más abundante la familia Chitinophagaceae, perteneciente al *phylum* Bacteroidetes, que supone el 13,1 % en CNB y el 4,9 % en CNA. Tanto la familia Plantomycetaceae como la familia Hyphomicrobiaceae, pertenecientes a los *phyla* Planctomyces y Proteobacteria respectivamente, son las segundas más abundantes en el caso de CNB con una abundancia relativa del 4,9 %. En cuanto a CNA, la familia Planctomycetaceae es la segunda más abundante con un 3,3 %, pero no se obtuvieron secuencias de este sitio que se clasificaran con las familias Hyphomicrobiaceae y Sphingomonadaceae, ambas de la clase α -Proteobacteria (Fig. R2-2). Aunque se observan diferencias entre ambos sitios, éstas no son estadísticamente significativas, según la herramienta Libcompare, debido al pequeño tamaño de muestra analizado.

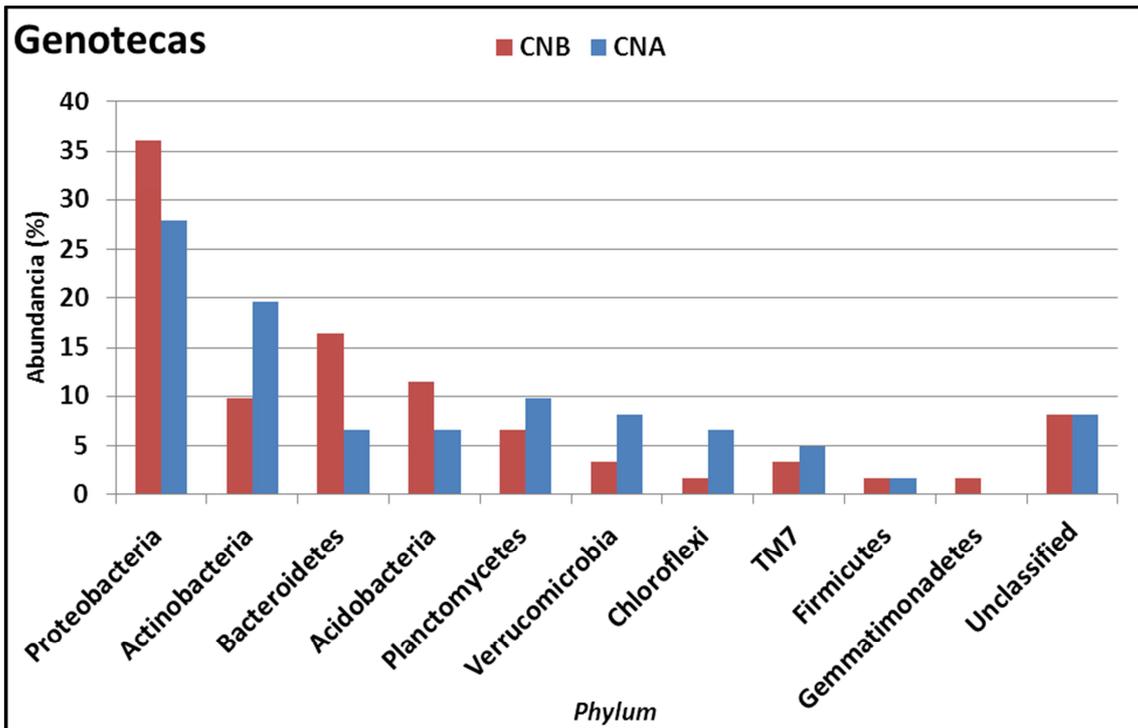


Figura R2-1. Distribución de todos los *phyla* de las genotecas clásicas de 16S *rRNA*.

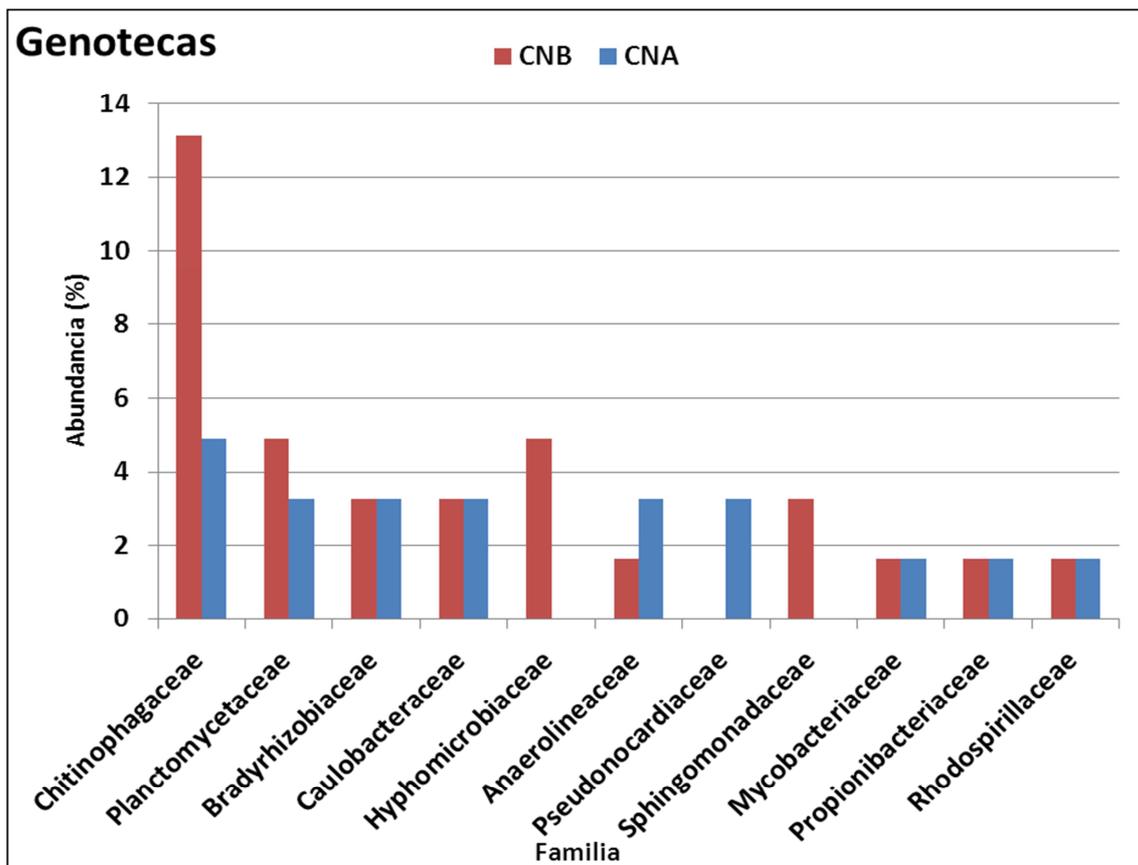


Figura R2-2. Distribución de las familias más abundantes de las genotecas clásicas de 16S *rRNA*. Solo se muestran aquellas familias que suponen al menos el 3,3 % de abundancia relativa en la suma de ambos sitios.

En el estudio de los metagenomas mediante pirosecuenciación se analizaron solo los sitios bajo y alto muestreados en 2008 al igual que en las genotecas clásicas. De ellos se obtuvieron 247.846 lecturas en CNB y 193.004 lecturas en CNA que tras filtrarlas, resultaron en 222.689 secuencias con longitud media de 371 bp y 167.219 secuencias con longitud media de 388 bp, respectivamente, únicas y de calidad. Dichas secuencias se han anotado con la base de datos M5NR de MG-RAST dando como resultado 420 secuencias en CNB y 277 secuencias en CNA homólogas a genes que transcriben algún ARN ribosómico (*rRNA*). Al enfrentar estas secuencias ribosómicas a la base de datos del gen 16S *rRNA* del RDP se clasificaron 66 y 45 secuencias, respectivamente (Tabla R2-1).

Al analizar la clasificación de las secuencias homólogas al gen 16S *rRNA* de CNB y CNA procedentes de pirosecuenciación vemos que, a nivel de familia, la distribución de las familias es muy diferente entre ambos, conseguimos clasificar el 78,8 % de las secuencias en CNB y el 60,0 % en CNA, siendo la familia Sphingomonadaceae la mayoritaria por suponer el 11,3 % en suma de ambos sitios, 9,1 % en CNB pero solo el 2,2 % en CNA. Sin embargo, la segunda familia más abundante procedente del *phylum* Bacteroidetes, Flavobacteriaceae que corresponde a un 8,2 % en suma, supone un 1,5 % en CNB y un 6,7 % en CNA. La tercera y cuarta familias más abundantes son Lachnospiraceae del *phylum* Firmicutes y Planctomycetaceae que suponen ambas un 3 % en CNB y un 4,4 % en CNA. En este estudio solo se encontraron secuencias de la familia Hyphomicrobiaceae en CNA con un 4,4 % de abundancia (Fig. R2-3), al contrario que en el estudio de genotecas clásicas en el que no se obtuvieron secuencias de dicha familia para CNA pero si para CNB suponiendo el 4,9 % de abundancia relativa para dicho sitio. Las familias Nitrosomonadaceae de la clase β -Proteobacteria con un 4,5 %, Microbacteriaceae y Micromonosporaceae ambas del *phylum* Actinobacteria, con un 3 % cada una, solo se observan en CNB (Fig. R2-3). A pesar de encontrar diferencias entre ambos sitios, éstas no son estadísticamente significativas, según las herramientas Libcompare y STAMP, debido al tamaño de muestra analizado.

Tabla R2-1. Recuento de secuencias antes y después del procesamiento de los sitios estudiados con secuenciación masiva "shotgun".

Muestra	Método de secuenciación	Lecturas	Secuencias	rRNA	16S RDP
CNB	454	247.846	222.689	420	66
CNA	454	193.004	167.219	277	45
CNB	Illumina	356.963.222	173.277.220	26.187.772	71.178
CNA	Illumina	365.590.198	171.236.955	26.947.483	65.286
CNX	Illumina	391.923.886	192.242.088	29.770.058	135.237

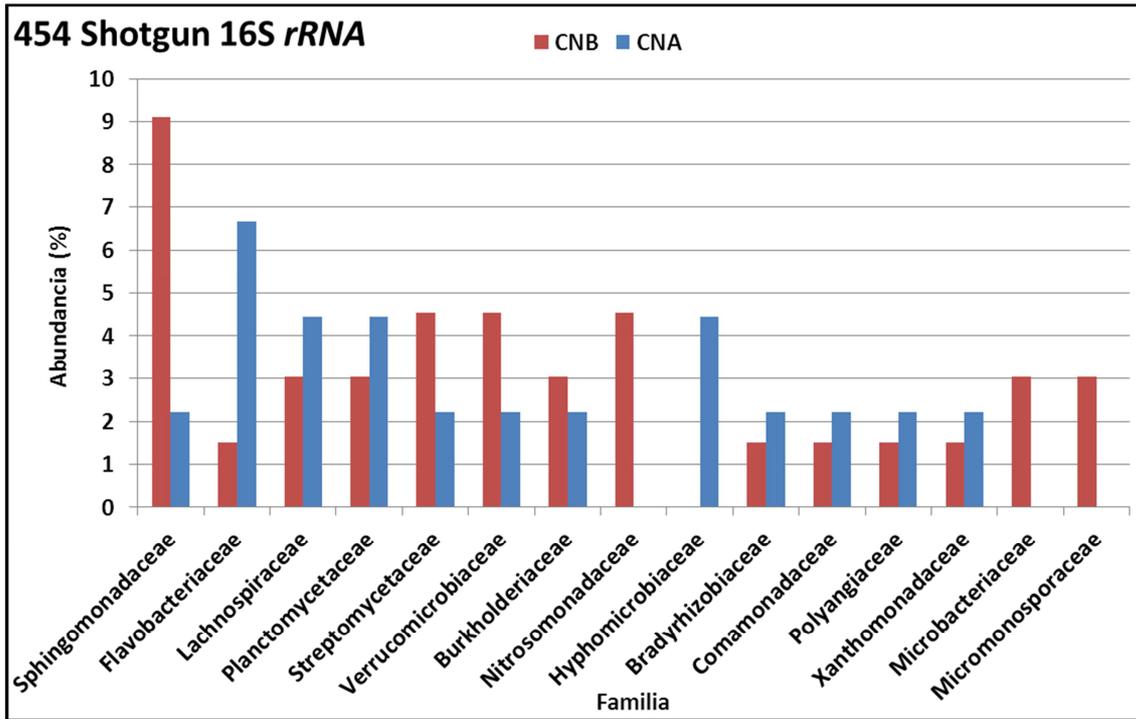


Figura R2-3. Distribución de las familias más abundantes obtenidas mediante pirosecuenciación. Solo se muestran aquellas familias que suponen al menos el 3 % de abundancia relativa en la suma de ambos sitios.

En el estudio de los metagenomas, mediante secuenciación Illumina, se analizaron los sitios bajo y alto más el sitio de expansión del robledal CNX muestreados en 2008. De ellos se obtuvieron 356.963.222 lecturas en CNB, 365.590.198 lecturas en CNA y 391.923.886 lecturas en CNX todos ellos con una longitud de 101 bp, que tras filtrarlas y solapar las pareadas (Paired-Ends reads), resultaron en 173.277.220 secuencias en CNB, 171.236.955 secuencias en CNA y 192.242.088 secuencias en CNX (Tabla R2-1) con un rango de tamaños entre 50 y 190 bp y un tamaño medio de 170 bp en todos ellos. Tras rastrear secuencias homólogas a genes *rRNA*, con un programa propio del JGI, se obtuvieron 26.187.772 secuencias en CNB, 26.947.483 secuencias en CNA y 29.770.058 secuencias en CNX. Se consiguieron clasificar, al menos a nivel de dominio, 1.389.054 secuencias en CNB, 1.293.856 secuencias en CNA y

Resultados Capítulo 2

1.908.773 secuencias en CNX. De éstas, se obtuvieron 71.178 secuencias en CNB, 65.286 secuencias en CNA y 135.237 secuencias en CNX pertenecientes al gen 16S *rRNA* que se clasificaron en RDP (Tabla R2-1).

Al analizar la clasificación de las secuencias homólogas al gen 16S *rRNA* de CNB, CNA y CNX procedentes de la secuenciación Illumina conseguimos clasificar el 65,7 % de las secuencias en CNB, el 59,6 % en CNA y el 64,4 % en CNX, distribuidas en 245 familias. De éstas, 88 familias muestran diferencias significativas entre algunos de los sitios comparados dos a dos. Además, 23 de estas familias tienen una abundancia relativa de, al menos, el 1 % en alguno de los sitios muestreados, mostrándose en la figura todas aquellas que presentan diferencias biológicamente relevantes (Fig. R2-4). Según Parks *et al.* (2010) se consideran diferencias biológicamente relevantes aquellas que, además de tener valores diferentes a los obtenidos por azar en una tabla de contingencia 2 x 2 a un intervalo de confianza del 95 %, presentan una diferencia de abundancia relativa mayor o igual al 1 % o tienen un ratio de abundancia mayor o igual a 2. Las familias mayoritarias en el robledal en expansión (CNX) fueron Flavobacteriaceae del *phylum* Bacteroidetes con un 11,3 %, Nitrospiraceae del *phylum* Nitrospirae con un 5,7 %, Solirubrobacteraceae del *phylum* Actinobacteria con un 3,2 %, Fusobacteriaceae del *phylum* Fusobacteria con un 2,5 % e Ignavibacteriaceae del *phylum* Chlorobi con un 2,4 %. Todas estas familias, excepto la familia Solirubrobacteraceae, son significativamente más abundantes en CNX, y en la Figura R2-4 aparecen marcadas con una estrella si no se encuentran diferencias entre CNB y CNA o con un triángulo con el vértice hacia arriba si los tres sitios son diferentes y la abundancia relativa aumenta con la altitud. Es decir, encontramos dos situaciones, una que depende del grado de desarrollo del robledal, maduro o en expansión, donde no vemos diferencias entre CNA y CNB, marcadas con una estrella. Otra situación menos marcada en el robledal maduro que depende del gradiente altitudinal, cuya máxima diferencia se aprecia en CNX, marcada con un triángulo con el vértice hacia arriba. La familia Solirubrobacteraceae es la tercera más abundante en CNX pero es significativamente menos abundante en este sitio respecto a los otros dos, en los que es la familia más abundante con un 5,1 % en CNB y 4,7 % en CNA, marcada con un círculo en la figura R2-4. Algo similar ocurre con las familias Hyphomicrobiaceae y Bradyrhizobiaceae de la clase α -Proteobacteria y del orden Rhizobiales, que son la cuarta y quinta familias más abundantes en suma de los tres sitios con un 3,2 % en CNB, un 3,5 % en CNA y un 1,9 % en CNX en el caso de la primera familia, Hyphomicrobiaceae. De esta familia no se encontraron secuencias en el sitio alto, CNA, con el estudio de genotecas clásicas; en el caso del estudio de secuencias obtenidas por pirosecuenciación no se encontraron secuencias de esta familia en el sitio bajo, CNB; sin

embargo, con la tecnología Illumina aparecen en ambos sitios y sin diferencias significativas. La familia Bradyrhizobiaceae presenta un 3,3 % tanto en CNB como en CNA y un 1,7 % en CNX. Las familias Nocardoidaceae del *phylum* Actinobacteria y Sphingomonadaceae del *phylum* Proteobacteria son significativamente diferentes en los tres sitios disminuyendo su abundancia relativa con la altitud con valores de 3,5 % en CNB, 2,7 % en CNA y 1,5 % en CNX para la primera (Nocardoidaceae) y 3,3 % en CNB, 2,9 % en CNA y 1,5 % en CNX para la segunda, Sphingomonadaceae. De esta familia no se encontraron secuencias en el sitio alto, CNA, con el estudio clásico; en el caso del estudio de secuencias obtenidas por pirosecuenciación se obtuvieron secuencias de ambos sitios, siendo más abundantes en el sitio bajo, CNB, pero sin significación estadística; mientras que con la tecnología Illumina se mantienen estas diferencias en la abundancia y sí son significativas en este estudio. Las familias Micromonosporaceae y Streptosporangiaceae ambas del *phylum* Actinobacteria son significativamente más abundantes en CNB 3,2 % frente a 1,6% en CNA y CNX en el primer caso y 1,1 % en CNB frente a 0,1 % en CNA y 0,3 % en CNX, marcadas con un cuadrado (Fig. R2-4).

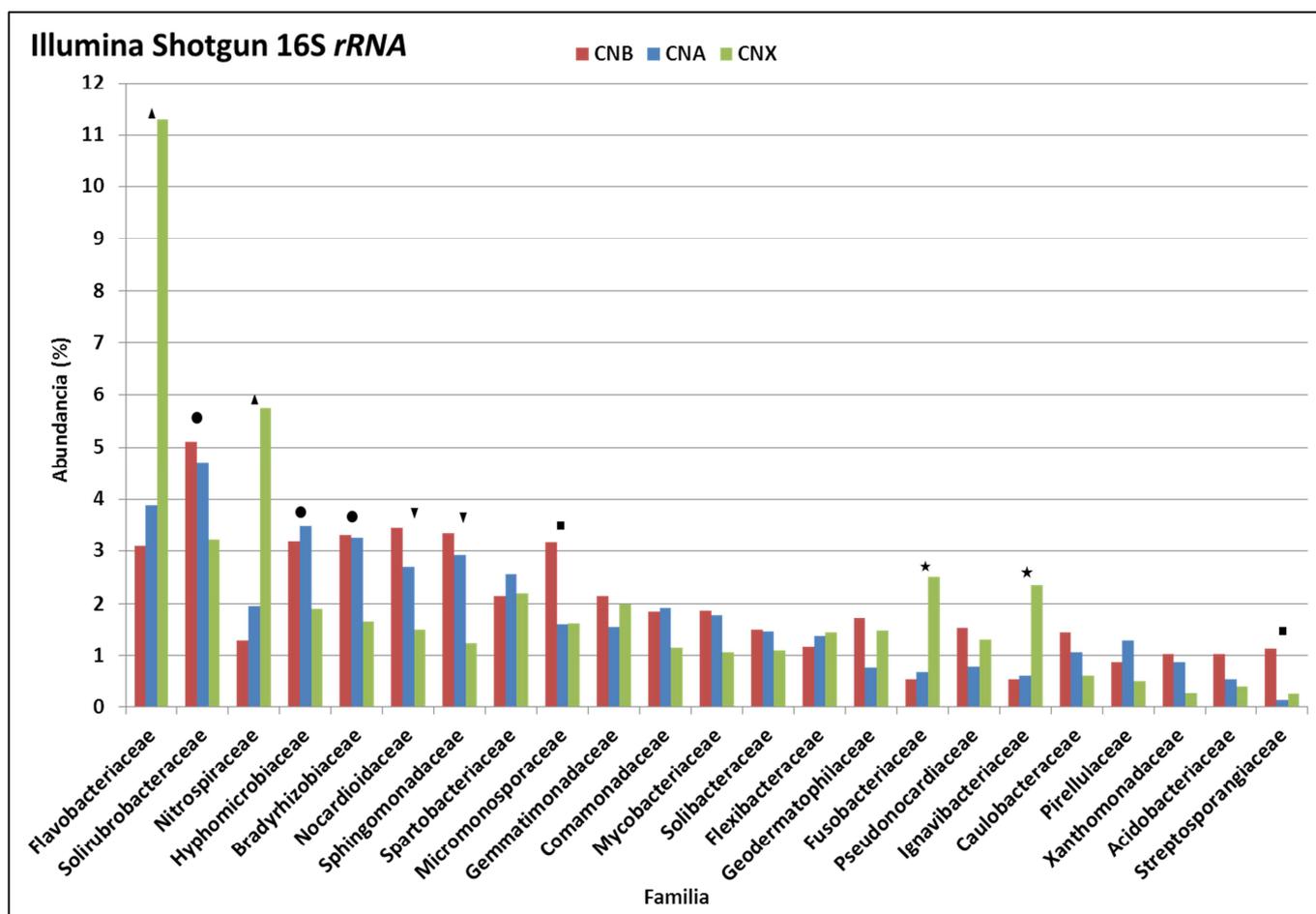


Figura R2-4. Distribución de las familias más abundantes de las muestras secuenciadas mediante tecnología Illumina. Solo se muestran aquellas familias que suponen más del 1 % de abundancia relativa en alguno de los tres sitios: CNB, CNA y CNX. Los símbolos marcan aquellas familias en las que se encontraron diferencias significativas biológicamente relevantes comparando los sitios dos a dos con el programa STAMP, tomando un p-valor corregido menor del 0,01. El significado de los símbolos es: ▲ Familias en las que la abundancia relativa aumenta significativamente con la altitud. ▼ Familias en las que la abundancia relativa disminuye significativamente con la altitud. ★ Familias significativamente más abundantes en CNX. ● Familias significativamente más abundantes en el robledal maduro. ■ Familias significativamente más abundantes en CNB.

En el estudio de las librerías NGS (amplicones) de la zona V6-V8 del gen 16S *rRNA* se analizaron las 3 parcelas de cada uno de los 4 sitios (CNB, CNA, CNX y CNR) que comprenden las 3 altitudes objeto de este análisis, para el muestreo de 2008. Pero en el año 2011 los sitios muestreados fueron CNB, CNA y CNX, con 3 parcelas en cada uno de los sitios. De este análisis se obtuvieron 483.072 lecturas con una longitud media de 408 bp distribuidas uniformemente entre las 21 muestras secuenciadas, con valores entre 18.228 lecturas en CNX1_11 y 26.082 lecturas en CNA3_08. Tras filtrar y descartar las lecturas con baja calidad, las posibles quimeras y las menores de 220 bp con el programa Pyrotagger, se obtuvieron 364.771 secuencias con una longitud media de 458 bp distribuidas uniformemente y con unos valores entre 14.122 secuencias en CNX1_11 y 19.995 secuencias en CNX2_11 (Tabla R2-2). Los cebadores utilizados

para obtener estos amplicones no solo amplifican la zona V6-V8 del gen 16S *rRNA* sino que, también amplifican de forma menos específica regiones del gen 18S *rRNA* de Eucariotas. Para filtrar las secuencias Procarióticas se alinearon las secuencias filtradas por Pyrotagger con la herramienta Aligner de RDP frente al gen 16S *rRNA* de *E. coli* J01695 y se eliminaron todas aquellas secuencias que no alinearon en la zona V6-V8, quedando un promedio de 14.560 secuencias por muestra, con un valor mínimo de 11.781 secuencias en CNA1_08 y un valor máximo de 18.205 secuencias en CNX2_11. Para realizar análisis que requieren normalización del número de secuencias de cada muestra, se normalizaron todas las muestras a 11.244 secuencias con el programa MOTHUR (Tabla R2-2).

Tabla R2-2. Recuento de secuencias obtenidas por pirosecuenciación de la zona V6-V8 de los amplicones del gen 16S *rRNA*.

* Secuencias obtenidas tras el filtro realizado en el JGI con la herramienta Pyrotagger.

** Secuencias obtenidas tras el filtro realizado al alinear las secuencias filtradas en el JGI con la herramienta Aligner del RDP. Las secuencias filtradas con RDP se normalizaron con MOTHUR.

Muestra	Lecturas	Secuencias JGI*	Secuencias RDP**	Normalizadas
CNB1_08	21.228	15.667	12.158	11.244
CNB1_11	24.279	18.903	17.044	11.244
CNB2_08	24.109	16.774	12.498	11.244
CNB2_11	23.850	18.439	16.548	11.244
CNB3_08	22.537	16.694	13.002	11.244
CNB3_11	24.852	19.177	17.467	11.244
CNA1_08	20.179	15.027	11.781	11.244
CNA1_11	20.073	15.661	14.287	11.244
CNA2_08	23.223	16.841	13.392	11.244
CNA2_11	23.255	17.742	15.985	11.244
CNA3_08	26.082	19.615	16.499	11.244
CNA3_11	24.214	18.975	17.173	11.244
CNX1_08	22.177	16.422	12.545	11.244
CNX1_11	18.228	14.122	12.961	11.244
CNX2_08	24.985	18.769	14.586	11.244
CNX2_11	25.035	19.995	18.205	11.244
CNX3_08	24.645	18.001	13.764	11.244
CNX3_11	23.851	18.569	16.894	11.244
CNR1	22.259	16.714	13.290	11.244
CNR2	21.692	16.072	12.587	11.244
CNR3	22.319	16.592	13.096	11.244

Al analizar la clasificación de las secuencias normalizadas homólogas al gen 16S *rRNA* obtenidas en el muestreo del año 2008 de los 4 sitios CNB, CNA, CNX y CNR [considerados para el análisis como 4 grupos con 3 réplicas biológicas (muestras) cada uno], conseguimos clasificar el 81,1 % de las secuencias en CNB, el 80,7 % en CNA, el 77,3 % en CNX y el 75,4 % en CNR a

Resultados Capítulo 2

nivel de *phylum*. A nivel de familia clasificamos el 52,1 % de las secuencias en CNB, el 50,4 % en CNA, el 45,3 % en CNX y el 44,8 % en CNR, distribuidas en 154 familias. De estas familias, 48 tienen diferencias estadísticamente significativas en la abundancia relativa promedio encontrada, al menos en uno de los sitios. En la Figura R2-5 se muestran las 20 familias más abundantes para los 4 sitios. Solo las familias Chitinophagaceae del *phylum* Bacteroidetes; Geodermatophilaceae, Micromonosporaceae y Pseudonocardiaceae del orden Actinomycetales; y las familias de las clases Gp4 y Gp1 del *phylum* Acidobacteria, no mostraron diferencias significativas en las medias de abundancia relativa para ningún sitio. Al igual que en los resultados de secuenciación Illumina (Fig. R2-4) las 3 familias de Actinomycetales son más abundantes en CNB que CNA y CNX con un 1,71 % en CNB, un 0,72 % en CNA y un 1,56 % en CNX para la familia Geodermatophilaceae, un 2,10 % en CNB, un 1,41 % en CNA y un 0,92 % en CNX para la familia Micromonosporaceae y un 1,59 % en CNB, un 0,98 % en CNA y un 1,01 % en CNX para la familia Pseudonocardiaceae. En este análisis con réplicas no podemos rechazar la hipótesis, estudiada mediante ANOVA, de que las medias para los 4 sitios estudiados en estas 4 familias sean diferentes significativamente aunque, en el caso de la familia Geodermatophilaceae las desviaciones estándar de las medias de CNX y CNR no solapan con la desviación estándar de la media de CNA y en el caso de la familia Chitinophagaceae las desviaciones estándar de las medias de CNA y CNX no solapan con la desviación estándar de la media de CNR. La familia más abundante es de la clase Gp4 del *phylum* Acidobacteria con un 2,82 % en CNB, un 2,71 % en CNA, un 2,90 % en CNX y un 4,07 % en CNR, siendo más abundante en suelo no rizosférico, pero dicha diferencia no es estadísticamente significativa. La segunda familia más abundante es Plantomycetaceae del *phylum* Plantomycetes con un 2,67 % en CNB, un 2,58 % en CNA, un 2,41 % en CNX y un 3,59 % en CNR, siendo significativamente más abundante en suelo no rizosférico. La tercera familia más abundante es Bradyrhizobiaceae que al igual que en el estudio de secuenciación Illumina (Fig. R2-4) es significativamente más abundante en el robleal maduro (3,44 % en CNB y 3,68 % en CNA) respecto del robleal en expansión (1,85 %) y del suelo suelto (1,11 %). Lo mismo ocurre con la familia Mycobacteriaceae del *phylum* Actinobacteria que no se ha considerado biológicamente relevante en el análisis Illumina pero aparece significativamente más abundante en el robleal maduro (1,35 % en CNB y 1,36 % en CNA) frente a la zona de expansión (0,46 % en CNX y 0,32 % en CNR). El caso de la familia Sphingomonadaceae de la clase α -Proteobacteria, que ocupa la séptima posición en abundancia, muestra que hay diferencias significativas entre el robleal maduro (2,92 % en CNB y 2,57 % en CNA) y el suelo suelto (1,07 % en CNR), siendo CNX con un 2,10 % el que tiene un valor intermedio de abundancia relativa pero no es significativo respecto a los demás en el análisis *a posteriori* (post-hoc), realizado por el método de Tukey-

Kramer con un intervalo de confianza del 95 %. El mismo efecto se observa en la familia Nocardioideae del *phylum* Actinobacteria con un 2,31 % en CNB, un 2,52 % en CNA, un 1,12 % en CNX y un 0,52 % en CNR; y en las familias Caulobacteraceae e Hyphomicrobiaceae ambas también de la clase α -Proteobacteria, con un 1,18 % en CNB, un 1,00 % en CNA, un 0,87 % en CNX y un 0,32 % en CNR para la familia Caulobacteraceae y un 0,87 % en CNB, un 0,92 % en CNA, un 0,54 % en CNX y un 0,21 % en CNR en la familia Hyphomicrobiaceae. Indicar que en esta última familia aparecían diferencias significativas entre el robledal maduro frente al robledal en expansión en los resultados obtenidos por secuenciación Illumina. En la figura R2-5 puede verse que no solapan las desviaciones estándar de las medias de CNB con CNX ni de CNA con CNX, para la familia Hyphomicrobiaceae, pero el análisis post-hoc solo confirma diferencias significativas entre el robledal maduro y el suelo suelto quedando el robledal en expansión con un valor de abundancia relativa intermedio entre ambas zonas, sin ser significativo frente a ninguna de ellas. En la octava familia, Solirubrobacteraceae del *phylum* Actinobacteria y del orden Solirubrobacterales, observamos no solo que en CNX con un 1,50 % es menos abundante significativamente que en CNB con un 2,26 % y que en CNA con un 2,42 %, al igual que ocurría en los datos de secuenciación Illumina (Fig. R2-4), si no que también es significativamente menos abundante que en CNR con un 2,25 %. La familia Conexibacteraceae, perteneciente también al orden Solirubrobacterales, presenta una distribución diferente puesto que en CNX pese a ser el menos abundante con un 0,93 % no muestra diferencias significativas con CNA con un 1,01 % ni con CNB con un 1,34 %, solo se observan diferencias significativas entre CNA y CNX con CNR donde constituye el 2,03 %, quedando el robledal de la zona baja en una posición intermedia sin diferencias significativas con ninguno de los otros sitios. La sexta familia más abundante, Gemmatimonadaceae del *phylum* Gemmatimonadetes, es significativamente más abundante en CNR con un 3,09 % respecto al robledal maduro (1,76 % en CNB y 1,36 % en CNA) quedando el robledal en expansión CNX con un valor de abundancia relativa intermedio (2,76 %) sin diferencias significativas con CNR ni CNB pero sí muestra diferencias significativas con CNA. El mismo perfil presentan las familias de las clases Spartobacteria y Subdivisión3 pertenecientes al *phylum* Verrucomicrobia, con un 1,30 % en CNB, un 1,54 % en CNA, un 2,37 % en CNX y un 3,76 % en CNR para la familia de la clase Spartobacteria y un 0,46 % en CNB, un 0,49 % en CNA, un 0,84 % en CNX y un 1,07 % en CNR en la familia de la clases Subdivisión3. La familia de la clase Gp7 del *phylum* Acidobacteria con un 0,50 % en CNB, un 0,75 % en CNA, un 0,80 % en CNX y un 1,26 % en CNR, es más abundante en suelo no rizosférico, pero la diferencia sólo es significativa entre CNR y CNB (Fig. R2-5).

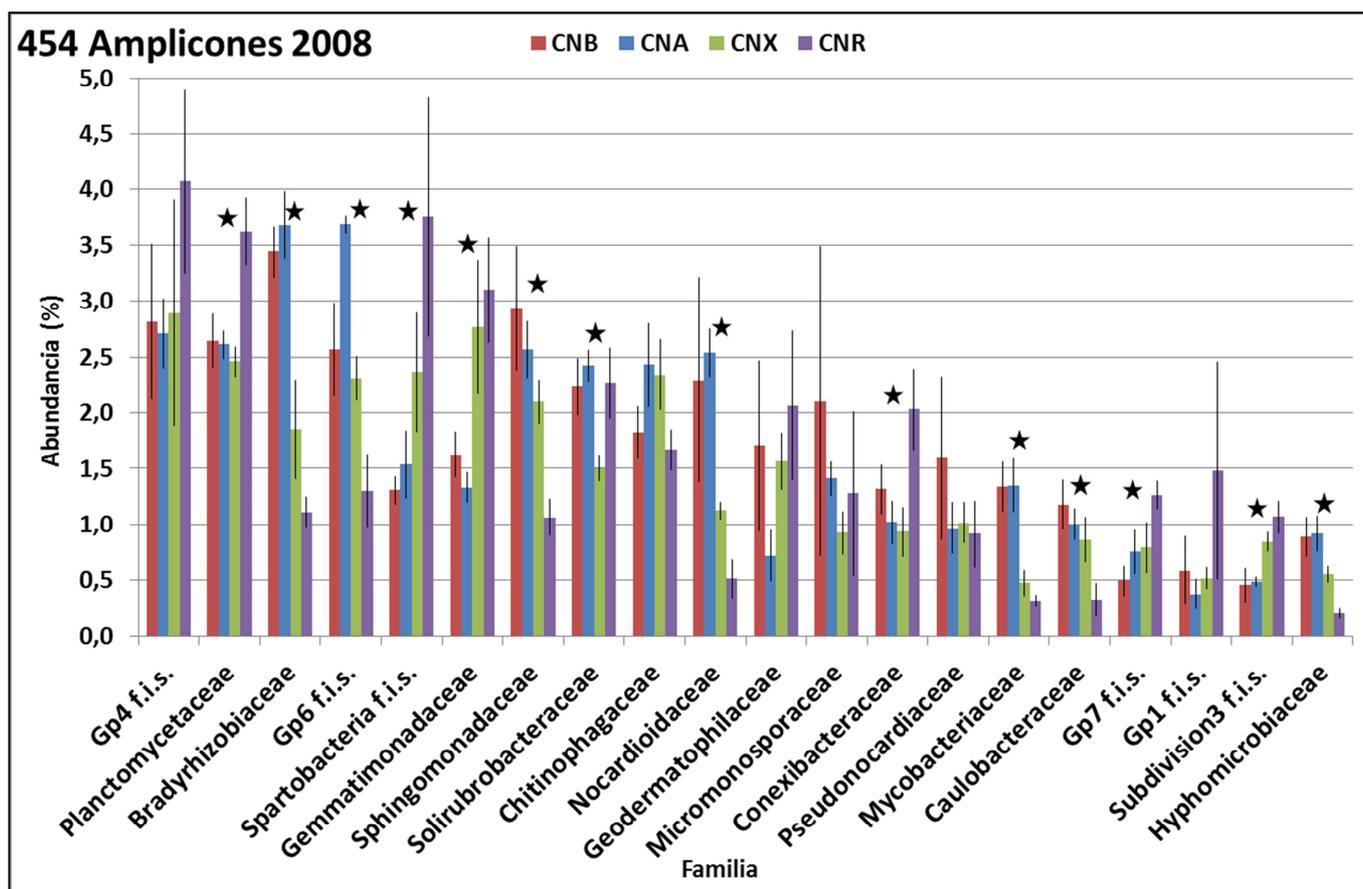


Figura R2-5. Distribución de las familias más abundantes de las muestras tomadas en el año 2008 y secuenciadas mediante pirosecuenciación de amplicones de la zona V6-V8 del gen 16S *rRNA*. Sólo se muestran aquellas familias que suponen al menos el 2,6 % de abundancia relativa en suma de los cuatro sitios: CNB, CNA, CNX y CNR. Las barras de error se corresponden con la desviación estándar del valor medio de la abundancia relativa de cada familia para cada sitio, calculada a partir de las 3 muestras de cada sitio. En las familias marcadas con una estrella se encontraron diferencias estadísticamente significativas entre las medias de al menos uno de los cuatro sitios analizados mediante ANOVA. Se analizó qué sitios mostraban diferencias significativas entre sus medias mediante el test post-hoc de Tukey-Kramer con un intervalo de confianza del 95 % en el programa STAMP. f.i.s. = Familia sin nomenclatura definitiva en la clasificación taxonómica.

Al analizar la clasificación de las secuencias normalizadas homólogas al gen 16S *rRNA* obtenidas en el muestreo del año 2011 de los sitios CNB, CNA y CNX conseguimos clasificar el 81,8 % de las secuencias en CNB, el 81,0 % en CNA y el 79,6 % en CNX a nivel de *phylum*. A nivel de familia clasificamos el 48,5 % de las secuencias en CNB, el 47,3 % en CNA y el 45,9 % en CNX, distribuidas en 145 familias. De éstas, la única familia con diferencias estadísticamente significativas en la abundancia relativa promedio encontrada, al menos en uno de los sitios, es la familia Anaerolineaceae del *phylum* Chloroflexi. Esta familia no es mayoritaria en las comunidades microbianas de los sitios evaluados, constituyendo un promedio del 0,23 % en CNB, un 0,43 % en CNA y un 0,34 % en CNX, siendo significativamente más abundante en CNA que en CNB con un p-valor corregido marginal de 0,044. En la Figura R2-6 se muestran las 19

familias más abundantes para los 3 sitios. Se observa que ninguna de ellas presenta diferencias significativas teniendo en cuenta el factor de corrección de falsos positivos para múltiples test (Storey FDR). Aunque la familia Mycobacteriaceae del *phylum* Actinobacteria, al igual que en el muestreo de 2008 (Fig. R2-5), es significativamente más abundante en el robleal maduro (1,05 % en CNB y 1,04 % en CNA) que en el robleal en expansión (0,36 % en CNX), en este caso con un p-valor sin corrección de 0,013 puede considerarse un falso negativo con el factor de corrección aplicado. En el caso de las familias Nocardoidaceae y Solirubrobacteraceae del *phylum* Actinobacteria con un 1,65 % en CNB, un 2,14 % en CNA y un 1,21 en CNX en la primera familia no solapan las desviaciones estándar de las medias de CNA y CNX. En el caso de la familia Solirubrobacteraceae con un 1,43 % en CNB, un 1,92 % en CNA y un 1,42 % en CNX tampoco solapa la desviación estándar de la media de CNA ni con la de CNB ni con la de CNX (Fig. R2-6). Algo similar ocurre en las familias Sphingomonadaceae y Bradyrhizobiaceae de la clase α -Proteobacteria siendo menos abundantes en el robleal en expansión, sin solapar la desviación estándar de las medias de CNB y CNX en la primera y las medias de CNA y CNX en la segunda familia. La familia Sphingomonadaceae presenta valores de abundancia de un 2,66 % en CNB, un 2,41 % en CNA y un 2,10 % en CNX; mientras que los porcentajes son del 2,29 % en CNB, un 2,23 % en CNA y un 1,64 % en CNX para la familia Bradyrhizobiaceae. La familia más abundante en este año de muestreo fue la perteneciente a la clase Gp4, perteneciente al *phylum* Acidobacteria, con valores de abundancia relativa de 3,05 % en CNB, 3,72 % en CNA y 2,83 % en CNX; siendo la familia mayoritaria también en el muestreo de 2008 con una abundancia relativa promedio del 2,75 % en estos tres mismos sitios (Fig. R2-5). La familia Gemmatimonadaceae ha aumentado su abundancia relativa en el robleal maduro entre los años de muestreo desde valores en torno al 1,5 % en 2008 a valores en torno al 1,9 % en 2011, junto con una disminución de la abundancia relativa promedio en el robleal en expansión con un valor de 2,77 % en 2008 (Fig. R2-5) y un valor de 2,09 % en 2011 (Fig. R2-6). Pese a las variaciones observadas en los valores promedio de abundancia relativa entre los distintos años de muestreo solo son significativas las disminuciones de abundancia en las muestras de robleal maduro para las familias Bradyrhizobiaceae del *phylum* Proteobacteria (con un 3,44 % en CNB y un 3,68 % en CNA en 2008 frente a un 2,29 % en CNB y un 2,23 % en CNA en 2011) y Solirubrobacteraceae del *phylum* Actinobacteria (con un 2,23 % en CNB y un 2,42 % en CNA en 2008 frente a un 1,43 % en CNB y un 1,92 % en CNA en 2011).

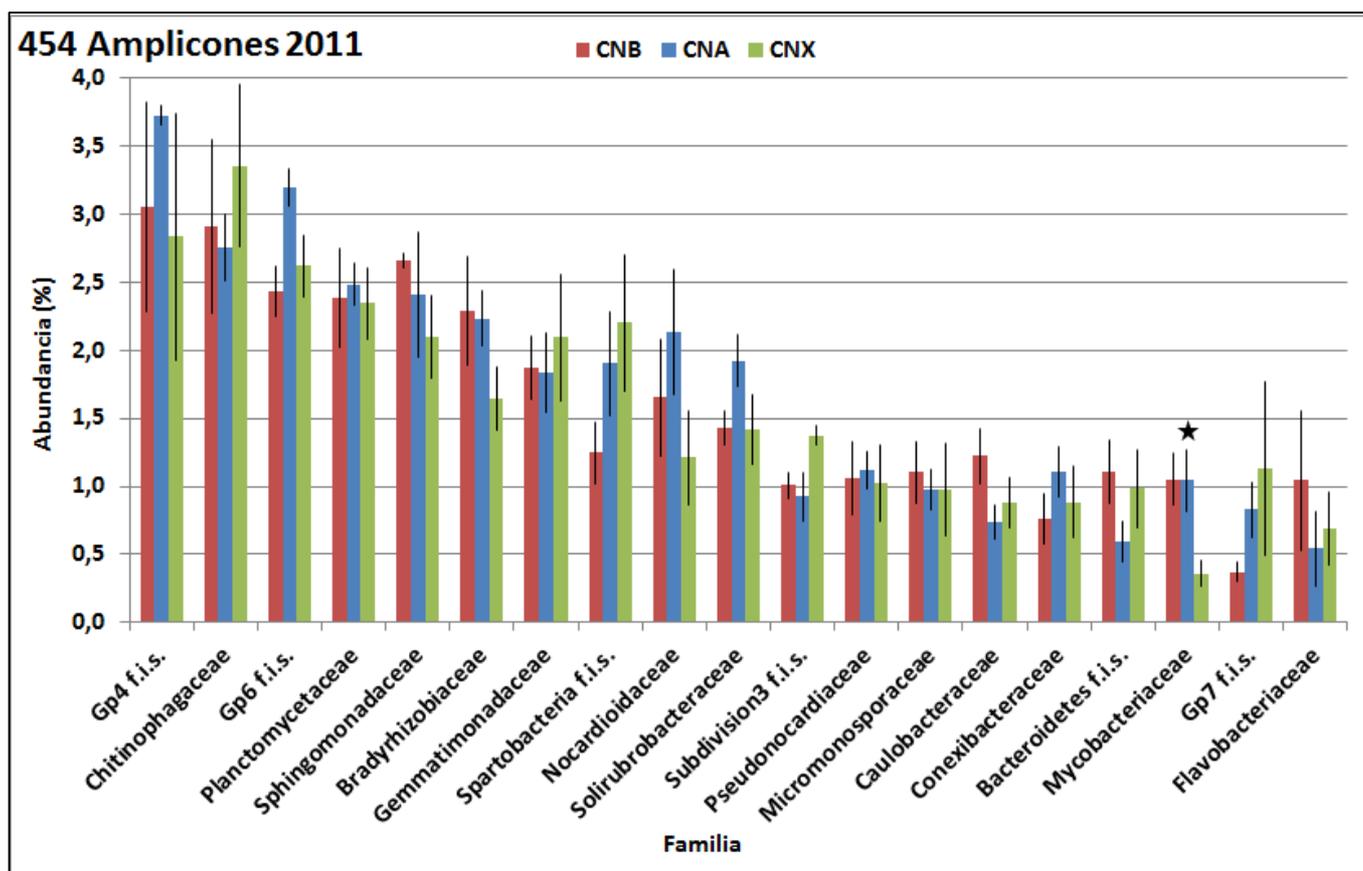


Figura R2-6. Distribución de las familias más abundantes de las muestras tomadas en el año 2011 y secuenciadas mediante pirosecuenciación de amplicones de la zona V6-V8 del gen 16S *rRNA*. Solo se muestran aquellas familias que suponen al menos el 1 % de abundancia relativa en alguno de los tres sitios: CNB, CNA y CNX. Las barras de error se corresponden con la desviación estándar del valor medio de la abundancia relativa de cada familia para cada sitio, calculada a partir de las 3 muestras de cada sitio. En la familia marcada con una estrella se encontraron diferencias estadísticamente significativas entre las medias de al menos uno de los tres sitios analizados mediante ANOVA. Se analizó qué sitios mostraban diferencias significativas entre sus medias mediante el test post-hoc de Tukey-Kramer con un intervalo de confianza del 95 % en el programa STAMP. f.i.s. = Familia sin nomenclatura definitiva en la clasificación taxonómica.

2 Estudio de la diversidad α mediante el gen 16S *rRNA*

Para analizar la diversidad α entendida como la riqueza y equidad observadas de cada muestra y estimar la riqueza del suelo de origen de cada una de ellas, se agruparon las secuencias del gen 16S *rRNA* de cada muestra en OTUs con una distancia génica máxima del 3 %, considerando a dicha distancia que cada OTU corresponde con una especie distinta. Con este tipo de análisis se consigue estudiar la diversidad α de una comunidad microbiana al nivel taxonómico más bajo permitido por el gen 16S *rRNA* y se evitan sesgos introducidos por los estudios que trabajan con las secuencias clasificadas, como es el caso del grupo artificial de “secuencias no clasificadas”. Debido a que la obtención de OTUs requiere trabajar con secuencias de la misma región del gen, para poder alinearlas, este análisis solo se ha llevado a cabo con las genotecas clásicas y con los amplicones 454 del gen 16S *rRNA*.

En las genotecas se observan los mismos resultados de diversidad α para los dos sitios analizados CNB y CNA, resultando en 59 OTUs la riqueza observada, 591 OTUs como valor medio de la riqueza esperada en el suelo de origen, que oscila entre 255 OTUs como mínimo y 1.501 OTUs como máximo. La distribución de las secuencias entre los distintos OTUs es casi la equidad total ($J' = 0,997$), siendo todos los OTUs singletons (con una única secuencia) excepto dos OTUs que están formados por 2 secuencias. El índice de diversidad de Shannon da un valor de 4,065 muy próximo al valor máximo de dicho índice ($H'_{\max} = 4,111$) para 61 secuencias evaluadas por muestra (Tabla R2-3). Con los resultados obtenidos de la curva de rarefacción, que está en la zona inicial de la misma, donde es recta (Fig. R2-7), y el valor de cobertura calculado con el índice de Good's coverage que da un valor de 6,56 % de los OTUs presentes en la comunidad microbiana original que se han muestreado (Tabla R2-3), observamos que 61 clones son insuficientes para tener muestras representativas.

Tabla R2-3. Diversidad α de las muestras estudiadas con genotecas clásicas del gen 16S *rRNA*. OTUs es el número de grupos obtenido a una distancia génica del 3 %. Chao1 es el valor medio de OTUs que se estima que hay en los suelos de origen, seguido de los valores mínimo y máximo para este índice. H' es el índice de Shannon seguido de su varianza. J' es el índice de equidad de Pielou. La cobertura se ha medido siguiendo el método de Good's coverage.

Muestra	Distancia	Secuencias	OTUs	Chao1	LCI95	UCI95	H'	varH	J'	Cobertura
CNB	0,03	61	59	591	255	1.501	4,065	0,008	0,997	6,56%
CNA	0,03	61	59	591	255	1.501	4,065	0,008	0,997	6,56%

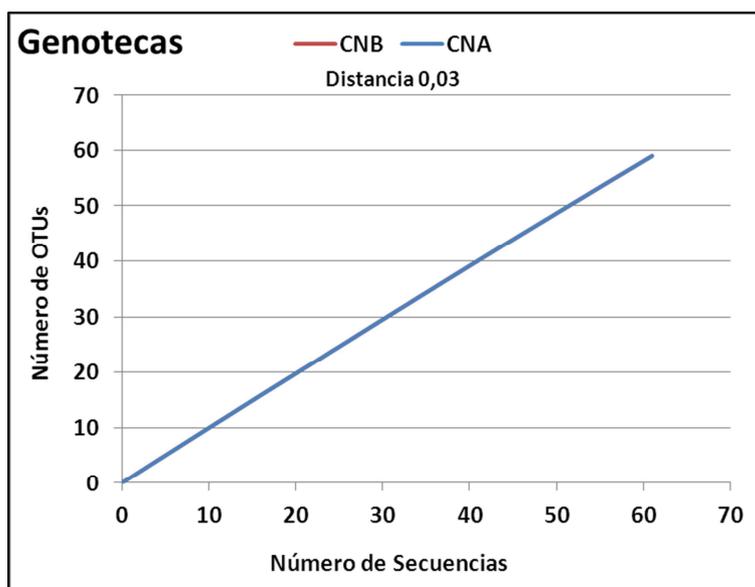


Figura R2-7. Curva de rarefacción de las muestras estudiadas con genotecas clásicas del gen 16S *rRNA*. Los OTUs se agruparon a una distancia génica máxima del 3 %.

En el análisis con amplicones 454 de la zona V6-V8, para las secuencias filtradas obtenidas de las 21 muestras estudiadas (muestreos de 2008 y 2011), la riqueza mínima observada es de 3.572 OTUs en CNB1_08 y la riqueza máxima observada es de 5.216 OTUs en CNX2_11 (Tabla R2.4). Según el índice de Chao1, la riqueza mínima esperada en los suelos de origen es de 5.188 OTUs en CNB1_08, mientras que la riqueza máxima esperada es de 8.872 OTUs en CNX2_11 (Tabla R2-4). Los valores de diversidad, medidos con el índice de Shannon, superaron el valor de 7 en todas las muestras, siendo el valor mínimo de 7,381 en la muestra CNA2_08 y el máximo de 7,885 en la muestra CNX2_08. En todas las muestras las secuencias están distribuidas entre los distintos OTUs de forma muy próxima a la equidad con un valor mínimo del índice de Pielou de 0,895 en la muestra CNX3_11 y un valor máximo de 0,929 en la muestra CNX1_08. Los valores de cobertura medidos con el índice de Good fueron superiores al 85 % en todas las muestras, siendo el valor mínimo de 86,23 % en la muestra CNX1_11 y el valor máximo de 90,88 % en la muestra CNA2_08 (Tabla R2-4).

Para hacer un análisis comparativo de los distintos índices de diversidad α entre todas las muestras estudiadas, éstas se normalizaron a 11.244 secuencias (Tablas R2-2 y R2-4). Se observa que la muestra con menor riqueza es CNA2_08 con 2.774 OTUs y la de mayor riqueza es CNX2_08 con 3.528 OTUs. Mientras que la muestra con menor riqueza esperada, según el índice de Chao1, es CNA2_08 con 4.489 OTUs y la muestra con mayor riqueza esperada en su suelo de origen es CNA2_11 con 6.396 OTUs (Tabla R2-4). Pese a existir diferencias en la riqueza observada entre las distintas muestras, estas diferencias no son significativas

estadísticamente entre los cuatro sitios estudiados ni entre los distintos años de muestreo. Pero sí aparecen diferencias estadísticamente significativas entre la riqueza estimada mediante el índice de Chao1 en el robledal bajo (CNB) entre los dos años de muestreo, siendo este valor mayor para todos los sitios en el muestreo de 2011 (Fig. R2-8). La diversidad medida mediante el índice de Shannon da valores superiores a 7 en las muestras normalizadas, siendo el valor mínimo de 7,173 en la muestra CNA2_08 y el máximo de 7,578 en la muestra CNX2_08 (Tabla R2-3). La diversidad es mayor en el muestreo del robledal maduro (CNB y CNA) en 2011 respecto al muestreo de 2008, mostrándose la tendencia inversa en el robledal en expansión, pero sin ser estadísticamente significativa en ningún caso. Tampoco son significativas las diferencias entre sitios aunque los valores son mayores en la zona de expansión (CNX y CNR) que en el robledal maduro (Fig. R2-9). Los valores de equidad medidos con el índice de Pielou varían entre 0,900 para la muestra CNX3_11 y 0,930 para la muestra CNX1_08 (Tabla R2-4), siendo el robledal en expansión (CNX) el que muestra mayor varianza entre las distintas réplicas estudiadas en ambos años de muestreo (Fig. R2-10). Pese a que todos los OTUs están distribuidos de forma muy equitativa en todas las muestras, se observa cierta dominancia en las muestras de robledal maduro (CNB y CNA) que es significativamente mayor que la encontrada en el suelo suelto (CNR) en el muestreo de 2008 (Fig. R2-10). Según los resultados obtenidos por el índice de cobertura de Good (Tabla R2-4) con un mínimo del 83,68 % en CNX2_11 y un máximo de 88,43 % en CNA2_08; y las curvas de rarefacción de los cuatro sitios estudiados (Fig. R2-11), con 11.244 secuencias por muestra no se consigue secuenciar el 100 % de las especies microbianas (OTUs a una distancia de 0,03) que habitan en los suelos forestales estudiados, pero el esfuerzo de muestreo realizado es suficiente para obtener muestras representativas de cada uno de ellos. Según las curvas de esfuerzo no hay diferencias significativas entre los cuatro sitios, aunque muestra que el robledal en expansión (CNX) es el que requiere un mayor esfuerzo de muestreo para obtener secuencias de todas las especies microbianas que habitan en su comunidad rizosférica (Fig. R2-11).

Resultados Capítulo 2

Tabla R2-4. Diversidad α de las muestras estudiadas con amplicones 454 del gen 16S *rRNA*. OTUs es el número de grupos obtenido a una distancia génica del 3 %. Chao1 es el valor medio de OTUs que se estima que hay en los suelos de origen, seguido de los valores mínimo y máximo para este índice. H' es el índice de Shannon seguido de su varianza. J' es el índice de equidad de Pielou. La cobertura se ha medido siguiendo el método de Good's coverage. La muestra es el número de la parcela de cada sitio de muestreo, seguido de un número que indica el año de muestreo.

Muestra	Estado	Distancia	Secuencias	OTUs	Chao1	LCI95	UCI95	H'	varH	J'	Cobertura
CNA1_08	Filtradas	0,03	15.027	3.702	5.568	5.340	5.828	7,481	0,00012	0,910	89,23%
	Normalizadas	0,03	11.244	3.068	5.120	4.854	5.426	7,327	0,00015	0,913	86,79%
CNA1_11	Filtradas	0,03	15.661	3.990	7.196	6.830	7.608	7,469	0,00013	0,901	87,01%
	Normalizadas	0,03	11.244	3.183	5.860	5.526	6.241	7,314	0,00016	0,907	85,01%
CNA2_08	Filtradas	0,03	16.841	3.624	5.429	5.203	5.688	7,381	0,00011	0,901	90,88%
	Normalizadas	0,03	11.244	2.774	4.489	4.254	4.762	7,173	0,00016	0,905	88,43%
CNA2_11	Filtradas	0,03	17.742	4.390	7.372	7.046	7.738	7,588	0,00011	0,905	87,93%
	Normalizadas	0,03	11.244	3.199	6.396	5.989	6.863	7,347	0,00016	0,910	84,79%
CNA3_08	Filtradas	0,03	19.615	4.840	8.126	7.779	8.514	7,692	0,00010	0,907	88,18%
	Normalizadas	0,03	11.244	3.183	6.127	5.754	6.553	7,327	0,00016	0,908	85,01%
CNA3_11	Filtradas	0,03	18.975	4.583	7.486	7.171	7.840	7,645	0,00010	0,907	88,63%
	Normalizadas	0,03	11.244	3.173	5.376	5.097	5.697	7,389	0,00015	0,916	86,10%
CNB1_08	Filtradas	0,03	15.667	3.572	5.188	4.984	5.420	7,412	0,00012	0,906	89,80%
	Normalizadas	0,03	11.244	2.841	4.590	4.353	4.863	7,204	0,00016	0,906	88,08%
CNB1_11	Filtradas	0,03	18.903	4.548	7.419	7.105	7.771	7,609	0,00010	0,903	88,72%
	Normalizadas	0,03	11.244	3.221	5.556	5.264	5.890	7,360	0,00016	0,911	85,55%
CNB2_08	Filtradas	0,03	16.774	4.091	6.001	5.776	6.257	7,609	0,00010	0,915	89,69%
	Normalizadas	0,03	11.244	3.169	5.317	5.042	5.632	7,374	0,00016	0,915	86,31%
CNB2_11	Filtradas	0,03	18.439	4.571	7.616	7.287	7.986	7,609	0,00011	0,903	88,10%
	Normalizadas	0,03	11.244	3.266	6.112	5.760	6.513	7,337	0,00017	0,907	84,53%
CNB3_08	Filtradas	0,03	16.694	3.779	5.583	5.361	5.837	7,467	0,00011	0,906	90,51%
	Normalizadas	0,03	11.244	2.887	4.671	4.429	4.951	7,260	0,00015	0,911	88,12%
CNB3_11	Filtradas	0,03	19.177	5.104	8.217	7.899	8.571	7,811	0,00009	0,915	87,33%
	Normalizadas	0,03	11.244	3.520	6.020	5.722	6.357	7,535	0,00014	0,923	84,06%
CNR1	Filtradas	0,03	16.714	4.102	5.937	5.719	6.185	7,673	0,00009	0,922	89,84%
	Normalizadas	0,03	11.244	3.149	5.209	4.941	5.516	7,442	0,00013	0,924	86,79%
CNR2	Filtradas	0,03	16.072	4.337	6.386	6.152	6.650	7,746	0,00010	0,925	88,44%
	Normalizadas	0,03	11.244	3.393	5.524	5.261	5.824	7,529	0,00013	0,926	85,50%
CNR3	Filtradas	0,03	16.592	4.029	5.729	5.524	5.963	7,622	0,00010	0,918	90,21%
	Normalizadas	0,03	11.244	3.079	4.771	4.547	5.028	7,385	0,00014	0,919	87,66%
CNX1_11	Filtradas	0,03	14.122	3.915	6.716	6.394	7.080	7,573	0,00012	0,915	86,23%
	Normalizadas	0,03	11.244	3.329	6.031	5.697	6.412	7,450	0,00014	0,919	84,75%
CNX2_11	Filtradas	0,03	19.995	5.216	8.872	8.500	9.286	7,817	0,00009	0,913	87,36%
	Normalizadas	0,03	11.244	3.526	6.364	6.025	6.748	7,526	0,00015	0,921	83,68%
CNX3_11	Filtradas	0,03	18.569	4.447	7.385	7.062	7.747	7,515	0,00012	0,895	88,57%
	Normalizadas	0,03	11.244	3.127	5.562	5.253	5.917	7,246	0,00019	0,900	85,79%
CNX1_08	Filtradas	0,03	16.422	4.330	6.536	6.280	6.825	7,776	0,00009	0,929	88,87%
	Normalizadas	0,03	11.244	3.320	5.274	5.027	5.556	7,537	0,00012	0,930	86,29%

CNX2_08	Filtradas	0,03	18.769	5.007	7.425	7.168	7.714	7,885	0,00008	0,926	88,53%
	Normalizadas	0,03	11.244	3.528	5.773	5.504	6.079	7,578	0,00013	0,928	84,61%
CNX3_08	Filtradas	0,03	18.001	4.387	6.555	6.308	6.834	7,647	0,00010	0,912	89,57%
	Normalizadas	0,03	11.244	3.185	5.262	4.998	5.565	7,380	0,00015	0,915	86,32%

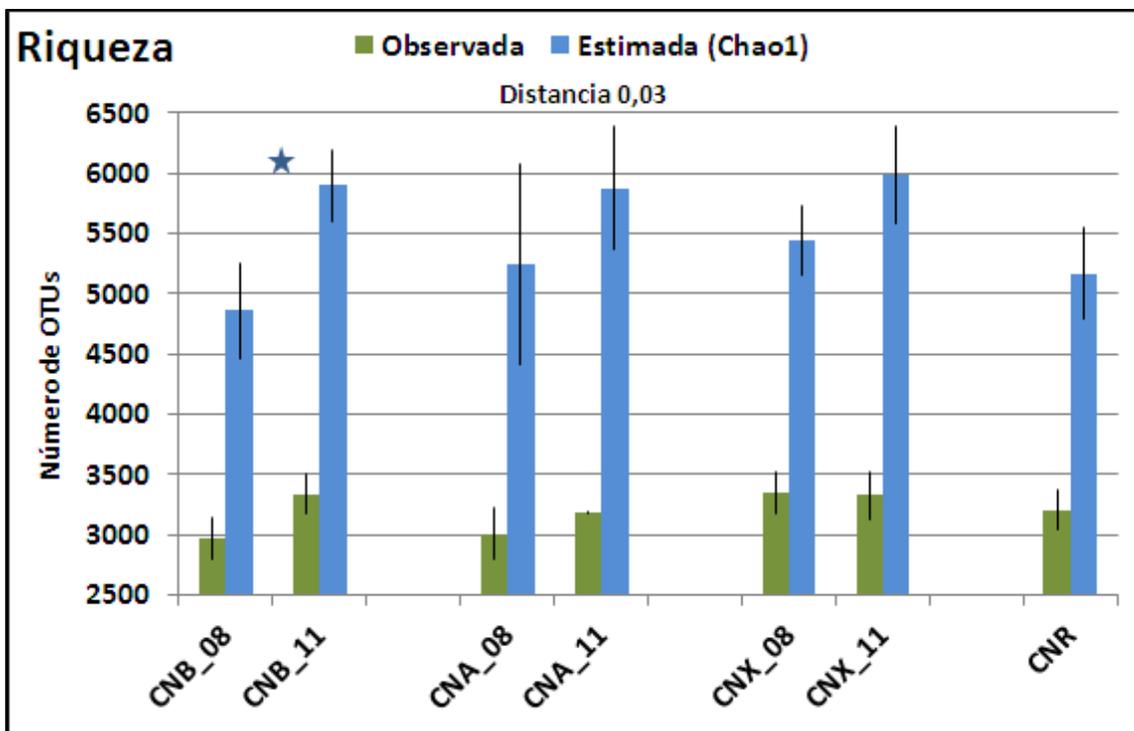


Figura R2-8. Riqueza observada y estimada de especies medidas como OTUs a una distancia del 3 % obtenidas tras la pirosecuenciación de amplicones del gen 16S *rRNA*. Las barras de error muestran la desviación estándar de la media observada entre las réplicas de cada sitio para cada año de muestreo. La estrella indica diferencias estadísticamente significativas en los valores medios de riqueza estimada comprobados mediante la prueba t de Student para dos muestras emparejadas.

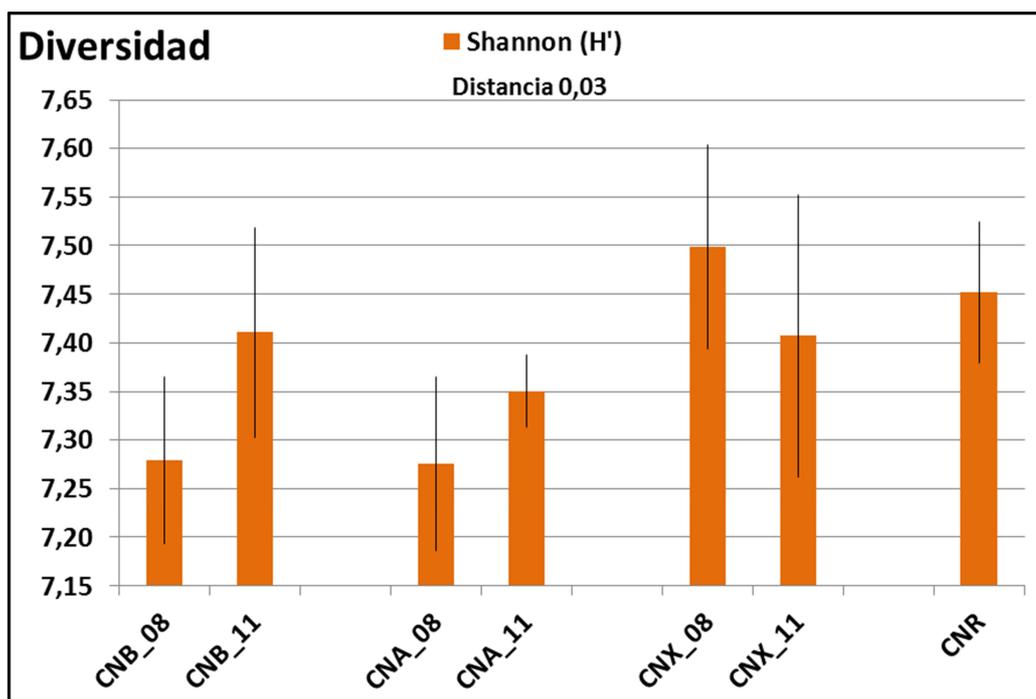


Figura R2-9. Diversidad de especies medidas como OTUs a una distancia del 3 %, analizada con el índice de Shannon tras la pirosecuenciación de amplicones del gen 16S *rRNA*. Las barras de error muestran la desviación estándar de la media observada entre las réplicas de cada sitio para cada año de muestreo.

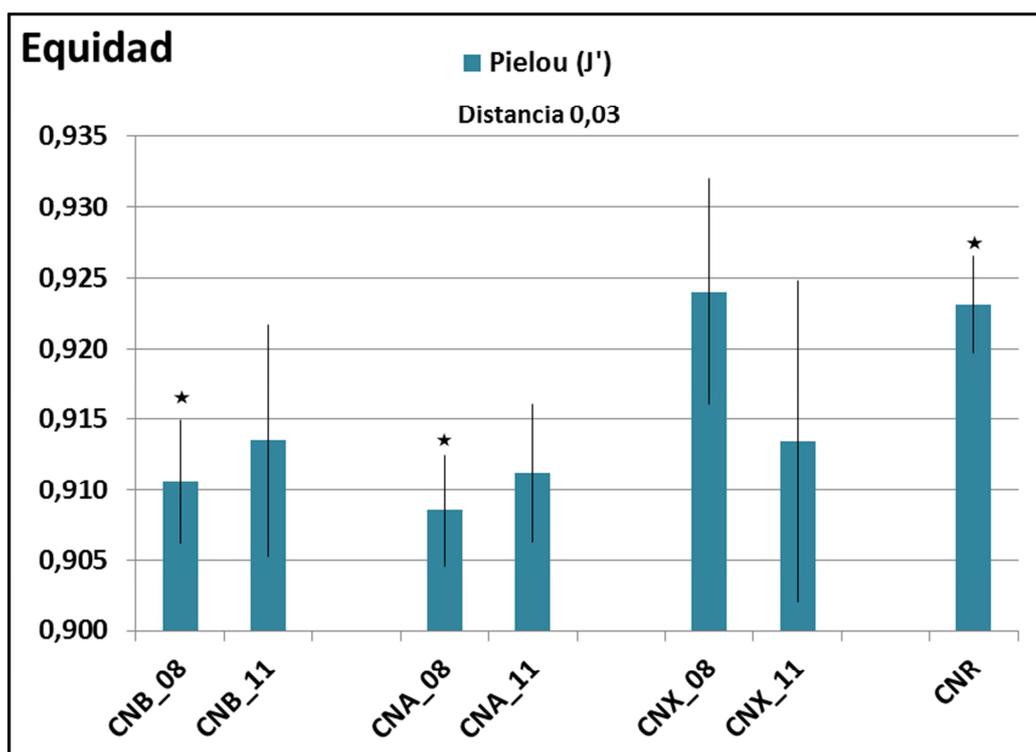


Figura R2-10. Equidad de la distribución de las distintas especies medidas como OTUs a una distancia del 3 % tras la pirosecuenciación de amplicones del gen 16S *rRNA*, analizada con el índice de Pielou. Las barras de error muestran la desviación estándar de la media observada entre las réplicas de cada sitio para cada año de muestreo. La estrella indica diferencias estadísticamente significativas en los valores medios de equidad comprobados mediante la prueba t de Student para dos muestras independientes suponiendo varianzas desiguales.

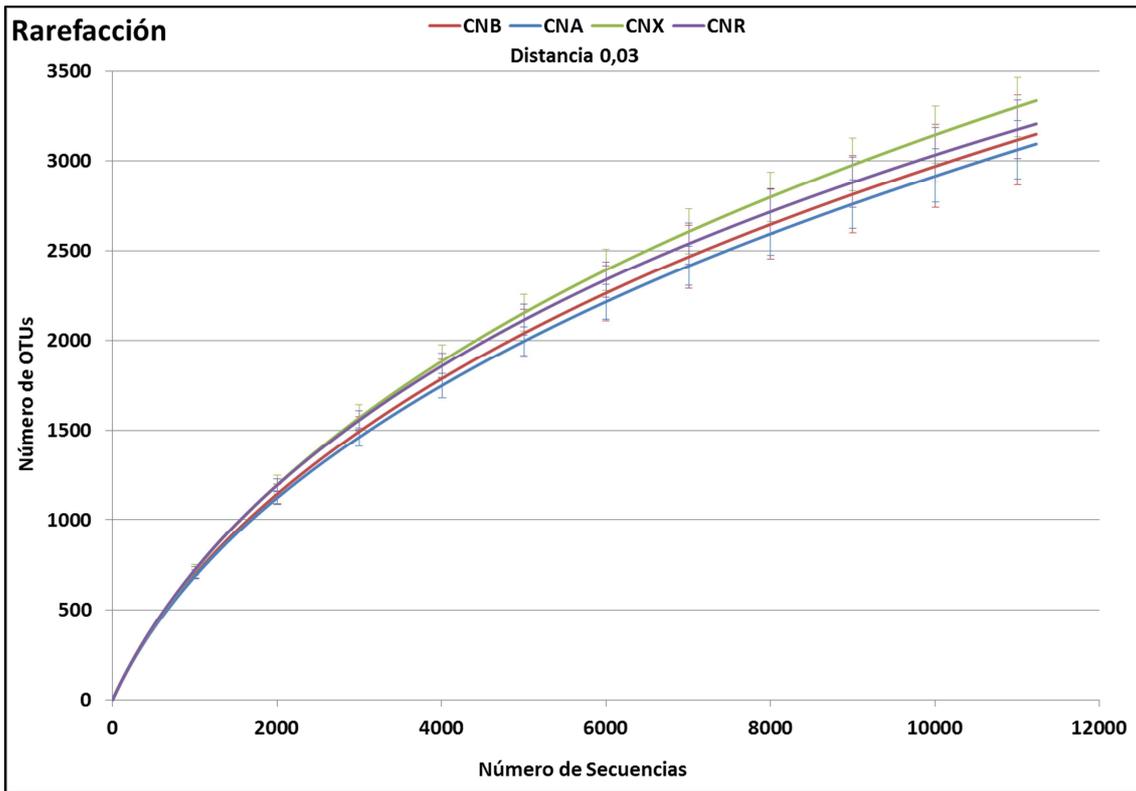


Figura R2-11. Curvas de rarefacción de especies medidas como OTUs a una distancia del 3 % tras la pirosecuenciación de amplicones del gen 16S *rRNA*. Las barras de error muestran la desviación estándar de la media observada entre las réplicas de cada sitio en ambos años de muestreo.

3 Estudio de la diversidad β mediante el gen 16S *rRNA*

Este estudio también se ha realizado con las secuencias obtenidas por secuenciación Sanger de las genotecas en *E.coli* y los amplicones 454 de la zona V6-V8 del gen 16S *rRNA* (Tablas R2-3 y R2-4). Entendiendo por diversidad β la diferencia de composición de las comunidades microbianas de las muestras comparadas dos a dos.

En el análisis con genotecas se observa que el índice de distancia (disimilitud) de Jaccard tiene un valor de 0,903 indicando que, a nivel de especie, ambos sitios son muy diferentes y se confirma en el clustering de los dos sitios juntos, ya que solo 4 OTUs están compartidos entre los dos sitios estudiados. El primero de estos OTUs pertenecería a una especie del género *Spartobacteria_genera_incertae_sedis* según la clasificación del RDP del *phylum* Verrucomicrobia, con 1 secuencia en CNB y 2 secuencias en CNA. El segundo OTU pertenecería a una especie del género *Ferruginibacter* y el tercer OTU pertenecería a una especie del género *Terrimonas* ambas de la familia Chitinophagaceae del *phylum* Bacteroidetes; el cuarto OTU pertenecería a una especie del género *Bradyrhizobium* del *phylum* Proteobacteria. Los 3 últimos OTUs tienen 1 secuencia por sitio. Pese a ser tan diferentes, debido a la baja cobertura de muestreo de este análisis mostrado por el índice de Good's (Tabla R2-3), el estudio estadístico de ambas comunidades mediante β -Libshuff indica que no se encuentran diferencias estadísticamente significativas entre ellas, con un p-valor de 0,78.

En el análisis con amplicones 454 de la zona V6-V8, se filtraron las secuencias después de ser alineadas en RDP y se normalizaron las muestras a 11.244 secuencias (Tablas R2-2 y R2-4). Se realizó el agrupamiento jerárquico aglomerativo mediante UPGMA, sin tener en cuenta la abundancia de los OTUs, analizados mediante Jaccard, ni la abundancia en los nodos del árbol filogenético, analizados mediante Unifrac (Fig. R2-13). Dicho análisis pone de manifiesto que, las muestras de suelo suelto (CNR) se separan del resto de muestras a un 79,6 % de distancia y que las muestras rizosféricas de roble se separan en dos grupos al 76,5 % de distancia. Uno de estos dos grupos solo está compuesto por muestras del robledal en expansión (CNX2 y CNX3 de ambos años de muestreo) y el otro grupo engloba al resto de muestras rizosféricas. La división en estos tres grupos de las 21 muestras, primero suelo suelto y luego la subdivisión expansión (dos de las tres réplicas) del resto de muestras, tiene una excelente robustez estadística ya que se da en el 100 % de los árboles creados para hacer dicho estudio. Todas las muestras están separadas a una distancia superior al 72 %, siendo las

muestras CNA2_11 y CNB2_08 las que tienen una estructura filogenética más parecida agrupándose a una distancia del 72,1 % (Fig. R2-13). Estos valores de distancia son semejantes a los del índice de Jaccard indicando que las comunidades microbianas (tanto a nivel de especie en Jaccard como teniendo en cuenta la pertenencia a distintas ramas del árbol filogenético en Unweighted-Unifrac) de cada muestra son muy distintas de las del resto. Este resultado se confirma con los análisis del test estadístico Unifrac y el AMOVA que dan p-valores inferiores a 0,0001 entre todas las muestras comparadas dos a dos. Esto nos indica además que las comunidades microbianas de cada muestra son significativamente diferentes del resto de muestras (tanto a nivel de especie en AMOVA como teniendo en cuenta la pertenencia a distintas ramas del árbol filogenético en Unweighted-Unifrac).

El análisis también se realizó mediante UPGMA teniendo en cuenta la abundancia de cada OTU para cada muestra (Fig. R2-14), analizado mediante el comando Weighted-Unifrac en MOTHUR y el peso (número de secuencias) de cada nodo en el árbol filogenético analizado mediante Unifrac. En este caso, las muestras de suelo suelto (CNR) son las que más se diferencian del resto, muestras rizosféricas de roble, con un 19,1 % de distancia. Las muestras de la rizosfera de roble se dividen en dos grandes grupos al 13,3 % de distancia, uno en el que mayoritariamente se agrupan las muestras de robledal muestreadas en 2008 y otro grupo que, a su vez, se subdivide en dos grupos al 12,6 %, el primero compuesto por muestras de robledal obtenidas en 2011 y un segundo grupo compuesto solo por muestras del robledal en expansión (CNX) tanto de 2008 como de 2011. La división en estos tres grandes grupos de las muestras rizosféricas de roble no tiene robustez estadística en el análisis Jackknife y corresponde con los resultados de los análisis del test estadístico Unifrac y el AMOVA en los que no aparecen diferencias significativas entre ninguna de las muestras rizosféricas comparadas dos a dos. Pese a mostrar una separación clara entre suelo suelto y rizosfera, las diferencias no son estadísticamente significativas (tanto a nivel de especie en AMOVA como teniendo en cuenta la estructura filogenética en Weighted-Unifrac) (Fig. R2-14). Por tanto vemos una imagen en cierta forma similar según tengamos en cuenta o no el número de secuencias de cada OTU, que en ambos casos nos muestra que la rizosfera de roble selecciona una comunidad microbiana diferente a la del suelo suelto. Además dentro de la situación de muestras rizosféricas encontramos diferencias en cuanto a la abundancia de los grupos taxonómicos que comparten, debido a las diferencias observadas en los factores abióticos (Temperatura y Humedad Relativa) en los dos años de muestreo.

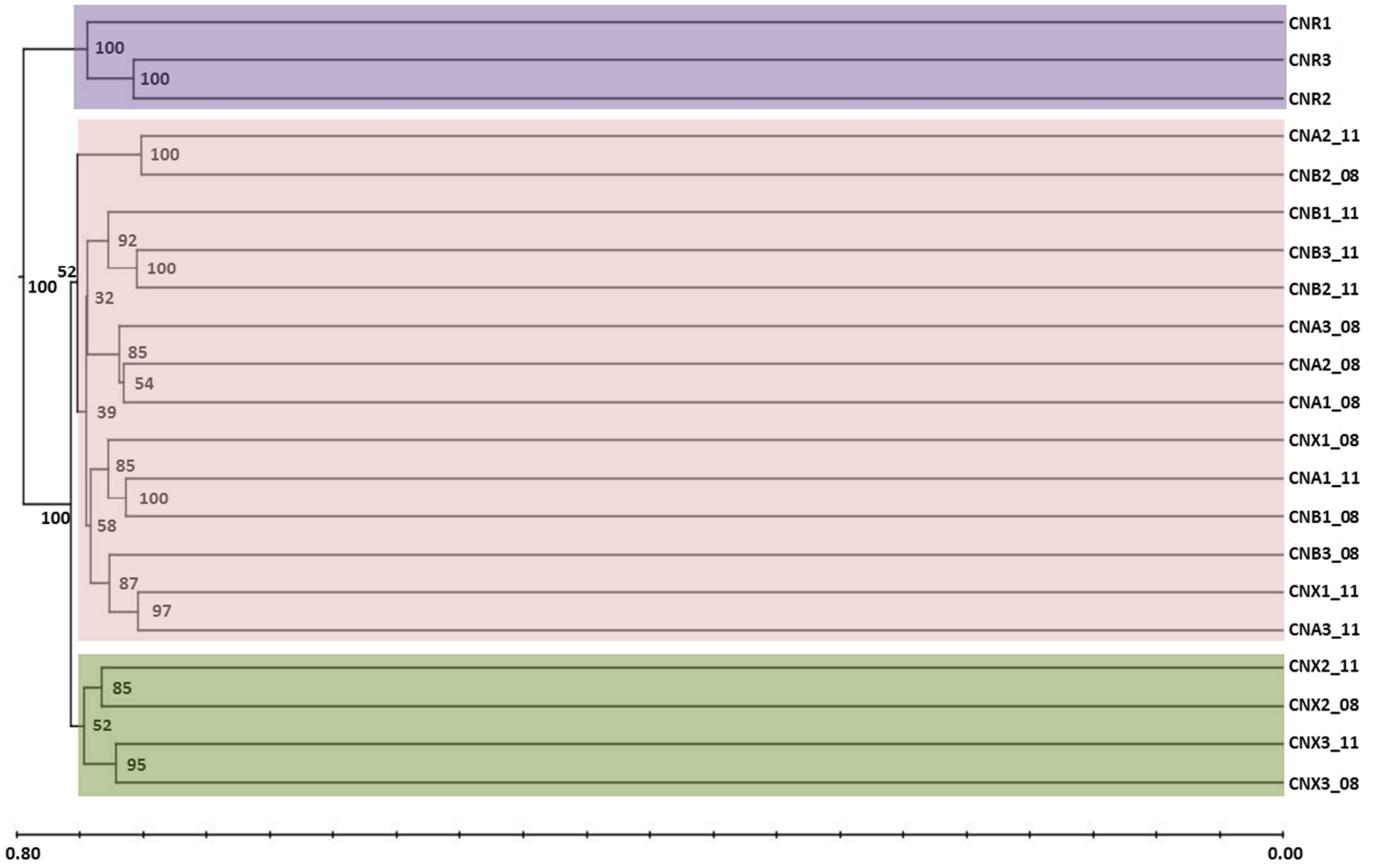


Figura R2-13. Agrupamiento jerárquico aglomerativo usando la matriz de distancias Unifrac y el algoritmo UPGMA teniendo en cuenta solo Presencia/Ausencia de secuencias de cada muestra en cada nodo del árbol filogenético. Para las secuencias normalizadas procedentes de amplicones 454 de la zona V6-V8 del gen 16S *rRNA*. Análisis realizado con las distancias filogenéticas encontradas entre las secuencias de cada muestra. Los valores de los nodos del árbol se obtuvieron por el método de Jackknife con 1.000 iteraciones de 8.000 secuencias seleccionadas al azar de cada muestra.

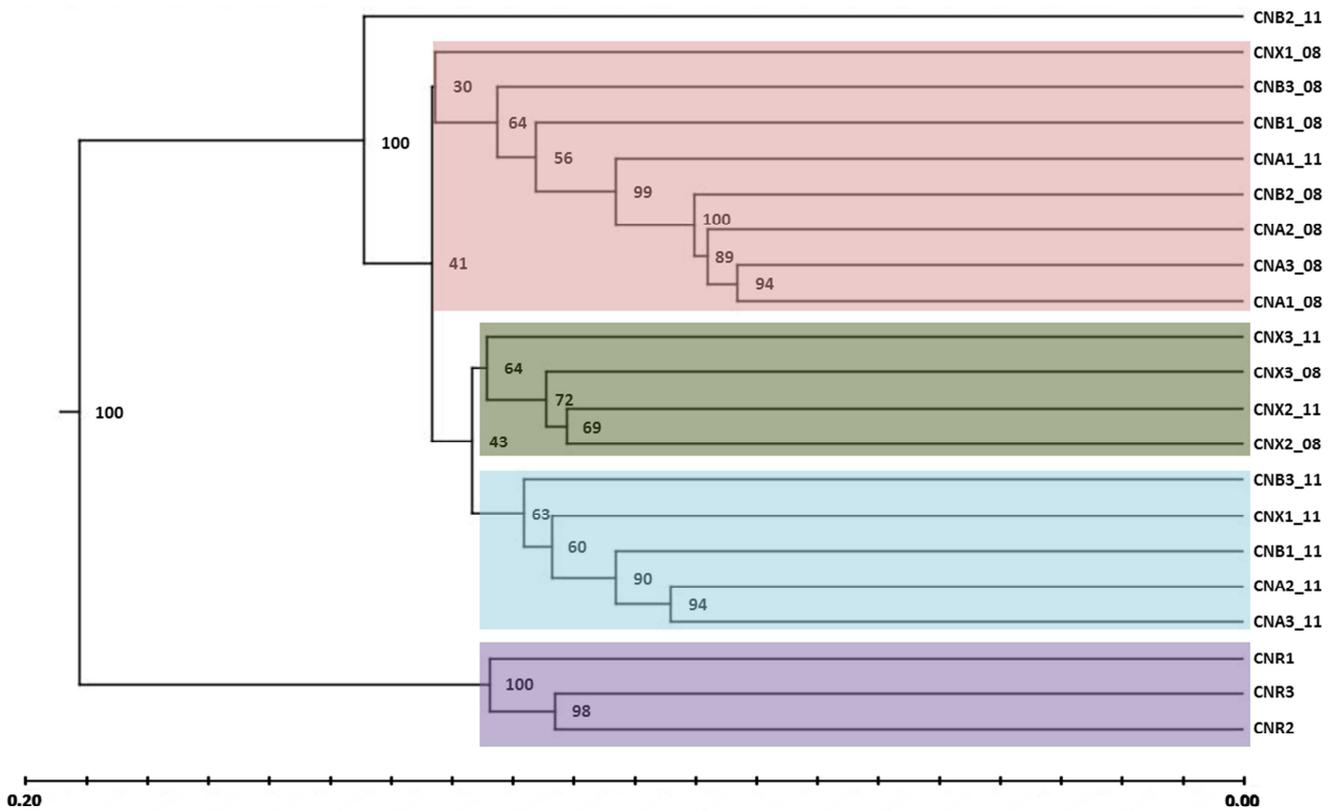


Figura R2-14. Agrupamiento jerárquico aglomerativo usando la matriz de distancias Unifrac y el algoritmo UPGMA teniendo en cuenta la abundancia de secuencias de cada muestra en cada nodo del árbol filogenético. Para las secuencias normalizadas procedentes de amplicones 454 de la zona V6-V8 del gen 16S *rRNA*. Análisis realizado con las distancias filogenéticas encontradas entre las secuencias de cada muestra. Los valores de los nodos del árbol se obtuvieron por el método de Jackknife con 1.000 iteraciones de 8.000 secuencias seleccionadas al azar de cada muestra.

Al comparar las distintas muestras, teniendo en cuenta las secuencias normalizadas que clasifican en RDP a nivel de familia, vemos que éstas se agrupan por su composición filogenética en 3 grupos, atendiendo a la PC1 (primera coordenada principal) que explica el 42,4 % de la varianza entre muestras. En el primer grupo están mezcladas las muestras del robleal maduro (CNB y CNA) con valores negativos en PC1; el segundo grupo, próximo al primero pero con valores positivos en PC1, está compuesto por las muestras del robleal en expansión (CNX); y el tercer grupo está formado por las muestras de suelo no rizosférico (CNR), con los valores positivos más altos en PC1 (Fig. R2-15). La coordenada principal 2 (PC2) explica el 25,8 % de la varianza y divide las muestras en dos grupos. El primero de ellos con valores negativos, siendo la muestra con el valor negativo más alto CNX3_11 y positivos cercanos a cero en el que están las muestras rizosféricas y CNR3 no rizosférica. El segundo grupo lo

Resultados Capítulo 2

forman las otras dos muestras no rizosféricas CNR1 y CNR2. Quedando la muestra del robledal de la zona baja CNB3_08 desagrupada en PC2 con el valor positivo más alto (Figs. R2-15 y R2-16).

En la figura R2-16 se observan las familias que más contribuyen a la varianza de las dos primeras coordenadas principales, que supone más del 68 % de la varianza total entre todas las muestras estudiadas. Siendo las familias Bradyrhizobiaceae, Nocardioideaceae, Sphingomonadaceae y Mycobacteriaceae las que más peso tienen en los valores negativos de la PC1. Por el contrario, la familia Gemmatimonadaceae junto con las familias Geodermatophilaceae, Planctomycetaceae y Conexibacteraceae son las que más peso tienen en los valores positivos de PC1. Las familias Chitinophagaceae, Flavobacteriaceae y Sphingomonadaceae son las que más contribuyen a los valores negativos de varianza de la PC2 y las familias Micromonosporaceae y Geodermatophilaceae las que más contribuyen a los valores positivos.

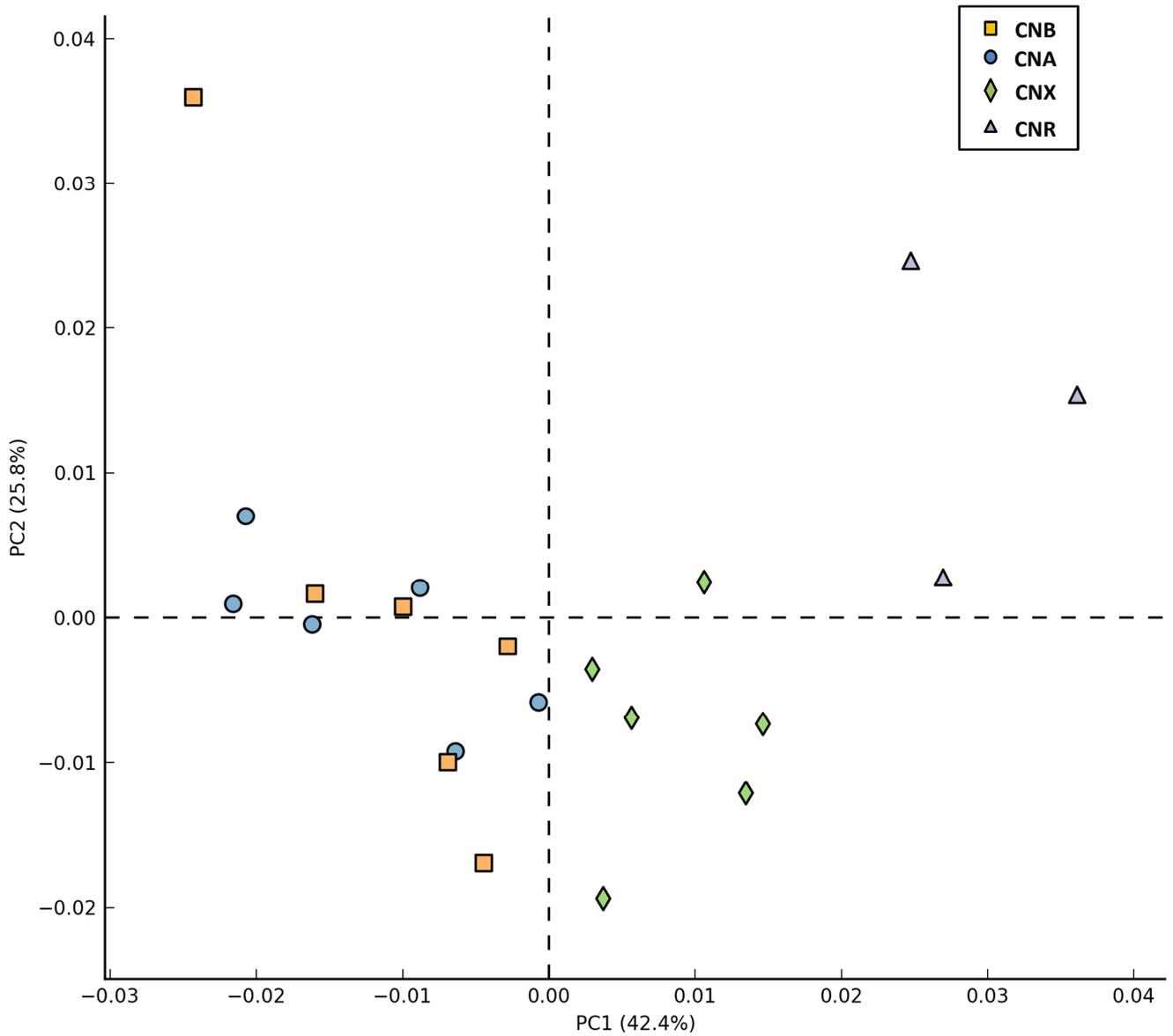


Figura R2-15. Análisis de Coordenadas Principales (PCoA). Realizado en el programa STAMP con las 21 muestras pertenecientes a los 4 sitios muestreados en ambos años de muestreo, teniendo en cuenta las secuencias clasificadas a nivel de familia.

Resultados Capítulo 2

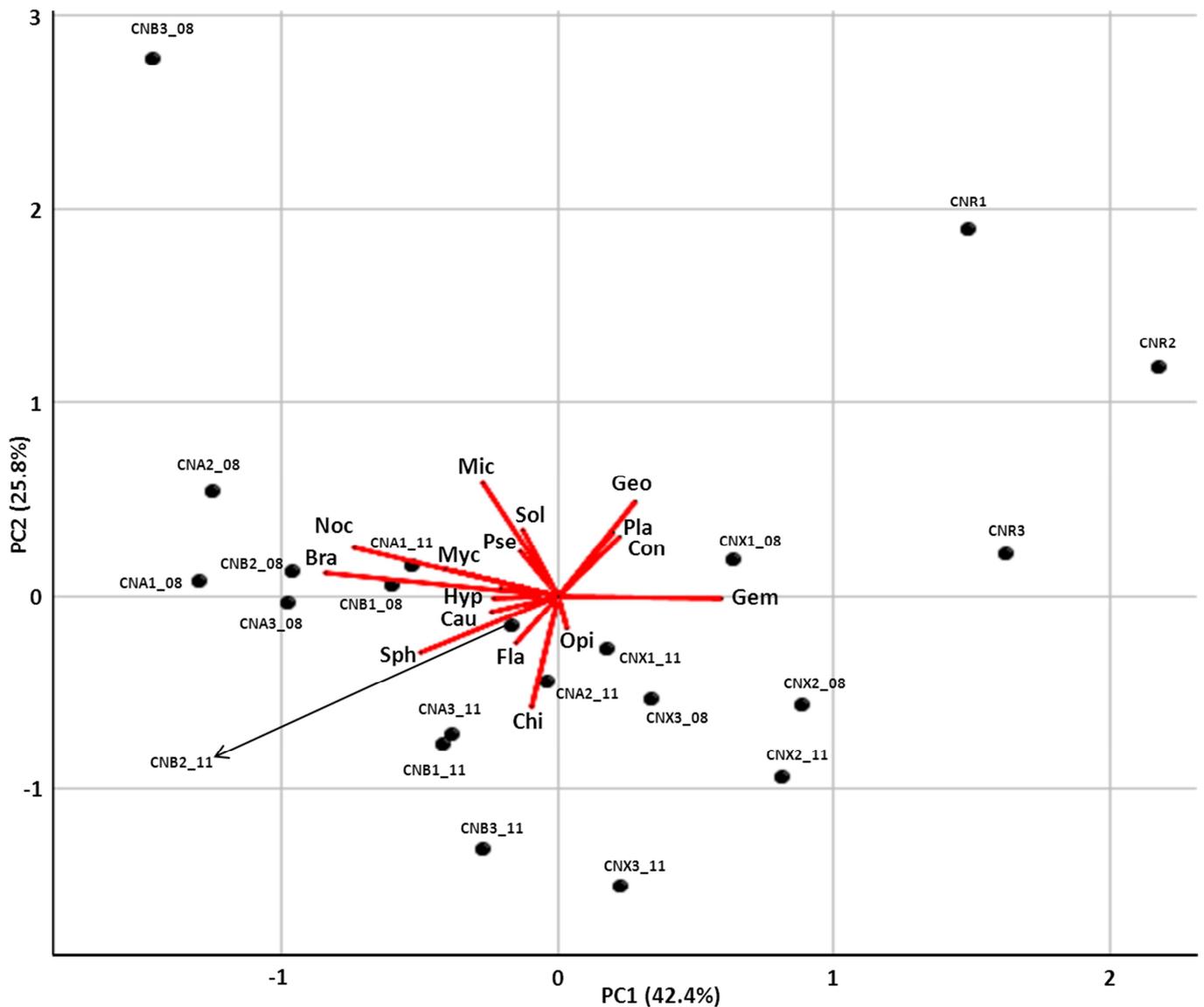


Figura R2-16. Análisis de Correspondencias (CA). Realizado en el programa Ginkgo con las 21 muestras pertenecientes a los 4 sitios muestreados en ambos años de muestreo, teniendo en cuenta las secuencias clasificadas a nivel de familia. Se muestran solo los vectores más relevantes para los 2 primeros componentes. Bra, Bradyrhizobiaceae; Noc, Nocardioideae; Myc, Mycobacteriaceae; Pse, Pseudonocardiaceae; Mic, Micromonosporaceae; Sol, Solirubrobacteraceae; Geo, Geodermatophilaceae; Pla, Planctomycetaceae; Con, Conexibacteraceae; Gem, Gemmatimonadaceae; Opi, Opitutaceae; Chi, Chitinophagaceae; Fla, Flavobacteriaceae; Sph, Sphingomonadaceae; Cau, Caulobacteraceae; Hyp, Hyphomicrobiaceae.

Ya que las familias Bradyrhizobiaceae, Nocardioideae y Mycobacteriaceae son las que tienen un mayor peso a la hora de marcar las diferencias en la rizosfera de los robles, se procedió a analizar en detalle los OTUs de las mismas en el gradiente altitudinal. Los resultados nos muestran que, pese a no encontrar diferencias significativas a nivel de familia, en el roblel maduro aparecen OTUs significativamente diferentes entre el sitio alto (CNA) y el sitio bajo (CNB) del roblel. En la familia Bradyrhizobiaceae obtenemos 12 OTUs pertenecientes al

género *Bradyrhizobium*, de los cuales 7 son significativamente más abundantes en CNA y los otros 5 son significativamente más abundantes en CNB. En la familia Nocardioideaceae obtenemos 9 OTUs pertenecientes al género *Nocardioides*, de los cuales 5 son significativamente más abundantes en CNA y los otros 4 son significativamente más abundantes en CNB. De esta misma familia se observan 4 OTUs del género *Kribbella*, de los cuales 1 es significativamente más abundante en CNA y los otros 3 son significativamente más abundantes en CNB. En la familia Mycobacteriaceae obtenemos 6 OTUs pertenecientes al género *Mycobacterium*, de los cuales 3 son significativamente más abundantes en CNA y los otros 3 son significativamente más abundantes en CNB (Tabla R2-5).

TablaR2-5. OTUs con abundancia relativa significativamente diferente entre los sitios alto (CNA) y bajo (CNB) del robledal maduro pertenecientes a las familias Bradyrhizobiaceae (*Bradyrhizobium*), Nocardioideaceae (*Nocardioides* y *Kribbella*) y Mycobacteriaceae (*Mycobacterium*). Recuento de secuencias obtenidas por pirosecuenciación de la zona V6-V8 de los amplicones del gen 16S *rRNA* en suma de las réplicas previamente normalizadas. p-valor obtenido mediante la prueba estadística utilizada por la herramienta Lib Compare de RDP. En azul se muestran los OTUs específicos de CNA y en rojo los específicos de CNB.

OTUs	CNA	CNB	p-valor	RDP
OTU_63162	177	0	2,61E-54	<i>Bradyrhizobium</i>
OTU_83528	109	0	7,70E-34	<i>Bradyrhizobium</i>
OTU_83527	93	0	5,05E-29	<i>Bradyrhizobium</i>
OTU_79435	41	0	2,27E-13	<i>Bradyrhizobium</i>
OTU_79418	40	0	4,55E-13	<i>Bradyrhizobium</i>
OTU_79456	33	0	5,82E-11	<i>Bradyrhizobium</i>
OTU_83541	31	0	2,33E-10	<i>Bradyrhizobium</i>
OTU_29448	5	60	1,12E-13	<i>Bradyrhizobium</i>
OTU_12406	4	34	1,34E-07	<i>Bradyrhizobium</i>
OTU_10661	2	118	2,69E-33	<i>Bradyrhizobium</i>
OTU_25417	0	80	4,14E-25	<i>Bradyrhizobium</i>
OTU_25462	0	55	1,39E-17	<i>Bradyrhizobium</i>
OTU_76578	65	0	1,36E-20	<i>Nocardioides</i>
OTU_61501	39	0	9,09E-13	<i>Nocardioides</i>
OTU_03942	34	0	2,91E-11	<i>Nocardioides</i>
OTU_80157	31	0	2,33E-10	<i>Nocardioides</i>
OTU_27524	26	8	5,28E-04	<i>Nocardioides</i>
OTU_27528	3	51	6,88E-13	<i>Nocardioides</i>
OTU_28848	0	31	2,33E-10	<i>Nocardioides</i>
OTU_24831	0	30	4,66E-10	<i>Nocardioides</i>
OTU_27548	0	30	4,66E-10	<i>Nocardioides</i>
OTU_78022	45	0	1,42E-14	<i>Kribbella</i>
OTU_13077	1	47	8,53E-14	<i>Kribbella</i>
OTU_07379	1	34	5,09E-10	<i>Kribbella</i>

Resultados Capítulo 2

OTU_24809	0	39	9,09E-13	<i>Kribbella</i>
OTU_26128	42	9	6,76E-07	<i>Mycobacterium</i>
OTU_73763	34	0	2,91E-11	<i>Mycobacterium</i>
OTU_73775	30	0	4,66E-10	<i>Mycobacterium</i>
OTU_25062	4	41	2,12E-09	<i>Mycobacterium</i>
OTU_25059	0	36	7,28E-12	<i>Mycobacterium</i>
OTU_25058	0	35	1,46E-11	<i>Mycobacterium</i>

Se han comparado los 4 sitios muestreados (CNB, CNA, CNX y CNR) teniendo en cuenta la abundancia relativa de las familias dominantes (Figs. R2-5 y R2-6) y los parámetros ambientales obtenidos de dichos sitios (Tablas R1-1, R1-2, R1-4 y R1-5). En el programa CANOCO se introdujeron como variables ambientales suplementarias el año de muestreo, la presencia o ausencia de rizosfera de roble y el sitio del que proceden según el gradiente altitudinal evaluado (bajo, alto y expansión). En la figura R2-17 se observa una distribución parecida a la obtenida en el PCoA (Fig. R2-15) con una clara separación entre robledal maduro (CNB y CNA), robledal en expansión (CNX) y sitio no rizosférico (CNR) en el Eje 1, el cual explica el 69,1 % de la varianza observada y muestra una correlación positiva entre el pH y el estado de maduración del robledal. Los sitios se separan en el Eje 2, el cual explica el 19,9 % de la varianza, principalmente por el año en que se muestrearon. También afectan en la ordenación la precipitación acumulada, el contenido en nitrógeno y la textura del suelo pero no son estadísticamente significativos. Se realizó el test de Monte Carlo para evaluar si los factores ambientales introducidos tienen relación con la ordenación obtenida de los sitios, debido a la abundancia relativa de las familias dominantes. Dicho test fue estadísticamente significativo (999 permutaciones, p-valor = 0,005) indicando que, al menos, uno de los factores ambientales está correlacionado con la ordenación encontrada. En este análisis sólo el pH tiene una correlación estadísticamente significativa (p-valor = 0,004).

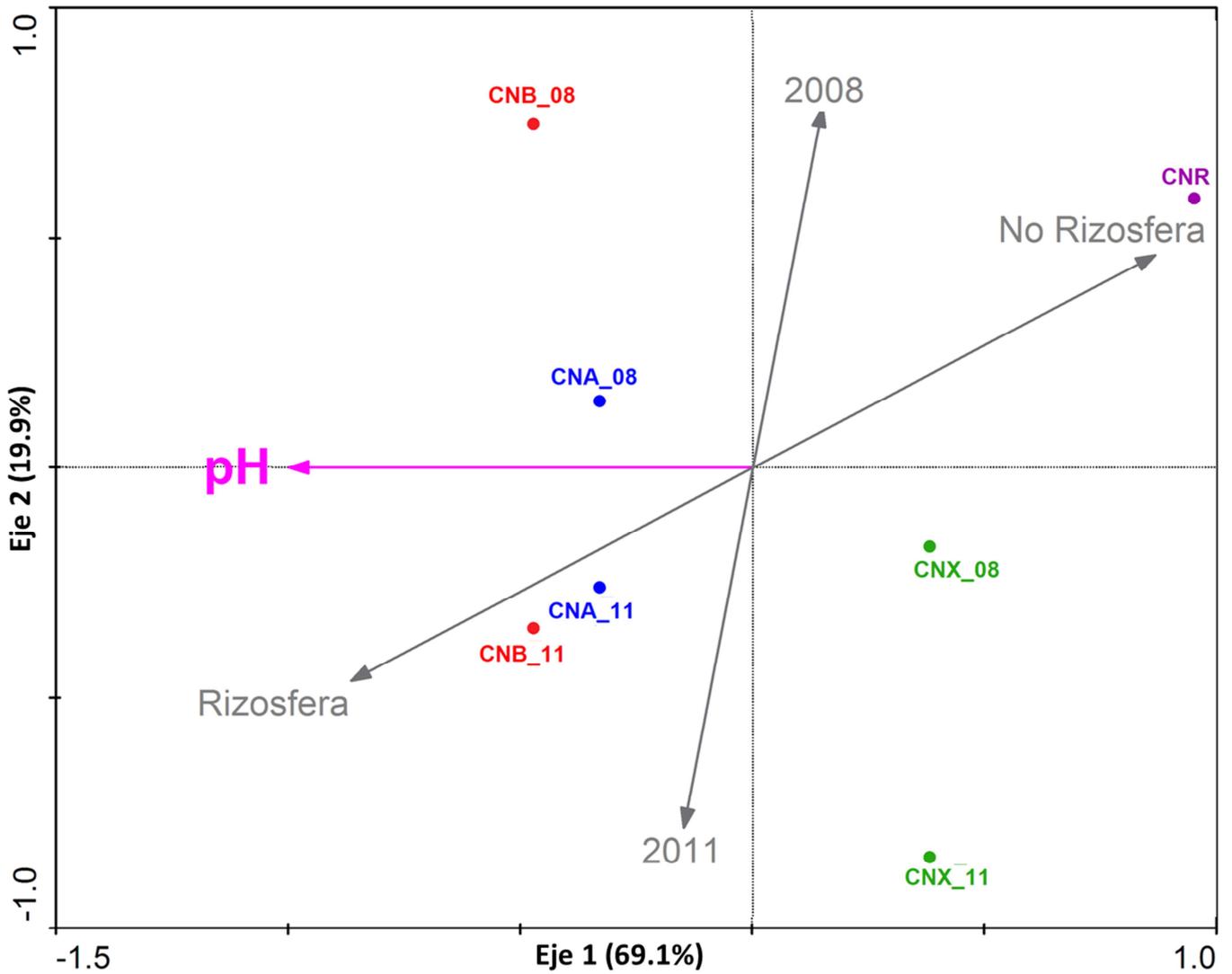


Figura R2-17. Análisis de Correspondencias Canónicas (CCA). Realizado en el programa CANOCO con los 4 sitios muestreados en ambos años de muestreo, teniendo en cuenta las secuencias promedio de cada sitio clasificadas a nivel de familia. En gris se muestran los vectores de las variables ambientales utilizadas como suplementarias. En fucsia se muestran solo los vectores de las variables ambientales que afectan significativamente a la ordenación de los sitios.

Capítulo 3: Influencia de la especie arbórea en la estructura y composición de las comunidades microbianas asociadas a la rizosfera de quercíneas

Se determinó la influencia de la especie arbórea en la estructura y composición de las comunidades microbianas analizando el DNA medioambiental extraído de la rizosfera de los robles procedentes del robledal maduro de la ladera Sur de Sierra Nevada (CNB y CNA) y del robledal muestreado en la ladera Norte (CTR). Estas comunidades microbianas se compararon con las de la rizosfera de las encinas procedentes del encinar no afectado por el incendio en la ladera Sur (LJN) y el encinar de la ladera Norte (CTE). Todos ellos fueron muestreados en la primavera de los años 2008 y 2011.

1 Clasificación taxonómica de las secuencias del gen 16S *rRNA*

En el estudio clásico de genotecas, se analizaron los 3 sitios de roble (CNB, CNA y CTR) y los 2 de encina (LJN y CTE), pero sólo en el muestreo de la primavera de 2008. De ellos se obtuvieron al menos 75 clones por sitio con un amplicón casi completo del gen 16S *rRNA* (≈ 1460 bp), de dichos clones se consiguieron un total de 305 secuencias de buena calidad y no quiméricas, 61 secuencias de cada sitio. Tras analizar nuestras secuencias frente a la base de datos de Bacteria y Archaea del RDP vemos que, a nivel de *phylum*, conseguimos clasificar el 91,8 % de las secuencias en CNB, CNA y CTE y el 95,1 % en CTR y LJN. El *phylum* Proteobacteria es el mayoritario en 4 de los 5 sitios; siendo el segundo *phylum* más abundante en el encinar de la ladera Norte (CTE). Las proporciones relativas son del 36,1 % en CNB, el 27,9 % en CNA y el 41,0 % en CTR en el caso de los robles y del 34,4 % en LJN y el 21,3 % en CTE para encinas. El segundo *phylum* más abundante es Actinobacteria con un 9,8 % en CNB, un 19,7 % en CNA y un 13,1 % en CTR en el caso de robles y un 24,6 % en LJN y un 18,0 % en CTE para encinas. El *phylum* Acidobacteria se muestra en tercera posición debido a ser el mayoritario para CTE con un 23,0 % pero con valores inferiores al 15 % en el resto de sitios. El cuarto *phylum* más abundante es Bacteroidetes siendo el segundo para CNB con un 16,4 % y para CTR con un 13,1 %, pero sin llegar al 9 % en los otros 3 sitios. Los demás *phyla* son menores al 10 % para todos los sitios (Fig. R3-1). A nivel de familia conseguimos clasificar el 52,5 % de las secuencias en CNB, el 42,6 % en CNA y el 54,1 % en CTR para robles y el 57,4 % en LJN y el 39,3 % en CTE para encinas. La familia más abundante es Chitinophagaceae, perteneciente al *phylum* Bacteroidetes, que supone el 13,1 % en CNB, el 8,2 % en CTR y el 4,9 % en CNA, por tanto mayoritaria en todos los robles, pero supone el 4,9 % en LJN y solo el 1,6 % en CTE para las encinas, es decir por debajo del tercer puesto. La familia Gemmatimonadaceae ocupa la segunda posición debido a su abundancia en la rizosfera de encinas con un 6,6 % en la ladera Sur (LJN) y un 8,2 % en la Norte (CTE), siendo además la familia más abundante para estas dos muestras; sin embargo, solo aparece una secuencia (CNB y CTR) o ninguna (CNA) de esta familia en la rizosfera de robles, siendo por tanto posible candidata como bioindicador diferencial. La tercera familia más abundante es Bradyrhizobiaceae debido a ser la más abundante en LJN (igual de abundante que Gemmatimonadaceae) con un 6,6 %, suponiendo un 3,3 % para los otros 4 sitios. La familia Caulobacteraceae (α -Proteobacteria) en sexta posición es más abundante en robledal, con un 3,3 % en los 3 sitios (CNB, CNA y CTR), que en encinar, con una secuencia en LJN (1,6 %) y ninguna en CTE (Fig. R3-2). Aunque se observan diferencias entre robles y encinas y diferencias propias de cada sitio muestreado, éstas no son

estadísticamente significativas, según la herramienta STAMP, debido al pequeño tamaño de muestra analizado.

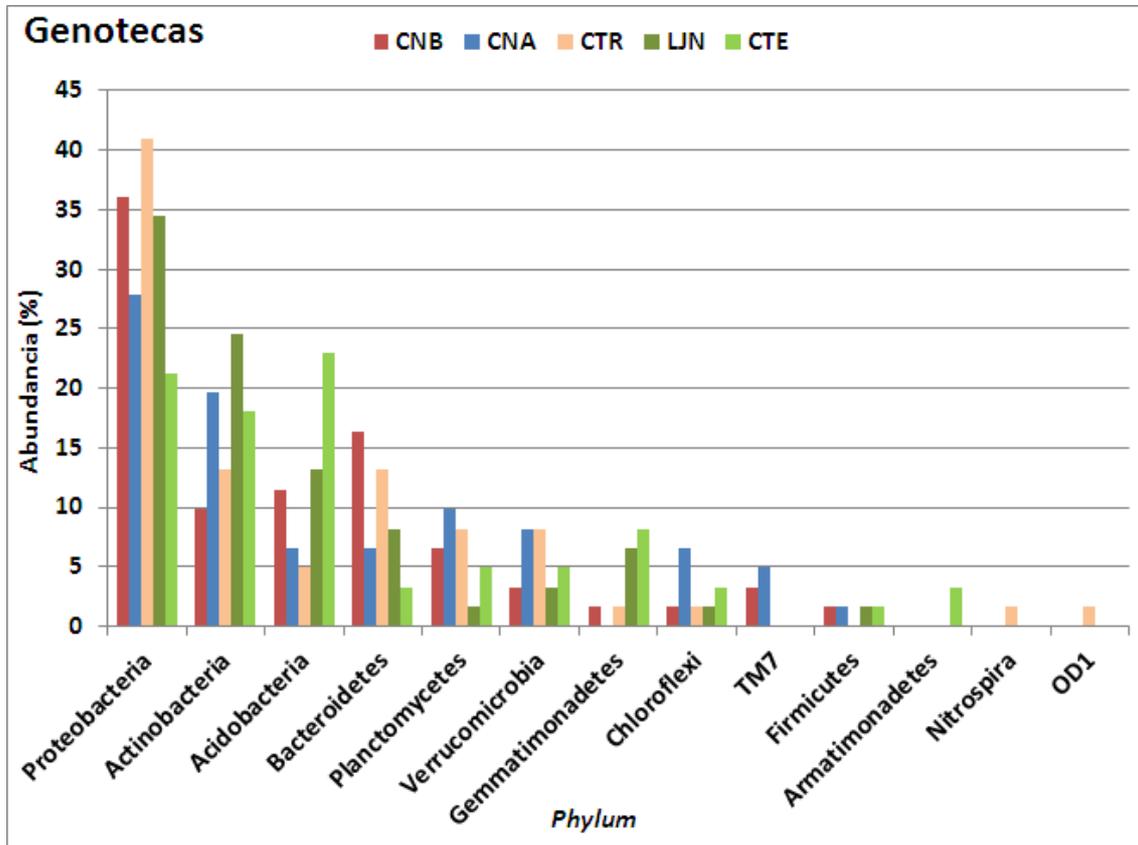


Figura R3-1. Distribución de todos los *phyla* de las genotecas clásicas de 16S *rRNA*.

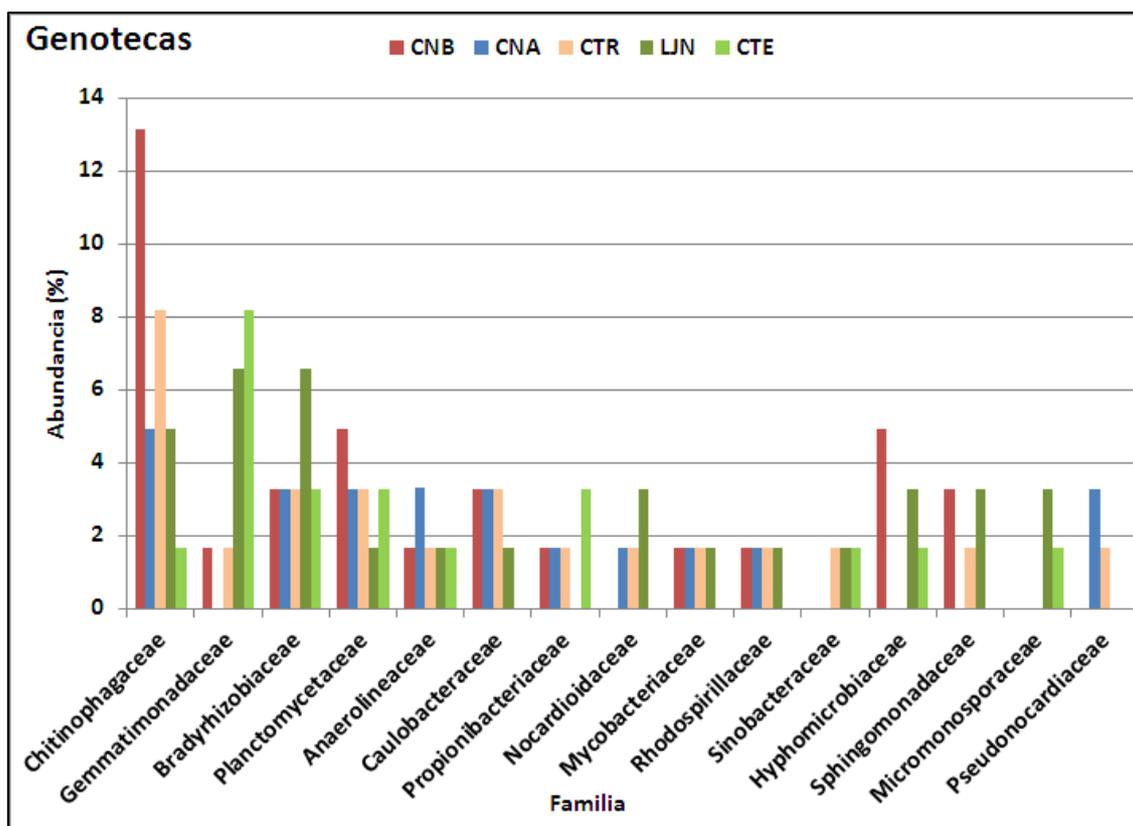


Figura R3-2. Distribución de las familias más abundantes de las genotecas clásicas de 16S *rRNA*. Solo se muestran aquellas familias que suponen al menos el 4,9 % de abundancia relativa en la suma de los 5 sitios.

En el estudio de los metagenomas mediante pirosecuenciación se analizaron solo los tres sitios de la ladera Sur de Sierra Nevada, es decir los dos sitios del roblel maduro de Cañar (CNB y CNA) y el encinar no quemado de Lanjarón (LNJ), todos ellos muestreados en 2008 al igual que en el caso de las genotecas clásicas. De ellos se obtuvieron 247.846 lecturas en CNB, 193.004 lecturas en CNA y 520.430 lecturas en LNJ que tras filtrarlas, resultaron en 222.689 secuencias con longitud media de 371 bp, 167.219 secuencias con longitud media de 388 bp y 412.302 secuencias con longitud media de 387 bp, respectivamente, únicas y de calidad. Dichas secuencias se han anotado con la base de datos M5NR de MG-RAST dando como resultado 420 secuencias en CNB, 277 secuencias en CNA y 1.071 secuencias en LNJ homólogas a genes que transcriben algún ARN ribosómico (*rRNA*). Al enfrentar estas secuencias ribosómicas a la base de datos del gen 16S *rRNA* del RDP se clasificaron 66, 45 y 172 secuencias, respectivamente (Tabla R3-1).

Al analizar la clasificación de las secuencias homólogas al gen 16S *rRNA* de CNB, CNA y LNJ procedentes de pirosecuenciación vemos que, la distribución de las familias es muy diferente entre ellas. Hemos conseguido clasificar el 78,8 % de las secuencias en CNB, el 60,0 %

en CNA y el 81,3 % en LNJ, siendo la familia Sphingomonadaceae la mayoritaria por suponer el 14,4 % en suma de los 3 sitios, 9,1 % en CNB pero solo el 2,2 % en CNA y 3,1 % en LNJ. Flavobacteriaceae, que pertenece al *phylum* Bacteroidetes, es la segunda familia más abundante a la que corresponde a un 11,9 % en suma, supone un 1,5 % en CNB, un 6,7 % en CNA y un 3,8 % en LNJ. La tercera y quinta familias más abundantes son Lachnospiraceae del *phylum* Firmicutes y Planctomycetaceae que suponen ambas un 3 % en CNB, un 4,4 % en CNA, mientras que en LNJ es la menos representada con un 2,5 % en la primera familia y un 1,9 % en la segunda. La familia Sphingobacteriaceae (*phylum* Bacteroidetes) ocupa la cuarta posición con un 9,7 % en suma, pero es la más abundante en el encinar (LNJ) con un 7,5 %, y solo se encontró una secuencia en el robledal, perteneciente al sitio alto (CNA), suponiendo un 2,2 % para dicho sitio. Algo parecido ocurre en la penúltima familia (Acidobacteriaceae) ya que solo se encontraron 7 secuencias procedentes del encinar, suponiendo un 4,4 % para dicho sitio (LNJ). Las familias Verrucomicrobiaceae, Streptomycetaceae y Burkholderiaceae muestran el mismo perfil entre ellas, siendo más abundantes en el sitio bajo del robledal y menos abundantes en el encinar. Las familias Comamonadaceae y Polyangiaceae (Beta- y Delta-Proteobacteria, respectivamente) muestran un perfil similar a las familias Lachnospiraceae y Planctomycetaceae, siendo el sitio alto del robledal (CNA) el más abundante y el encinar (LNJ) el menos abundante (Fig. R3-3). A pesar de encontrar diferencias entre ambos sitios, éstas no son estadísticamente significativas, según la herramienta STAMP, debido al pequeño tamaño de muestra analizado.

Tabla R3-1. Recuento de secuencias antes y después del procesamiento de los sitios estudiados con secuenciación masiva "shotgun".

Muestra	Método de secuenciación	Lecturas	Secuencias	rRNA	RDP
CNB	454	247.846	222.689	420	66
CNA	454	193.004	167.219	277	45
LNJ	454	520.430	412.302	1.071	172
CNB	Illumina	356.963.222	173.277.220	26.187.772	71.178
CNA	Illumina	365.590.198	171.236.955	26.947.483	65.286
LNJ	Illumina	374.512.840	180.589.099	27.515.446	65.828

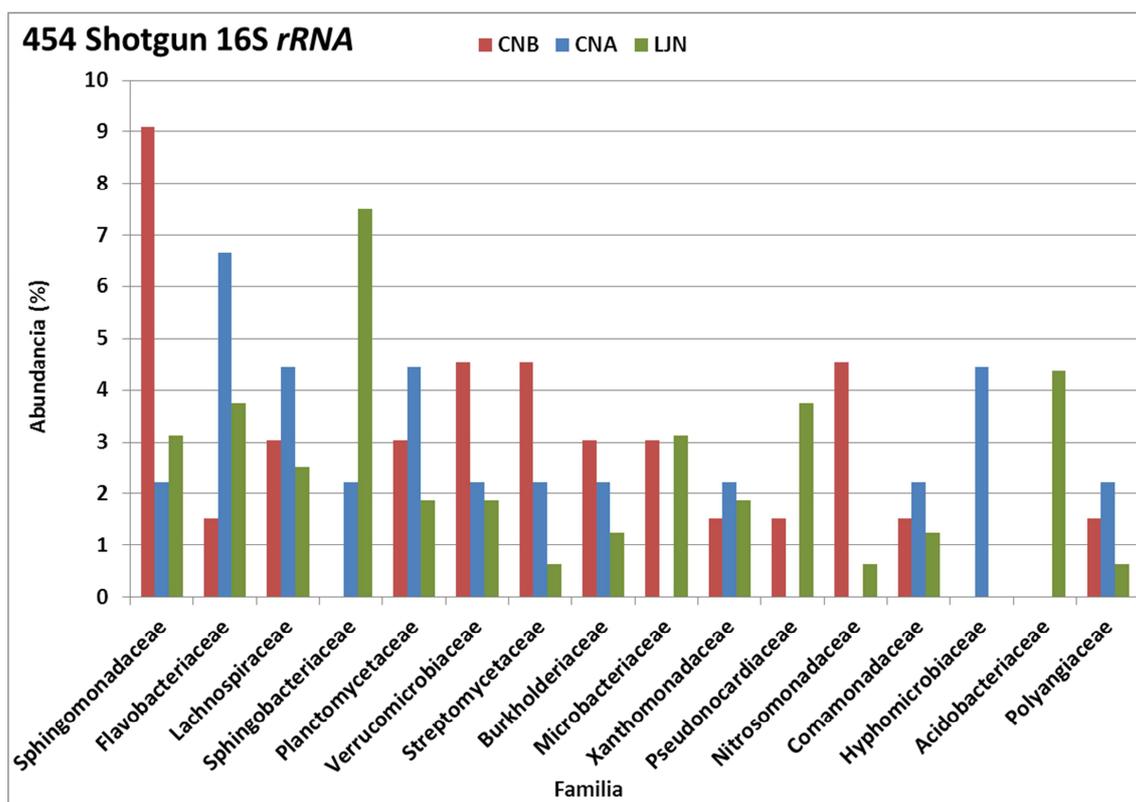


Figura R3-3. Distribución de las familias más abundantes de las muestras pirosecuenciadas. Solo se muestran aquellas familias que suponen al menos el 4,4 % de abundancia relativa en la suma de los 3 sitios.

En el estudio de los metagenomas mediante secuenciación Illumina se analizaron los mismos sitios que en la técnica anterior (Tabla R3-1), robleal bajo (CNB), robleal alto (CNA) y encinar no quemado (LNJ); todos ellos de la ladera Sur de Sierra Nevada y muestreados en la primavera de 2008. De ellos se obtuvieron 356.963.222 lecturas en CNB, 365.590.198 lecturas en CNA y 374.512.840 lecturas en LNJ, todas ellas con una longitud de 101 bp. Tras filtrar y solapar las lecturas pareadas (Paired-Ends reads), resultaron en 173.277.220 secuencias en CNB, 171.236.955 secuencias en CNA y 180.589.099 secuencias en LNJ (Tabla R3-1) con un rango de tamaños entre 50 y 190 bp y un tamaño medio de 170 bp en todos ellos. Tras rastrear secuencias homólogas a genes *rRNA* se obtuvieron 26.187.772 secuencias en la muestra de CNB, 26.947.483 secuencias en CNA y 27.515.446 secuencias en LNJ. Se consiguieron clasificar, al menos a nivel de dominio, 1.389.054 secuencias en CNB, 1.293.856 secuencias en CNA y 1.328.764 secuencias en LNJ. De éstas, se obtuvieron 71.178 secuencias en CNB, 65.286 secuencias en CNA y 65.828 secuencias en LNJ pertenecientes al gen 16S *rRNA* que se clasificaron en RDP (Tabla R3-1).

Al analizar la clasificación de las secuencias homólogas al gen 16S *rRNA* de CNB, CNA y LNJ procedentes de la secuenciación Illumina conseguimos clasificar el 65,7 % de las secuencias en CNB, el 59,6 % en CNA y el 65,8 % en CNX, distribuidas en 238 familias, de las que 53 muestran diferencias significativas entre algunos de los sitios comparados dos a dos. De estas familias, 24 tienen una abundancia relativa de, al menos, el 1 % en alguno de los sitios muestreados y se han marcado (en la Fig. R3-4) todas aquellas que presentan diferencias biológicamente relevantes entre robles y encinas. Las familias mayoritarias en el encinar (LJN) fueron Solirubrobacteraceae del *phylum* Actinobacteria con un 5,7 % en LJN, un 5,1 % en CNB y un 4,7 % en CNA; Bradyrhizobiaceae del *phylum* Proteobacteria con un 4,2 % en LJN y un 3,3 % en CNB y CNA; Solibacteraceae del *phylum* Acidobacteria con un 2,0 % en LJN y un 1,5 % en CNB y CNA; y la familia Oxalobacteraceae del *phylum* Proteobacteria con un 1,1 % en LJN, un 0,6 % en CNB y un 0,5 % en CNA. Estas 4 familias no presentaron diferencias significativas entre los dos sitios del robledal maduro. También fueron mayoritarias en LJN las familias Pseudonocardiaceae (*phylum* Actinobacteria) con un 2,1 % en LJN, un 1,5 % en CNB y un 0,8 % en CNA; y Acidobacteriaceae con un 2,4 % en LJN, un 1,0 % en CNB y un 0,6 % en CNA, ambas familias fueron también significativamente menos abundantes en el sitio alto del robledal maduro (CNA) respecto del sitio bajo (CNB) (Fig. R3-4). La mayor proporción de Acidobacterias en el encinar de la ladera Sur (LJN) también se reflejó en los datos obtenidos por genotecas clásicas a nivel de *phylum* (Fig. R3-1) y mediante pirosecuenciación (Fig. R3-3). También se observó una mayor proporción de la familia Bradyrhizobiaceae en encinar, en los datos obtenidos por genotecas clásicas (Fig. R3-2), pero no se detectó en los estudios de pirosecuenciación (en dicho estudio solo se obtuvo 1 secuencia de esta familia en cada sitio del robledal maduro y ninguna en el encinar). Las familias Flavobacteraceae (*phylum* Bacteroidetes) con un 3,1 % en CNB, un 3,9 % en CNA y un 2,5 % en LJN y Nitrospiraceae (*phylum* Nitrospirae) con un 1,3 % en CNB, un 2,0 % en CNA y un 1,0 % en LJN, muestran el mismo perfil, siendo significativamente menos abundantes en el encinar que en el robledal y además son significativamente más abundantes en el sitio alto del robledal maduro (CNA) respecto del sitio bajo (CNB). Solo 5 familias de las mostradas en la Figura R3-4 no mostraron diferencias significativas entre ninguno de los 3 sitios estudiados y son la familia Hyphomicrobiaceae [α -Proteobacteria característica del bosque maduro (Fig. R2-16)], la familia Comamonadaceae (β -Proteobacteria), la familia Mycobacteriaceae [Actinobacteria característica del bosque maduro (Fig. R2-16)], la familia Flexibacteraceae (Bacteroidetes) y la familia Microbacteriaceae del *phylum* Actinobacteria (Fig. R3-4).

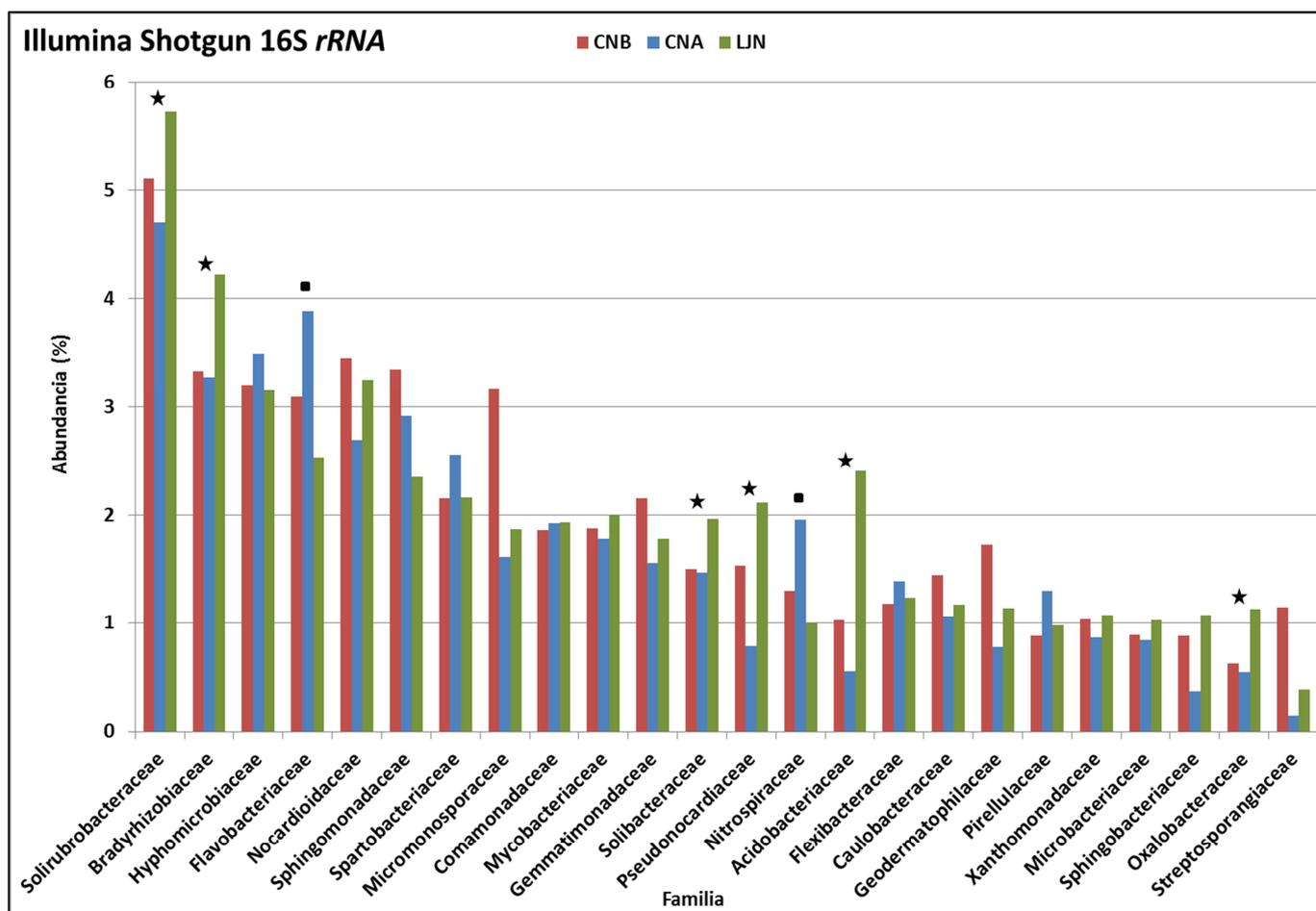


Figura R3-4. Distribución de las familias más abundantes de las muestras secuenciadas mediante tecnología Illumina. Solo se muestran aquellas familias que suponen más del 1 % de abundancia relativa en alguno de los tres sitios. Los símbolos marcan aquellas familias en las que se encontraron diferencias significativas biológicamente relevantes comparando los sitios dos a dos con el programa STAMP, tomando un p-valor corregido menor del 0,01. Siendo: ★ Familias significativamente más abundantes en LNJ. ■ Familias significativamente menos abundantes en LNJ.

En el estudio de las librerías NGS (amplicones) de la zona V6-V8 del gen 16S *rRNA* se analizaron las 3 parcelas (muestras) de cada uno de los 5 sitios, que comprenden todas las situaciones de robleal bien desarrollado CNB, CNA y CTR; y encinar sano LNJ y CTE de este análisis, tanto en el muestreo del año 2008 como en el muestreo de 2011. De ellos se obtuvieron 656.298 lecturas con una longitud media de 408 bp distribuidas uniformemente entre las 30 muestras secuenciadas, con valores entre 10.852 lecturas en LNJ2_11 y 26.082 lecturas en CNA3_08. Tras filtrar y descartar las lecturas con baja calidad, las posibles quimeras y las menores de 220 bp con el programa Pyrotagger, se obtuvieron 497.195 secuencias con una longitud media de 458 bp distribuidas uniformemente y con unos valores entre 8.493 secuencias en LNJ2_11 y 19.615 secuencias en CNA3_08 (Tabla R3-2). Los cebadores utilizados para obtener estos amplicones no solo amplifican la zona V6-V8 del gen 16S *rRNA* sino que,

también amplifican de forma menos específica regiones del gen 18S *rRNA* de Eucariotas. Para filtrar las secuencias procarióticas se alinearon las secuencias filtradas por Pyrotagger con la herramienta Aligner de RDP frente al gen 16S *rRNA* de *E. coli* J01695 y se eliminaron todas aquellas secuencias que no alinearon en la zona V6-V8, quedando un promedio de 14.001 secuencias por muestra, con un valor mínimo de 7.762 secuencias en LNJ2_11 y un valor máximo de 17.639 secuencias en CTR3_11. Para realizar análisis que requieren normalización del número de secuencias de cada muestra, se normalizaron todas las muestras a 7.762 secuencias (Tabla R3-2).

Resultados Capítulo 3

Tabla R3-2. Recuento de secuencias obtenidas por pirosecuenciación de la zona V6-V8 de los amplicones del gen 16S *rRNA*.

* Secuencias obtenidas tras el filtro realizado en el JGI con la herramienta Pyrotagger.

** Secuencias obtenidas tras el filtro realizado al alinear las secuencias filtradas en el JGI con la herramienta Aligner del RDP. Las secuencias filtradas con RDP se normalizaron con MOTHUR.

Muestra	Lecturas	Secuencias JGI*	Secuencias RDP**	Normalizadas
CNA1_08	20.179	15.027	11.781	7.762
CNA1_11	20.073	15.661	14.287	7.762
CNA2_08	23.223	16.841	13.392	7.762
CNA2_11	23.255	17.742	15.985	7.762
CNA3_08	26.082	19.615	16.499	7.762
CNA3_11	24.214	18.975	17.173	7.762
CNB1_08	21.228	15.667	12.158	7.762
CNB1_11	24.279	18.903	17.044	7.762
CNB2_08	24.109	16.774	12.498	7.762
CNB2_11	23.850	18.439	16.548	7.762
CNB3_08	22.537	16.694	13.002	7.762
CNB3_11	24.852	19.177	17.467	7.762
CTR1_08	21.809	16.233	12.773	7.762
CTR1_11	23.457	18.221	16.527	7.762
CTR2_08	19.544	14.208	11.244	7.762
CTR2_11	23.629	18.658	17.033	7.762
CTR3_08	20.746	15.492	12.412	7.762
CTR3_11	25.588	19.424	17.639	7.762
LJN1_08	20.737	15.362	11.517	7.762
LJN1_11	19.499	15.663	14.026	7.762
LJN2_08	19.143	13.984	10.802	7.762
LJN2_11	10.852	8.493	7.762	7.762
LJN3_08	16.799	12.210	9.241	7.762
LJN3_11	23.181	18.139	16.138	7.762
CTE1_08	23.334	17.259	13.774	7.762
CTE1_11	22.925	17.872	16.306	7.762
CTE2_08	22.868	17.663	13.388	7.762
CTE2_11	20.326	15.621	14.103	7.762
CTE3_08	25.674	19.156	14.977	7.762
CTE3_11	18.306	14.022	12.543	7.762

Al analizar la clasificación de las secuencias normalizadas homólogas al gen 16S *rRNA* obtenidas en el muestreo del año 2008 de los 5 sitios CNB, CNA, CTR, LJN y CTE conseguimos clasificar el 52,3 % de las secuencias en CNB, el 50,6 % en CNA, el 47,6 % en CTR, el 53,9 % en LJN y el 51,1 % en CTE, distribuidas en 153 familias. De estas familias, ninguna tiene diferencias estadísticamente significativas en la abundancia relativa promedio encontrada entre los sitios. En la Figura R3-5 se muestran las 15 familias más abundantes para los 5 sitios. Solo la familia

Bradyrhizobiaceae del *phylum* Proteobacteria mostró diferencias significativas en las medias de abundancia relativa en la comparación ladera Sur (CNB, CNA y LNJ) frente a la ladera Norte (CTR y CTE) con un p-valor corregido de 0,014, siendo más abundante en el Sur con una abundancia relativa promedio de los 3 sitios del 3,95 % y menos abundante en la Norte con un 2,15 %. Tampoco aparecieron diferencias significativas tras corregir los p-valores para múltiples tests (Storey FDR) al comparar las muestras de robledal (CNB, CNA y CTR) frente a las de encinar (LJN y CTE). Al contrario que en los estudios anteriores, las familias pertenecientes al *phylum* Acidobacteria (familias de las clases Gp6 y Gp4), en este análisis, son menos abundantes en las muestras del encinar de la ladera Sur (LJN), aunque no es significativo estadísticamente. La familia perteneciente a la clase Gp1 es más abundante en LJN debido a la réplica LJN1_08, pero las otras dos réplicas muestran unos valores de abundancia similares al resto de muestras, por ello, la desviación estándar es tan grande en LJN para esta familia (Fig. R3-5).

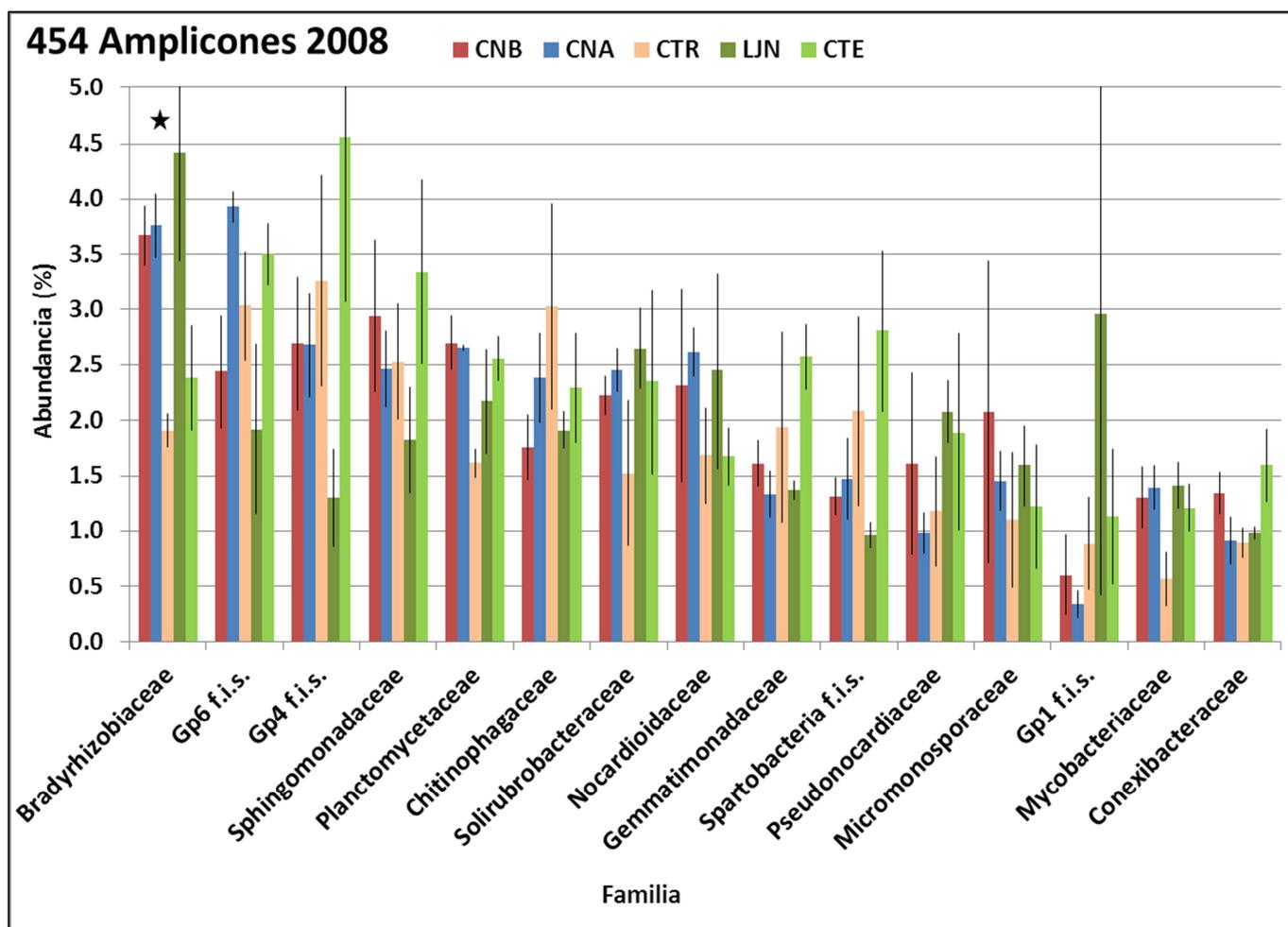


Figura R3-5. Distribución de las familias más abundantes de las muestras tomadas en el año 2008 y secuenciadas mediante pirosecuenciación de amplicones de la zona V6-V8 del gen 16S *rRNA*. Sólo se muestran aquellas familias que suponen al menos el 5,7 % de abundancia relativa en suma de los cinco sitios. Las barras de error se corresponden con la desviación estándar del valor medio de la abundancia relativa de cada familia para cada sitio, calculada a partir de las 3 muestras de cada sitio. La familia marcada con una estrella mostró diferencias estadísticamente significativas entre las medias de los sitios de la ladera Sur frente a los de la ladera Norte, analizados mediante ANOVA y con el factor de corrección para tests múltiples Storey FDR. f.i.s. = Familias sin nomenclatura definitiva en la clasificación taxonómica.

Al analizar la clasificación de las secuencias normalizadas homólogas al gen 16S *rRNA* obtenidas en el muestreo del año 2011 de los sitios CNB, CNA, CTR, LNJ y CTE conseguimos clasificar el 48,8 % de las secuencias en CNB, el 47,4 % en CNA, el 47,3 % en CTR, el 48,0 % en LNJ y el 48,2 % en CTE, distribuidas en 166 familias. De éstas, las únicas dos familias con diferencias estadísticamente significativas en la abundancia relativa promedio encontrada, al menos en uno de los sitios son la familia Nitrospiraceae del *phylum* Nitrospira y la familia Burkholderiaceae del *phylum* Proteobacteria. Estas familias no se consideraron mayoritarias en las comunidades microbianas de los sitios evaluados. La familia Nitrospiraceae supone un 0,05 % en CNA, un 0,1 % en CNB y LNJ, un 0,2 % en CTE y un 0,5 % en CTR, suponiendo 45

secuencias en la muestra más abundante de CTR. La familia Burkholderiaceae supone un 0,03 % en CNA, un 0,1 % en CTR y CTE, un 0,2 % en CNB y un 0,4 % en LNJ, suponiendo 34 secuencias en la muestra más abundante de LNJ. En la Figura R3-6 se muestran las 14 familias más abundantes para los 5 sitios. Ninguna familia tiene diferencias significativas en la abundancia relativa promedio encontrada entre los sitios. La distribución de las distintas familias entre ambos años de muestreo (2008 y 2011), resultó en perfiles muy similares en las familias mayoritarias (Figuras R3-5 y R3-6), apareciendo diferencias significativas solo en la familia Bradyrhizobiaceae (p-valor corregido de 0,024). Esta familia disminuyó significativamente en los 5 sitios muestreados, desde un promedio del 3,2 % en el año 2008 a un promedio del 2,1 % en 2011. También se observó que al comparar los 5 sitios en ambos años de muestreo, sólo una familia mostró diferencias estadísticamente significativas entre el robleal y el encinar, la familia de la clase Gp22 del *phylum* Acidobacteria. Esta familia con una abundancia promedio del 0,014 % en encinar y un 0,052 % en robleal solo fue significativamente más abundante en el robleal al tener en cuenta las muestras de ambos años (2008 y 2011). Dicha familia de la clase Gp22 se consideró minoritaria (y por ello no aparece en las Figs. R3-5 y R3-6) ya que resultó en 86 secuencias asignadas a la misma en suma de las 232.860 secuencias de las 30 muestras analizadas, suponiendo un 0,037 % del total. La familia de Acidobacteria perteneciente a la clase Gp1 es más abundante en las muestras de encinar (LJN y CTE) siendo esta diferencia más pronunciada en la ladera Sur (LJN), al igual que ocurría en las muestras de 2008 (Fig. R3-5). En el muestreo de 2011 (Fig. R3-6) la desviación estándar no es tan grande como en el muestreo de 2008, pero es suficiente para considerar que los valores promedio de la abundancia relativa de esta familia en los 5 sitios no es estadísticamente diferente.

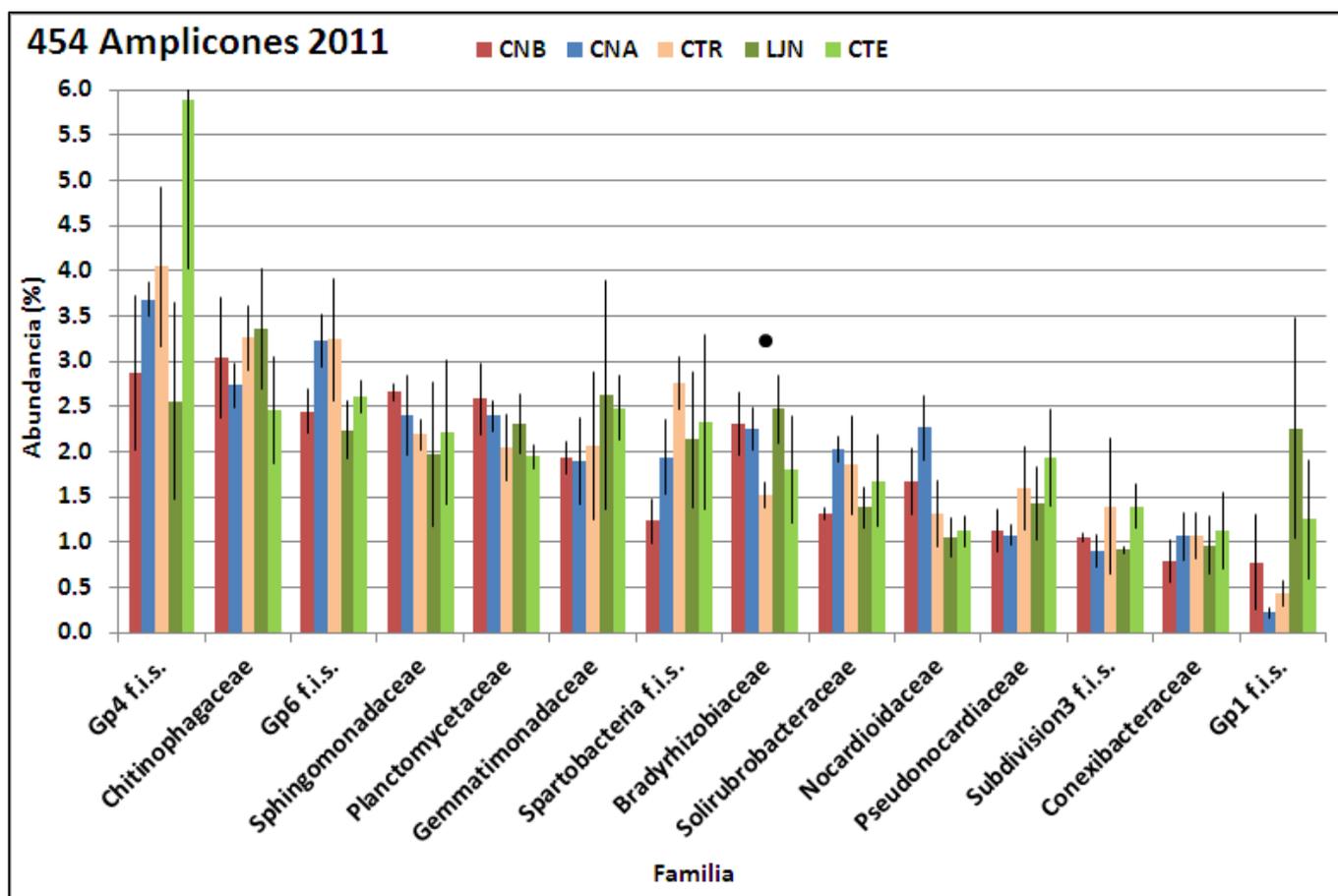


Figura R3-6. Distribución de las familias más abundantes de las muestras tomadas en el año 2011 y secuenciadas mediante pirosecuenciación de amplicones de la zona V6-V8 del gen 16S *rRNA*. Sólo se muestran aquellas familias que suponen al menos el 4,9 % de abundancia relativa en suma de los cinco sitios. Las barras de error se corresponden con la desviación estándar del valor medio de la abundancia relativa de cada familia para cada sitio, calculada a partir de las 3 muestras de cada sitio. La familia marcada con un círculo mostró diferencias estadísticamente significativas entre las medias de los 5 sitios comparando los dos años de muestreo, analizados mediante ANOVA y con el factor de corrección para tests múltiples Storey FDR. f.i.s. = Familias sin nomenclatura definitiva en la clasificación taxonómica.

2 Estudio de la diversidad α mediante el empleo del gen 16S *rRNA*

Para analizar la diversidad α , entendida como la riqueza y equidad observadas de cada muestra y estimar la riqueza del suelo de origen de cada una de ellas, se agruparon las secuencias del gen 16S *rRNA* de cada muestra en OTUs con una distancia génica máxima del 3 %, considerando que a dicha distancia cada OTU corresponde con una especie distinta.

En las genotecas se observan distintos resultados de diversidad α entre robles y encinas muestreados en 2008. La riqueza observada fue menor en las encinas, con 56 y 58 OTUs en LNJ y CTE, respectivamente. El mayor valor de riqueza observada se obtuvo en el robledal de la ladera Norte (CTR) con 60 OTUs. El mismo resultado se observó en la riqueza estimada mediante el índice de Chao1 con un máximo de 916 OTUs en CTR y un mínimo de 429 en CTE. La distribución de las secuencias entre los distintos OTUs es casi la equidad total en robles con $J' = 0,998$ en CTR siendo todos los OTUs singletons (con una única secuencia) excepto un OTU que está formado por 2 secuencias y $J' = 0,997$ en CNB y CNA con 2 OTUs con 2 secuencias más 57 OTUs singletons. Para las encinas la equidad fue menor, siendo el menor de ellos el encinar de la ladera Sur (LJN) con $J' = 0,987$ y con un OTU formado por 4 secuencias pertenecientes al género *Bradyrhizobium* (Proteobacteria de la familia Bradyrhizobiaceae) y 2 OTUs con 2 secuencias cada uno, el resto fueron singletons. El índice de diversidad de Shannon da un valor máximo de 4,088 en CTR y un mínimo de 3,975 en LJN, en ambos casos resultan valores muy próximos al valor máximo de dicho índice para 61 secuencias evaluadas por muestra ($H'_{\max} = 4,111$). También se observan valores de riqueza y diversidad mayores en la ladera Norte que en la ladera Sur (Tabla R3-3). Con los resultados obtenidos de la curva de rarefacción, que está en la zona inicial de la misma donde es recta (Fig. R3-7), y el valor de cobertura calculado con el índice de Good's coverage que da un valor máximo de 13,11 % en LJN y un valor mínimo de 3,28 % en CTR (Tabla R3-3), observamos que 61 clones son insuficientes para tener muestras representativas de las comunidades microbianas originales objeto de este estudio.

Resultados Capítulo 3

Tabla R3-3. Diversidad α de las muestras estudiadas con genotecas clásicas del gen 16S *rRNA*. OTUs es el número de grupos obtenido a una distancia génica del 3 %. Chao1 es el valor medio de OTUs que se estima que hay en los suelos de origen, seguido de los valores mínimo y máximo para este índice. H' es el índice de Shannon seguido de su varianza. J' es el índice de equidad de Pielou. La cobertura se ha medido siguiendo el método de Good's coverage.

Muestra	Distancia	Secuencias	OTUs	Chao1	LCI95	UCI95	H'	varH	J'	Cobertura
CNB	0,03	61	59	591	255	1.501	4,065	0,008	0,997	6,56%
CNA	0,03	61	59	591	255	1.501	4,065	0,008	0,997	6,56%
CTR	0,03	61	60	915	364	2.463	4,088	0,008	0,998	3,28%
LJN	0,03	61	56	515	223	1.311	3,974	0,009	0,987	13,11%
CTE	0,03	61	58	429	201	1.017	4,040	0,008	0,996	9,84%

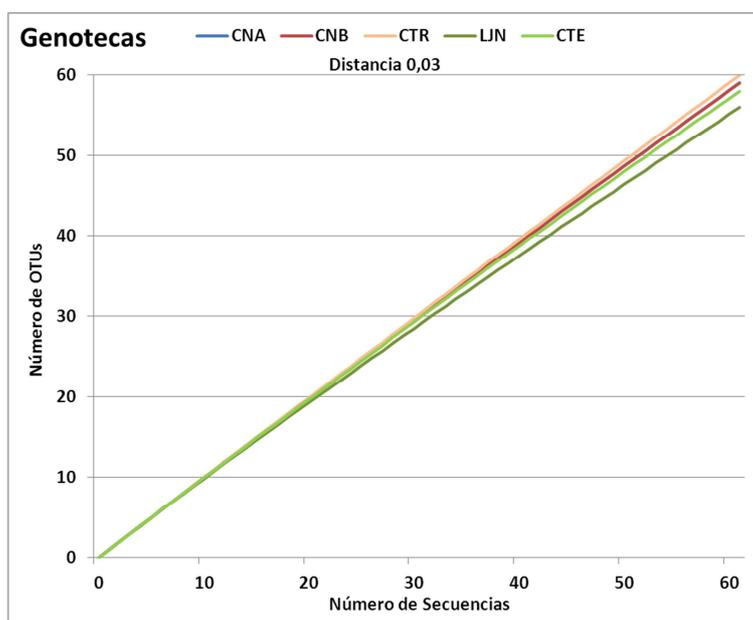


Figura R3-7. Curva de rarefacción de las muestras estudiadas con genotecas clásicas del gen 16S *rRNA*. Los OTUs se agruparon a una distancia génica máxima del 3 %.

En el análisis con amplicones 454 de la zona V6-V8, para las secuencias filtradas obtenidas de las 30 muestras resultantes de las 3 parcelas de cada uno de los cinco sitios estudiados en los muestreos de 2008 y 2011, la riqueza mínima observada es de 2.750 OTUs en la segunda parcela del encinar de la ladera Sur en el muestreo de 2011 (LJN2_11) y la riqueza máxima observada es de 5.104 OTUs (Tabla R3.4) en la tercera parcela del roble bajo en el muestreo de 2011 (CNB3_11). Por otra parte, la riqueza mínima esperada en los suelos de origen según el índice de Chao1 es de 4.982 OTUs en LJN2_11 y la riqueza máxima esperada en los suelos de origen según el índice de Chao1 es de 8.847 OTUs en LJN3_11 (Tabla R3-4). Los valores de diversidad medidos con el índice de Shannon superaron el valor de 7 en todas las muestras, siendo el valor mínimo de 7,279 en la muestra LJN2_11 y el máximo de

7,864 en la muestra CTE2_08. En todas las muestras las secuencias están distribuidas entre los distintos OTUs de forma muy próxima a la equidad con un valor mínimo del índice de Pielou de 0,887 en la muestra CTE3_11 y un valor máximo de 0,922 en la muestra CTE2_08. Los valores de cobertura medidos con el índice de Good fueron superiores al 82 % en todas las muestras, siendo el valor mínimo de 82,63 % en la muestra LNJ2_11 y el valor máximo de 90,88 % en la muestra CNA2_08 (Tabla R3-4).

Para hacer un análisis comparativo de los distintos índices de diversidad α entre todas las muestras estudiadas, éstas se normalizaron a 7.762 secuencias (Tablas R3-2 y R3-4). Se observa que la muestra con menor riqueza es LNJ1_08 con 2.053 OTUs y la de mayor riqueza es CNB3_11 con 2.863 OTUs, la muestra con menor riqueza esperada en su suelo de origen según el índice de Chao1 es LNJ2_08 con 3.633 OTUs y la muestra con mayor riqueza esperada en su suelo de origen según el índice de Chao1 es CNB3_11 con 5.470 OTUs (Tabla R3-4). Pese a existir diferencias en la riqueza observada entre las distintas muestras, ésta solo es significativa al comparar los resultados obtenidos en el muestreo de 2008 frente al de 2011. En 2011 la riqueza obtenida es mayor en los cinco sitios (Fig. R3-8) pero más acusada en el caso del encinar de la ladera Sur (LJN). La riqueza estimada mediante el índice de Chao1 muestra el mismo resultado que la riqueza observada, significativamente más abundante en el año de muestreo 2011. Respecto a la diversidad medida con el índice de Shannon se observa que el valor mínimo es de 6,719 en la muestra LNJ1_08 y el máximo 7,420 en la muestra CNB3_11 (Tabla R3-4). En este caso, aparecen diferencias significativas entre los años de muestreo y entre robles y encinas. Al comparar la diversidad de los cinco sitios mediante ANOVA también aparecen diferencias significativas y según el test post hoc de la Diferencia Honestamente Significativa (HSD de sus siglas en inglés) de Tukey, la diversidad del sitio LJN es significativamente menor a la de los sitios CNB y CTR (Fig. R3-9). En cuanto a la equidad medida mediante el índice de Pielou vemos que, el valor mínimo es de 0,881 en LNJ1_08 y el máximo es de 0,932 en CTR1_11. Las diferencias encontradas en este índice respecto al año de muestreo no son estadísticamente significativas, pero sí es significativa la diferencia de equidad entre robledal y encinar, siendo este último grupo el que presenta menores valores. Al comparar la equidad de los cinco sitios mediante ANOVA también aparecen diferencias significativas y según el test post hoc HSD de Tukey, el sitio LJN es significativamente menos equitativo que los sitios CNB y CTR (Fig. R3-10). Según los resultados obtenidos por el índice de cobertura de Good (Tabla R3-4) con un mínimo del 79,36 % en CNB3_11 y un máximo de 86,18 % en LNJ1_08, y las curvas de rarefacción de los cinco sitios estudiados (Fig. R3-11), con 7.762 secuencias por muestra no se consigue secuenciar el 100 % de las especies microbianas (OTUs

Resultados Capítulo 3

al 0,03) que habitan en los suelos forestales estudiados. Aunque, a pesar de ello, el esfuerzo de muestreo realizado es suficiente para obtener muestras representativas de cada uno de los sitios de estudio. Según las curvas de esfuerzo no hay diferencias significativas entre los cinco sitios, pero sí se presentan diferencias estadísticamente significativas entre los años de muestreo para el encinar de la ladera Sur (LJN). Además LJN de 2008 es el que requiere un menor esfuerzo de muestreo para obtener secuencias de todas las especies microbianas que habitan en su comunidad rizosférica (Fig. R3-11).

Tabla R3-4. Diversidad α de las muestras estudiadas con amplicones 454 del gen 16S *rRNA*. OTUs es el número de grupos obtenido a una distancia genética del 3 %. Chao1 es el valor medio de OTUs que se estima que hay en los suelos de origen, seguido de los valores mínimo y máximo para este índice. H' es el índice de Shannon seguido de su varianza. J' es el índice de equidad de Pielou. La cobertura se ha medido siguiendo el método de Good's coverage. La muestra es el número de la parcela de cada sitio de muestreo, seguido de un número que indica el año de muestreo.

Muestra	Estado	Distancia	Secuencias	OTUs	Chao1	LCI95	UCI95	H'	varH	J'	Cobertura
CNA1_08	Filtradas	0,03	15.027	3702	5568	5340	5828	7,481	0,00012	0,910	89,23%
	Normalizadas	0,03	7.762	2490	4325	4070	4620	7,194	0,00021	0,920	83,33%
CNA1_11	Filtradas	0,03	15.661	3990	7196	6830	7608	7,469	0,00013	0,901	87,01%
	Normalizadas	0,03	7.762	2535	4826	4513	5189	7,190	0,00021	0,917	82,01%
CNA2_08	Filtradas	0,03	16.841	3624	5429	5203	5688	7,381	0,00011	0,901	90,88%
	Normalizadas	0,03	7.762	2207	3737	3507	4009	7,025	0,00021	0,912	86,10%
CNA2_11	Filtradas	0,03	17.742	4390	7372	7046	7738	7,588	0,00011	0,905	87,93%
	Normalizadas	0,03	7.762	2563	4667	4380	4998	7,238	0,0002	0,922	82,21%
CNA3_08	Filtradas	0,03	19.615	4840	8126	7779	8514	7,692	0,00010	0,907	88,18%
	Normalizadas	0,03	7.762	2489	4644	4344	4993	7,201	0,0002	0,921	82,76%
CNA3_11	Filtradas	0,03	18.975	4583	7486	7171	7840	7,645	0,00010	0,907	88,63%
	Normalizadas	0,03	7.762	2593	4605	4332	4922	7,285	0,00019	0,927	82,35%
CNB1_08	Filtradas	0,03	15.667	3572	5188	4984	5420	7,412	0,00012	0,906	89,80%
	Normalizadas	0,03	7.762	2311	3952	3714	4231	7,093	0,00021	0,916	84,94%
CNB1_11	Filtradas	0,03	18.903	4548	7419	7105	7771	7,609	0,00010	0,903	88,72%
	Normalizadas	0,03	7.762	2569	4584	4309	4902	7,251	0,0002	0,924	82,45%
CNB2_08	Filtradas	0,03	16.774	4091	6001	5776	6257	7,609	0,00010	0,915	89,69%
	Normalizadas	0,03	7.762	2549	4553	4276	4875	7,237	0,00021	0,923	82,90%
CNB2_11	Filtradas	0,03	18.439	4571	7616	7287	7986	7,609	0,00011	0,903	88,10%
	Normalizadas	0,03	7.762	2642	4864	4569	5205	7,247	0,00022	0,920	81,29%
CNB3_08	Filtradas	0,03	16.694	3779	5583	5361	5837	7,467	0,00011	0,906	90,51%
	Normalizadas	0,03	7.762	2348	3859	3640	4115	7,134	0,0002	0,919	85,17%
CNB3_11	Filtradas	0,03	19.177	5104	8217	7899	8571	7,811	0,00009	0,915	87,33%
	Normalizadas	0,03	7.762	2863	5470	5135	5853	7,420	0,00019	0,932	79,36%
CTR1_08	Filtradas	0,03	16.233	3738	5477	5262	5721	7,416	0,00013	0,902	90,30%
	Normalizadas	0,03	7.762	2336	3932	3700	4203	7,094	0,00023	0,915	85,04%
CTR1_11	Filtradas	0,03	18.221	4706	7816	7482	8190	7,748	0,00009	0,916	87,74%
	Normalizadas	0,03	7.762	2746	5183	4861	5553	7,377	0,00018	0,932	80,66%

Resultados Capítulo 3

CTR2_08	Filtradas	0,03	14.208	3638	5355	5144	5597	7,503	0,00012	0,915	88,94%
	Normalizadas	0,03	7.762	2460	4362	4093	4675	7,197	0,00020	0,922	83,73%
CTR2_11	Filtradas	0,03	18.658	4566	7587	7259	7954	7,640	0,00010	0,907	88,31%
	Normalizadas	0,03	7.762	2555	4803	4494	5162	7,238	0,00020	0,922	82,25%
CTR3_08	Filtradas	0,03	15.492	3884	5960	5710	6244	7,594	0,00010	0,919	89,07%
	Normalizadas	0,03	7.762	2471	4160	3922	4437	7,224	0,00019	0,925	84,04%
CTR3_11	Filtradas	0,03	19.424	4496	7280	6977	7620	7,576	0,00010	0,901	89,02%
	Normalizadas	0,03	7.762	2502	4779	4463	5146	7,202	0,00020	0,920	82,45%
LJN1_08	Filtradas	0,03	15.362	3592	5611	5363	5894	7,308	0,00015	0,893	89,41%
	Normalizadas	0,03	7.762	2053	3746	3485	4055	6,719	0,00033	0,881	86,18%
LJN1_11	Filtradas	0,03	15.663	4087	6928	6606	7291	7,534	0,00013	0,906	87,22%
	Normalizadas	0,03	7.762	2404	4389	4107	4718	7,088	0,00024	0,911	83,56%
LJN2_08	Filtradas	0,03	13.984	3588	5688	5432	5979	7,454	0,00013	0,911	88,14%
	Normalizadas	0,03	7.762	2179	3633	3415	3889	6,958	0,00024	0,905	85,96%
LJN2_11	Filtradas	0,03	8.493	2750	4982	4688	5321	7,279	0,00020	0,919	82,63%
	Normalizadas	0,03	7.762	2435	4527	4234	4869	7,092	0,00025	0,910	82,99%
LJN3_08	Filtradas	0,03	12.210	3292	5183	4943	5457	7,403	0,00014	0,914	87,53%
	Normalizadas	0,03	7.762	2219	3849	3604	4136	7,001	0,00023	0,909	85,70%
LJN3_11	Filtradas	0,03	18.139	4982	8847	8452	9287	7,714	0,00012	0,906	85,99%
	Normalizadas	0,03	7.762	2600	4951	4632	5320	7,176	0,00026	0,913	81,56%
CTE1_08	Filtradas	0,03	17.259	4293	6696	6423	7004	7,644	0,00010	0,914	88,86%
	Normalizadas	0,03	7.762	2367	4268	3992	4591	7,133	0,00021	0,918	84,28%
CTE1_11	Filtradas	0,03	17.872	4713	7911	7575	8285	7,670	0,00011	0,907	86,95%
	Normalizadas	0,03	7.762	2588	4661	4383	4983	7,201	0,00023	0,916	81,86%
CTE2_08	Filtradas	0,03	17.663	5076	8545	8192	8938	7,864	0,00009	0,922	85,86%
	Normalizadas	0,03	7.762	2518	4876	4551	5254	7,189	0,00021	0,918	82,08%
CTE2_11	Filtradas	0,03	15.621	4595	8062	7700	8467	7,719	0,00012	0,915	84,71%
	Normalizadas	0,03	7.762	2680	5091	4771	5460	7,263	0,00023	0,920	80,74%
CTE3_08	Filtradas	0,03	19.156	4393	6708	6444	7006	7,617	0,00010	0,908	90,10%
	Normalizadas	0,03	7.762	2267	3797	3571	4063	7,060	0,00022	0,914	85,51%
CTE3_11	Filtradas	0,03	14.022	3952	6929	6598	7302	7,346	0,00020	0,887	85,16%
	Normalizadas	0,03	7.762	2469	4792	4473	5161	6,892	0,00039	0,882	82,00%

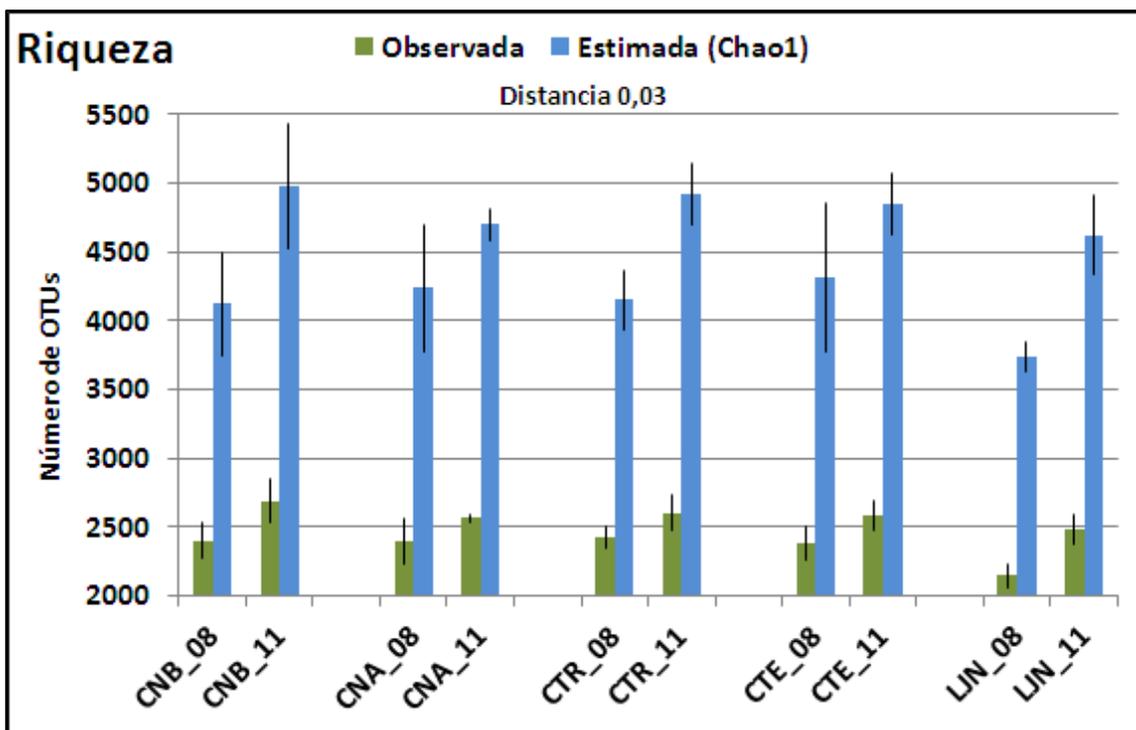


Figura R3-8. Riqueza observada y estimada de especies medidas como OTUs a una distancia del 3 % obtenidas tras la pirosecuenciación de amplicones del gen 16S *rRNA*. Las barras de error muestran la desviación estándar de la media observada entre las réplicas de cada sitio para cada año de muestreo.

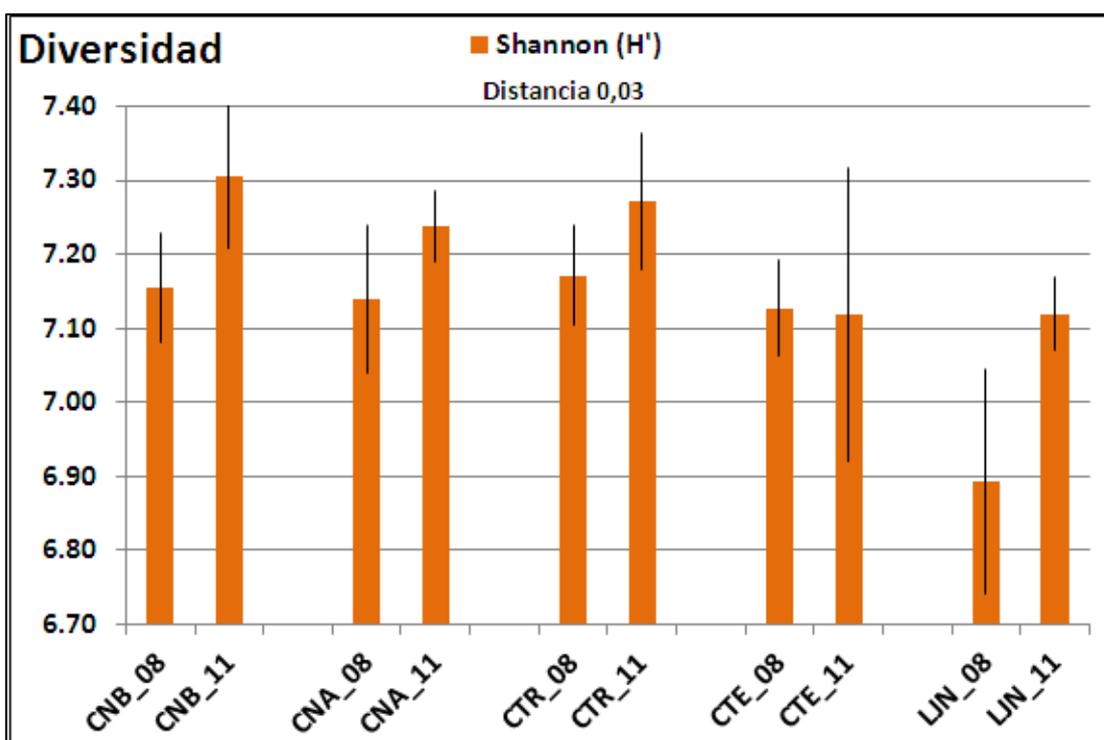


Figura R3-9. Diversidad de especies medidas como OTUs a una distancia del 3 %, analizada con el índice de Shannon tras la pirosecuenciación de amplicones del gen 16S *rRNA*. Las barras de error muestran la desviación estándar de la media observada entre las réplicas de cada sitio para cada año de muestreo.

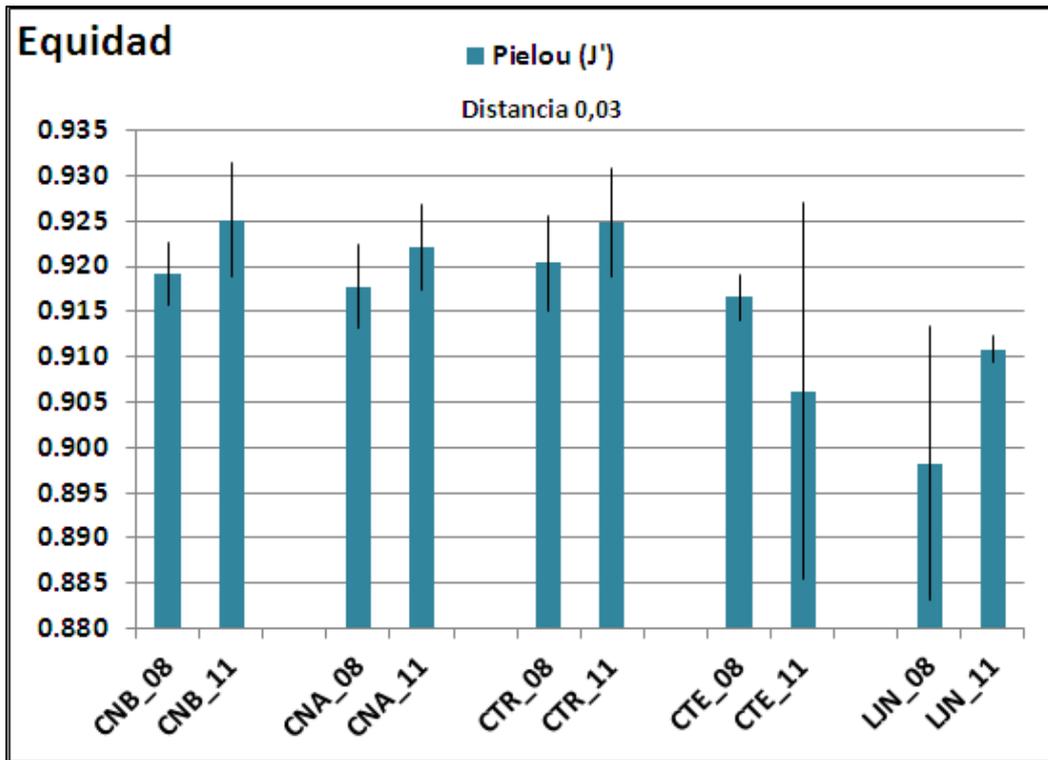


Figura R3-10. Equidad de la distribución de las distintas especies medidas como OTUs a una distancia del 3 % tras la pirosecuenciación de amplicones del gen 16S *rRNA*, analizada con el índice de Pielou. Las barras de error muestran la desviación estándar de la media observada entre las réplicas de cada sitio para cada año de muestreo.

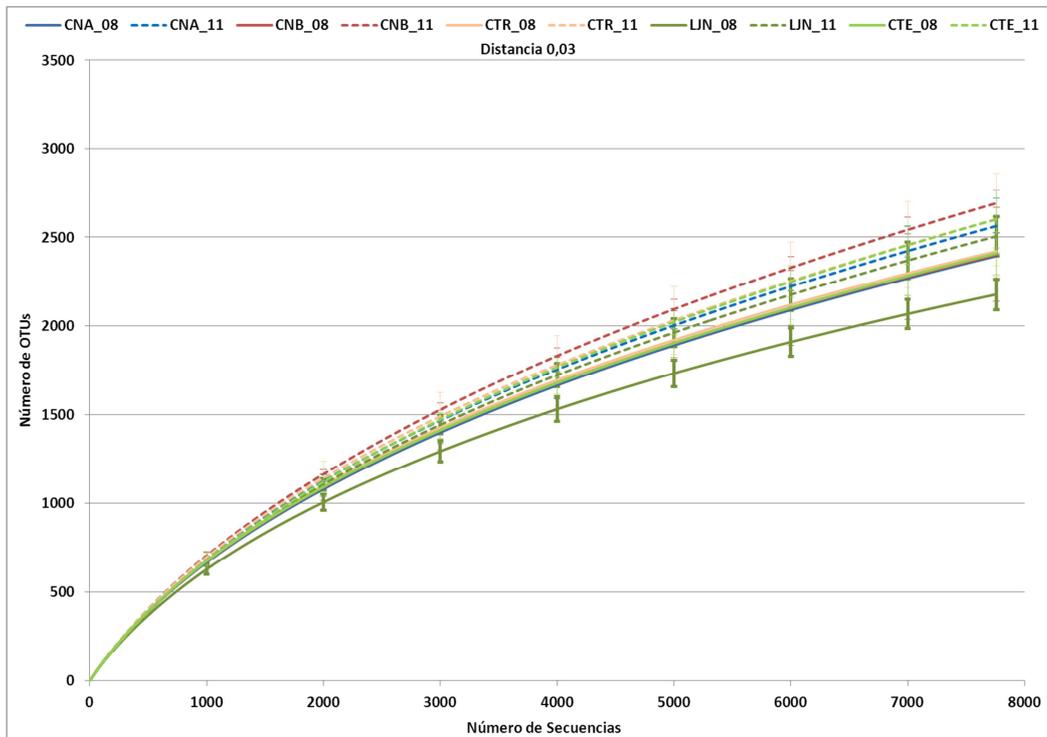


Figura R3-11. Curvas de rarefacción de especies medidas como OTUs a una distancia del 3 % tras la pirosecuenciación de amplicones del gen 16S *rRNA*. Las barras de error muestran la desviación estándar de la media observada entre las réplicas de cada sitio para cada año de muestreo, engrosadas en los sitios con diferencias significativas. — Muestras de 2008. Muestras de 2011.

3 Estudio de la diversidad β mediante empleo del gen 16S *rRNA*

Este estudio también se ha realizado con las secuencias obtenidas por secuenciación Sanger de las genotecas en *E.coli* y los amplicones 454 de la zona V6-V8 del gen 16S *rRNA* (Tablas R3-3 y R3-4). Entendiendo por diversidad β la diferencia de composición de las comunidades microbianas de las muestras comparadas dos a dos.

En el análisis con genotecas se observa que el sitio con más diferencias en su comunidad microbiana es el encinar de la ladera Norte (CTE), ya que está alejado de los demás en el 100 % de los casos. La distancia de CTE con el resto de sitios, medido con el índice de disimilitud de Unifrac, es del 26,6 %. Los sitios más parecidos son el robledal bajo de la ladera Sur (CNB) y el robledal de la ladera Norte (CTR), con un índice de disimilitud del 20,5 % (Fig. R3-12). Pese a que la mayoría de OTUs están representados con una única secuencia, resultando solo 3 OTUs con secuencias de, al menos, dos de los sitios comparados, todos ellos tienen comunidades microbianas filogenéticamente muy próximas (Fig. R3-12). Este mismo resultado se observa al comparar las comunidades mediante β -Libshuff, ya que ninguno de los sitios es estadísticamente diferente a los otros. Teniendo en cuenta que el p-valor corregido para múltiples test mediante el factor de corrección de Bonferroni es de 0,005 para este análisis y los sitios con mayores diferencias CNB y CTE tienen un p-valor de 0,0189 (Tabla R3-5).

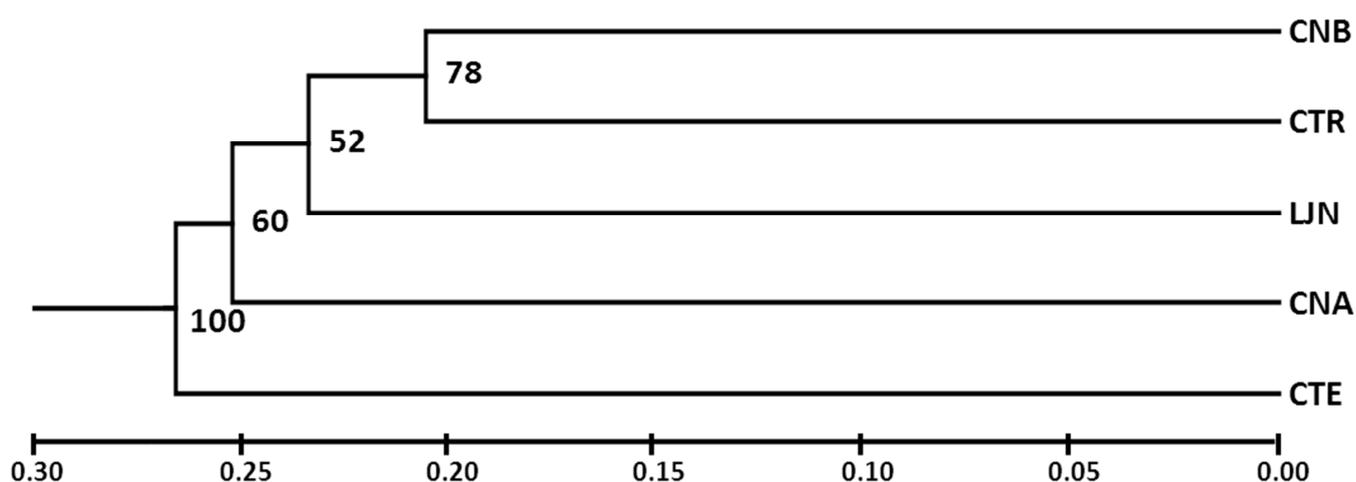


Figura R3-12. Agrupamiento jerárquico aglomerativo usando la matriz de distancias Unifrac y el algoritmo UPGMA teniendo en cuenta la abundancia de secuencias de cada muestra en cada nodo del árbol filogenético. Para las secuencias de los sitios estudiados con genotecas clásicas del gen 16S *rRNA*. Análisis realizado con las distancias filogenéticas encontradas entre las secuencias de cada muestra. Los valores de los nodos del árbol se obtuvieron por el método de Jackknife con 1.000 iteraciones de 50 secuencias seleccionadas al azar de cada muestra.

Tabla R3-5. Comparación dos a dos de los 5 sitios estudiados con genotecas del gen 16S *rRNA* mediante el test χ^2 -Libshuff. Intervalo de confianza del 95 % y p-valor crítico corregido por el factor de Bonferroni de 0,005.

Comparación	p-valor
CNB-CNA	0,7662
CTE-CNA	0,0734
CTR-CNA	0,4658
LJN-CNA	0,0192
CTE-CNB	0,0189
CTR-CNB	0,4313
LJN-CNB	0,2750
CTE-CTR	0,2502
CTE-LJN	0,1431
LJN-CTR	0,2138

En el análisis con amplicones 454 de la zona V6-V8, para las secuencias filtradas después de ser alineadas en RDP y normalizar las muestras a 7.762 secuencias (Tablas R3-2 y R3-4), el agrupamiento jerárquico aglomerativo mediante UPGMA y sin tener en cuenta la abundancia de los OTUs, analizados mediante Jaccard ni la abundancia en los nodos del árbol filogenético, analizados mediante Unifrac, pone de manifiesto que las muestras no se agrupan respecto a los sitios de los que proceden. Siendo las más próximas entre si dos réplicas del encinar de la ladera Sur muestreadas en 2011 (LJN1.11 y LJN2.11) unidas en el 95 % de los árboles obtenidos al 66 % de distancia (Fig. R3-13). Todas las muestras están agrupadas al 79 % de distancia, pero se separan a una distancia inferior al 66 % indicando que las comunidades microbianas (tanto a nivel de especie en Jaccard como teniendo en cuenta la pertenencia a distintas ramas del árbol filogenético en Unweighted-Unifrac) de cada muestra son muy distintas de las del resto (Fig. R3-13). Este resultado se confirma con los análisis del test estadístico Unifrac y el AMOVA que dan p-valores inferiores a 0,0001 entre todas las muestras comparadas dos a dos, indicando además que las comunidades microbianas de cada muestra son significativamente diferentes del resto de muestras (tanto a nivel de especie en AMOVA como teniendo en cuenta la pertenencia a distintas ramas del árbol filogenético en Unweighted-Unifrac).

En el caso de las muestras agrupadas mediante UPGMA teniendo en cuenta la abundancia de cada OTU para cada muestra, analizado mediante el comando Weighted-Unifrac en MOTHUR y el peso (número de secuencias) de cada nodo en el árbol filogenético analizado mediante Unifrac, aparecen dos grupos separados en el 100 % de los árboles creados

a un 17 % de distancia (Fig. R3-14). En ambos grupos aparecen muestras procedentes de robledal y encinar, al igual que se agrupan muestras de la ladera Norte y la ladera Sur de Sierra Nevada y tampoco se separan las muestras por el año en el que se realizó el muestreo. Todas las muestras están agrupadas a una distancia superior al 17 % y las muestras más próximas, procedentes del robledal de la ladera Norte, en la zona alta y muestreadas en 2008 (CNA1.08, CNA2.08 y CNA3.08) se agrupan al 9,3 % de distancia. Quedando las réplicas CNA1.08 y CNA3.08 agrupadas al 8,9 % de distancia en el 71 % de los árboles creados (Fig. R3-14). Estos resultados se corresponden con los resultados de los análisis del test estadístico Unifrac y el AMOVA en los que no aparecen diferencias significativas entre ninguna de las muestras rizosféricas comparadas dos a dos. De modo que las diferencias encontradas entre los 5 sitios no son mayores que las encontradas entre las 3 réplicas de cada sitio entre sí. Así, según este análisis, las comunidades microbianas de los 5 sitios estudiados no pueden considerarse diferentes ni respecto a su localización geográfica (ladera Norte o Sur), ni al año en el que se muestrearon (2008 o 2011), ni tampoco por estar asociadas a distintas especies de quercíneas (*Q. pyrenaica* Willd. o *Q. ilex* subsp. *ballota*)

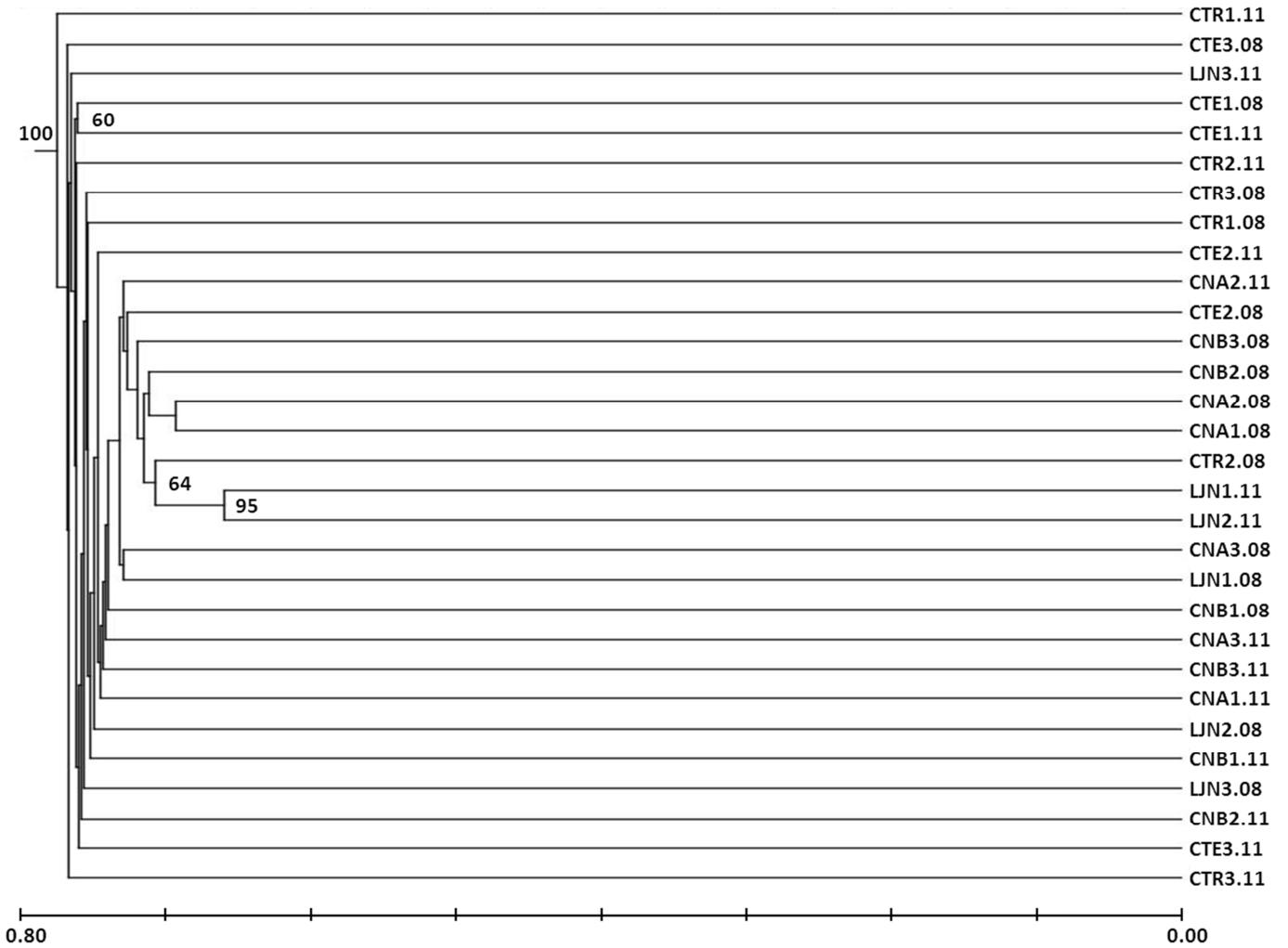


Figura R3-13. Agrupamiento jerárquico aglomerativo usando la matriz de distancias Unifrac y el algoritmo UPGMA teniendo en cuenta solo Presencia/Ausencia de secuencias de cada muestra en cada nodo del árbol filogenético. Para las secuencias normalizadas procedentes de amplicones 454 de la zona V6-V8 del gen 16S *rRNA*. Análisis realizado con las distancias filogenéticas encontradas entre las secuencias de cada muestra. Se muestran los valores de los nodos obtenidos por el método de Jackknife con 1.000 iteraciones de 6.000 secuencias seleccionadas al azar de cada muestra con un valor igual o superior al 60 %.

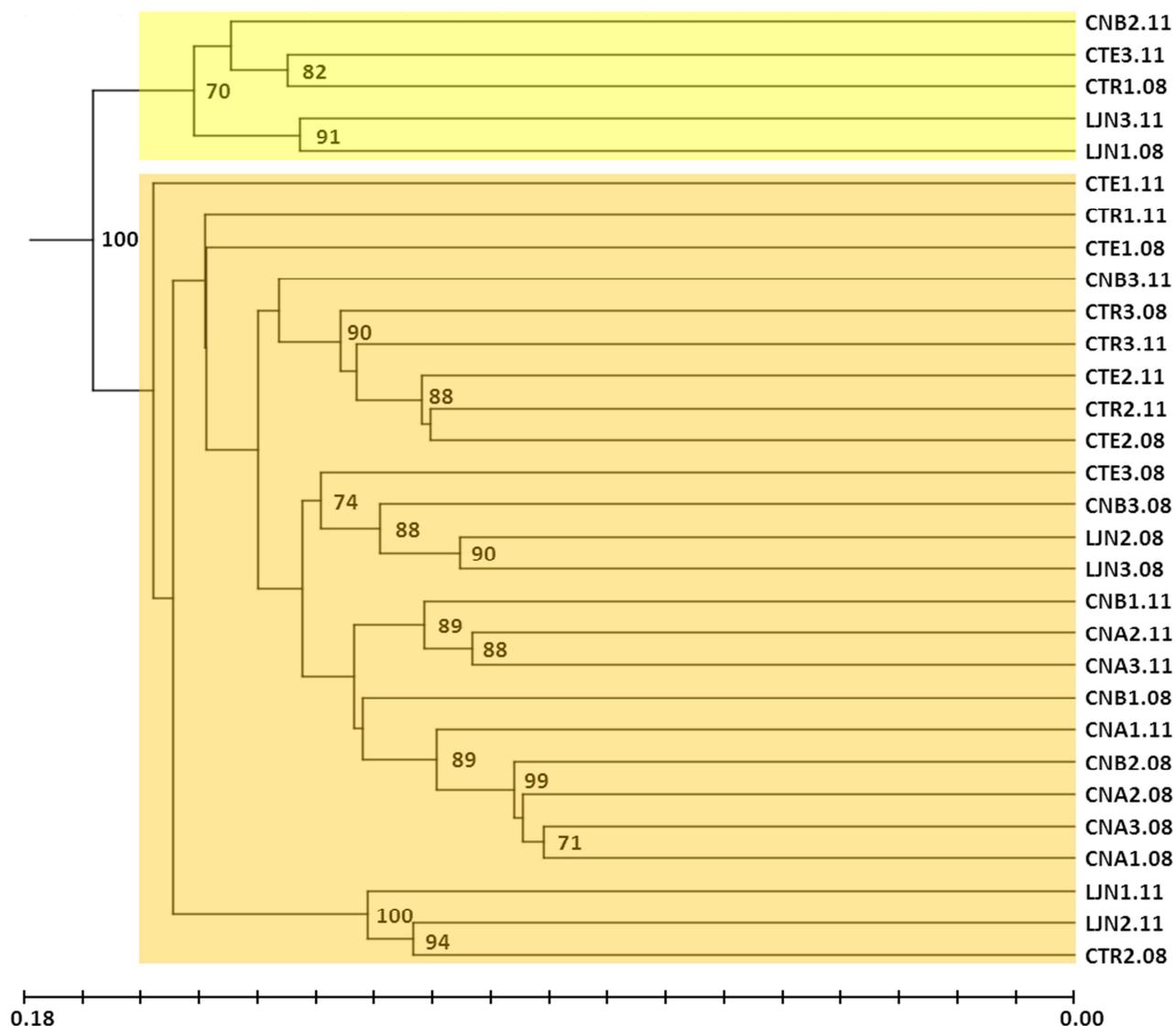


Figura R3-14. Agrupamiento jerárquico aglomerativo usando la matriz de distancias Unifrac y el algoritmo UPGMA teniendo en cuenta la abundancia de secuencias de cada muestra en cada nodo del árbol filogenético. Para las secuencias normalizadas procedentes de amplicones 454 de la zona V6-V8 del gen 16S *rRNA*. Análisis realizado con las distancias filogenéticas encontradas entre las secuencias de cada muestra. Se muestran los valores de los nodos obtenidos por el método de Jackknife con 1.000 iteraciones de 6.000 secuencias seleccionadas al azar de cada muestra con un valor igual o superior al 60 %.

Al comparar las distintas muestras, teniendo en cuenta las secuencias normalizadas que clasifican en RDP a nivel de familia, no vemos que éstas se agrupen por su composición filogenética en grupos bien diferenciados. Atendiendo a la PC1 (primera coordenada principal) que explica el 36,4 % de la varianza entre muestras. Están mezcladas todas las muestras, tanto de robleal (CNB, CNA y CTR) como de encinar (LJN y CTE), estando estas últimas en los valores negativos más altos (LJN) y los valores positivos más altos (CTE) de dicha coordenada (Fig. R3-15). La coordenada principal 2 (PC2) explica el 18,3 % de la varianza entre muestras. La muestra del encinar de la ladera Sur (LJN) ocupa los valores positivos más altos de esta

coordinada así como, los valores negativos más altos junto con las muestras del robledal de la ladera Sur (CNB y CNA). Se observan mayores distancias entre las muestras de la ladera Sur (CNB, CNA y LNJ), en los dos años de muestreo realizados (2008 y 2011), que en las muestras de la ladera Norte (CTR y CTE), aunque éstas no forman grupos discernibles en el PCoA (Fig. R3-15).

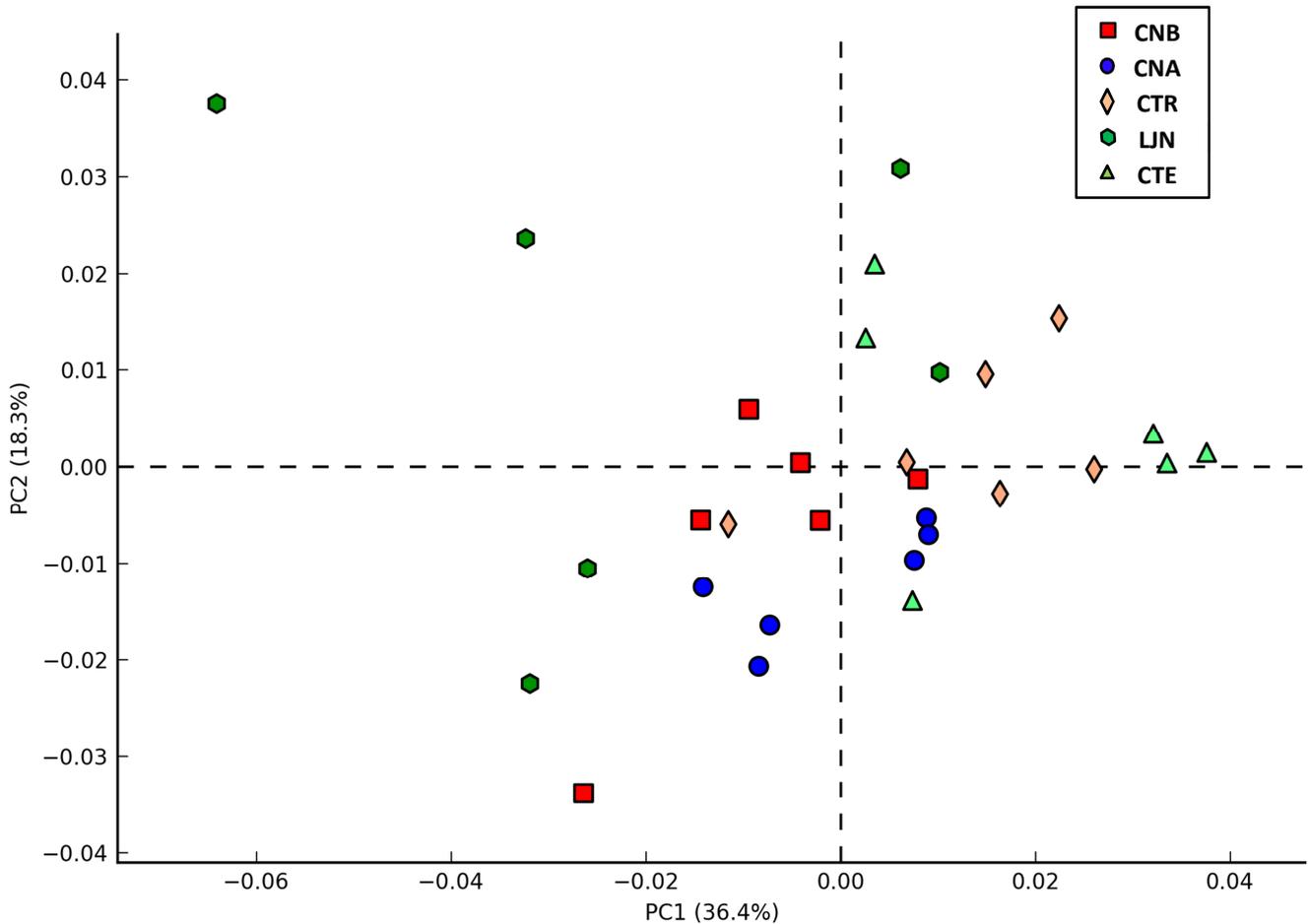


Figura R3-15. Análisis de Coordenadas Principales (PCoA). Realizado en el programa STAMP con las 30 muestras pertenecientes a los 5 sitios muestreados en ambos años de muestreo, teniendo en cuenta las secuencias clasificadas a nivel de familia.

Se han comparado los 5 sitios muestreados (CNB, CNA, CTR, LNJ y CTE) teniendo en cuenta la abundancia relativa de las familias dominantes (Figs. R3-5 y R3-6) y los parámetros ambientales obtenidos de dichos sitios (Tablas R1-1, R1-2, R1-4 y R1-5). En el programa CANOCO se introdujeron como variables ambientales suplementarias el año de muestreo, la rizosfera de la que fueron tomadas (roble o encina) y la ladera de la que proceden según su localización geográfica (Norte o Sur). En la figura R3-16 se observa una distribución parecida a la obtenida en el PCoA (Fig. R3-15) para todas las muestras excepto LNJ que queda en la

Resultados Capítulo 3

posición completamente opuesta según el Eje principal 1, el cual explica el 43,4 % de la varianza observada. Esta diferencia en la ordenación de las muestras del sitio LJM al introducir las variables ambientales no está correlacionada con el pH. Pero el pH es la única variable ambiental que aparece estadísticamente significativa (test de Monte Carlo con 999 permutaciones da un p-valor de 0,043), ya que las otras 4 muestras (CNB, CNA, CTR y CTE) si tienen una ordenación correlacionada entre la abundancia de las familias dominantes y el pH de sus suelos, aunque con un p-valor marginal de 0,051 teniendo en cuenta las muestras de los 5 sitios (Fig. R3-16). Los sitios se separan en el Eje 2, el cual explica el 40,0 % de la varianza, principalmente por el año en que se muestrearon. También afectan en la ordenación el contenido en materia orgánica, la salinidad y la textura de los suelos pero no son estadísticamente significativos.

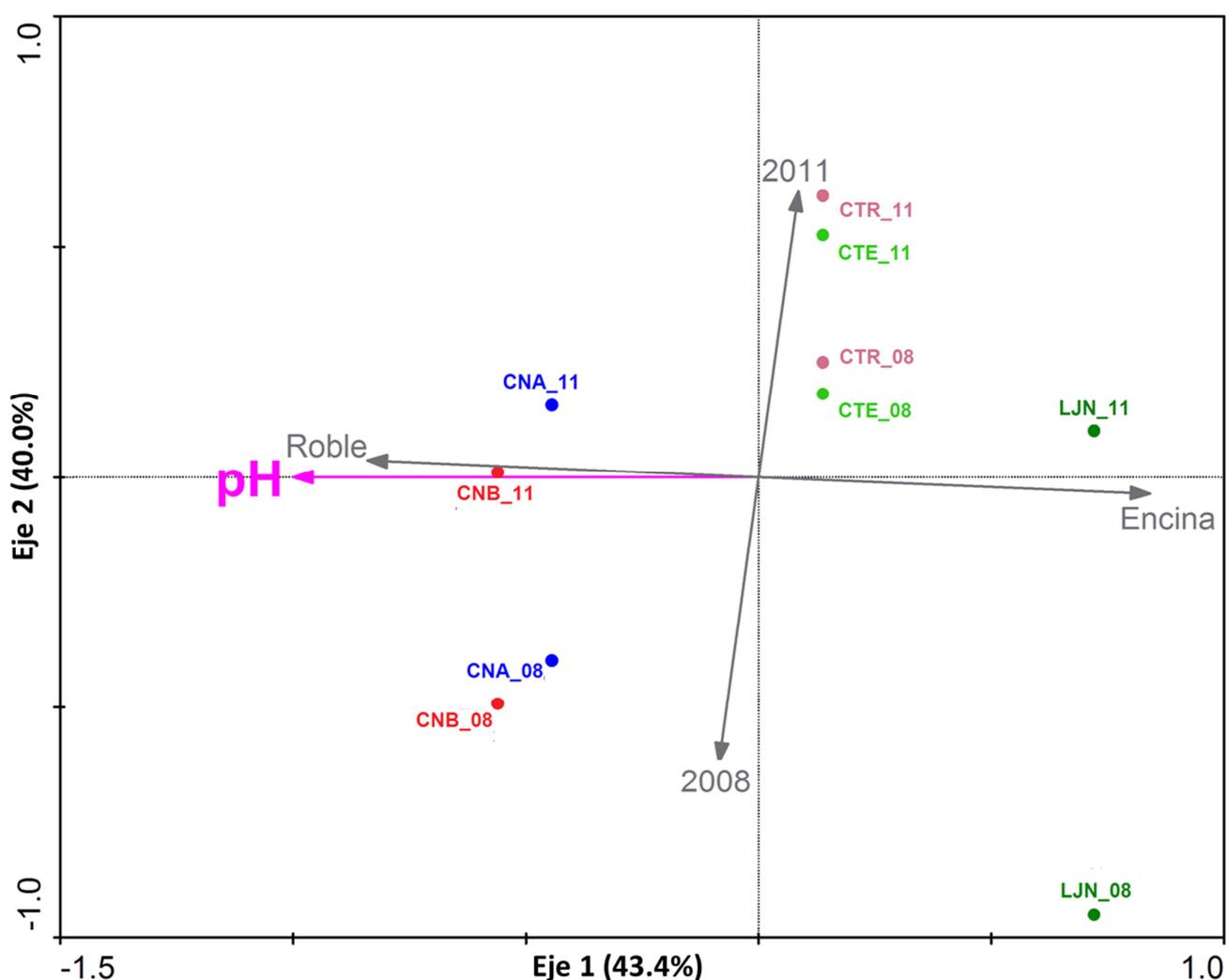


Figura R3-16. Análisis de Correspondencias Canónicas (CCA). Realizado en el programa CANOCO con los 5 sitios muestreados en ambos años de muestreo, teniendo en cuenta las secuencias promedio de cada sitio clasificadas a nivel de familia. En gris se muestran los vectores de las variables ambientales utilizadas como suplementarias. En fucsia se muestran solo los vectores de las variables ambientales que afectan significativamente a la ordenación de los sitios.

Capítulo 4: Efecto del incendio en la estructura y composición de las comunidades microbianas asociadas a la rizosfera de encina (*Q. ilex* subsp. *ballota*)

Se determinaron los efectos del incendio en la estructura y composición de las comunidades microbianas analizando el DNA medioambiental extraído de la rizosfera de un encinar quemado (LQ), de un encinar próximo no afectado (LJN) y de suelo no rizosférico de un sitio de matorral afectado por el mismo incendio (LJR) situados en la ladera sur de Sierra Nevada. Todos ellos fueron muestreados en la primavera de los años 2008 y 2011. En el matorral afectado por el incendio, también se muestreó la rizosfera de encinas germinadas después del incendio (LJRE), en la primavera del año 2011.

1 Clasificación taxonómica de las secuencias del gen 16S *rRNA*

En el estudio clásico de genotecas, se analizaron solo los sitios quemado (LJQ) y no quemado (LJN) muestreados en el año 2008. De ellos se obtuvieron al menos 75 clones por sitio con un amplicón casi completo del gen 16S *rRNA* (≈ 1460 bp). De estos clones se consiguieron un total de 122 secuencias de buena calidad y no quiméricas, 61 secuencias en cada sitio. Tras analizar nuestras secuencias frente a la base de datos de Bacteria y Archaea del RDP vemos que, a nivel de *phylum*, conseguimos clasificar el 95,1 % de las secuencias en ambos sitios, siendo el *phylum* Proeobacteria el mayoritario con una abundancia relativa del 34,4 % en ambos sitios. El segundo *phylum* con mayor abundancia es Actinobacteria con un 24,6 % en LJN y un 27,9 % en LJQ. El tercer *phylum*, Bacteroidetes, es más abundante en LJQ con un 16,4 % que en LJN con un 8,2 %, al igual que se observa en Actinobacteria. Los *phyla* Acidobacteria y Gemmatimonadetes son más abundantes en el encinar no quemado; teniendo en el primero una abundancia relativa de un 13,1 % en LJN y un 8,2 % en LJQ; mientras que el segundo *phylum* tiene una abundancia relativa de un 6,6 % en LJN y un 3,3 % en LJQ. Los demás *phyla* son menores al 4 % para ambos sitios (Fig. R4-1). A nivel de familia conseguimos clasificar el 57,4 % de las secuencias en LJN y el 67,2 % en LJQ, siendo la más abundante la familia Micrococcaceae, perteneciente al orden Actinomycetales, *phylum* Actinobacteria, que supone el 19,7 % en LJQ y sin secuencias clasificadas para esta familia en LJN. La segunda familia con mayor abundancia relativa es Chitinophagaceae, perteneciente al *phylum* Bacteroidetes, que supone el 11,5 % en LJQ y el 4,9 % en LJN. Las familias Bradyrhizobiaceae (clase α -Proteobacteria) y Gemmatimonadaceae (*phylum* Gemmatimonadetes) son las más abundantes en el encinar no quemado (LJN) con un 6,6 % de abundancia en este sitio y un 3,3 % en LJQ (Fig. R4-2). Se observan diferencias muy acusadas en las seis familias más abundantes entre ambos sitios, pero debido al pequeño tamaño de muestra analizado sólo es estadísticamente significativa, según la herramienta Libcompare, la familia Micrococcaceae con un p-valor de 0,009.

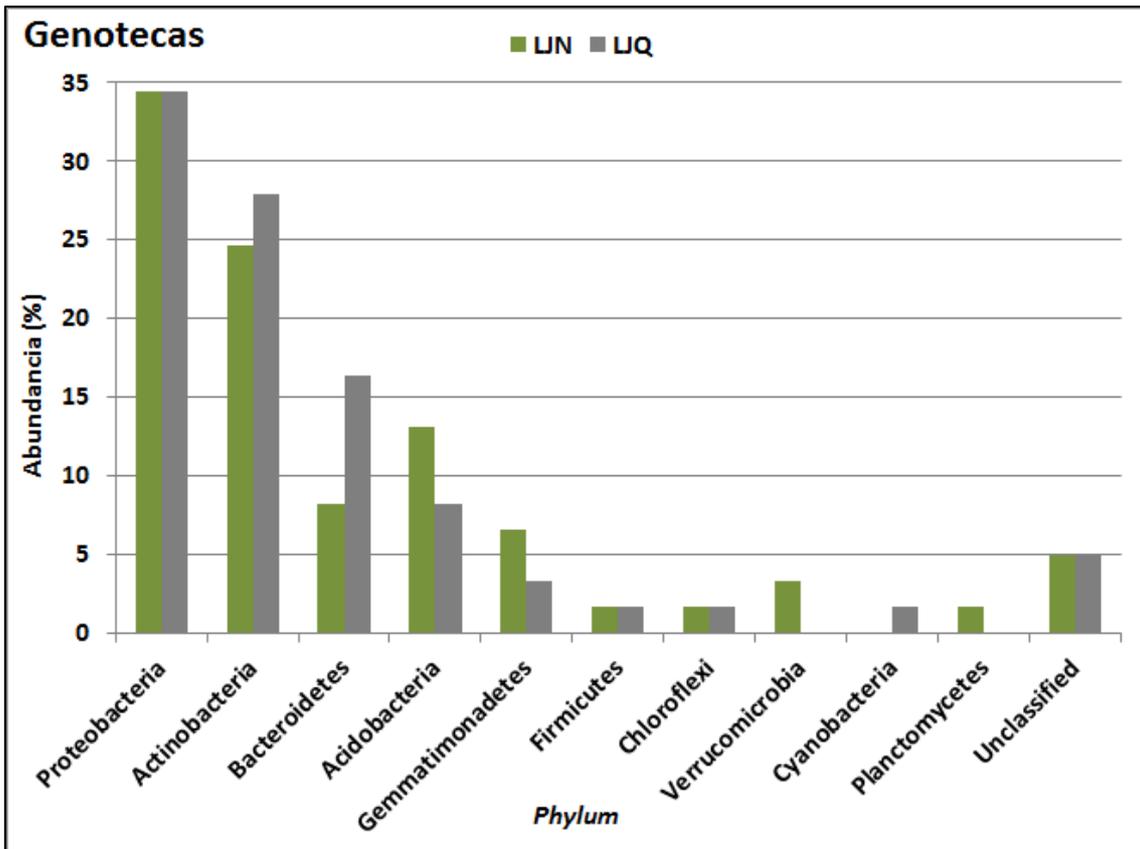


Figura R4-1. Distribución de todos los *phyla* de las genotecas clásicas de 16S *rRNA*.

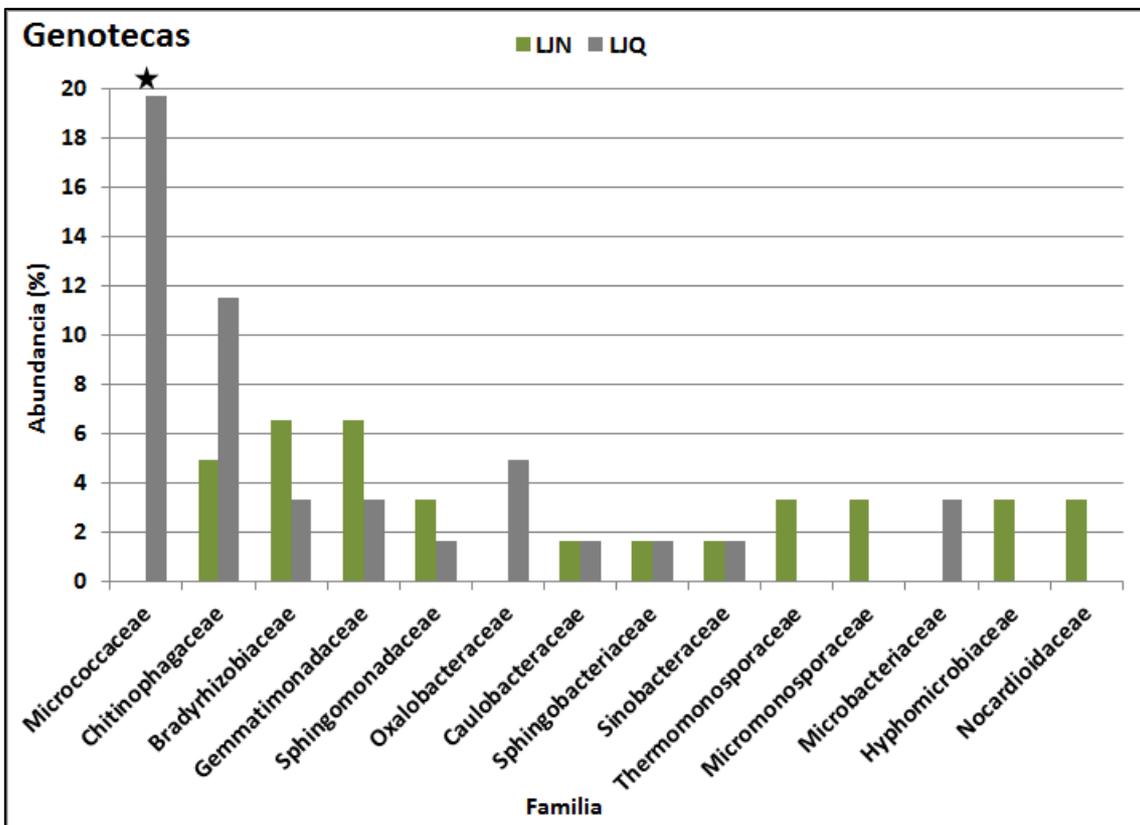


Figura R4-2. Distribución de las familias más abundantes de las genotecas clásicas de 16S *rRNA*. Solo se muestran aquellas familias que suponen al menos el 3,3 % de abundancia relativa en la suma de ambos sitios.

Resultados Capítulo 4

En el estudio de los metagenomas mediante pirosecuenciación se analizaron sólo los sitios quemado y no quemado muestreados en 2008 al igual que en las genotecas clásicas. De ellos se obtuvieron 520.430 lecturas en LJN y 316.973 lecturas en LJQ que tras filtrarlas, resultaron en 412.302 secuencias con longitud media de 371 bp y 253.660 secuencias con longitud media de 388 bp, respectivamente, únicas y de calidad. Dichas secuencias se han anotado con la base de datos M5NR de MG-RAST dando como resultado 1.071 secuencias en LJN y 912 secuencias en LJQ homólogas a genes que transcriben algún ARN ribosómico (*rRNA*). Al enfrentar estas secuencias ribosómicas a la base de datos del gen 16S *rRNA* del RDP se clasificaron 172 y 85 secuencias, en LJN y LJQ respectivamente (Tabla R4-1).

Al analizar la clasificación de las secuencias homólogas al gen 16S *rRNA* de LJN y LJQ procedentes de pirosecuenciación vemos que la distribución de las familias es muy diferente entre ambas muestras. Se ha conseguido clasificar el 81,9 % de las secuencias en LJN y el 85,9 % en LJQ, siendo la familia Micrococcaceae (*phylum* Actinobacteria) la mayoritaria por suponer el 16,9 % en suma de ambos sitios, y teniendo en cuenta que no aparece ninguna secuencia de esta familia en LJN. Las familias Oxalobacteraceae (clase β -Proteobacteria) con un 9,9 % en LJQ y un 0,6 % en LJN; Xanthomonadaceae (clase γ -Proteobacteria) con un 7,0 % en LJQ y un 1,9 % en LJN y la familia Bradyrhizobiaceae (clase α -Proteobacteria) con un 7,0 % en LJQ y sin secuencias clasificadas en LJN; junto con la familia Micrococcaceae son las más abundantes en el encinar quemado. La familia mayoritaria en el encinar no quemado es Sphingobacteriaceae, perteneciente al *phylum* Bacteroidetes, con un 7,5 % en LJN y sin secuencias clasificadas en LJQ (Fig. R4-3). El resto de familias no superan el 5 % de abundancia relativa en ninguno de los sitios evaluados. Pese a las diferencias tan acusadas observadas entre ambos sitios, según la herramienta STAMP y debido al tamaño de muestra analizado, estas diferencias sólo son estadísticamente significativas en la familia Micrococcaceae con un p-valor corregido de $3,23 \times 10^{-5}$.

Tabla R4-1. Recuento de secuencias antes y después del procesamiento de los sitios estudiados con secuenciación masiva "shotgun".

Muestra	Método de secuenciación	Lecturas	Secuencias	rRNA	16S RDP
LJQ	454	316.973	253.660	912	85
LJN	454	520.430	412.302	1.071	172
LJQ	Illumina	386.720.378	186.654.779	30.032.562	106.501
LJN	Illumina	374.512.840	180.589.099	27.515.446	65.828

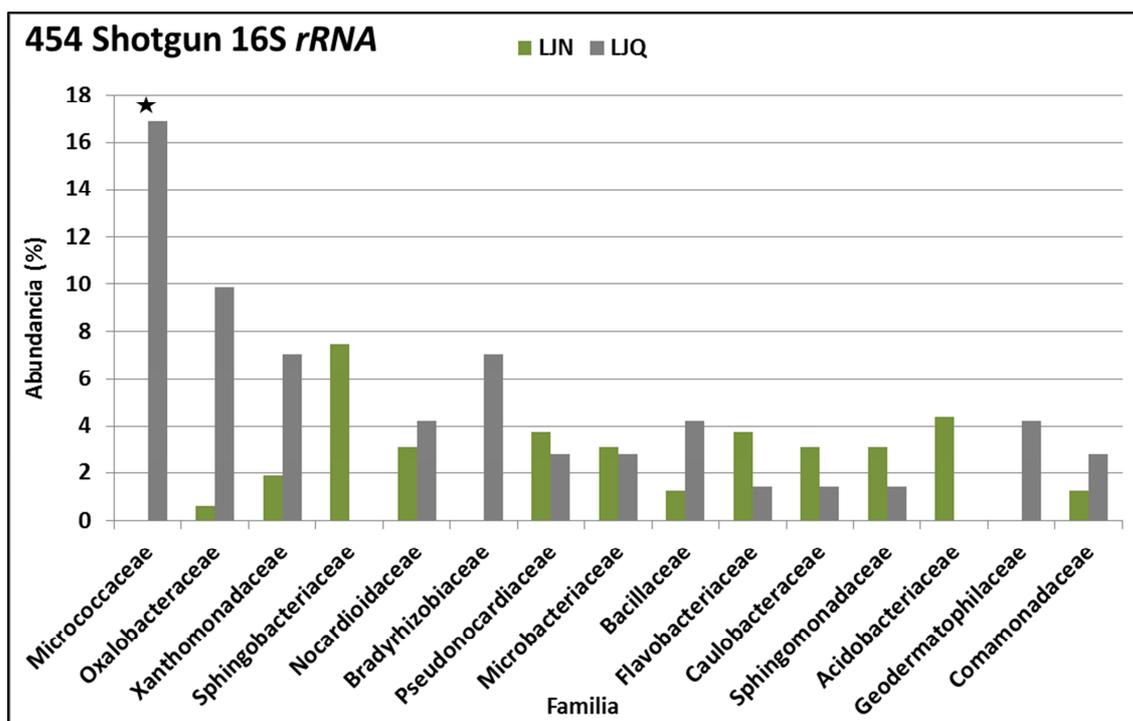


Figura R4-3. Distribución de las familias más abundantes obtenidas mediante pirosecuenciación. Solo se muestran aquellas familias que suponen al menos el 4 % de abundancia relativa en la suma de ambos sitios.

En el estudio de los metagenomas, mediante secuenciación Illumina, también se analizaron los sitios LNJ y LJQ muestreados en 2008. De ellos se obtuvieron 374.512.840 lecturas en LNJ y 386.720.378 lecturas en LJQ todas ellas con una longitud de 101 bp, que tras filtrarlas y solapar las lecturas pareadas (Paired-Ends reads), resultaron en 180.589.099 secuencias en LNJ y 186.654.779 secuencias en LJQ (Tabla R4-1) con un rango de tamaños entre 50 y 190 bp y un tamaño medio de 170 bp en ambos sitios. Tras rastrear secuencias homólogas a genes *rRNA* se obtuvieron 27.515.446 secuencias en LNJ y 30.032.562 secuencias en LJQ. Se consiguieron clasificar, a nivel de dominio, 1.328.764 secuencias en LNJ y 1.443.109 secuencias en LJQ. De éstas, se obtuvieron 65.828 secuencias en LNJ y 106.501 secuencias en LJQ pertenecientes al gen 16S *rRNA* que se clasificaron en RDP (Tabla R4-1).

Al analizar la clasificación de las secuencias homólogas al gen 16S *rRNA* de LNJ y LJQ procedentes de la secuenciación Illumina conseguimos clasificar el 65,8 % de las secuencias en LNJ y el 73,1 % en LJQ, distribuidas en 237 familias. De éstas, 70 familias muestran diferencias estadísticamente significativas entre ambos sitios. Además, 22 de estas familias tienen una abundancia relativa de, al menos, el 1 % en alguno de los sitios muestreados, mostrándose en la figura todas aquellas que presentan diferencias biológicamente relevantes (Fig. R4-4). La familia mayoritaria, como en las técnicas anteriores, es Micrococcaceae con un

Resultados Capítulo 4

13,7 % en LJQ y un 0,2 % en LNJ, dicha familia es significativamente más abundante en el encinar quemado (LJQ), marcada con una estrella en la Fig. R4-4. Lo mismo sucede con la tercera familia más abundante, Geodermatophilaceae (también del *phylum* Actinobacteria) con un 8,0 % en LJQ y un 1,1 % en LNJ; y la familia Oxalobacteraceae (clase β -Proteobacteria) con un 2,5 % en LJQ y un 1,1 % en LNJ, ambas marcadas también con una estrella (Fig. R4-4). Estas tres familias presentan el mismo perfil en el estudio de los metagenomas mediante pirosecuenciación (Fig. R4-3), aunque en dicho estudio solo es significativamente diferente la familia Micrococcaceae. Por otra parte, las familias significativamente más abundantes en el encinar no quemado (LJN), marcadas con un círculo en la Fig. R4-4, son: Solirubrobacteraceae (*phylum* Actinobacteria) con un 5,7 % en LNJ y un 4,3 % en LJQ; Bradyrhizobiaceae (clase α -Proteobacteria) con un 4,5 % en LNJ y un 1,5 % en LJQ y con un perfil contrario al mostrado en el estudio mediante pirosecuenciación (Fig. R4-3); Spartobacteriaceae (*phylum* Verrucomicrobia) con un 2,2 % en LNJ y un 1,1 % en LJQ; Acidobacteriaceae (*phylum* Acidobacteria) con un 2,4 % en LNJ y un 0,3 % en LJQ; Solibacteraceae (*phylum* Acidobacteria) con un 2,0 % en LNJ y un 0,5 % en LJQ; y Mycobacteriaceae (*phylum* Actinobacteria) también con un 2,0 % en LNJ y un 0,5 % en LJQ. Las familias Nocardioideaceae y Micromonosporaceae, ambas pertenecientes al *phylum* Actinobacteria, son las únicas, de las 22 mostradas (Fig. R4-4), sin diferencias estadísticamente significativas entre ambas muestras.

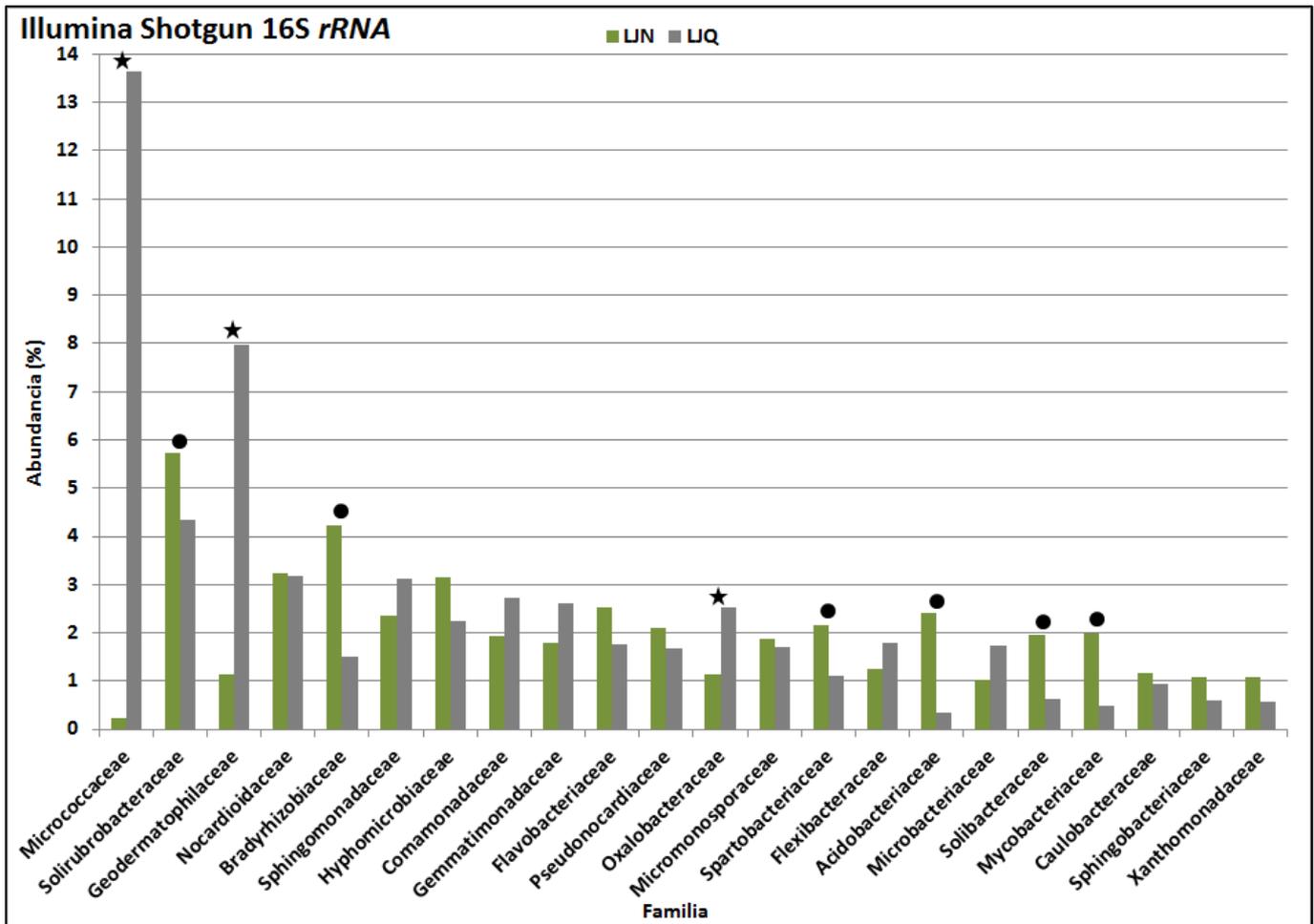


Figura R4-4. Distribución de las familias más abundantes de las muestras secuenciadas mediante tecnología Illumina. Solo se muestran aquellas familias que suponen más del 1 % de abundancia relativa en alguno de los dos sitios. Los símbolos marcan aquellas familias en las que se encontraron diferencias significativas biológicamente relevantes comparando ambos sitios con el programa STAMP, tomando un p-valor corregido menor del 0,01. El significado de los símbolos es: ★ Familias significativamente más abundantes en LJQ. ● Familias significativamente más abundantes en LNJ.

En el estudio de las librerías NGS (amplicones) de la zona V6-V8 del gen 16S *rRNA* se analizaron las 3 parcelas de cada uno de los 3 sitios (LNJ, LJQ y LJR) que comprenden las 3 situaciones objeto de este análisis, para el muestreo del año 2008. Pero en el año 2011 los sitios muestreados fueron LNJ, LJQ, LJR y LJRE, compartiendo los dos últimos las mismas 3 parcelas. De este análisis se obtuvieron 440.567 lecturas con una longitud media de 408 bp distribuidas uniformemente entre las 21 muestras secuenciadas, con valores entre 10.852 lecturas en LNJ2_11 y 27.627 lecturas en LJR2_11. Tras filtrar y descartar las lecturas con baja calidad, las posibles quimeras y las menores de 220 bp con el programa Pyrotagger, se obtuvieron 340.203 secuencias con una longitud media de 458 bp distribuidas uniformemente y con unos valores entre 8.493 secuencias en LNJ2_11 y 24.004 secuencias en LJR2_11 (Tabla

Resultados Capítulo 4

R4-2). Los cebadores utilizados para obtener estos amplicones no solo amplifican la zona V6-V8 del gen 16S *rRNA* sino que, también amplifican de forma menos específica regiones del gen 18S *rRNA* de eucariotas. Para filtrar las secuencias procarióticas se alinearon las secuencias filtradas por Pyrotagger con la herramienta Aligner de RDP frente al gen 16S *rRNA* de *E. coli* J01695 y se eliminaron todas aquellas secuencias que no alinearon en la zona V6-V8, quedando un promedio de 13.839 secuencias por muestra, con un valor mínimo de 7.762 secuencias en LJN2_11 y un valor máximo de 20.253 secuencias en LJR2_11. Para realizar análisis que requieren normalización del número de secuencias de cada muestra, se normalizaron todas las muestras a 7.762 secuencias con el programa MOTHUR (Tabla R4-2).

Tabla R4-2. Recuento de secuencias obtenidas por pirosecuenciación de la zona V6-V8 de los amplicones del gen 16S *rRNA*.

* Secuencias obtenidas tras el filtro realizado en el JGI con la herramienta Pyrotagger.

** Secuencias obtenidas tras el filtro realizado al alinear las secuencias filtradas en el JGI con la herramienta Aligner del RDP. Las secuencias filtradas con RDP se normalizaron con MOTHUR.

Muestra	Lecturas	Secuencias JGI*	Secuencias RDP**	Normalizadas
LJN1_08	20.737	15.362	11.517	7.762
LJN1_11	19.499	15.663	14.026	7.762
LJN2_08	19.143	13.984	10.802	7.762
LJN2_11	10.852	8.493	7.762	7.762
LJN3_08	16.799	12.210	9.241	7.762
LJN3_11	23.181	18.139	16.138	7.762
LJQ1_08	22.478	17.278	14.367	7.762
LJQ1_11	22.331	17.469	15.698	7.762
LJQ2_08	18.026	12.009	9.744	7.762
LJQ2_11	19.223	14.883	13.557	7.762
LJQ3_08	17.797	12.809	9.887	7.762
LJQ3_11	23.225	18.271	16.611	7.762
LJR1_08	20.387	15.967	14.438	7.762
LJR1_11	23.921	19.602	16.488	7.762
LJR2_08	24.744	18.895	17.231	7.762
LJR2_11	27.627	24.004	20.253	7.762
LJR3_08	17.485	12.621	10.493	7.762
LJR3_11	18.739	14.355	12.994	7.762
LJRE1	25.737	20.216	18.233	7.762
LJRE2	22.672	17.815	13.156	7.762
LJRE3	25.964	20.158	17.992	7.762

Al analizar la clasificación de las secuencias normalizadas homólogas al gen 16S *rRNA* obtenidas en el muestreo del año 2008 de los 3 sitios LJN, LJQ y LJR conseguimos clasificar el 81,8 % de las secuencias en LJN, el 85,7 % en LJQ y el 80,1 % en LJR a nivel de *phylum*. A nivel

de familia clasificamos el 53,9 % de las secuencias en LNJ, el 64,9 % en LJQ y el 53,2 % en LJR, distribuidas en 142 familias. De estas familias, 16 tienen diferencias estadísticamente significativas en la abundancia relativa promedio encontrada, al menos en uno de los sitios. En la Figura R4-5 se muestran las 21 familias que tienen una abundancia relativa promedio de, al menos, el 1 % en alguno de los sitios muestreados. La familia más abundante es Micrococcaceae (Actinobacteria) con un 0,2 % en LNJ, un 14,2 % en LJQ y un 2,1 % en LJR, al igual que se observó en todas las técnicas anteriores. Dicha familia es significativamente más abundante en el encinar quemado (LJQ) que en el encinar no quemado (LJN), quedando el suelo quemado (LJR) en una proporción intermedia significativa frente a los otros dos sitios. La segunda familia más abundante es Geodermatophilaceae (Actinobacteria) con un 1,3 % en LNJ, un 8,2 % en LJQ y un 2,8 % en LJR. Esta familia es significativamente más abundante en LJQ que en LNJ y LJR, no existiendo diferencias estadísticamente significativas entre los dos últimos. La familia Chitinophagaceae (*phylum* Bacteroidetes) ocupa la cuarta posición con un 1,9 % en LNJ, un 2,1 % en LJQ y un 3,3 % en LJR, siendo significativamente más abundante en el suelo quemado (LJR). La familia Bradyrhizobiaceae (clase α -Proteobacteria) ocupa la sexta posición con un 4,4 % en LNJ, un 1,1 % en LJQ y un 1,6 % en LJR, siendo significativamente más abundante en el encinar no quemado (LJN) que en los sitios afectados por el incendio. El mismo perfil se observa en las familias Solirubrobacteraceae y Mycobacteriaceae, ambas pertenecientes al *phylum* Actinobacteria, con un 2,6 % en LNJ, un 1,6 % en LJQ y un 1,1 % en LJR, en la primera familia y con un 1,4 % en LNJ y un 0,5 % en LJQ y en LJR para Mycobacteriaceae. La familia perteneciente a la clase Subdivisión3 del *phylum* Verrucomicrobia, ocupa la última posición con un 0,3 % en LNJ, un 0,2 % en LJQ y un 1,7 % en LJR, siendo significativamente más abundante en suelo no rizosférico afectado por incendio. Las familias pertenecientes al *phylum* Acidobacteria no muestran diferencias significativas en este estudio al tener en cuenta la varianza en la abundancia relativa de las 3 réplicas de cada sitio (Fig. R4-5).

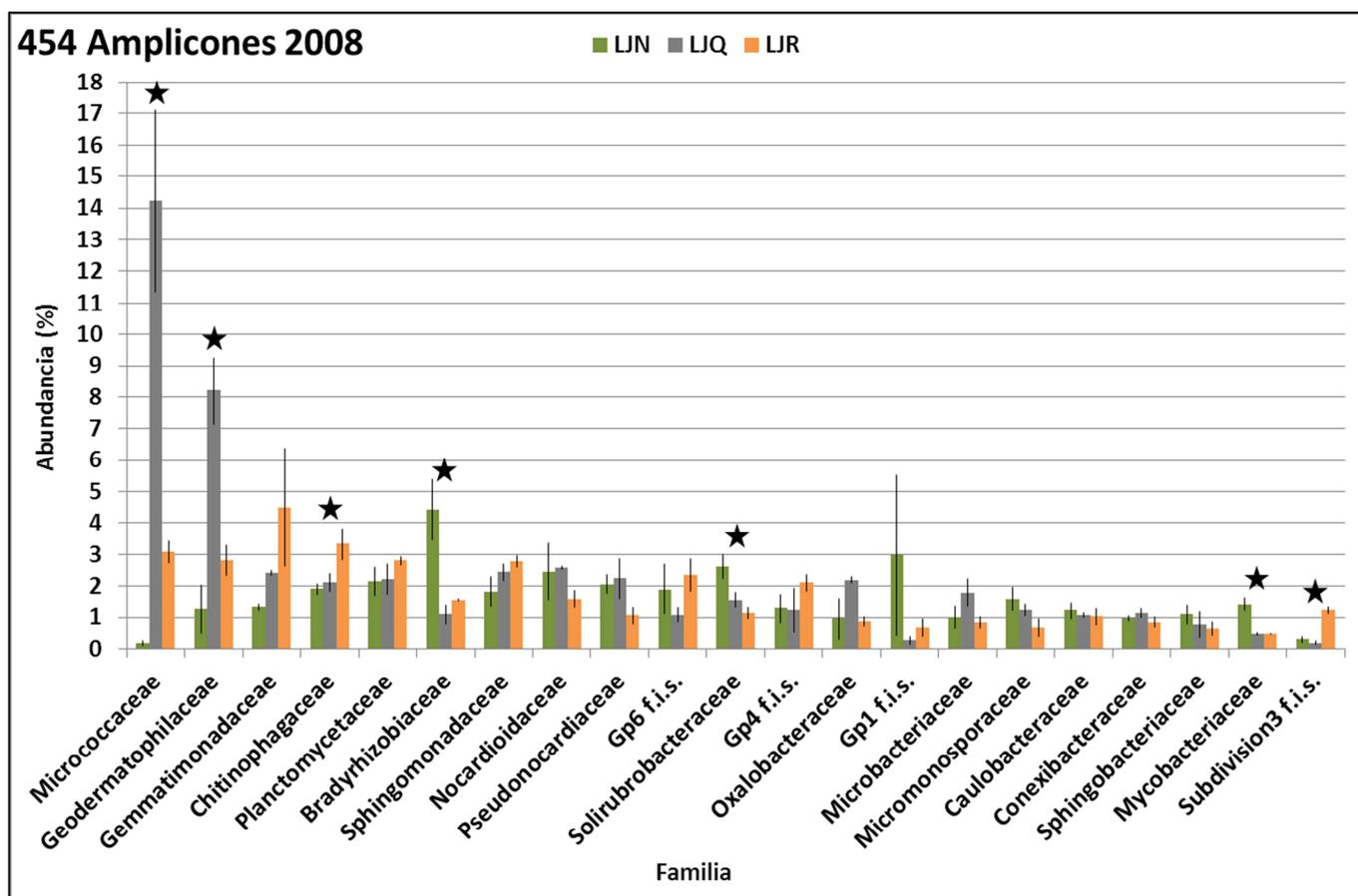


Figura R4-5. Distribución de las familias más abundantes de las muestras tomadas en el año 2008 y secuenciadas mediante pirosecuenciación de amplicones de la zona V6-V8 del gen 16S *rRNA*. Solo se muestran aquellas familias que suponen más del 1 % de abundancia relativa promedio en alguno de los tres sitios. Las barras de error se corresponden con la desviación estándar del valor medio de la abundancia relativa de cada familia para cada sitio, calculada a partir de las 3 muestras de cada sitio. En las familias marcadas con una estrella se encontraron diferencias estadísticamente significativas entre las medias de al menos uno de los tres sitios analizados mediante ANOVA. Se analizó qué sitios mostraban diferencias significativas entre sus medias mediante el test post-hoc de Tukey-Kramer con un intervalo de confianza del 95 % en el programa STAMP. f.i.s. = Familia sin nomenclatura definitiva en la clasificación taxonómica.

Al analizar la clasificación de las secuencias normalizadas homólogas al gen 16S *rRNA* obtenidas en el muestreo del año 2011 de los sitios LJN, LJQ, LJR y LJRE conseguimos clasificar el 80,7 % de las secuencias en LJN, el 82,9 % en LJQ, el 84,7 % en LJR y el 82,5 % en LJRE a nivel de *phylum*. A nivel de familia clasificamos el 48,0 % de las secuencias en LJN, el 51,7 % en LJQ, el 54,2 % en LJR y el 52,7 % en LJRE, distribuidas en 144 familias. De éstas, las 2 únicas familias con diferencias estadísticamente significativas en la abundancia relativa promedio encontrada, al menos en uno de los sitios, son la familia Geodermatophilaceae del *phylum* Actinobacteria y la familia perteneciente a la clase Gp2 del *phylum* Acidobacteria. En la Figura R4-6 se muestran las 19 familias que tienen una abundancia relativa promedio de, al menos, el 1 % en alguno de los sitios muestreados. La familia perteneciente a la clase Gp2 del *phylum* Acidobacteria no se

considera mayoritaria con una abundancia relativa promedio del 0,8 % en LJN, el 0,03 en LJQ, el 0,02 en LJR y el 0,19 en LJRE. Dicha familia, es significativamente más abundante en encinas no quemadas, ya provengan de suelo afectado por el incendio (LJRE) o no (LJN). La familia Geodermatophilaceae es la tercera más abundante con un 0,6 % en LJN, un 5,2 % en LJQ, un 4,5 % en LJR y un 2,5 % en LJRE. Dicha familia, es significativamente más abundante en encinar y suelo quemado (LJQ y LJR) que en encinas no quemadas (LJN y LJRE); pero a pesar de ello, no se observan diferencias estadísticamente significativas entre el encinar no quemado (LJN) y las encinas germinadas en suelo quemado (LJRE). Seis años después del incendio, la familia Micrococcaceae sigue siendo la más abundante en los sitios afectados por el mismo (5,4 % en LJQ, 3,1 % en LJR, 1,8 % en LJRE y 0,1 % en LJN). Sin embargo, el test ANOVA da un p-valor corregido marginal de 0,06, debido a la gran varianza entre réplicas de cada sitio, que queda reflejada en las barras de error (desviación estándar) de la figura R4-6, y debido a que la abundancia relativa de esta familia se ha reducido casi un tercio respecto a la observada en el muestreo de 2008 (Fig. R4-5). De igual forma, la familia Bradyrhizobiaceae (2,5 % en LJN, 1,1 % en LJQ, 1,9 % en LJR y 2,4 % en LJRE) sigue siendo, 6 años después del incendio, menos abundante en el encinar quemado (LJQ) que en el encinar no quemado (LJN), pero el test ANOVA da un p-valor corregido marginal de 0,06, debido al aumento de esta familia en los sitios afectados por el incendio (Fig. R4-6). Esto podría suponer una recuperación de las comunidades microbianas después del incendio.

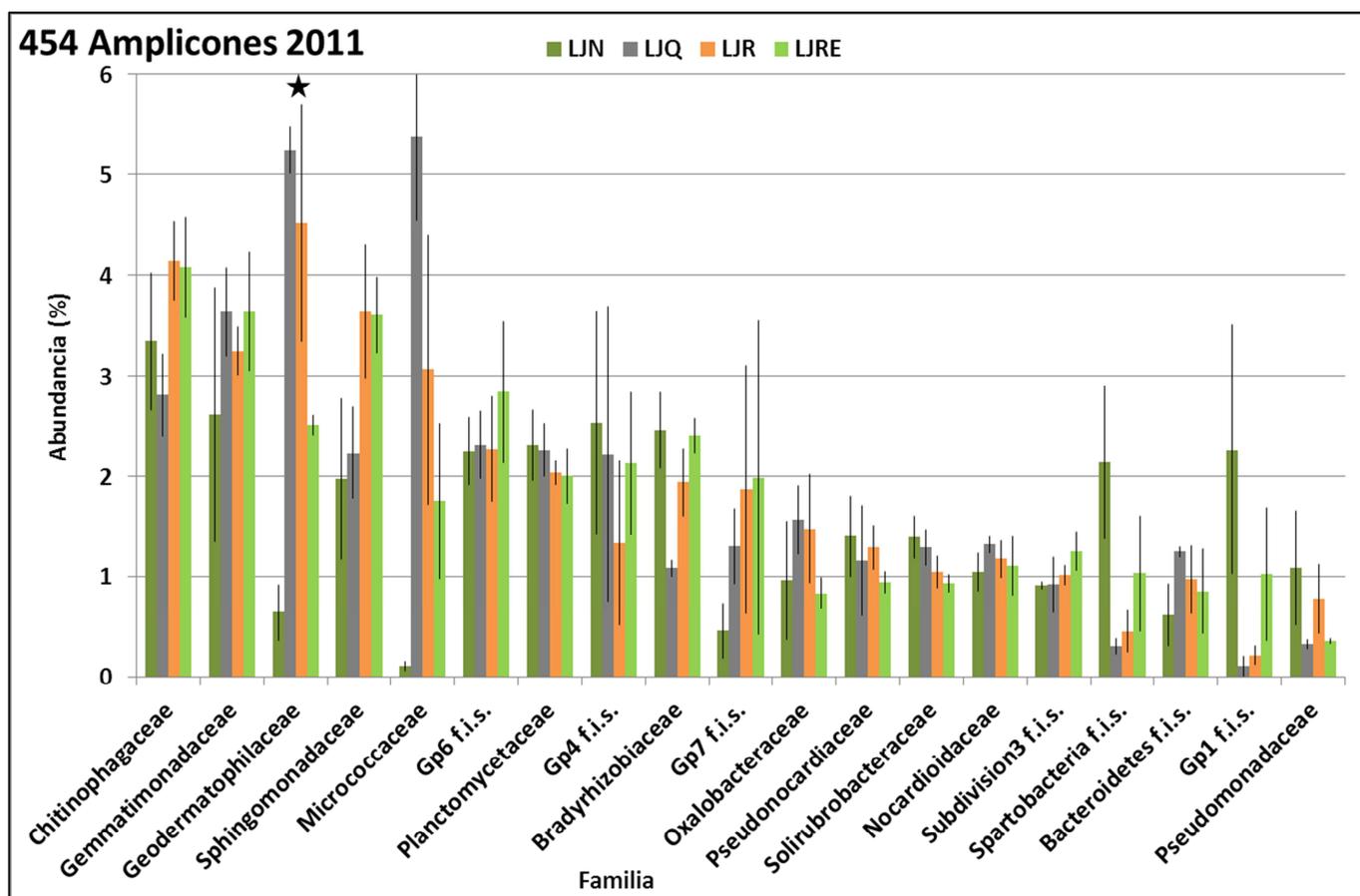


Figura R4-6. Distribución de las familias más abundantes de las muestras tomadas en el año 2011 y secuenciadas mediante pirosecuenciación de amplicones de la zona V6-V8 del gen 16S *rRNA*. Solo se muestran aquellas familias que suponen más del 1 % de abundancia relativa promedio en alguno de los cuatro sitios. Las barras de error se corresponden con la desviación estándar del valor medio de la abundancia relativa de cada familia para cada sitio, calculada a partir de las 3 muestras de cada sitio. En la familia marcada con una estrella se encontraron diferencias estadísticamente significativas entre las medias de al menos uno de los cuatro sitios analizados mediante ANOVA. Se analizó qué sitios mostraban diferencias significativas entre sus medias mediante el test post-hoc de Tukey-Kramer con un intervalo de confianza del 95 % en el programa STAMP. f.i.s. = Familia sin nomenclatura definitiva en la clasificación taxonómica.

2 Estudio de la diversidad α mediante el gen 16S *rRNA*

Para analizar la diversidad α entendida como la riqueza y equidad observadas de cada muestra y estimar la riqueza del suelo de origen de cada una de ellas, se agruparon las secuencias del gen 16S *rRNA* de cada muestra en OTUs con una distancia génica máxima del 3 %, considerando a dicha distancia que cada OTU corresponde con una especie distinta.

En las genotecas se observan distintos resultados de diversidad α entre encinar quemado (LJQ) y no quemado (LJN) muestreados en 2008. La riqueza fue menor en LJQ con 51 OTUs que en LJN, con 56 OTUs. El mismo resultado se observó en la riqueza estimada mediante el índice de Chao1 con 427 OTUs en LJQ y 515 OTUs en LJN. La distribución de las secuencias entre los distintos OTUs está próxima a la equidad total en ambos sitios. El encinar no quemado (LJN) con $J' = 0,987$ está formado por 1 OTU con 4 secuencias pertenecientes al género *Bradyrhizobium* (Proteobacteria de la familia Bradyrhizobiaceae), 2 OTUs con 2 secuencias cada uno y 53 OTUs singletons. El encinar quemado (LJQ) con $J' = 0,952$ está formado por 1 OTU con 9 secuencias pertenecientes al género *Arthrobacter* (Actinobacteria de la familia Micrococcaceae), 2 OTUs con 2 secuencias y 48 OTUs singletons. El índice de diversidad de Shannon da un valor de 3,974 en LJN y un valor de 3,741 en LJQ, siendo ambos muy próximos al valor máximo de dicho índice ($H'_{\max} = 4,111$) para 61 secuencias evaluadas por muestra (Tabla R4-3). Con los resultados obtenidos de la curva de rarefacción, que está en la zona inicial de la misma, donde es recta (Fig. R4-7), y el valor de cobertura calculado con el índice de Good's coverage que da un valor de 13,11 % en LJN y un valor de 21,31 % en LJQ (Tabla R4-3), observamos que 61 clones son insuficientes para tener muestras representativas.

Tabla R4-3. Diversidad α de las muestras estudiadas con genotecas clásicas del gen 16S *rRNA*. OTUs es el número de grupos obtenido a una distancia génica del 3 %. Chao1 es el valor medio de OTUs que se estima que hay en los suelos de origen, seguido de los valores mínimo y máximo para este índice. H' es el índice de Shannon seguido de su varianza. J' es el índice de equidad de Pielou. La cobertura se ha medido siguiendo el método de Good's coverage.

Muestra	Distancia	Secuencias	OTUs	Chao1	LCI95	UCI95	H'	varH	J'	Cobertura
LJN	0,03	61	56	515	224	1.312	3,974	0,010	0,987	13,11%
LJQ	0,03	61	51	427	187	1.092	3,741	0,017	0,952	21,31%

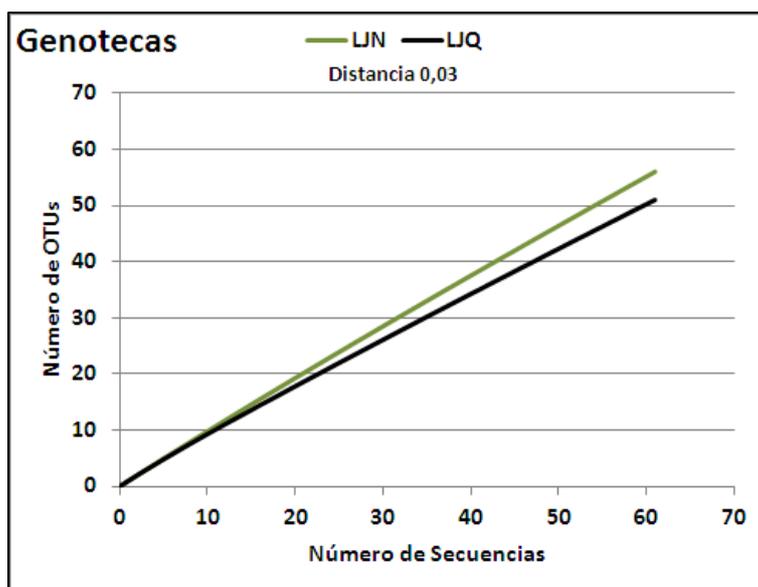


Figura R4-7. Curva de rarefacción de los sitios estudiados con genotecas clásicas del gen 16S *rRNA*. Los OTUs se agruparon a una distancia génica máxima del 3 %.

En el análisis con amplicones 454 de la zona V6-V8, para las secuencias filtradas obtenidas de las 21 muestras estudiadas (muestreos de 2008 y 2011), la riqueza mínima observada es de 2.750 OTUs en LQN2_11 y la riqueza máxima observada es de 5.986 OTUs en LJR2_11 (Tabla R4-4). Según el índice de Chao1, la riqueza mínima esperada en los suelos de origen es de 4.405 OTUs en LQJ3_08, mientras que la riqueza máxima esperada es de 10.089 OTUs en LJR2_11 (Tabla R4-5). Los valores de diversidad, medidos con el índice de Shannon, superaron el valor de 6,8 en todas las muestras, siendo el valor mínimo de 6,829 en la muestra LQJ3_08 y el máximo de 7,894 en la muestra LJR2_11. En todas las muestras las secuencias están distribuidas entre los distintos OTUs de forma muy próxima a la equidad con un valor mínimo del índice de Pielou de 0,847 en la muestra LQJ1_08 y un valor máximo de 0,923 en la muestra LJR3_08. Los valores de cobertura medidos con el índice de Good fueron superiores al 82 % en todas las muestras, siendo el valor mínimo de 82,63 % en la muestra LQN2_11 y el valor máximo de 90,02 % en la muestra LQJ3_08 (Tabla R4-4).

Para hacer un análisis comparativo de los distintos índices de diversidad α entre todas las muestras estudiadas, éstas se normalizaron a 7.762 secuencias (Tablas R4-2 y R4-4). Se observa que la muestra con menor riqueza es LQJ3_08 con 1.884 OTUs y la de mayor riqueza es LJR3_08 con 2.616 OTUs. Mientras que la muestra con menor riqueza esperada, según el índice de Chao1, es LQJ3_08 con 2.996 OTUs, la muestra con mayor riqueza esperada en su suelo de origen es LQN3_11 con 6.396 OTUs (Tabla R4-4). Pese a existir diferencias en la riqueza observada entre las distintas muestras, estas diferencias sólo son significativas estadísticamente entre el encinar quemado (LQJ) y el no quemado (LQN) en el muestreo de

2008. A su vez, se refleja un aumento de la riqueza observada entre ambos años de muestreo (2008 y 2011) en los encinares (LJN y LJQ), pero esta diferencia solo es estadísticamente significativa en el encinar no quemado (LJN). Del mismo modo se observan estas diferencias en los encinares en cuanto a la riqueza estimada mediante el índice de Chao1 (Fig. R4-8). Respecto a la diversidad, medida mediante el índice de Shannon, se observa que el valor mínimo es de 6,294 en la muestra LJQ1_08 y el máximo 7,247 en la muestra LJR3_08 (Tabla R4-4). Al comparar la diversidad de los cuatro sitios mediante ANOVA se observan diferencias estadísticamente significativas y según el test post hoc HSD de Tukey el sitio LJQ es significativamente el menos diverso de todos (Fig. R4-9). En cuanto a la equidad, medida mediante el índice de Pielou vemos que, el valor mínimo es de 0,832 en LJQ1_08 y el máximo es de 0,921 en LJR3_08. El estudio de la equidad nos da un resultado similar al observado en el estudio de la diversidad, siendo LJQ el sitio con menor equidad (Fig. R4-10). Ambos índices muestran una diferencia en sus valores promedio más acusada en el muestreo de 2008 (Figs. R4-9 y R4-10), debido a OTUs predominantes en el encinar quemado. Según los resultados obtenidos por el índice de cobertura de Good (Tabla R4-4) con un mínimo del 81,56 % en LJN3_11 y un máximo de 88,25 % en LJQ_08; y las curvas de rarefacción de los cuatro sitios estudiados (Fig. R4-11), con 7.762 secuencias por muestra no se consigue secuenciar el 100 % de las especies microbianas (OTUs al 0,03) que habitan en los suelos forestales estudiados. Sin embargo, el esfuerzo de muestreo realizado es suficiente para obtener muestras representativas de cada uno de ellos. Según las curvas de esfuerzo hay diferencias estadísticamente significativas entre años de muestreo para los encinares (LJN y LJQ). Pese a estas diferencias debidas al año de muestreo en los encinares estudiados, también se observan diferencias estadísticamente significativas entre el encinar quemado (LJQ) y el no quemado (LJN), pero solo en el muestreo de 2008, al igual que ocurre con la riqueza y la diversidad. Además, LJQ de 2008 es el que requiere un menor esfuerzo de muestreo para obtener secuencias de todas las especies microbianas que habitan en su comunidad rizosférica (Fig. R4-11).

Tabla R4-4. Diversidad α de las muestras estudiadas con amplicones 454 de la zona V6-V8 del gen 16S *rRNA*. OTUs es el número de grupos obtenido a una distancia génica del 3 %. Chao1 es el valor medio de OTUs que se estima que hay en los suelos de origen, seguido de los valores mínimo y máximo para este índice. H' es el índice de Shannon seguido de su varianza. J' es el índice de equidad de Pielou. La cobertura se ha medido siguiendo el método de Good's coverage. La muestra es el número de la parcela de cada sitio de muestreo, seguido de un número que indica el año de muestreo.

Muestra	Estado	Distancia	Secuencias	OTUs	Chao1	LCI95	UCI95	H'	varH	J'	Cobertura
LJN1_08	Filtradas	0,03	15.362	3.592	5.611	5.363	5.894	7,308	0,00015	0,893	89,41%
	Normalizadas	0,03	7.762	2.053	3.746	3.485	4.055	6,719	0,00033	0,881	86,18%

Resultados Capítulo 4

LJN1_11	Filtradas	0,03	15.663	4.087	6.928	6.606	7.291	7,534	0,00013	0,906	87,22%
	Normalizadas	0,03	7.762	2.404	4.389	4.107	4.718	7,088	0,00024	0,911	83,56%
LJN2_08	Filtradas	0,03	13.984	3.588	5.688	5.432	5.979	7,454	0,00013	0,911	88,14%
	Normalizadas	0,03	7.762	2.179	3.633	3.415	3.889	6,958	0,00024	0,905	85,96%
LJN2_11	Filtradas	0,03	8.493	2.750	4.982	4.688	5.321	7,279	0,00020	0,919	82,63%
	Normalizadas	0,03	7.762	2.435	4.527	4.234	4.869	7,092	0,00025	0,910	82,99%
LJN3_08	Filtradas	0,03	12.210	3.292	5.183	4.943	5.457	7,403	0,00014	0,914	87,53%
	Normalizadas	0,03	7.762	2.219	3.849	3.604	4.136	7,001	0,00023	0,909	85,70%
LJN3_11	Filtradas	0,03	18.139	4.982	8.847	8.452	9.287	7,714	0,00012	0,906	85,99%
	Normalizadas	0,03	7.762	2.600	4.951	4.632	5.320	7,176	0,00026	0,913	81,56%
LJQ1_08	Filtradas	0,03	17.278	3.723	6.139	5.852	6.465	6,964	0,00023	0,847	89,67%
	Normalizadas	0,03	7.762	1.929	3.790	3.498	4.137	6,294	0,00052	0,832	86,18%
LJQ1_11	Filtradas	0,03	17.469	4.166	7.288	6.937	7.682	7,321	0,00016	0,878	88,12%
	Normalizadas	0,03	7.762	2.280	4.196	3.919	4.519	6,826	0,00035	0,883	84,19%
LJQ2_08	Filtradas	0,03	12.009	3.288	5.404	5.145	5.699	7,162	0,00023	0,884	86,26%
	Normalizadas	0,03	7.762	2.041	3.741	3.483	4.044	6,536	0,00041	0,858	85,71%
LJQ2_11	Filtradas	0,03	14.883	3.456	5.857	5.563	6.190	7,105	0,00019	0,872	88,56%
	Normalizadas	0,03	7.762	2.106	3.874	3.608	4.188	6,708	0,00033	0,877	85,44%
LJQ3_08	Filtradas	0,03	12.809	2.874	4.405	4.195	4.648	6,829	0,00028	0,858	90,02%
	Normalizadas	0,03	7.762	1.884	2.996	2.817	3.209	6,352	0,00049	0,842	88,25%
LJQ3_11	Filtradas	0,03	18.271	4.510	7.848	7.489	8.251	7,456	0,00014	0,886	87,64%
	Normalizadas	0,03	7.762	2.430	4.356	4.090	4.665	6,953	0,00033	0,892	82,94%
LJR1_08	Filtradas	0,03	15.967	3.552	5.916	5.626	6.247	7,275	0,00014	0,890	89,47%
	Normalizadas	0,03	7.762	2.080	3.649	3.406	3.937	6,853	0,00027	0,897	86,58%
LJR1_11	Filtradas	0,03	19.602	5.134	8.385	8.053	8.755	7,774	0,00010	0,910	87,53%
	Normalizadas	0,03	7.762	2.446	4.509	4.219	4.847	7,129	0,00023	0,914	83,17%
LJR2_08	Filtradas	0,03	18.895	4.807	8.049	7.708	8.431	7,690	0,00010	0,907	87,69%
	Normalizadas	0,03	7.762	2.534	4.610	4.324	4.943	7,214	0,00021	0,920	82,65%
LJR2_11	Filtradas	0,03	24.004	5.986	10.089	9.697	10.523	7,894	0,00008	0,908	88,04%
	Normalizadas	0,03	7.762	2.356	4.292	4.015	4.615	7,054	0,00024	0,908	83,81%
LJR3_08	Filtradas	0,03	12.621	3.991	7.032	6.690	7.418	7,653	0,00013	0,923	83,72%
	Normalizadas	0,03	7.762	2.616	4.643	4.370	4.958	7,247	0,00021	0,921	81,90%
LJR3_11	Filtradas	0,03	14.355	3.581	6.271	5.949	6.636	7,205	0,00019	0,880	87,42%
	Normalizadas	0,03	7.762	2.230	3.895	3.649	4.183	6,860	0,00032	0,890	85,24%
LJRE1	Filtradas	0,03	20.216	5.078	8.721	8.347	9.137	7,707	0,00010	0,903	87,69%
	Normalizadas	0,03	7.762	2.521	4.728	4.420	5.085	7,201	0,00021	0,919	82,61%
LJRE2	Filtradas	0,03	17.815	4.713	7.685	7.372	8.035	7,583	0,00014	0,897	87,17%
	Normalizadas	0,03	7.762	2.274	4.234	3.948	4.570	6,896	0,00031	0,892	84,39%
LJRE3	Filtradas	0,03	20.158	4.969	8.206	7.867	8.584	7,721	0,00009	0,907	88,35%
	Normalizadas	0,03	7.762	2.478	4.443	4.168	4.764	7,164	0,00022	0,917	83,26%

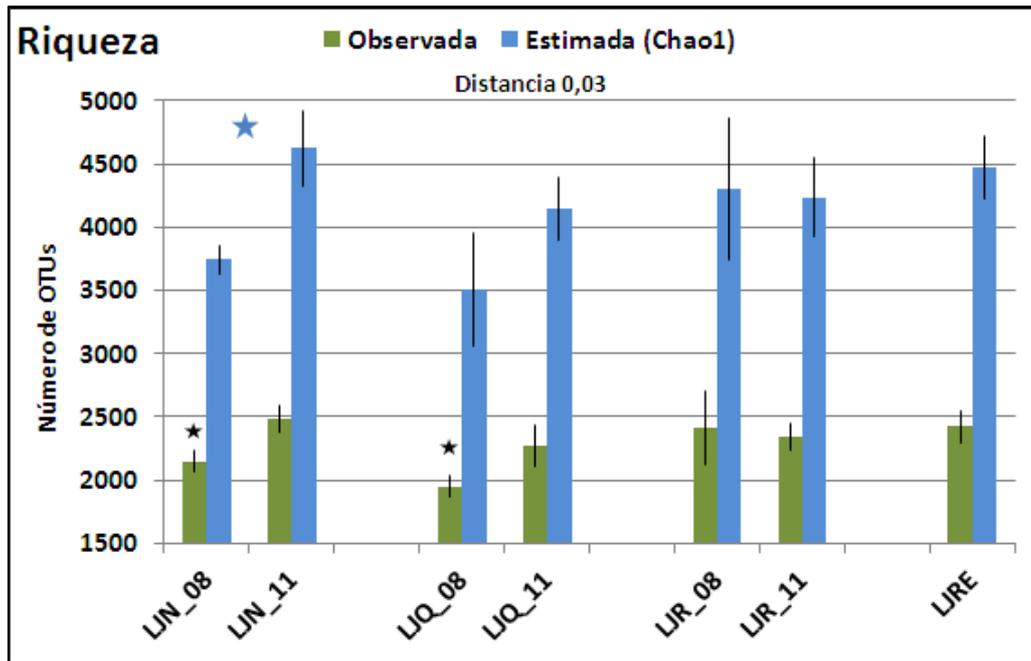


Figura R4-8. Riqueza observada y estimada de especies medidas como OTUs a una distancia del 3 % obtenidas tras la pirosecuenciación de amplicones de la zona V6-V8 del gen 16S *rRNA*. Las barras de error muestran la desviación estándar de la media observada entre las réplicas de cada sitio para cada año de muestreo. ★ Diferencias estadísticamente significativas en los valores medios de riqueza observada y estimada comprobados mediante la prueba t de Student para dos muestras emparejadas. ★ Diferencias estadísticamente significativas en los valores medios de riqueza observada comprobados mediante la prueba t de Student para dos muestras independientes suponiendo varianzas iguales.

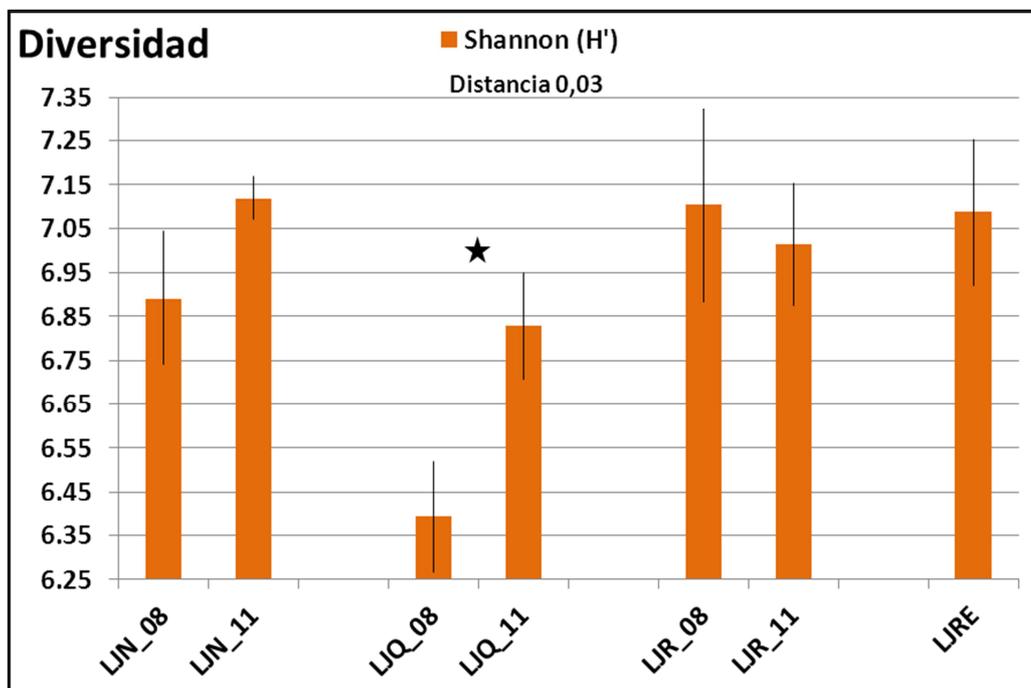


Figura R4-9. Diversidad de especies medidas como OTUs a una distancia del 3 %, analizada con el índice de Shannon tras la pirosecuenciación de amplicones de la zona V6-V8 del gen 16S *rRNA*. Las barras de error muestran la desviación estándar de la media observada entre las réplicas de cada sitio para cada año de muestreo. La estrella indica diferencias estadísticamente significativas en los valores medios de diversidad comprobados mediante ANOVA y test post hoc HSD de Tukey.

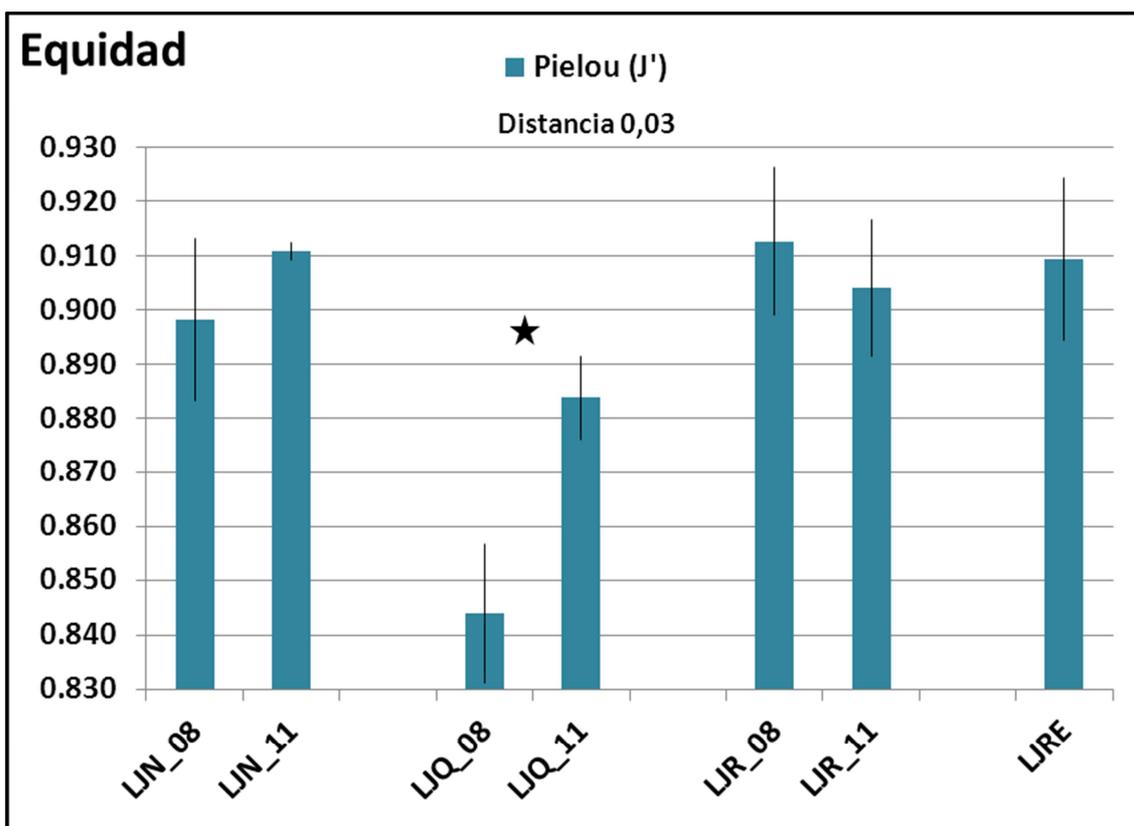


Figura R4-10. Equidad de la distribución de las distintas especies medidas como OTUs a una distancia del 3 % tras la pirosecuenciación de amplicones del gen 16S *rRNA*, analizada con el índice de Pielou. Las barras de error muestran la desviación estándar de la media observada entre las réplicas de cada sitio para cada año de muestreo. La estrella indica diferencias estadísticamente significativas en los valores medios de equidad comprobados ANOVA y test post hoc HSD de Tukey.

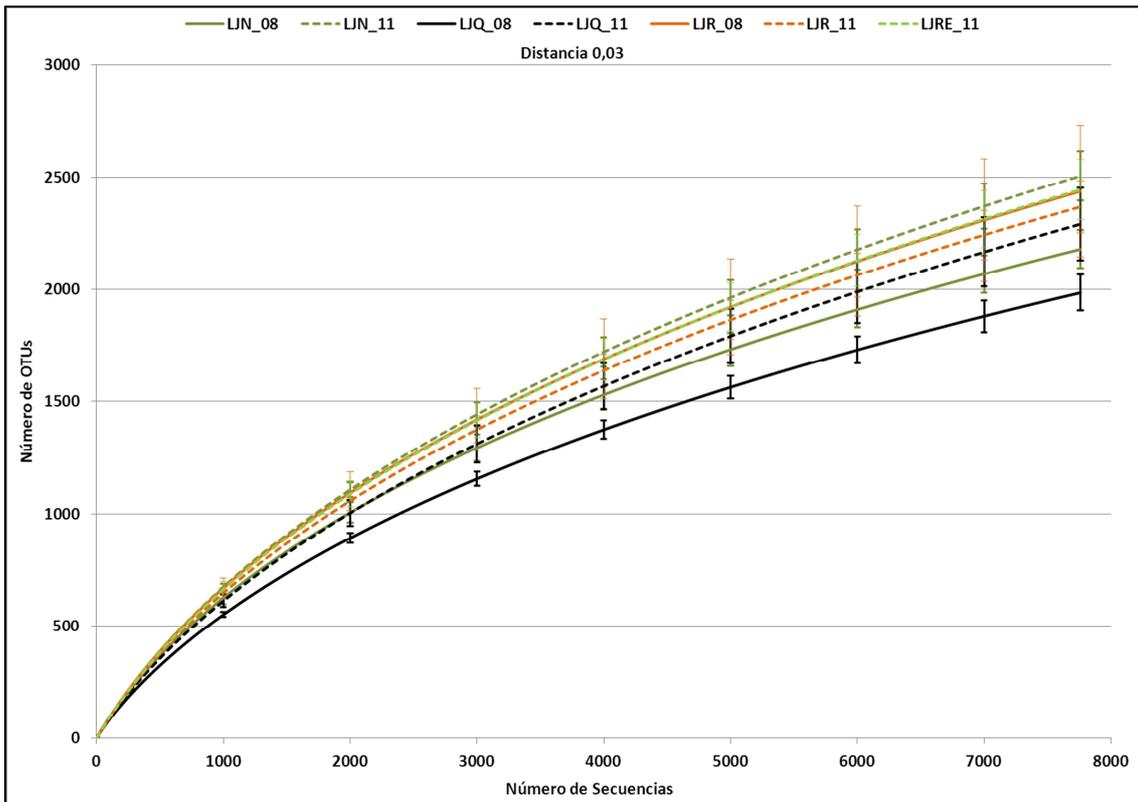


Figura R4-11. Curvas de rarefacción de especies medidas como OTUs a una distancia del 3 % tras la pirosecuenciación de amplicones del gen 16S *rRNA*. Las barras de error muestran la desviación estándar de la media observada entre las réplicas de cada sitio para cada año de muestreo, engrosadas en los sitios con diferencias significativas. — Muestras de 2008. Muestras de 2011.

3 Estudio de la diversidad β mediante empleo del gen 16S *rRNA*

Este estudio también se ha realizado con las secuencias obtenidas por secuenciación Sanger de las genotecas en *E.coli* y los amplicones 454 de la zona V6-V8 del gen 16S *rRNA* (Tablas R4-3 y R4-4). Entendiendo por diversidad β la diferencia de composición de las comunidades microbianas de las muestras comparadas dos a dos.

En el análisis con genotecas se observa que el índice de disimilitud de Jaccard tiene un valor de 0,935 indicando que, a nivel de especie, ambos sitios son muy diferentes y se confirma en el clustering de los dos sitios juntos, ya que solo 3 OTUs están compartidos entre los dos sitios estudiados. El primero de estos OTUs pertenecería a una especie del género *Bradyrhizobium* del *phylum* Proteobacteria, con 4 secuencias en LJN y 1 secuencia en LJQ. El segundo OTU pertenecería a una especie del género *Gp6* del *phylum* Acidobacteria. El tercer OTU pertenecería a una especie del género *Gemmatimonas* del *phylum* Gemmatimonadetes. Estos 2 últimos OTUs compartidos tienen 1 secuencia por sitio. Pese a la baja cobertura de muestreo de este análisis mostrado por el índice de Good's (Tabla R4-3), el estudio estadístico de ambas comunidades mediante β -Libshuff indica que se encuentran diferencias estadísticamente significativas entre ellas, con un p-valor de 0,0079.

En el análisis con amplicones 454 de la zona V6-V8, se filtraron las secuencias después de ser alineadas en RDP y se normalizaron las muestras a 7.762 secuencias (Tablas R4-2 y R4-4). Se realizó el agrupamiento jerárquico aglomerativo mediante UPGMA, sin tener en cuenta la abundancia de los OTUs, analizados mediante Jaccard, ni la abundancia en los nodos del árbol filogenético, analizados mediante Unifrac (Fig. R4-13). Dicho análisis pone de manifiesto que no hay una separación clara entre los distintos sitios estudiados. Siendo las más próximas entre si dos réplicas del encinar quemado muestreado en 2008 (LJQ1.08 y LJQ2.08) unidas en el 64 % de los árboles obtenidos al 63,8 % de distancia (Fig. R4-13). Todas las muestras están agrupadas al 79 % de distancia, pero se separan a una distancia inferior al 63,8 % indicando que las comunidades microbianas (tanto a nivel de especie en Jaccard como teniendo en cuenta la pertenencia a distintas ramas del árbol filogenético en Unweighted-Unifrac) de cada muestra son muy distintas de las del resto (Fig. R4-13). Este resultado se confirma con los análisis del test estadístico Unifrac y el AMOVA que dan p-valores inferiores a 0,0001 entre todas las muestras comparadas dos a dos. Esto nos indica además que las comunidades microbianas de cada muestra son significativamente diferentes del resto de muestras (tanto a

nivel de especie en AMOVA como teniendo en cuenta la pertenencia a distintas ramas del árbol filogenético en Unweighted-Unifrac).

El análisis también se realizó mediante UPGMA teniendo en cuenta la abundancia de cada OTU para cada muestra (Figura R4-14), analizado mediante el comando Weighted-Unifrac en MOTHUR y el peso (número de secuencias) de cada nodo en el árbol filogenético analizado mediante Unifrac. En este caso, las muestras se dividen en 2 grupos con un soporte del 100 % de los árboles creados. Un grupo está constituido en su mayoría por muestras de encinar no quemado (LJN) muestreadas en 2008 (Fig.R4-14). El otro grupo se subdivide en el 97 % de los árboles creados en 2 subgrupos. Uno de estos subgrupos contiene las muestras del encinar no quemado (LJN) muestreadas en 2011 y las de las encinas germinadas en suelo afectado por el incendio (LJRE). El otro subgrupo contiene todas las muestras del encinar quemado (LJQ) y el suelo no rizosférico afectado por el incendio (LJR). En este subgrupo de muestras afectadas por el incendio (Fig. R4-14), las 3 réplicas del encinar quemado muestreado en 2008 se separan del resto de muestras en el 88 % de los árboles creados al 14,2 % de distancia. Dichas réplicas (LJQ1.08, LJQ2.08 y LJQ3.08) son las muestras más próximas, agrupadas las 3 al 9,2 % y LJQ2.08 y LJQ3.08 al 7 % de distancia en el 100 % de los árboles creados (Fig. R4-14). La diferencia entre las encinas no afectadas por el incendio (LJN y LJRE) y las muestras quemadas (LJQ y LJR) tiene robustez estadística en el análisis Jackknife, pero todas las muestras están agrupadas al 18,3 % de distancia (Fig. R4-14). La similitud de la estructura filogenética de todas las muestras se corresponde con los resultados de los análisis del test estadístico Unifrac y el AMOVA en los que no aparecen diferencias significativas entre ninguna de las muestras rizosféricas comparadas dos a dos. Este tipo de análisis nos indica que se trata de comunidades microbianas muy parecidas, aunque permite distinguir claramente las muestras afectadas por el incendio, al menos, 6 años después de que éste tuviera lugar.

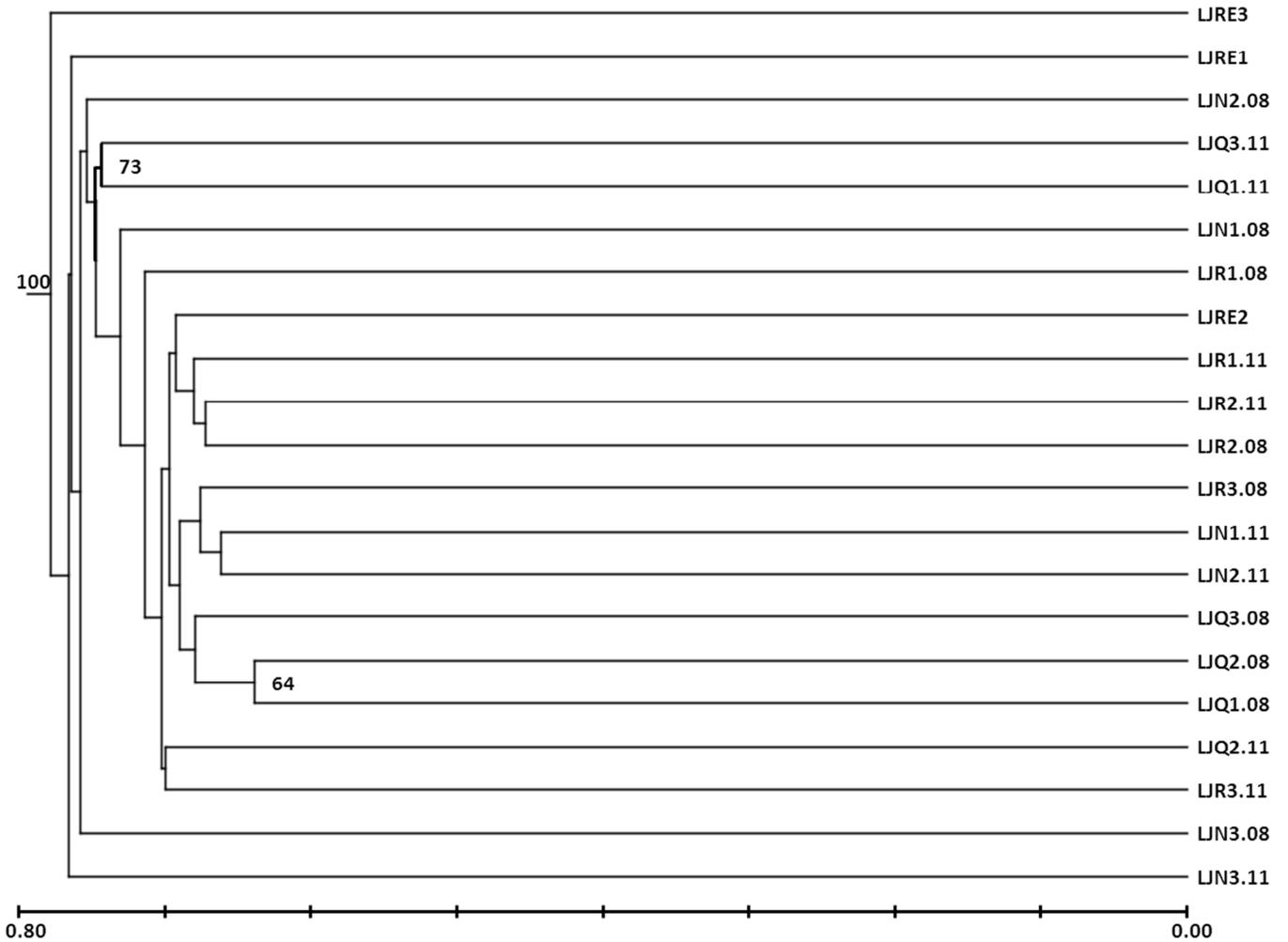


Figura R4-13. Agrupamiento jerárquico aglomerativo usando la matriz de distancias Unifrac y el algoritmo UPGMA teniendo en cuenta solo Presencia/Ausencia de secuencias de cada muestra en cada nodo del árbol filogenético. Para las secuencias normalizadas procedentes de amplicones 454 de la zona V6-V8 del gen 16S *rRNA*. Análisis realizado con las distancias filogenéticas encontradas entre las secuencias de cada muestra. Se muestran los valores de los nodos obtenidos por el método de Jackknife con 1.000 iteraciones de 6.000 secuencias seleccionadas al azar de cada muestra con un valor igual o superior al 60 %.

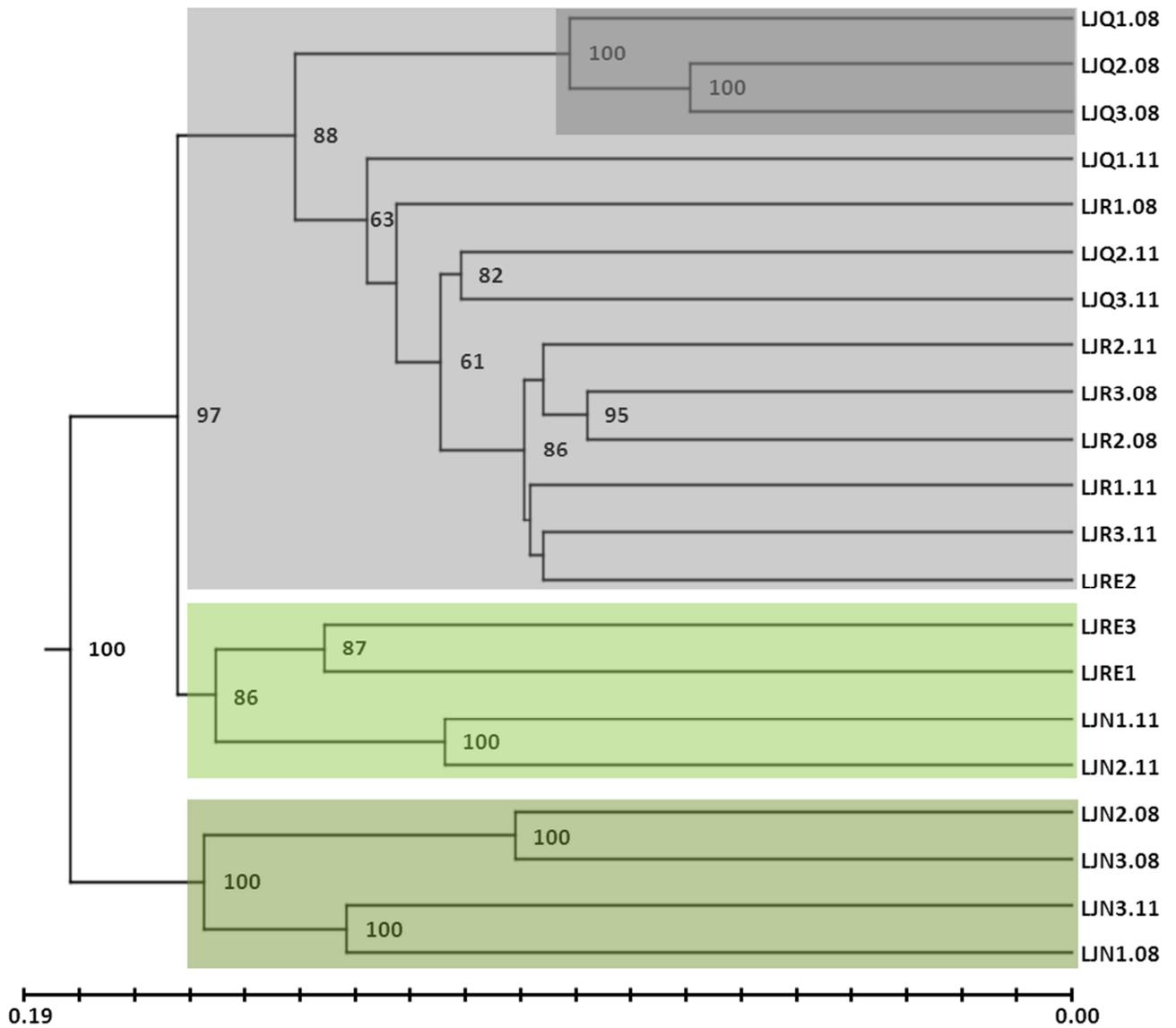


Figura R4-14. Agrupamiento jerárquico aglomerativo usando la matriz de distancias Unifrac y el algoritmo UPGMA teniendo en cuenta la abundancia de secuencias de cada muestra en cada nodo del árbol filogenético. Para las secuencias normalizadas procedentes de amplicones 454 de la zona V6-V8 del gen 16S *rRNA*. Análisis realizado con las distancias filogenéticas encontradas entre las secuencias de cada muestra. Se muestran los valores de los nodos obtenidos por el método de Jackknife con 1.000 iteraciones de 6.000 secuencias seleccionadas al azar de cada muestra con un valor igual o superior al 60 %.

Al comparar las distintas muestras teniendo en cuenta las secuencias normalizadas que clasifican en RDP a nivel de familia, vemos que éstas se agrupan por su composición filogenética en 3 grupos, atendiendo a la PC1 (primera coordenada principal) que explica el 68,3 % de la varianza entre muestras (Fig. R4-15). El primer grupo está constituido por las muestras del encinar quemado (LJQ) muestreado en 2008, con valores negativos más altos en PC1. El segundo grupo, próximo al tercero pero con valores negativos en PC1, está compuesto por las muestras del encinar quemado (LJQ) muestreado en 2011 y una muestra del suelo no

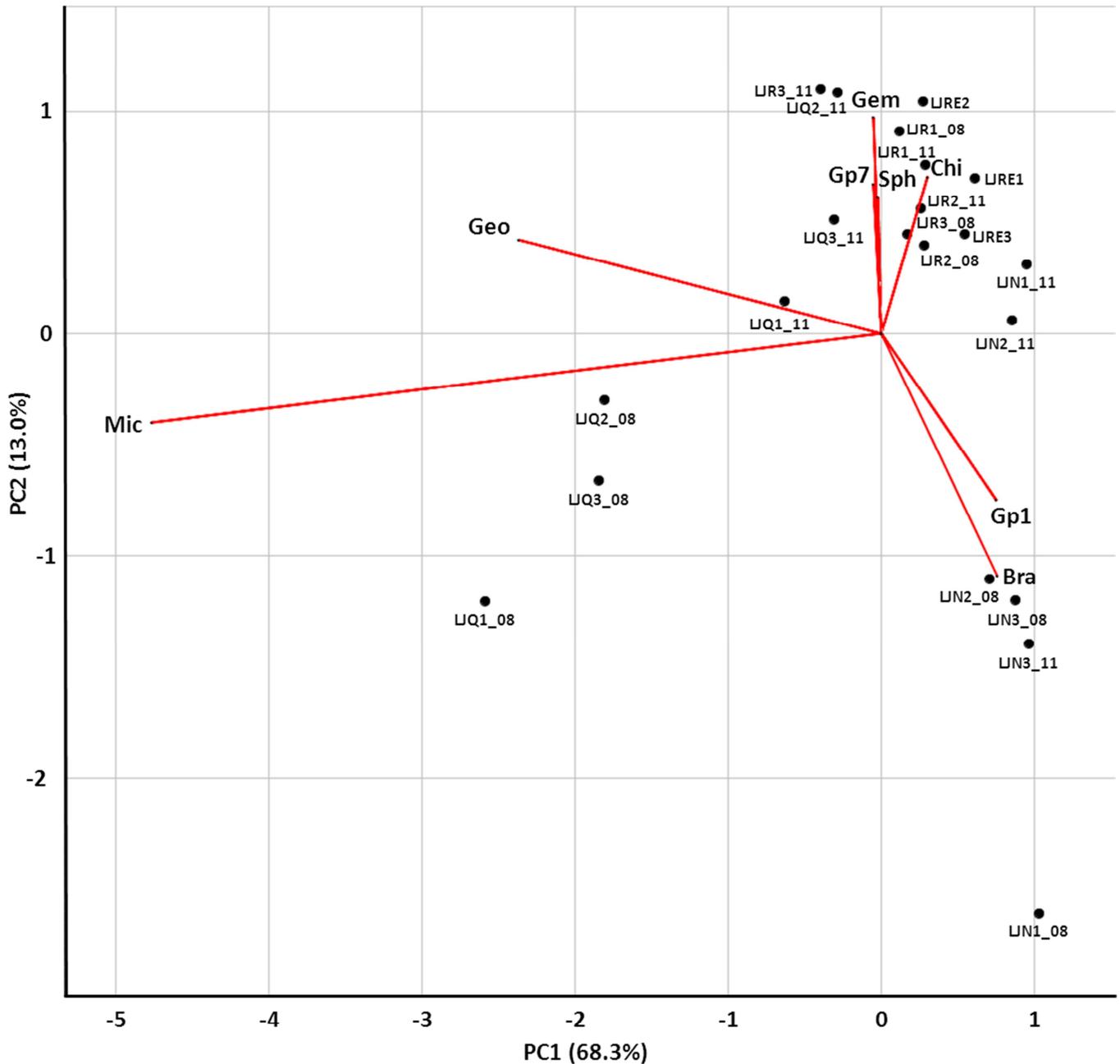


Figura R4-16. Análisis de Correlación Canónica (CCA). Realizado en el programa Ginkgo con las 21 muestras pertenecientes a los 4 sitios muestreados en ambos años de muestreo, teniendo en cuenta las secuencias clasificadas a nivel de familia. Se muestran solo los vectores más relevantes para los 2 primeros componentes. Mic, Micrococcaceae; Geo, Geodermatophilaceae; Gp7, Gp7 familia_ *incertae_sedis*; Gem, Gemmatimonadaceae; Sph, Sphingomonadaceae; Chi, Chitinophagaceae; Gp1, Gp1 familia_ *incertae_sedis*; Bra, Bradyrhizobiaceae.

Se han comparado los 4 sitios muestreados (LJN, LJQ, LJR y LJRE) teniendo en cuenta la abundancia relativa de las familias dominantes (Figs. R4-5 y R4-6) y los parámetros ambientales obtenidos de dichos sitios (Tablas R1-1, R1-2, R1-4 y R1-5). En el programa CANOCO se introdujeron como variables ambientales suplementarias el año de muestreo, la

Resultados Capítulo 4

presencia o ausencia de rizosfera de encina y si proceden de un sitio afectado (quemado) o no por el incendio (no quemado). En la figura R4-17 se observa una distribución parecida a la obtenida en el PCoA (Fig. R4-15) con una clara separación entre encinar quemado (LJQ), suelo quemado sin y con encinas germinadas posteriormente (LJR y LJRE) y encinar no quemado (LJN) en el Eje 1, el cual explica el 69,9 % de la varianza observada y muestra una correlación positiva entre el pH y la alteración de las comunidades microbianas provocada por el incendio. Los sitios se separan en el Eje 2, el cual explica el 26,9 % de la varianza, principalmente por el año en que se muestrearon y si son o no muestras asociadas a la rizosfera de encina. También afecta en la ordenación la textura del suelo pero no es estadísticamente significativa. Se realizó el test de Monte Carlo para evaluar si los factores ambientales introducidos tienen relación con la ordenación obtenida de los sitios, debido a la abundancia relativa de las familias dominantes. Dicho test fue estadísticamente significativo (999 permutaciones, p-valor = 0,003) indicando que, al menos, uno de los factores ambientales está correlacionado con la ordenación encontrada. En este análisis sólo el pH tiene una correlación estadísticamente significativa (p-valor = 0,002).

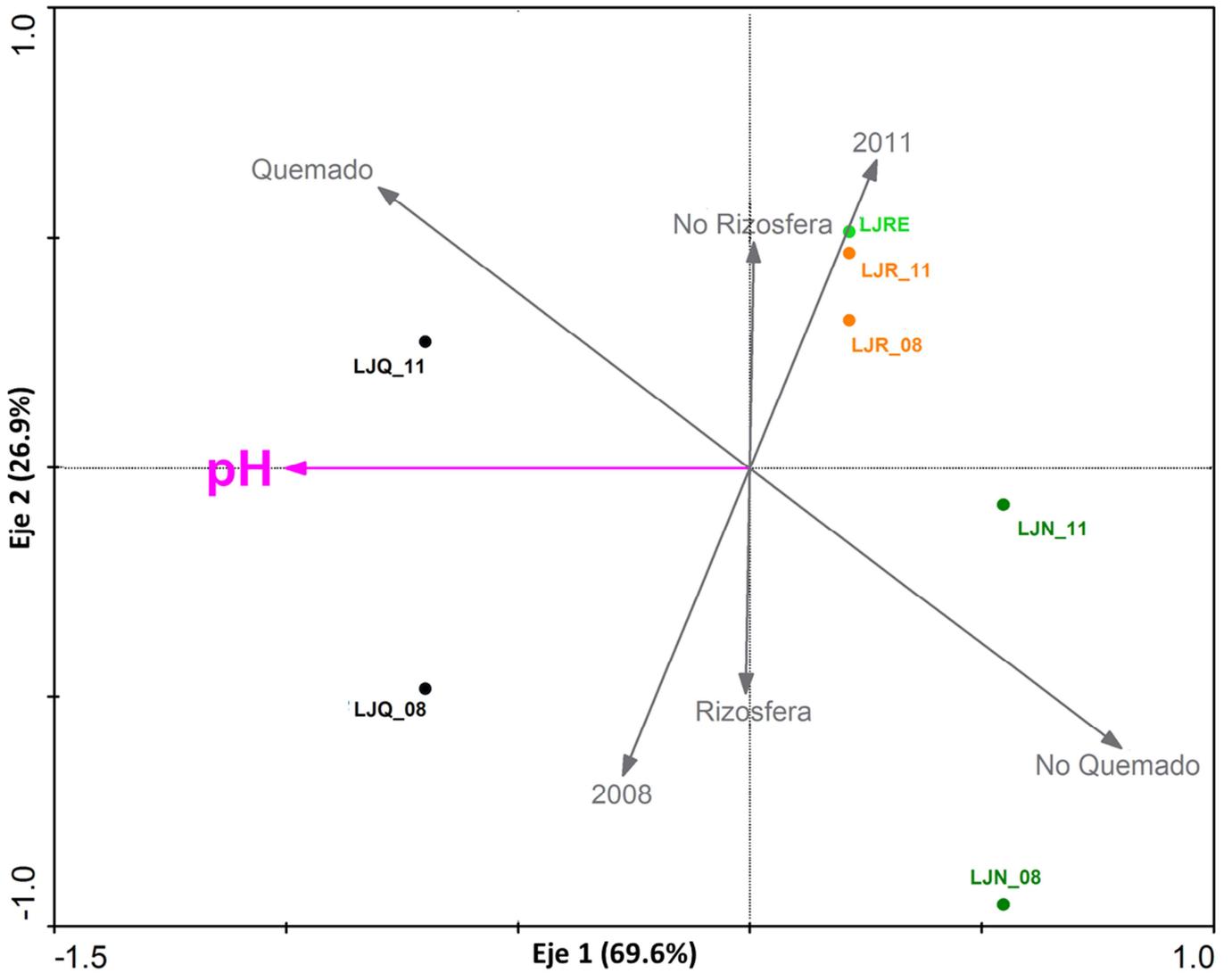


Figura R4-17. Análisis de Correspondencias Canónicas (CCA). Realizado en el programa CANOCO con los 4 sitios muestreados en ambos años de muestreo, teniendo en cuenta las secuencias promedio de cada sitio clasificadas a nivel de familia. En gris se muestran los vectores de las variables ambientales utilizadas como suplementarias. En fucsia se muestran solo los vectores de las variables ambientales que afectan significativamente a la ordenación de los sitios.

Capítulo 5: Dominancia del género *Arthrobacter* en la rizosfera de encinas quemadas

Dados los resultados obtenidos en el capítulo anterior, hemos realizado un análisis más detallado de la familia Micrococcaceae, pues supone más del 12 % de las secuencias clasificadas en el encinar quemado (LJQ) en el muestreo del año 2008, y más del 5 % para el muestreo del año 2011 (Figs. R4-2 a R4-6 y R4-16). Observamos que el género más abundante de dicha familia es *Arthrobacter*, ya que el 93,6 % de las secuencias pertenecientes a esta familia en el sitio LJQ muestreado en 2008, obtenidas mediante pirosecuenciación de los amplicones del gen 16S *rRNA* de la regiones hipervariables V6-V8, clasifican en dicho género, mostrando una abundancia relativa del 13,3 % (Fig. R5-1) en el total de la comunidad microbiana. En la Figura R5-1 se muestran los 12 géneros que tienen una abundancia relativa promedio de, al menos, el 1 % en alguno de los sitios muestreados. El género *Blastococcus* de la familia Geodermatophilaceae presenta el mismo perfil que el género *Arthrobacter* y ocupa la segunda posición en abundancia con un 5,4 % en LJQ y un 0,7 % en LJN. Además ambos géneros pertenecen al *phylum* Actinobacteria, lo que explica la predominancia del mismo en las muestras de la rizosfera quemada. El resto de géneros no alcanzan el 5 % de abundancia relativa promedio en suma de ambos sitios estudiados.

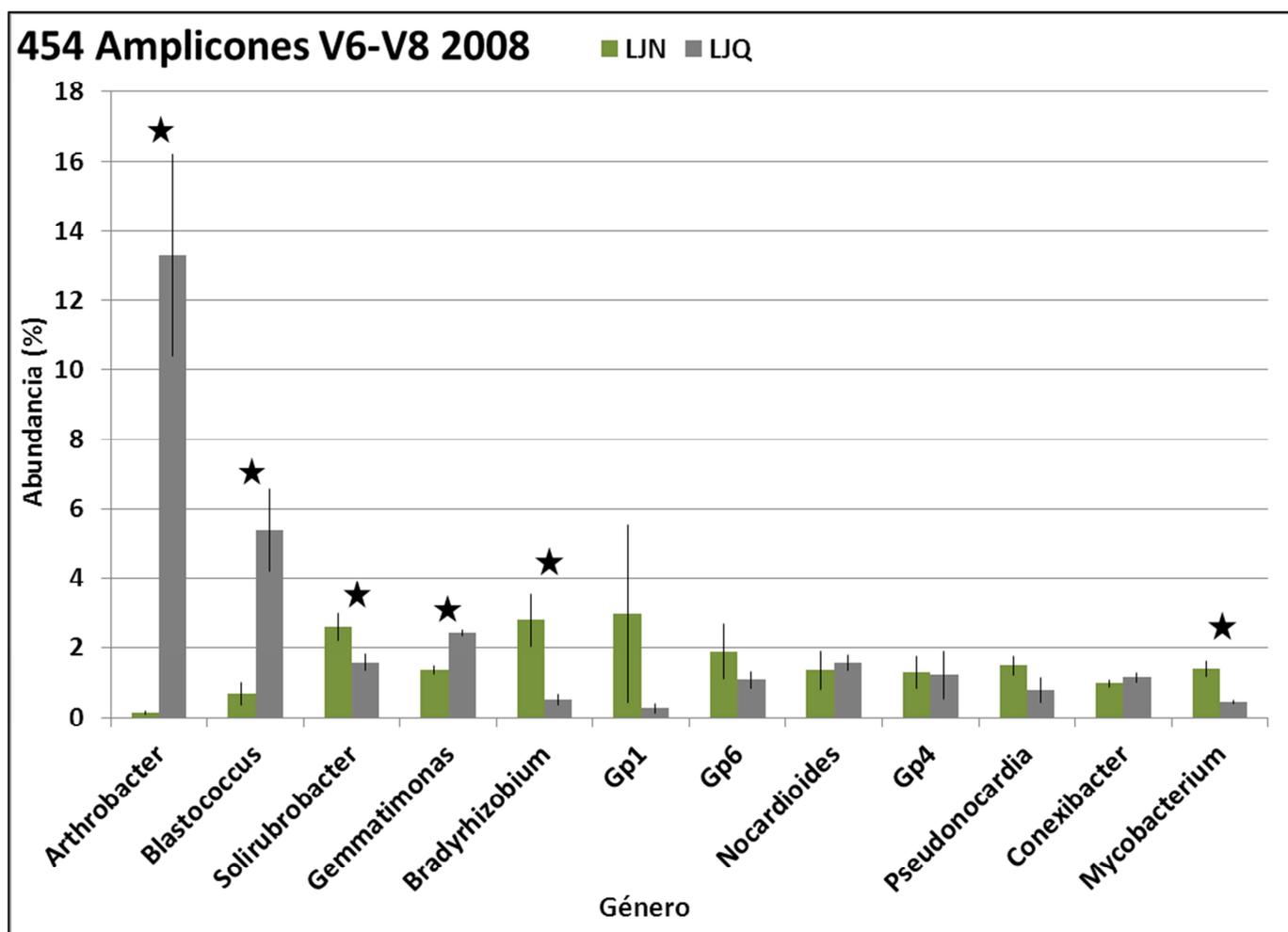


Figura R5-1. Distribución de los géneros más abundantes de las muestras de encinar quemado (LQJ) y control (LQN) tomadas en el año 2008 y secuenciadas mediante pirosecuenciación de amplicones de la zona V6-V8 del gen 16S *rRNA*. Solo se muestran aquellos géneros que suponen más del 1 % de abundancia relativa promedio en alguno de los dos sitios. Las barras de error se corresponden con la desviación estándar del valor medio de la abundancia relativa de cada familia para cada sitio, calculada a partir de las 3 muestras de cada sitio. En los géneros marcados con una estrella se encontraron diferencias estadísticamente significativas entre las medias de los dos sitios analizados mediante el test de Welch con un intervalo de confianza del 95 % en el programa STAMP.

1 Relevancia del género *Arthrobacter* en las comunidades estudiadas

En primer lugar, tratamos de confirmar que los datos obtenidos en la Figura R5-1, respecto a la dominancia en el encinar quemado (LJQ) del género *Arthrobacter*, no son debidos a un sesgo en la PCR de los cebadores utilizados para amplificar la zona V6-V8 del gen 16S *rRNA*. Para ello, se analizaron amplicones de la zona V4-V5 de dicho gen para los mismos sitios del encinar, quemado (LJQ) y no quemado (LJN) muestreados en 2008. De este análisis se obtuvieron 28.599 lecturas con una longitud media de 429 bp distribuidas uniformemente entre los 2 sitios secuenciados, con valores de 15.012 lecturas en LJN y 13.587 en LJQ. Tras filtrar y descartar las lecturas de baja calidad, las posibles quimeras (paso realizado con el programa MOTHUR), las menores de 150 bp y las mayores de 650 bp con la herramienta Pipeline Initial Process de RDP, se obtuvieron 23.589 secuencias con una longitud media de 408 bp distribuidas uniformemente y con unos valores de 12.373 secuencias en LJN y 11.216 secuencias en LJQ. Para realizar análisis que requieren normalización del número de secuencias de cada muestra, se normalizaron las muestras a 9.613 secuencias con el programa MOTHUR.

Al analizar la clasificación de las secuencias normalizadas homólogas al gen 16S *rRNA* obtenidas de los 2 sitios LJN y LJQ conseguimos clasificar el 92,1 % de las secuencias en LJN y el 92,8 % en LJQ a nivel de *phylum*. A nivel de género clasificamos el 50,2 % de las secuencias en LJN y el 54,4 % en LJQ distribuidas en 281 géneros. De éstos, 39 géneros muestran diferencias estadísticamente significativas entre ambos sitios. Además, 19 de estos géneros tienen una abundancia relativa de, al menos, el 1 % en alguno de los sitios muestreados; en la figura se presentan todos aquellos géneros que muestran diferencias biológicamente relevantes (Fig. R5-2). Al igual que en la clasificación de secuencias del gen 16S *rRNA* de la zona V6-V8 (Fig. R5-1), el género mayoritario es *Arthrobacter* con un 10,2 % en LJQ y un 0,3 % en LJN. En este caso el género *Blastococcus* (*phylum* Actinobacteria) ocupa la sexta posición en abundancia relativa con un 3,6 % en LJQ y un 0,5 % en LJN. Al comparar las dos zonas secuenciadas del gen 16S *rRNA*, se observa una disminución en la abundancia relativa de los géneros pertenecientes al *phylum* Actinobacteria acompañada de un aumento de los géneros del *phylum* Acidobacteria ocupando *Gp4* y *Gp1* la segunda y tercera posición respectivamente, teniendo ambos además, una abundancia relativa mayor al 5 % en suma de los dos sitios estudiados (Fig. R5-2).

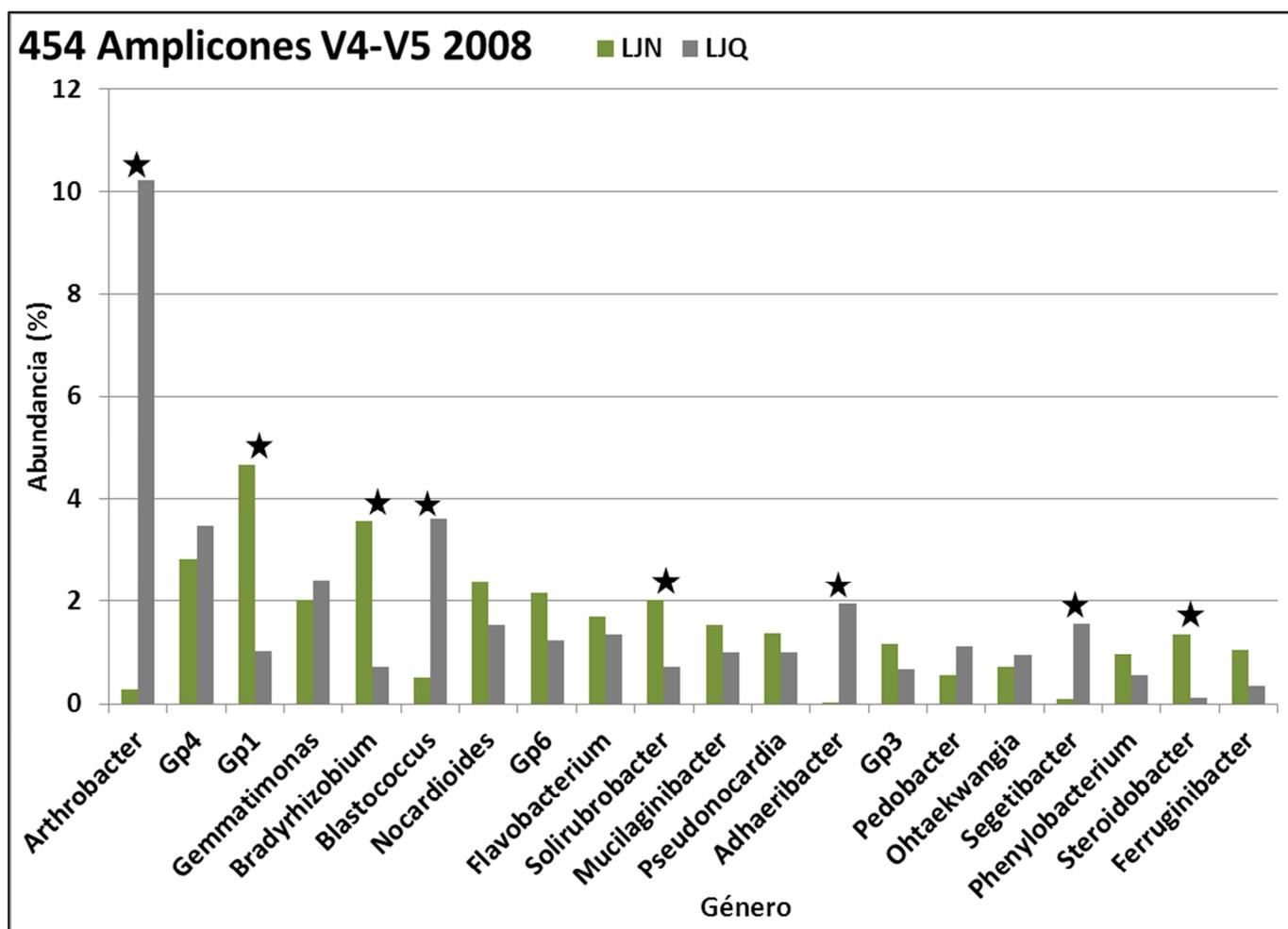


Figura R5-2. Distribución de los géneros más abundantes de las muestras de encinar quemado (LJQ) y control (LJN) tomadas en el año 2008 y secuenciadas mediante pirosecuenciación de amplicones de la zona V4-V5 del gen 16S *rRNA*. Solo se muestran aquellos géneros que suponen más del 1 % de abundancia relativa en alguno de los dos sitios. En los géneros marcados con una estrella se encontraron diferencias significativas biológicamente relevantes comparando ambos sitios con el programa STAMP, tomando un p-valor corregido menor del 0,01.

Para conocer la relevancia del género *Arthrobacter* en las muestras que provienen del área afectada por el incendio se ha llevado a cabo el análisis comparativo de la abundancia relativa de dicho género en todas las muestras estudiadas y mediante todas las técnicas utilizadas en este trabajo, excepto los resultados obtenidos con amplicones de la zona V4-V5 del gen 16S *rRNA* que han sido detallados anteriormente (Fig. R5-2).

En el estudio clásico de genotecas el 18,03 % de las secuencias pertenecientes al encinar quemado (LJQ) clasifican con el género *Arthrobacter* y el 4,92 % de las secuencias pertenecientes al suelo quemado (LJR), pero no hay ninguna secuencia homóloga a dicho género en el resto de muestras (Tabla R5-1).

En el estudio de los amplicones de la zona V6-V8 del gen 16S *rRNA* de las muestras tomadas en el año 2008 el 13,32 % de las secuencias obtenidas en LJQ y el 2,89 % de las secuencias obtenidas en LJR son homólogas al género *Arthrobacter*, constituyendo dicho género entre el 0,07 % y el 0,27 % de las secuencias en el resto de muestras independientemente de si provienen de suelo no rizosférico o suelo rizosférico de robles o encinas (Tabla R5-1). En el estudio de las muestras tomadas en el año 2011 se observa una abundancia relativa con valores entre 0,03 % y 0,15 % en las muestras no afectadas por el incendio; sin embargo, en LJQ supone el 5,12 %, en LJR el 2,84 % y en LJRE el 1,64 %, valores que pese a ser significativamente menores que en el muestreo de 2008 (en el caso de LJQ) siguen siendo significativamente más altos que en las muestras de los sitios no quemados (Tabla R5-1).

En el estudio de los metagenomas mediante pirosecuenciación se analizaron solo los sitios del encinar quemado (LJQ) y no quemado (LJN) y del robledal alto (CNA) y bajo (CNB) muestreados en 2008. Analizando las secuencias homólogas al gen 16S *rRNA*, de cada una de las cuatro muestras, observamos que pertenecen al género *Arthrobacter* el 16,90 % de las secuencias de LJQ y el 1,49 % de las secuencias en CNB, sin obtener secuencias homólogas a dicho género en los otros dos sitios (LJN y CNA). Analizando las secuencias anotadas en MG-RAST como genes de cada comunidad microbiana (metagenoma) de cada una de las muestras de estos cuatro sitios, observamos que pertenecen al género *Arthrobacter* el 1,69 % de los genes anotados en LJQ y entre el 0,29 % y el 0,35 % de los genes anotados en las tres muestras no quemadas (Tabla R5-1).

En el estudio de los metagenomas, mediante secuenciación Illumina, también se analizaron los sitios del encinar LJQ y LJN y del robledal CNA y CNB, además del sitio de expansión del robledal (CNX) muestreados en 2008. Analizando las secuencias homólogas al gen 16S *rRNA*, de cada una de las cinco muestras, observamos que pertenecen al género *Arthrobacter* el 7,45 % de las secuencias de LJQ y entre el 0,03 % y el 0,22 % de las secuencias de las cuatro muestras no quemadas. Analizando los contigs anotados en MG-RAST como genes de cada comunidad microbiana (metagenoma) de cada una de las muestras de estos cinco sitios, observamos que pertenecen al género *Arthrobacter* el 15,47 % de los genes anotados en LJQ y entre el 0,24 % y el 0,64 % de los genes anotados en las cuatro muestras no quemadas (Tabla R5-1).

La Tabla R5-1 muestra la dominancia del género *Arthrobacter* en las muestras tomadas en los sitios afectados por el incendio (LJQ, LJR y LJRE) con las distintas técnicas empleadas. Siendo además, el género más abundante para todas ellas, excepto en el metagenoma

Resultados Capítulo 5

obtenido mediante pirosecuenciación, pues con esta técnica resultó estar en la novena posición en abundancia relativa quedando el género *Streptomyces* (familia Streptomycetaceae también perteneciente al *phylum* Actinobacteria) con un 4,10 % de los genes anotados.

Tabla R5-1. Abundancia relativa del género *Arthrobacter* en las comunidades microbianas de todos los sitios estudiados mediante todas las técnicas utilizadas.

	LJN	LJQ	LJR	LJRE	CTE	CNB	CNA	CNX	CNR	CTR
Clones	0	18,03	4,92		0	0	0		0	0
Amplicones V6-V8 2008	0,14	13,32	2,89		0,13	0,27	0,07	0,22	0,07	0,15
Amplicones V6-V8 2011	0,10	5,12	2,84	1,64	0,03	0,07	0,09	0,15		0,16
454 16S rRNA	0	16,90				1,49	0			
454 Metagenoma	0,31	1,69				0,35	0,29			
Illumina 16S rRNA	0,13	7,45				0,22	0,03	0,18		
Illumina Metagenoma	0,64	15,47				0,44	0,44	0,24		

2 Aislamiento de cepas del género *Arthrobacter*

A partir de 1 gramo de suelo procedente del encinar quemado (LJQ) tomado en 2008 se obtuvieron 3×10^6 ufc (unidades formadoras de colonias) (ver apartado 4 de Material y Métodos). Mediante el empleo del medio de cultivo HH' se aislaron un total de 231 colonias, de las que posteriormente se descartaron 144 por ser de crecimiento rápido y estar constituidas por cepas homólogas a los géneros *Bacillus*, *Pseudomonas* y *Erwinia*. Las 87 colonias restantes tardaron en crecer entre 3 y 5 días, se secuenció el gen 16S rRNA casi completo (≈ 1460 bp) de todas ellas, se descartaron 29 por ser homólogas a los géneros *Streptomyces* y *Stenotrophomonas*, quedando 58 colonias que fueron homólogas al género *Arthrobacter*, de éstas se seleccionaron 55, ya que las otras 3 presentaban problemas a la hora de crecerlas en placa.

Con el fin de confirmar si los aislados homólogos a *Arthrobacter* corresponden con los OTUs mayoritarios obtenidos mediante pirosecuenciación, se compararon las 1.593 secuencias procedentes de los amplicones de la zona V4-V5 de LJQ, agrupadas en un mismo OTU con al menos una similitud del 97 % homólogas a dicho género, con las secuencias obtenidas mediante secuenciación Sanger de las colonias. Las 1.648 secuencias de 408 bp formaron 1 OTU al 97 % de similitud y 46 OTUs al 100 % de similitud. Los 46 OTUs se ordenaron de mayor a menor según el número de secuencias procedentes de pirosecuenciación que contienen y en la Tabla R5-2 se muestran los 11 OTUs que están formados por más de 1 de estas secuencias. El OTU mayoritario constituido por 1.032 secuencias 100 % idénticas contiene 41 aislados de los cuales 20 son homólogos a la especie *Arthrobacter globiformis*, 13 a la especie *A. oxydans*, 3 a la especie *A. sulfonivorans*, 2 a la especie *A. humicola* y las especies *A. oryzae*, *A. pascens* y *A. scleromae* están constituidas por 1 único aislado (Tabla R5-2). El segundo OTU más abundante está formado por 468 secuencias, todas ellas procedentes de la pirosecuenciación pues no se obtuvo ningún aislado cuyo gen 16S rRNA fuera 100 % similar. El tercer OTU más abundante está formado por 17 secuencias de las cuales 3 pertenecen a cepas aisladas homólogas a la especie *A. sulfureus* (Tabla R5-2). El cuarto OTU más abundante está constituido por 18 secuencias y contiene 11 aislados de los cuales 10 son homólogos a la especie *A. nitroguajacolicus* y 1 a la especie *A. nicotinovorans*. el resto de OTUs no contienen aislados (Tabla R5-2).

Resultados Capítulo 5

Tabla R5-2. OTUs homólogos al género *Arthrobacter* con más de una secuencia 100 % idéntica procedentes de la pirosecuenciación y de la secuenciación de los amplicones de los 55 aislados del gen 16S *rRNA* de la zona V4-V5 de la muestra LJQ. # Número de secuencias. * Especies a las que son homólogas las cepas aisladas con una identidad superior al 99 % alineando el amplicón del gen 16S *rRNA* casi completo mediante BLASTn.

OTU	# PIROSECUENCIACIÓN	# AISLADOS	ESPECIES*
1	991	41	<i>A. globiformis</i> 20 <i>A. oxydans</i> 13 <i>A. sulfonivorans</i> 3 <i>A. humicola</i> 2 <i>A. oryzae</i> 1 <i>A. pascens</i> 1 <i>A. scleromae</i> 1
2	468	0	-
3	14	3	<i>A. sulfureus</i> 3
4	7	11	<i>A. nitroguajacolicus</i> 10 <i>A. nicotinovorans</i> 1
5	6	0	-
6	6	0	-
7	3	0	-
8	2	0	-
9	2	0	-
10	2	0	-
11	2	0	-

Los 55 amplicones de aproximadamente 1460 bp del gen 16S *rRNA* fueron alineados con las secuencias del gen 16S *rRNA* completo de las 19 cepas tipo que dieron mayor identidad según BLASTn. Se utilizó *Arthrobacter sanguinis* como grupo externo (outgroup) por ser la cepa más alejada filogenéticamente de los 55 aislados al utilizar las 63 cepas tipo del género *Arthrobacter* contenidas en la base de datos RDP. El árbol filogenético de las 75 secuencias (Fig. R5-3) confirma que hay un grupo mayoritario de 20 aislados (Tabla R5-2) muy próximos a la especie *A. globiformis* (72 % bootstrapping) aunque no 100 % idénticos a ésta y tampoco entre todos ellos. Todos los aislados están próximos a las cepas tipo obtenidas mediante BLASTn y se confirman los resultados de identidad obtenidos con esta herramienta, pues, pese a estar muy próximas a las cepas tipo sólo una secuencia de las cepas aisladas (AFG25) es 100 % idéntica a la secuencia del gen 16S *rRNA* de la cepa tipo a la que es homóloga (*A. humicola* KV-635^T) (Fig. R5-3). A pesar de ello, en las 55 cepas aisladas, sí aparecen secuencias del gen 16S *rRNA* 100 % idénticas entre sí, de modo que entre todas ellas dan lugar a 25 secuencias únicas (Fig. R5-3).

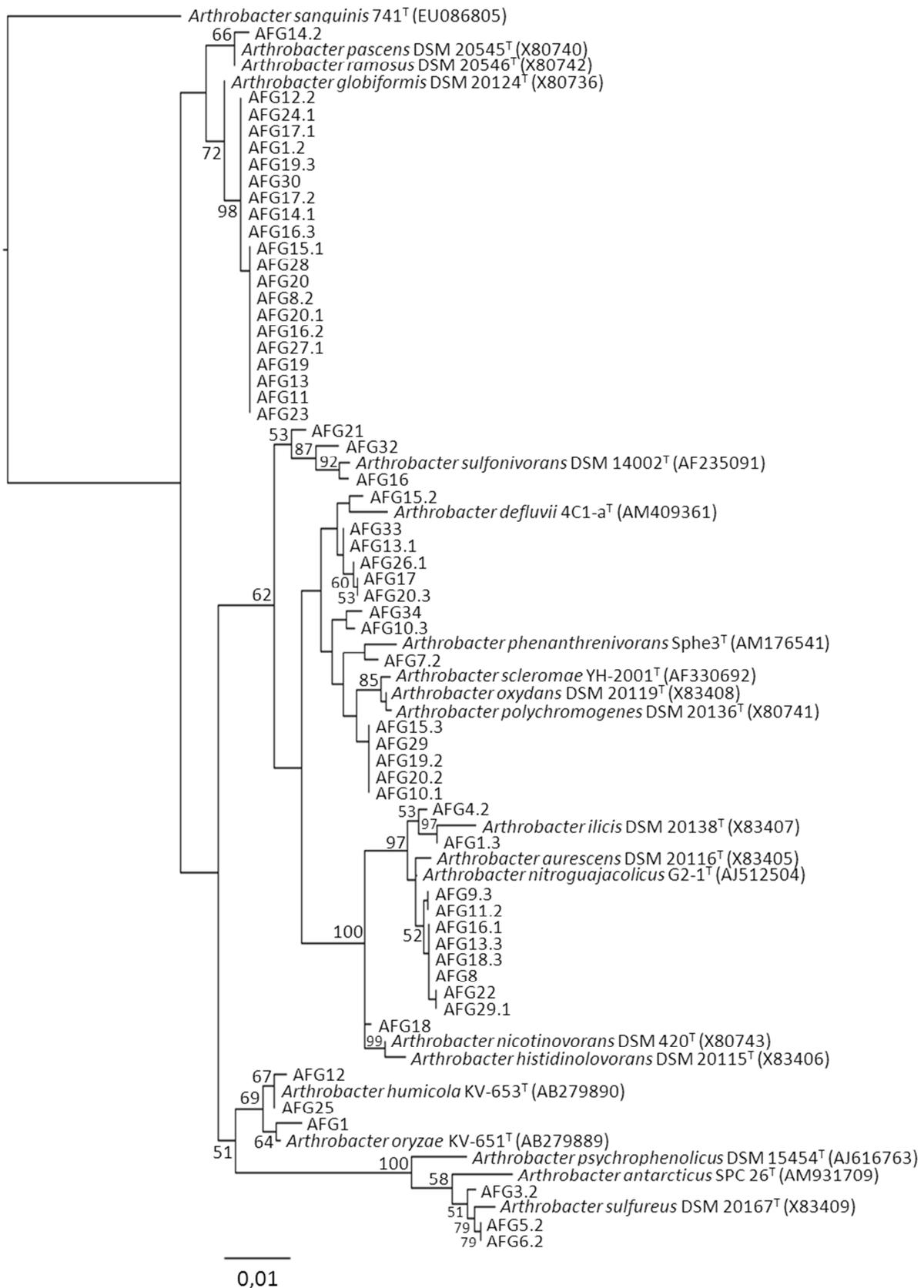


Figura R5-3. Árbol filogenético de Máxima Verosimilitud basado en las secuencias del gen 16S *rRNA*. Se indican los valores de bootstrap calculados para 1.000 réplicas. La barra muestra la distancia de la sustitución de 1 nucleótido de cada 100. Entre paréntesis se indican los números de acceso para el Genbank.

3 Reconstrucción del pan-genoma de *Arthrobacter*

El pan-genoma (pan proviene de la palabra griega *παν*, que significa todo) está constituido por los genes que están presentes en el conjunto de las cepas de una especie. Este pan-genoma estaría formado por los genes considerados indispensables (core-genome) más los genes que no aparecen en todas las cepas incluyendo aquellos que son únicos en alguna de ellas, conocido como genoma accesorio (dispensable-genome) (Tettelin *et al.* 2005, Medini *et al.* 2005). El estudio de metagenómica comparada se ha realizado comparando el metagenoma del encinar quemado (LJQ) frente a los genomas de cepas completamente secuenciadas del género *Arthrobacter* (obtenidas de la base de datos de genomas bacterianos completos del NCBI, <ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/>). De esta forma, elevamos el término pan-genoma a nivel de género y hablamos, por lo tanto, del core-genome del género *Arthrobacter*, siendo genes accesorios aquellos pertenecientes a las cepas de diferentes especies de este género con homología a los genes accesorios de las especies parcialmente secuenciadas en nuestro metagenoma. El término pan-genoma ya ha sido utilizado a nivel de género en un estudio de genómica comparada con el género bacteriano *Listeria* (den Bakker *et al.* 2010).

Ya que el género *Arthrobacter* parece dominar en la comunidad microbiana del suelo del encinar quemado (LJQ) (ver apartado 1), a fin de conocer las funciones que desempeñan los miembros de dicha comunidad y en particular las distintas especies del género *Arthrobacter*, se secuenció una muestra de DNA de LJQ mediante la tecnología Illumina PE101. De dicha muestra se obtuvieron 386.720.378 lecturas todas ellas con una longitud de 101 bp, que tras filtrarlas y solapar las lecturas pareadas (Paired-Ends reads), resultaron en 186.654.779 secuencias con un rango de tamaños entre 50 y 190 bp y un tamaño medio de 170 bp (Tabla R4-1). Posteriormente las secuencias se ensamblaron con el programa SOAPdenovo resultando en 383.979 contigs con un tamaño medio de 392 bp siendo el contig mayor de un tamaño de 53.753 bp.

Tras procesar los 383.979 contigs (que suponen un total de 150,7 Mbp) en el servidor MG-RAST, se eliminaron 913 contigs (el 0,2 %) ya que fueron detectados como duplicaciones artificiales, 2.270 contigs (el 0,6 %) tienen genes que transcriben RNA ribosómicos, 293.463 contigs (el 76,4 %) tienen genes que se anotaron como ORFs que codifican proteínas con función conocida, 86.188 contigs (el 22,4 %) tienen genes que se anotaron como ORFs que codifican proteínas con función desconocida y 1.139 contigs (el 0,3 %) no tienen ORFs (Fig. R5-4a).

Tras la anotación de los posibles genes en MG-RAST, se obtuvo que 31.554 contigs (el 8,2 %) son homólogos al género *Arthrobacter* con un tamaño medio de 567 bp, siendo el contig mayor de un tamaño de 48.568 bp. En total disponemos de 17,8 Mbp de información perteneciente a *Arthrobacter*, cuando el tamaño medio del genoma de las especies de este género es de 4,5 Mbp. Se encontraron 32.188 genes de los cuales 3.737 (el 11,6 %) transcriben RNA ribosómicos y 28.451 (el 88,4 %) corresponden con regiones codificantes (CDR). De éstas, 26.325 CDR (el 92,5 %) se anotaron en la base de datos M5NR de MG-RAST y 2.126 CDR (el 7,5 %) no se anotaron al no presentar homología (ORFans) (Fig. R5-4b).

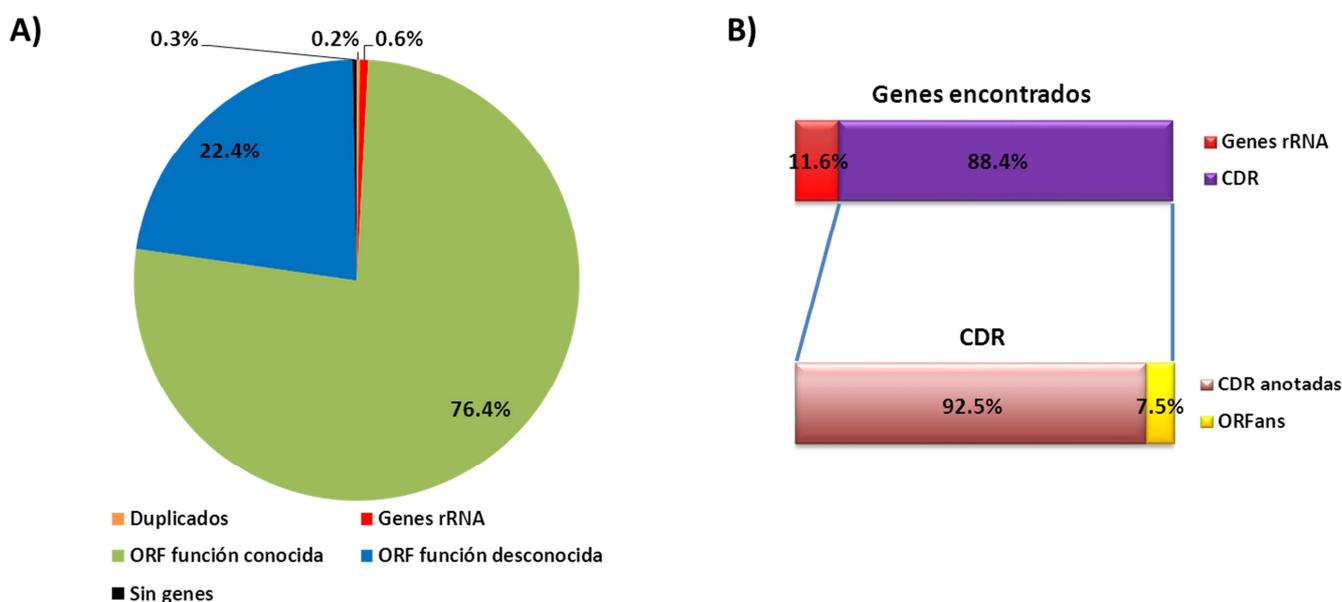


Figura R5-4. Procesamiento de las secuencias del metagenoma de una muestra de DNA de LJQ mediante la tecnología Illumina HiSeq2000 PE101 a través del servidor web MG-RAST. Procesamiento de los 383.979 contigs ensamblados con SOAPdenovo (A) y de los 31.445 contigs homólogos al género *Arthrobacter* (B).

Puesto que la especie mayoritaria en nuestro metagenoma es *Arthrobacter globiformis* (Tabla R5-2 y Fig. R5-3) se mapearon los 31.554 contigs (pan-genoma) homólogos a *Arthrobacter*, según MG-RAST, al único genoma de esta especie completamente secuenciado y disponible actualmente en las bases de datos, el genoma de la cepa *A. globiformis* NBRC12137 (subido a la base de datos del NCBI como “draft genome” por Fujita, N. del National Institute of Technology and Evaluation de Japón). Este genoma no está totalmente ensamblado, formando los replicones (cromosoma y plásmidos) de la bacteria, sino que está en la fase inicial de ensamblado resultando en un total de 125 contigs.

Resultados Capítulo 5

Al enfrentar los 31.554 contigs del pan-genoma a la cepa *A. globiformis* NBRC12137, 17.612 contigs mapearon en 78 de los contigs del genoma de referencia con una identidad promedio del 90,7 % y una identidad mínima del 77,5 % (Fig. R5-5). A la derecha de la Figura R5-5 quedan los 47 contigs del genoma de referencia que no se mapearon y se observa que son los de menor tamaño.

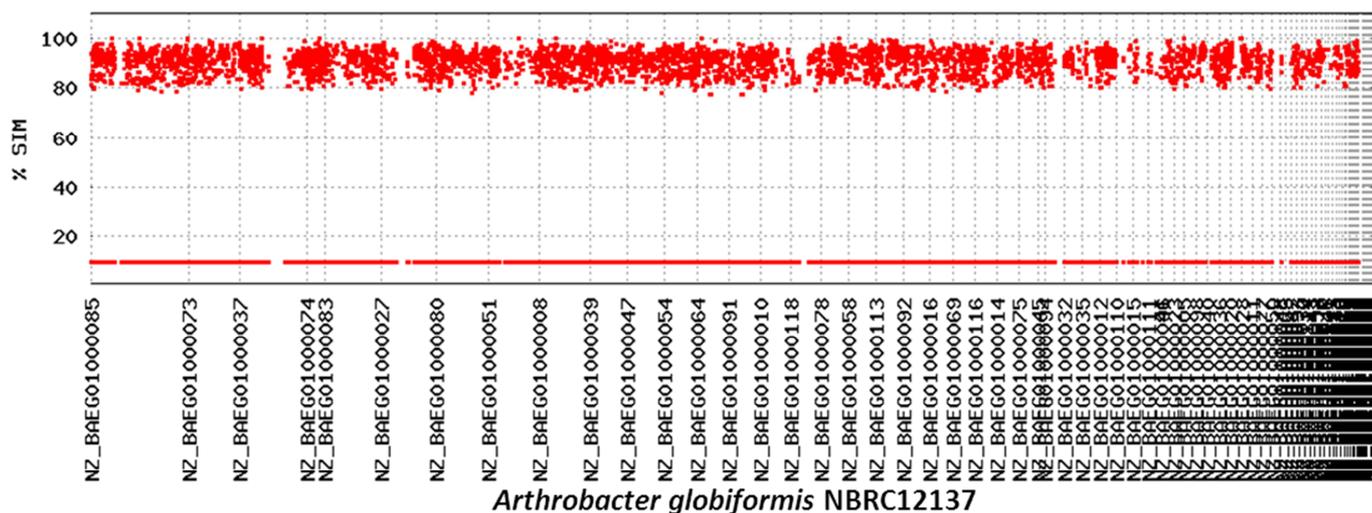


Figura R5-5. Reclutamiento genómico de los 31.554 contigs del metagenoma de LIQ homólogos al género *Arthrobacter* frente a la cepa *A. globiformis* NBRC12137. Alineamiento de secuencias de DNA realizado con la herramienta NUCmer con un tamaño mínimo de alineamiento de 65 bp. Rojo: Alineamiento de la hebra directa de las secuencias del metagenoma frente a la hebra directa del genoma de referencia. Azul: Alineamiento de la hebra inversa complementaria de las secuencias del metagenoma frente a la hebra directa del genoma de referencia. Las secuencias mapeadas se colapsan al 10 % de similitud para ver las regiones del genoma de referencia cubiertas.

De los 6 genomas completamente secuenciados y cerrados (ensamblaje definitivo en replicones) del género *Arthrobacter* que hay disponibles en las bases de datos se seleccionó como genoma de referencia la cepa *Arthrobacter* sp. FB24 por ser aquella a la que más genes, de los 31.554 contigs, del pan-genoma mapean (17.261 CDR según MG-RAST) (Tabla R5-3). De los 31.554 contigs enfrentados al genoma de referencia, 9.660 contigs mapearon en los 4 replicones (cromosoma y 3 plásmidos) con una identidad promedio del 87,5 % y una identidad mínima del 73,4 % (Fig. R5-6).

Tabla R5-3. Genomas secuenciados y cerrados del género *Arthrobacter* disponibles en las bases de datos del NCBI. CDR: regiones codificantes de los 31.554 contigs del metagenoma de LJQ anotadas en la base de datos M5NR de MG-RAST.

Genoma	CDR mapeadas
<i>Arthrobacter sp.</i> FB24	17.261
<i>Arthrobacter chlorophenolicus</i> A6	8.859
<i>Arthrobacter aurescens</i> TC1	3.86
<i>Arthrobacter arilaitensis</i> Re117	96
<i>Arthrobacter phenanthrenivorans</i> Sphe3	12
<i>Arthrobacter nitroguajacolicus</i> Rue61a	4

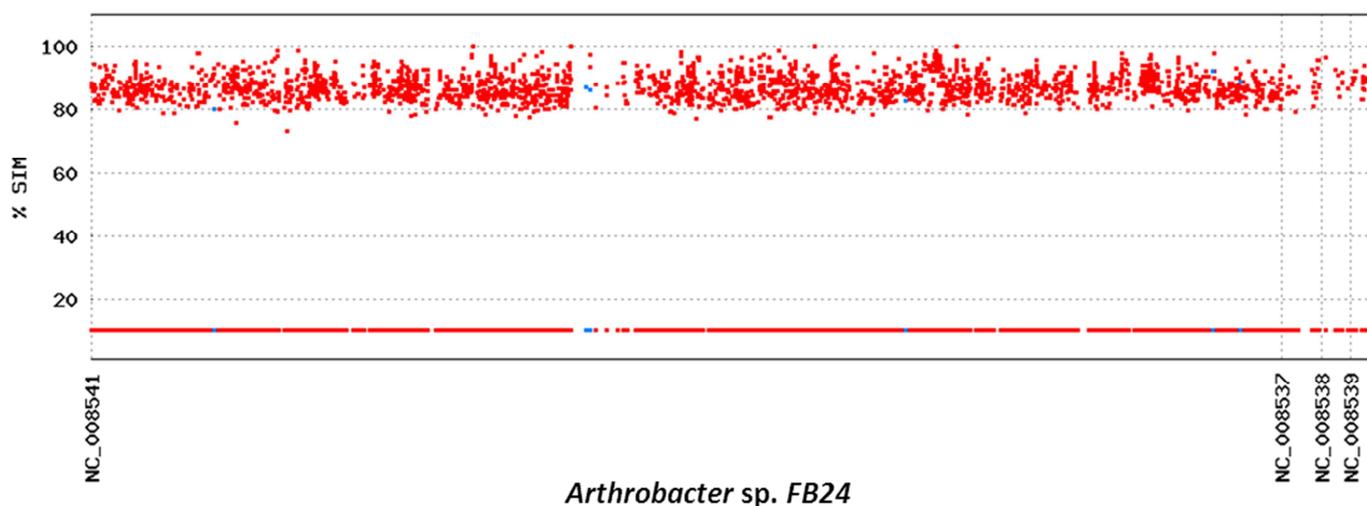


Figura R5-6. Reclutamiento genómico de los 31.554 contigs del metagenoma de LJQ homólogos al género *Arthrobacter* frente a la cepa *A. sp.* FB24. Alineamiento de secuencias de DNA realizado con la herramienta NUCmer con un tamaño mínimo de alineamiento de 65 bp. Rojo: Alineamiento de la hebra directa de las secuencias del metagenoma frente a la hebra directa del genoma de referencia. Azul: Alineamiento de la hebra inversa complementaria de las secuencias del metagenoma frente a la hebra directa del genoma de referencia. Las secuencias mapeadas se colapsan al 10 % de similitud para ver las regiones del genoma de referencia cubiertas.

Tras enfrentar los 32.188 genes encontrados en los 31.554 contigs del pan-genoma del género *Arthrobacter* a la cepa *A. sp.* FB24, mapearon 23.999 genes (74,6 %). De éstos, 17.261 genes están en regiones codificantes (CDR) homólogas a la cepa de referencia. Dicha cepa tiene 4.523 CDR únicas anotadas en la base de datos M5NR de MG-RAST de las cuales, 3.196 CDR están mapeadas por el metagenoma de LJQ (Fig. R5-7). La CDR más abundante tiene 109 secuencias mapeadas en un gen anotado, en el cromosoma del genoma de referencia, como una hemo peroxidasa (YP_830737.1). En la Figura R5-7 se observa la región del genoma de referencia cubierta por el metagenoma (el círculo interno) que supone el 70,7 % de las CDR de las cuales, el 50,4 % están en la cadena - y el 49,6 % en la cadena + de los replicones del genoma de *A. sp.* FB24.

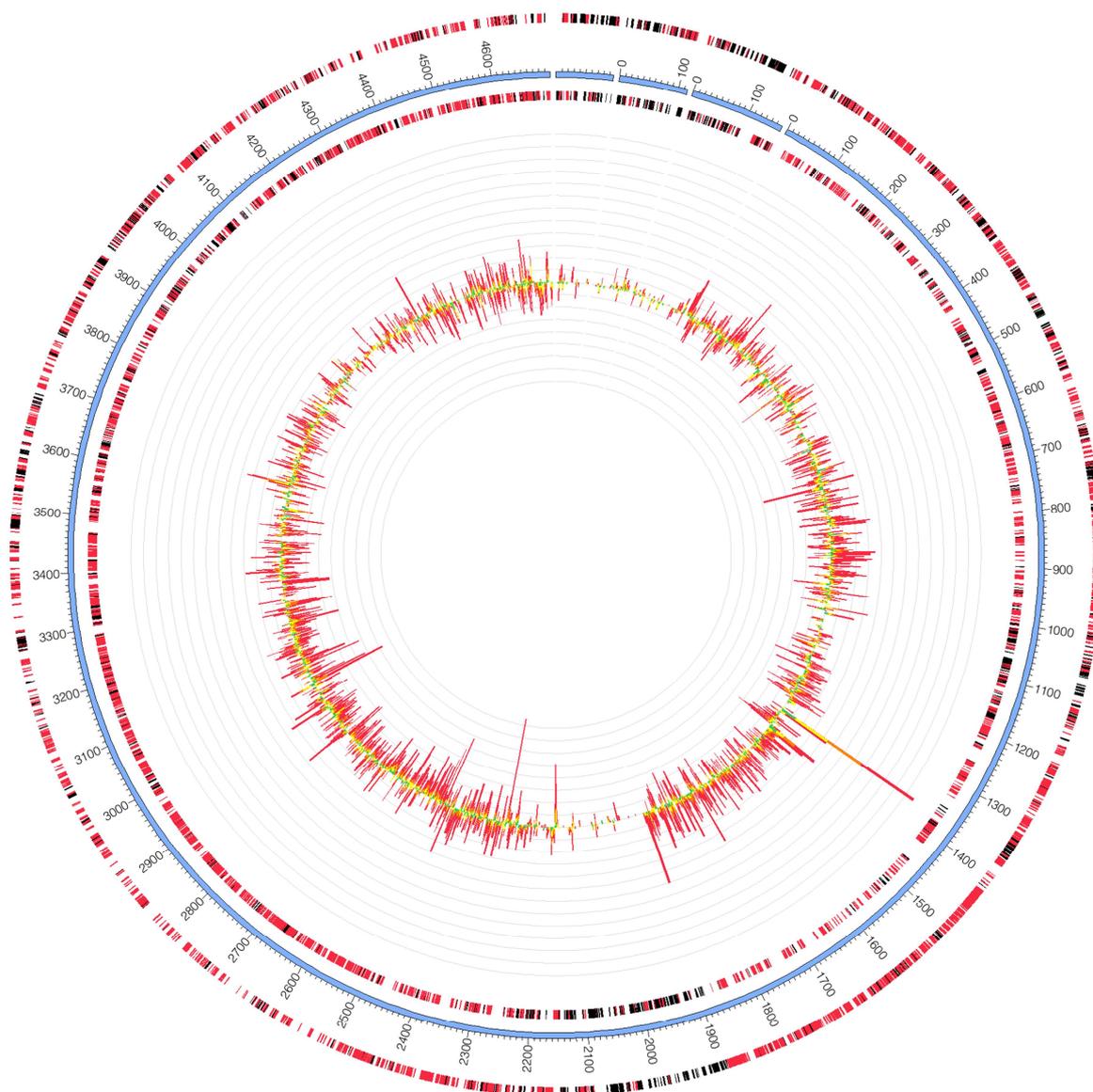


Figura R5-7. Reclutamiento genómico de las 17.261 CDR pertenecientes a los 31.554 contigs del metagenoma de LQ homólogos al género *Arthrobacter* frente a los CDR de la cepa *A. sp. FB24*. Círculo externo: Genes del genoma de referencia anotados en la cadena + mostrando en rojo CDR y en negro RNA no codificante. 2º Círculo: Replicones del genoma de referencia, donde se pueden apreciar el cromosoma y 3 plásmidos. 3º Círculo: Genes del genoma de referencia anotados en la cadena - mostrando en rojo CDR y en negro RNA no codificante. Círculo interno: Abundancia medida como número de secuencias del metagenoma que mapean frente al genoma de referencia. Hacia el exterior las que mapean en la cadena + y hacia el interior las que mapean en la cadena - y el color representa el e-valor del alineamiento siendo, verde de 10^{-5} a 10^{-10} , amarillo de 10^{-10} a 10^{-20} , naranja de 10^{-20} a 10^{-30} y rojo $< 10^{-30}$.

Los mismos 32.188 genes enfrentados a la cepa de referencia se introdujeron en la base de datos de genes ortólogos KO del KEGG (Kyoto Encyclopedia of Genes and Genomes) de los cuales 17.579 genes (54,6 %) fueron anotados en 5 categorías funcionales de nivel 1 de la base de datos. De modo que, el 61 % de los genes se anotaron en ortólogos relacionados con el metabolismo celular, el 18 % anotaron en ortólogos relacionados con el procesamiento de la

información que llega del exterior de la célula, el 16 % se anotaron con ortólogos relacionados con el procesamiento de la información genética, el 4 % se anotaron con ortólogos relacionados con los procesos celulares, el 0,9 % se anotaron con ortólogos relacionados con enfermedades humanas y el 0,06 % se anotaron con ortólogos relacionados con otros sistemas del organismo bacteriano (Fig. R5-8). Bajando al segundo nivel vemos que las funciones relativamente más abundantes en el pan-genoma son el metabolismo de aminoácidos (21 %) y carbohidratos (16 %) y los transportadores de membrana (14 %) (Fig. R5-8).

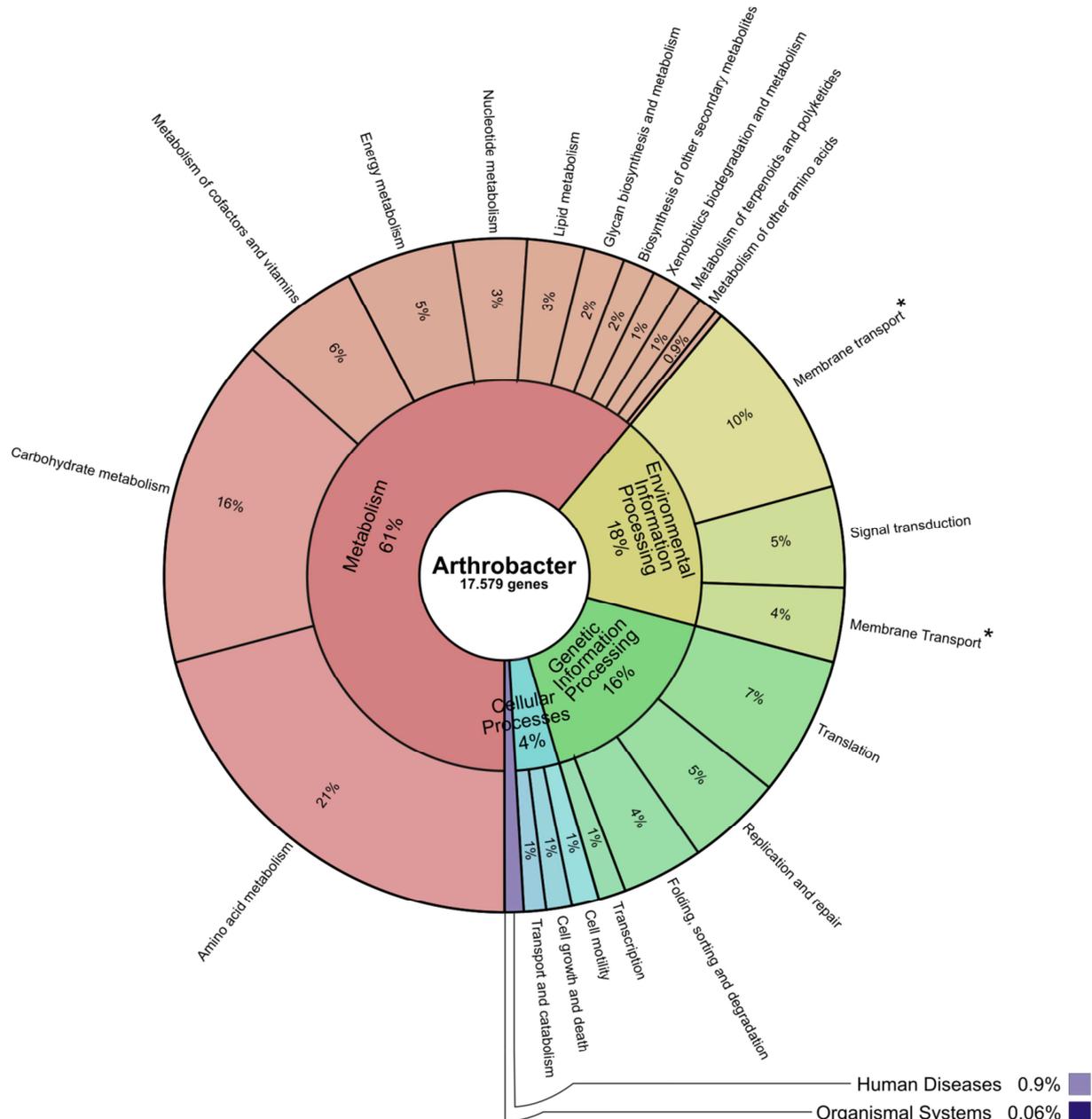


Figura R5-8. Caracterización funcional de los genes encontrados en el pan-genoma del género *Arthrobacter* hasta el 2º nivel del sistema jerárquico de categorías funcionales de la base de datos KO de KEGG. Los parámetros utilizados fueron: e-valor como máximo de 10^{-5} , una identidad como mínimo del 60 % y un tamaño de alineamiento como mínimo de 50 bp. Gráfico realizado con la herramienta Krona en el servidor web MG-RAST.

* La misma categoría anotada de dos maneras diferentes en MG-RAST.

Discusión

1. Consideraciones previas

En el presente trabajo se han estudiado las comunidades microbianas de la rizosfera de quercíneas forestales teniendo en cuenta el efecto del gradiente altitudinal, la especie vegetal del género *Quercus* con la que interaccionan (*Q. ilex* subsp. *ballota* o *Q. pyrenaica* Willd.) y su proceso de recuperación a medio y largo plazo, tres años y seis años, tras un incendio forestal. Con la idea de ver el efecto rizosférico de las quercíneas en las comunidades microbianas estudiadas a distintos niveles del gradiente altitudinal y su grado de recuperación tras la alteración provocada por el fuego, éstas se han comparado con comunidades microbianas no rizosféricas de las mismas zonas. Para llevar a cabo dichos estudios se han evaluado también las características físico-químicas de los suelos de origen de estas comunidades y la vegetación propia de cada sitio muestreado. Al comparar las comunidades microbianas a distintas altitudes también se tuvieron en cuenta los factores climáticos de cada sitio, la temperatura del aire y la precipitación, así como la temperatura y humedad relativa de los suelos, y se ha evaluado tanto el efecto de estos factores sobre las comunidades microbianas como sus diferencias en los dos años de muestreo (2008 y 2011). En el estudio comparativo de las comunidades rizosféricas de ambas quercíneas se han tenido en cuenta la localización geográfica de las muestras (ladera Norte y ladera sur de Sierra Nevada) y el año en el que se han muestreado (2008 y 2011).

2. Efectos del gradiente altitudinal en las comunidades microbianas

Nuestros resultados sugieren que el efecto de la cobertura vegetal enmascara los posibles efectos del gradiente altitudinal en las comunidades microbianas estudiadas en el robleal de Cañar, siendo la presencia o ausencia de rizosfera uno de los factores más importantes en la estructura y composición de las poblaciones microbianas, seguido del efecto producido por la densidad de robles (Figs. R2-15 y R2-17). Sin embargo, hay una clara correspondencia entre la distribución de la vegetación propia de los sitios estudiados (Fig. R1-1) y la estructura de las comunidades microbianas observadas (Figs. R1-3A, R1-4A, R2-13 y R2-14), siendo clara la separación del bosque maduro (CNA y CNB) en el que los arbustos dominantes son *Adenocarpus decorticans*, *Festuca elegans* y *Ulex parviflorus* en los pisos supra y mesomediterráneo, y el piornal dominado por *Festuca indigesta* propia del piso oromediterráneo. Aunque esta correspondencia desaparece entre la muestra no rizosférica (CNR) y la muestra rizosférica de mayor altitud (CNX), con una composición florística muy similar y una composición microbiana muy diferente debido al efecto rizosférico. En el caso del robleal maduro se observaron distintos OTUs (especies) de los géneros *Bradyrhizobium* (α -Proteobacteria), *Nocardioides*, *Kribbella* (familia Nocardioideaceae) y *Mycobacterium* (los 3 últimos pertenecientes al *phylum* Actinobacteria) con una distribución significativamente diferente entre el sitio bajo (CNB) y el alto (CNA). Es decir, hay OTUs pertenecientes al mismo género bacteriano que se distribuyen de forma diferente en función del gradiente altitudinal. Estas diferencias encontradas al nivel taxonómico más profundo parecen coincidir con la presencia de la leguminosa *Ulex parviflorus*, propia del melojar degradado, en CNB y de la gramínea *Festuca elegans*, propia de los melojares característicos, en CNA, aunque no se encontró correlación entre estos hechos (Fig. R2-17). Además el efecto de la rizosfera de los robles siempre deberá ser predominante sobre la mera presencia de un tipo u otro de arbusto en la zona de muestreo.

Observamos un gradiente térmico e hídrico entre los sitios analizados y pese a que ninguno presenta diferencias estadísticamente significativas (Tabla R1-5 y Fig. R1-7), las diferencias entre las comunidades microbianas del robleal maduro (CNB y CNA), a menor altitud y los pies aislados (CNX), a mayor altitud, se reducen ante unas condiciones hídricas más favorables, dadas en el año 2011 (Fig. R2-14). Este efecto solo fue significativo en la disminución de la abundancia relativa de las familias Bradyrhizobiaceae (*phylum* Proteobacteria) y Solirubrobacteraceae (*phylum* Actinobacteria) en el robleal maduro hasta

niveles similares a los del robledal en expansión, entre los dos años de muestreo (2008 y 2011) (Figs. R2-5 y R2-6).

Aunque algunos trabajos han mostrado que la comunidad microbiana responde al gradiente altitudinal y que en alta montaña el patrón más característico se basa en la disminución de la diversidad al aumentar la altitud (Brown, 2001; Lomolino, 2001; McCain, 2005), en este trabajo se observa una tendencia contraria entre los 3 sitios con muestras rizosféricas (CNA, CNB y CNX), aunque no es estadísticamente significativa (Figs. R2-8, R2-9 y R2-11). Pero sí aparece una mayor dominancia de OTUs pertenecientes a las familias Bradyrhizobiaceae, Sphingomonadaceae, Hyphomicrobiaceae, Caulobacteraceae (clase α -Proteobacteria), Nocardiodaceae y Mycobacteriaceae (clase Actinobacteria), mejor adaptadas a las condiciones rizosféricas en las muestras de robledal maduro (CNB y CNA) respecto a las muestras no rizosféricas (CNR) (Fig. R2-10), debido al llamado efecto rizosférico descrito por varios grupos de investigación para robles entre los que podemos destacar a Uroz *et al.* (2010). Por lo que podemos considerar que se trata de OTUs que exhiben un comportamiento copiotrófico (Fierer *et al.* 2007; Yuste *et al.* 2012) gracias a la mayor disponibilidad de carbono aportado por los exudados radicales de los melojares. Por otro lado, los OTUs pertenecientes a las familias Gemmatimonadaceae (*phylum* Gemmatimonadetes), Planctomycetaceae (*phylum* Planctomycetes), Conexibacteraceae, Geodermatophilaceae (ambos del *phylum* Actinobacteria) y Gp7 (*phylum* Acidobacteria) son significativamente más abundantes en las muestras no rizosféricas, posiblemente exhibiendo un comportamiento oligotrófico, descrito hasta ahora solo en el último de estos *phyla*, Acidobacteria (Fierer *et al.* 2007).

Si bien los factores físico-químicos de los suelos contribuyen a la estructura y composición de las comunidades microbianas edáficas (Uroz *et al.* 2013), el pH junto con la relación C/N es uno de los factores que más influencia tienen en la biomasa y estructura de dichas comunidades (Shen *et al.* 2013). En nuestro estudio el pH es el único factor con un efecto significativo en la composición de las comunidades microbianas con una relación directa con la cobertura vegetal y el efecto rizosférico, siendo las familias Bradyrhizobiaceae (Proteobacteria) y Nocardiodaceae (Actinobacteria) las que mayor correlación positiva tienen y la familia Gemmatimonadaceae la que mayor correlación negativa tiene con el pH. Sin embargo, en nuestro trabajo el pH y la diversidad microbiana no están correlacionados, mostrando una ligera tendencia a la disminución de la diversidad con el aumento del pH, al contrario de lo que ocurre en el trabajo de Shen y colaboradores (2013) en el que se observó una correlación positiva entre ambos parámetros. Posiblemente, en nuestro caso, se debe a un efecto más intenso del factor rizosfera que el del factor pH.

3. Influencia de la especie arbórea en las comunidades microbianas de la rizosfera de quercíneas

La diferencia más notable entre las dos especies del género *Quercus* evaluadas en este trabajo es que la encina (*Q. ilex* subsp. *ballota*) es un árbol de hoja perenne mientras que el roble melojo (*Q. pyrenaica* Willd) posee hojas marcescentes, lo cual genera un gran aporte puntual de nutrientes y materia orgánica provocado por la caída de la hoja en otoño en el caso del roble y un aporte menor, pero continuo en el caso de la encina. Este factor diferencial se suma a la selección de la comunidad microbiana rizosférica provocada por las diferencias encontradas en los exudados radicales propios de cada especie arbórea, demostrado con diferentes especies arbóreas de los géneros *Octoea*, *Millinedia* y *Eugenia*, en bosque Atlántico (Lambais *et al.* 2014). En nuestro caso particular, al tratarse de 2 especies arbóreas del mismo género vegetal, cabría la posibilidad de que los exudados radiculares fueran muy similares en su composición.

Según los resultados obtenidos en el capítulo 3 no podemos distinguir, al menos con los métodos utilizados en este trabajo, comunidades microbianas propias de encina o de roble melojo (Figs. R3-13, R3-14 y R3-15), al contrario que en los estudios de Lambais y colaboradores (2014). Tampoco observamos efecto en dichas comunidades provocado por las diferencias de composición florística que acompaña a cada especie de *Quercus* (Tablas R1-2 y R1-3). Del mismo modo, que no se han observado diferencias asociadas a la localización geográfica, ladera Norte o Sur, en las comunidades microbianas de las muestras estudiadas.

Las comunidades microbianas de las quercíneas estudiadas son muy diversas (Tabla R3-4 y Fig. R3-9) y están representadas por más de 100 familias procarióticas diferentes perteneciendo las más abundantes a los *phyla* Proteobacteria, Actinobacteria, Acidobacteria, Bacteroidetes, Verrucomicrobia y Planctomycetes, nombrados en orden de abundancia relativa. Al igual que en el trabajo de Uroz *et al.* (2010) realizado con robles, el *phylum* más abundante en la rizosfera de quercíneas es Proteobacteria; sin embargo, en su trabajo el segundo *phylum* más abundante es Acidobacteria y no Actinobacteria, como es nuestro caso, hecho que parece estar directamente relacionado con el pH más ácido (4,3) de los bosques estudiados por el grupo de Uroz, comparado con un pH entre 6,1 y 7,2 en los bosques de los sitios comparados en este trabajo (Tabla R1-1). Aunque en nuestro estudio solo hemos comparado las familias con homología a la base de datos del RDP (Figs. R3-5 y R3-6), también coincide con el trabajo de Uroz en que un alto porcentaje de secuencias pertenecientes al

dominio Bacteria no se consiguen clasificar a nivel de *phylum*. Esto nos vuelve a demostrar que el suelo sigue teniendo una enorme biodiversidad que sigue siendo desconocida por la comunidad científica y que requiere de una mejora en los métodos de cultivo bacterianos (Pham y Kim, 2012). Además también es necesaria una mayor profundidad en la secuenciación y mejora del ensamblaje de genomas de estas comunidades, mediante técnicas moleculares independientes de cultivo como la metagenómica (Choffnes *et al.* 2013) o el single cell genomics (Rinke *et al.* 2014) y una mayor profundidad de secuenciación del gen 16S *rRNA* para reducir la biosfera rara asociada a un alto porcentaje de secuencias no clasificadas (Sogin *et al.* 2006; Bartram *et al.* 2011; Lynch *et al.* 2012). Todo esto nos ayudará a clasificar esta microbiosfera actualmente desconocida y nos permitirá también ampliar las bases de datos taxonómicas actuales.

Las únicas diferencias observadas en la riqueza y diversidad de las comunidades microbianas de la rizosfera de las quercíneas estudiadas se deben a las diferentes condiciones climáticas, principalmente el régimen de precipitaciones, que se dieron en los dos años de muestreo (2008 y 2011). Estas diferencias fueron más pronunciadas en las muestras de la ladera Sur (CNB, CNA y LNJ) que en las de la ladera Norte (CTR y CTE) pero solo fueron significativas en el encinar de la ladera Sur (LJN) (Figs. de R3-8 a R3-11). Las familias Bradyrhizobiaceae y Solirubrobacteraceae son las únicas que muestran una clara diferencia en su abundancia relativa entre los dos años de muestreo, siendo relativamente más abundantes en el año 2008 que en el año 2011, aunque solo es estadísticamente significativa en el caso de la familia Bradyrhizobiaceae (Figs. R3-5 y R3-6). Además esta familia del orden Rhizobiales fue la única significativamente más abundante en las muestras de la ladera Sur que en las de la ladera Norte en el muestreo de 2008. La mayor dominancia de la familia Bradyrhizobiaceae en las muestras CNA, CNB y LNJ (Figs. R3-5 y R3-10) parece estar relacionada con una mejor adaptación de algunos miembros de esta familia a las condiciones de xericidad estival, propias del clima mediterráneo, más pronunciadas en la ladera Sur de Sierra Nevada en el año 2008.

Entre los factores físico-químicos estudiados en este trabajo, el pH es el único con un efecto marginalmente significativo en la composición de las comunidades microbianas de la rizosfera de las quercíneas muestreadas (Fig. R3-16), quedando enmascarado por el efecto de las diferentes condiciones climáticas dadas en los años 2008 y 2011 que parecen ser las causantes de las diferencias encontradas entre las familias más abundantes observadas en ambos años, principalmente en el encinar de la ladera Sur (LJN).

4. Efecto del incendio en las comunidades microbianas

Nuestros resultados muestran una alteración en la composición de las comunidades microbianas que estuvieron sometidas a un incendio forestal en el año 2005 en el municipio de Lanjarón. Los efectos fueron más patentes tres años después del mismo (2008) (Fig. R4-5), y pese a encontrar un menor efecto en las comunidades microbianas seis años después del incendio (2011) (Fig. R4-6), éstas no han tenido tiempo suficiente para volver a las condiciones naturales previas al incendio. Según Xiang *et al.* (2014) en un año las comunidades microbianas sometidas a un incendio forestal son substancialmente diferentes de las comunidades control y once años son suficientes para la recuperación de las mismas a la situación pre-incendio. En nuestro caso nos encontramos en una situación intermedia en el que las comunidades microbianas afectadas por el incendio (LJQ, LJR y LJRE) no son significativamente diferentes de las comunidades control (LJN) pero todavía no han alcanzado las condiciones pre-incendio (Fig. R4-14).

La vegetación arbustiva dominante en las muestras de Lanjarón también presenta un patrón diferencial entre los sitios quemados (LJQ y LJR) y el no quemado (LJN) (Fig. R1-2). Encontrando una menor diversidad vegetal en el encinar quemado (LJQ) con la gramínea *Bromus tectorum* como la única especie dominante de los sitios quemados que no se encontró en el encinar sano (LJN). Siendo las gramíneas *Festuca scariosa* y *Dactylis glomerata* propias del encinar control. Las leguminosas *Adenocarpus decorticans* y *Vicia lathyroides* presentaron una mayor plasticidad que las gramíneas desarrollándose tanto en encinar sano como en suelos quemados (LJR), pero no se encontraron en el encinar quemado (LJQ) debido posiblemente a la alteración más intensa a la que están sometidas las zonas boscosas (Cobo-Díaz, 2013) en un incendio forestal (Tabla R1-3).

Los parámetros físico-químicos del suelo también se ven profundamente alterados por el efecto del incendio forestal (Certini, 2005; Goberna *et al.* 2012) afectando también tanto directa como indirectamente a las comunidades microbianas que habitan en él. En nuestro trabajo se observó una disminución de la materia orgánica del suelo (SOM) y un incremento del pH en los suelos afectados por el incendio; aunque solo el pH mostró una correlación estadísticamente significativa y directa con la intensidad del incendio, mostrando el valor máximo en el sitio más afectado por el incendio, el encinar quemado (LJQ) (Tabla R1-1 y Fig. R4-17).

Los efectos indirectos del incendio forestal, entre los que se encuentran el aumento del pH, la disminución de la SOM, el aumento de compuestos recalcitrantes derivados de la combustión de la SOM y los organismos vivos atrapados en el mismo, la pérdida de agua por evaporación y la erosión del suelo, tienen un efecto más duradero que los efectos directos del mismo. Tanto los efectos directos como los indirectos favorecen la proliferación y la colonización de los nichos que quedan vacíos por los microorganismos más resistentes a estas condiciones de estrés. Las bacterias, con mayor capacidad de resistencia que el resto de microorganismos del suelo (protozoos, hongos, arqueas y virus), con mejores capacidades de adaptación y proliferación en suelos alterados por incendios forestales pertenecen a los *phyla* Firmicutes y Actinobacteria (Gram positivas), entre ellas algunas muy conocidas por sus capacidades de producir esporas de resistencia al estrés térmico (Smith *et al.* 2008; Bárcenas-Moreno *et al.* 2011; Khodada *et al.* 2011). En nuestro trabajo varias familias de estos dos *phyla* han mostrado ser significativamente más abundantes en los sitios quemados, pero solo el *phylum* Actinobacteria tiene representantes entre las familias dominantes. Siendo las familias Micrococcaceae y Geodermatophilaceae las más relevantes de este *phylum* y las más abundantes en las muestras más afectadas por el incendio, pertenecientes al encinar quemado en el muestreo de 2008 (Figs. R4-5 y R4-16). Las familias Chitinophagaceae del *phylum* Bacteroidetes y Subdivisión3 del *phylum* Verrucomicrobia son más abundantes en las muestras no rizosféricas afectadas por el incendio 3 años después del mismo, posiblemente debido al incremento del pH, efecto descrito en el *phylum* Bacteroidetes (Lauber *et al.* 2009). Por otro lado, se observa una disminución de las familias del *phylum* Acidobacteria debido posiblemente al aumento de pH (Lauber *et al.* 2009), aunque no es estadísticamente significativa (Fig. R4-5).

Las familias Bradyrhizobiaceae (perteneciente a la clase α -Proteobacteria), Mycobacteriaceae y Solirubrobacteraceae (ambas del *phylum* Actinobacteria) han sido descritas en el apartado anterior como familias dominantes de la rizosfera de quercíneas sanas propias del clima mediterráneo, por lo que no es sorprendente que sean significativamente menos abundantes en las muestras quemadas. Aunque solo se observa este efecto 3 años después del incendio ya que en el muestreo de 2011, seis años tras el incendio, sólo la familia Bradyrhizobiaceae seguía siendo relativamente menos abundante en la muestra más afectada, el encinar quemado (LJQ), pero estas diferencias no fueron estadísticamente significativas (Figs. R4-5 y R4-6) lo que denota una recuperación substancial de las comunidades microbianas de los sitios quemados. Siendo el sitio quemado en el que prosperaron encinas germinadas después del incendio (LJRE) el que presenta una recuperación más rápida,

mostrando una comunidad microbiana más parecida a la del encinar sano (LJN) (Fig. R4-15). Aunque se trata del establecimiento de una nueva comunidad rizosférica, la menor intensidad del incendio en esta zona puede haber contribuido a una recuperación más rápida.

Por lo tanto, podemos considerar una abundancia relativamente baja de las familias Solirubrobacteraceae y Mycobacteriaceae como un indicador de la alteración producida por un incendio forestal en las comunidades microbianas de la rizosfera de quercíneas tras un período relativamente corto, tres años, de que ocurriera el mismo. Y una abundancia relativamente alta de las familias Geodermatophilaceae y Micrococcaceae junto con una abundancia relativamente baja de la familia Bradyrhizobiaceae como un indicador de esta misma situación pero tras un período relativamente largo, al menos seis años. El estudio del *phylum* Actinobacteria es el más adecuado para conocer el estado y la evolución de las comunidades microbianas, rizosféricas o no, tras un desastre natural de estas características. (Fig. R1-6), ya que tiene individuos tanto muy resistentes como muy sensibles al efecto de los incendios (Figs. R4-5, R4-6, R4-16 y R5-1).

5. El género *Arthrobacter*

El género *Arthrobacter* fue propuesto por Conn y Dimmick (1947), tomando el nombre propuesto por Fischer (1895) para aquellas bacterias productoras de “arthrosporas” o esporas articuladas (del griego arthros, articulado). Conn y Dimmick clasificaron tres especies del género *Arthrobacter*, las especies *Arthrobacter globiforme*, *Arthrobacter helvolum* y *Arthrobacter tumescens*, siendo *A. globiforme* la especie tipo de dicho género. Skerman *et al.* (1980) cambiaron el nombre de *A. globiforme* por *A. globiformis*, siendo a partir de entonces el nombre utilizado para designar a la especie tipo. Éste género pertenece al *phylum* Actinobacteria, incluidas como bacterias corineformes Gram positivas. La mayoría presentan un metabolismo aerobio obligado, utilizando aquellas que tienen metabolismo anaerobio el nitrato (NO_3^-) en la cadena respiratoria como aceptor final de electrones. Sólo se ha observado metabolismo fermentativo en *A. globiformis* en condiciones anaerobias y en ausencia de nitrato y nitrito (NO_2^-), utilizando glucosa y piruvato como fuentes de carbono. Presentan un contenido G+C entre 59 y 66 mol% (Stackebrandt & Woese 1981; Keddie *et al.* 1986; Eschbach *et al.* 2003). La mayoría de las especies del género *Arthrobacter* exhiben un ciclo de vida complejo en el que pasan de forma de bacilo (Gram negativo) a forma de coco (Gram positivo) cuando se encuentran en medios pobres en nutrientes (Keddie y Jones, 1981). De modo que un cultivo en fase estacionaria (generalmente 2-7 días) está compuesto por células en forma de coco con un diámetro de 0,6-1,0 μm , aunque algunas especies muestran sólo células esféricas a lo largo de todo el ciclo de vida (Goodfellow *et al.* 2012).

Pese a estar descrito como un género ubicuo en muestras de suelo (Conn, 1928), su abundancia relativa es muy baja en los suelos no alterados estudiados en este trabajo y no se ha considerado la familia a la que pertenece, Micrococcaceae, como una de las dominantes de dichas comunidades microbianas (Tabla R5-1). Sin embargo, en las muestras rizosféricas tomadas tres años después del incendio, *Arthrobacter* es el género dominante junto con el género *Blastococcus* (perteneciente a la familia Geodermatophilaceae) ya que suponen más del 14 % del total de secuencias obtenidas en encinar quemado (Figs. R5-1, R5-2 y Tabla R5-1). Esta dominancia de *Arthrobacter* parece estar relacionada con sus capacidades de vivir en condiciones adversas como las provocadas en el suelo por el incendio forestal, con escasez de agua y nutrientes y acumulación de compuestos tóxicos y recalcitrantes derivados del mismo (Hagedorn y Holt, 1975a; Keddie *et al.* 1986; Megharaj *et al.* 2003; Nordin *et al.* 2005; SantaCruz-Calvo, 2014), así como a un aumento de su actividad metabólica favorecido por el aumento del pH (Qiu *et al.* 2009).

Pese a la relativa facilidad de aislar individuos del género *Arthrobacter* usando medios de cultivo estándar (Hagedorn y Holt, 1975b), no todas las especies se aíslan con la misma eficiencia, resultando algunas de ellas incapaces de crecer en un medios de cultivo adecuados para este género (Tabla R5-2), como el medio HH' (apartado 3.1 del Material y Métodos). De modo que, no se obtuvo ningún aislado del OTU2, segundo en abundancia relativa, del género *Arthrobacter* en muestras de suelo LJQ. Por el contrario, los aislados del OTU4 son los más fáciles de obtener, consiguiendo 11 aislados de éste último pese a tener una relación de secuencias obtenidas mediante pirosecuenciación de 1/66 respecto al OTU2. Aun así la mayoría de aislados (41 de los 55 clones) pertenecieron al OTU más abundante, el OTU1. La clasificación de los distintos aislados pone de manifiesto la falta de precisión de la asignación de OTUs con una similitud al menos del 97 % al valor taxonómico de especie, pues en nuestro caso, la utilización del 100 % de similitud del amplicón de 408 bp como condición para agrupar los distintos OTUs no fue suficiente para separar los amplicones entre cada una de las especies a las que son homólogos (Tabla R5-2). Así la mayoría de aislados están muy próximos filogenéticamente a las especies *A. globiformis*, *A. oxydans*, agrupadas en el OTU1 y *A. nitroguajacolicus*, agrupada en el OTU4 (Tabla R5-2 y Fig. R5-3). De todos modos, no podemos asumir que nuestros aislados pertenezcan a dichas especies ya que el único método taxonómico utilizado en este trabajo ha sido el estudio de ≈ 1.460 bp del gen 16S *rRNA* de éstos y ninguno de ellos fue 100 % idéntico al gen 16S *rRNA* de las cepas tipo (Fig. R5-3).

Mediante reclutamiento genómico (Rusch *et al.* 2007; Rodríguez-Valera *et al.* 2009) se confirmó la dominancia del género *Arthrobacter* en el metagenoma del encinar quemado (LJQ). El 8,2 % de los contigs del metagenoma de LJQ formaron parte del pan-genoma de *Arthrobacter* y al menos el 30 % de dichos contigs reclutaron con las cepas de referencia (Figs. R5-5 y R5-6). Se observó una identidad media del 90,7 % en el reclutamiento frente a *A. globiformis* NBRC12137 y un 87,5 % en el reclutamiento frente a *A. sp.* FB24, por lo que no pudimos confirmar que se tratara de las mismas especies de dicho género, pues para ello habría sido necesaria una identidad media mínima del 94 % (Konstantinidis y Tiedje, 2005; Belda-Ferre *et al.* 2012). Tras un reclutamiento de casi el doble de contigs frente a *A. globiformis* NBRC12137 que frente a *A. sp.* FB24 y la mejor identidad media podemos confirmar que la cepa *A. globiformis* NBRC12137 es la más parecida, de todas las disponibles en la base de datos, a nuestro pan-genoma (Figs. R5-5 y R5-6).

De los más de 380.000 genes predichos en el metagenoma de LJQ, el 8,4 % fueron homólogos al género *Arthrobacter*. De los genes predichos homólogos a *Arthrobacter* quedó

una fracción residual de más de 2.000 genes (el 7,5 % de las CDR) incapaz de ser anotada con las bases de datos actuales (Fig. R5-4).

Aunque desconocemos la cobertura con la que se secuenció el pan-genoma de *Arthrobacter*, se consiguieron mapear más del 70 % de las regiones codificantes (CDR) de la cepa *A. sp.* FB24, hecho que confirma la dominancia en LJQ de un grupo bacteriano muy próximo filogenéticamente a esta cepa (Fig. R5-7). Siendo la CDR más abundante en el metagenoma de LJQ una hemo peroxidasa homóloga a la de la cepa de referencia (Fig. R5-7). Esta observación sugiere la posible adaptación de los aislados de *Arthrobacter* de la rizosfera del encinar quemado a la disminución de nutrientes post-incendio mediante el uso de compuestos lignocelulósicos, acumulados por dicho incendio, como fuente alternativa de carbono (Brown *et al.* 2011). Parece que los individuos del género *Arthrobacter* que habitan la rizosfera del encinar quemado están muy activos, dado que la mayor parte de los genes de estos individuos encontrados y anotados pertenecen al metabolismo, siendo el metabolismo aminoacídico el predominante (Fig. R5-8). Aunque hay que ser precavidos al sacar conclusiones de los datos obtenidos de las bases de datos como MG-RAST, ya que se trata de funciones potencialmente activas (genes) y no de genes realmente activos (cDNA del RNA mensajero), al trabajar con metagenomas y no con metatranscriptomas y por la anotación a veces errónea de las mismas bases de datos (función Membrane transport de la Fig. R5-8).

6. Consideraciones finales y generales

Tras el uso de distintas técnicas para el estudio de la diversidad microbiana concluimos que las técnicas clásicas basadas en la TGGE y el uso de genotecas del gen 16S *rRNA* no son útiles para evaluar la diversidad microbiana de ambientes tan complejos como los suelos de los bosques de quercíneas del Espacio Natural de Sierra Nevada. Si bien la aproximación mediante TGGE nos permite evaluar la diversidad β , no es útil para estudios taxonómicos ni de diversidad α de esta envergadura (Figs. R1-3 a R1-6). Siendo el uso de genotecas la técnica menos adecuada por la falta de representación de las comunidades microbianas originales debido a los errores de muestreo de la misma (Tablas R2-3, R3-3 y R4-3 y Figs. R2-7, R3-7 y R4-7).

Aunque la pirosecuenciación directa del DNA ambiental (Edna) es una técnica muy utilizada en el estudio de la diversidad taxonómica y funcional de las comunidades microbianas (Daniel, 2005; Hugenholtz y Tyson, 2008; Sleator *et al.* 2008; Simon y Daniel, 2011) en este estudio se ha puesto de manifiesto que para comunidades complejas no es la técnica más apropiada. Utilizando media placa del 454 GS FLX Titanium en la muestra pirosecuenciada a más profundidad de lectura, la muestra del encinar control (LJN), no se consiguió obtener un número de secuencias del gen 16S *rRNA* útiles significativamente mayor que con el uso de genotecas clásicas, por lo que comparte los problemas de esta aproximación (Figs. R2-3, R3-3 y R4-3). Aunque esta técnica genera secuencias de mayor tamaño que la secuenciación mediante Illumina, el número de lecturas conseguidas y el precio por nucleótido están promoviendo un cambio rápido en los laboratorios del uso de esta tecnología por la tecnología Illumina (comunicación personal de la Dr. Susannah Green Tringe).

Un hecho paradójico que no supimos explicar es la altísima proporción de genes ribosómicos homólogos al dominio Eukarya (más del 90 % de los genes *rRNA* encontrados) que se obtuvieron en los metagenomas secuenciados con la tecnología Illumina, pese a que los genes predichos en los contigs ensamblados de los mismos son esencialmente bacterianos (más del 90 % de los genes anotados). Siendo el resto de genes anotados (menos del 10 %) eucarióticos, arqueanos, víricos y otros sin clasificación principalmente debido a artefactos en el ensamblaje.

Uno de los métodos más empleados, pese a sus limitaciones, para estudiar la diversidad β entre distintas comunidades microbianas es el cálculo de las distancias filogenéticas de los amplicones por el método Unifrac (Schloss, 2008; Lozupone *et al.* 2011). En

nuestro estudio todos los análisis realizados solo con Presencia/Ausencia de un filotipo u OTU concreto (Unweighted-Unifrac) llevaron a la conclusión de que se trata de comunidades significativamente diferentes, pero estas diferencias se obtuvieron tanto entre sitios como entre las réplicas de un mismo sitio (Figs. R2-13, R3-13 y R4-13), lo que sugiere que este método da mucho peso a la “biosfera rara”, es decir, a los taxones u OTUs singletons o doubletons. En comunidades tan complejas como las de este trabajo sería necesaria una cobertura aún mayor para poder utilizar este método como herramienta discriminadora. Sin embargo, los análisis realizados teniendo en cuenta la abundancia de cada taxón (Weighted-Unifrac) no muestran diferencias significativas entre ninguna de estas comunidades microbianas (Figs. R2-14, R3-14 y R4-14), ni siquiera en el caso más extremo, el del incendio forestal (Fig. R4-14). Este último análisis da mucho peso a los taxones dominantes, siendo una técnica muy conservadora, aunque fue capaz de distinguir y agrupar de forma separada la mayoría de muestras de los distintos sitios estudiados (Figs. R2-14, R3-14 y R4-14), al contrario que el método Unweighted-Unifrac incapaz de hacer una distinción clara entre los distintos sitios (Figs. R3-13 y R4-13). Dando los métodos de ordenación (PCoA y CCA) una visión más clara de la relación entre las distintas comunidades (Figs. R2-15, R2-16, R3-15, R4-15 y R4-16).

Conclusions

1. Both classical techniques such as fingerprinting or libraries of the 16S *rRNA* gene, and newer techniques like massive sequencing of metagenomes, are not useful to reflect the taxonomic diversity of ecosystems as complex as the rhizosphere unless done at an appropriate depth of sequencing.
2. The rhizosphere of forest species in natural ecosystems presents a highly diverse microbial community, with an estimated richness of over 8,000 bacterial species and dominance of certain families. In the bulk soil, richness is similar but families are distributed more evenly.
3. Microbial communities in the rhizosphere of undisturbed oaks can be considered similar despite being from different plant species (*Q. pyrenaica* Willd. or *Q. ilex* subsp. *ballota*), different geographical locations (North or South side) and different sampling years.
4. In an altitudinal gradient, from 1,400 to 2,000 meters above sea level, the differences in microbial communities are due to the vegetation with which they are associated and their degree of development, but not to climate type factors.
5. Potential biomarkers of a 600 meters altitudinal gradient are at or below the species level, so it is necessary to use a technology with greater discrimination or specificity in order to follow the evolution of these biomarkers.
6. The families *Bradyrhizobiaceae* (belonging to the α -Proteobacteria class), *Mycobacteriaceae* and *Solirubrobacteraceae* (both from the *phylum* Actinobacteria) are biomarkers of natural conditions in the rhizosphere of oaks in a Mediterranean climate.
7. The genera *Arthrobacter* and *Blastococcus*, from the *phylum* Actinobacteria, predominate in the rhizosphere of holm oaks three years after a forest fire and are biomarkers of choice for monitoring the recovery of the soil ecosystem.
8. Reconstructing the pan-genome of the genus *Arthrobacter*, predominant in burned holm oak rhizosphere, indicates an adaptation to the post-fire decreasing nutrients through the use of lignocellulosic substrates as alternative carbon sources.
9. The lack of proportionality between the number of 16S *rRNA* gene sequences of each OTU and the number of isolates of the genus *Arthrobacter* reflects the different growth capacity of different *Arthrobacter* strains in defined media.

Bibliografía

- Acea, M.J. and Carballas, T.** (1996). Changes in physiological groups of microorganisms in soil following wildfire. *FEMS Microbiol Ecol* **20**: 33-39.
- Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V. and Polz, M.F.** (2005). PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. *Appl Environ Microbiol* **71**(12): 8966-8969.
- Acosta-Martínez, V., Dowd, S., Sun, Y. and Allen, V.** (2008). Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. *Soil Biology and Biochemistry* **40**(11): 2762-2770.
- Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W. and Nielsen, P.H.** (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**(6): 533-538.
- Alexandrian, D., Esnault, F. and Calabri, G.** (1999). Forest fires in the Mediterranean area. *Unasylva* **197**(50): 35-41.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J.** (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403-410.
- Amann, R.I., Lin, C., Key, R., Montgomery, L. and Stahl, D.A.** (1992). Diversity among Fibrobacter isolates: towards a phylogenetic classification. *Syst Appl Microbiol* **15**: 23–31.
- Amann, R.I., Ludwig, W. and Schleifer, K.H.** (1995). Phylogenetic Identification and In Situ Detection of Individual Microbial cells without Cultivation. *Microbiol Rev* **59**: 143-169.
- Baker, G.C., Smith, J.J. and Cowan, D.A.** (2003). Review and re-analysis of domain-specific 16S primers. *J Microbiol Meth* **55**(3): 541-555.
- Barcenas-Moreno, G., García-Orenes, F., Mataix-Solera, J., Mataix-Beneyto, J. and Bååth, E.** (2011). Soil microbial recolonisation after a fire in a Mediterranean forest. *Biol Fert Soils* **47**(3): 261-272.
- Bartram, A.K., Lynch, M.D., Stearns, J.C., Moreno-Hagelsieb, G. and Neufeld, J.D.** (2011). Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl Environ Microbiol* **77**(11): 3846-3852.
- Belda-Ferre, P., Alcaraz, L.D., Cabrera-Rubio, R., Romero, H., Simon-Soro, A., Pignatelli, M., et al.** (2012). The oral metagenome in health and disease. *ISME J* **6**(1): 46-56.
- Berendsen, R.L., Pieterse, C.M. and Bakker, P.A.** (2012). The rhizosphere microbiome and plant health. *Trends Plant Sci* **17**(8): 478-486.
- Berg, G. and Smalla, K.** (2009). Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol Ecol* **68**(1): 1-13.
- Bergmann, G.T., Bates, S.T., Eilers, K.G., Lauber, C.L., Caporaso, J.G., Walters, W.A., et al.** (2011). The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biol Biochem* **43**(7): 1450-1455.

Bilofsky, H.S. and Burks, C. (1988). The GenBank® genetic sequence data bank. *Nucleic Acids Res* **16**(5): 1861-1863.

Brown, M.E., Walker, M.C., Nakashige, T.G., Iavarone, A.T. and Chang, M.C. (2011). Discovery and characterization of heme enzymes from unsequenced bacteria: application to microbial lignin degradation. *J Am Chem Soc* **133**(45): 18006-18009.

Bryant, J.A., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B.J. and Green, J.L. (2008). Colloquium paper: microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *P Natl Acad Sci USA* **105 Suppl 1**: 11505-11511.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C., Turnbaugh, P.J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *P Natl Acad Sci USA* **108**: 4516-4522.

Cardinale, M., Brusetti, L., Lanza, A., Orlando, S., Daffonchio, D., Puglia, A.M., et al. (2010). Rehabilitation of Mediterranean anthropogenic soils using symbiotic wild legume shrubs: Plant establishment and impact on the soil bacterial community structure. *App Soil Ecol* **46**(1): 1-8.

Certini, G. (2005) Effects of fire on properties of forest soils: a review. *Oecologia* **143**: 1–10.

Chao, A., Chazdon, R.L., Colwell, R.K. and Shen, T.J. (2006). Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* **62**(2): 361-371.

Chien, A., Edgar, D.B. and Trela, J.M. (1976). Deoxyribonucleic Acid Polymerase from the Extreme Thermophile *Thermus aquaticus*. *J Bacteriol* **127**(3): 1550-1557.

Clark, S.C. (1996). Mediterranean ecology and an ecological synthesis of the field sites. In: *Mediterranean desertification and land use*. Brandt, C.J. and Thornes, J.B. (Eds.). ISBN 978-0-471-94250-4, 271-301. John Wiley and sons, Ltd.

Clarke, K.R. and Ainsworth, M. (1993). A method of linking multivariate community structure to environmental variables. *Mar Ecol Prog Ser* **92**: 205-219.

Cobo-Díaz, J.F. (2013). **Análisis del metabolismo procariótico del nitrógeno en la rizosfera de quercíneas del Espacio Protegido de Sierra Nevada mediante técnicas moleculares y metagenómicas.** Tesis Doctoral. Universidad de Granada, Departamento de Microbiología. Estación Experimental del Zaidín - CSIC.

Conn, H.J. (1928). A type of bacteria abundant in productive soils, but apparently lacking in certain soils of low productivity. *N.Y. State Agric. Exp. stn. Tech Bull* **138**: 3-26.

Conn, H.J. and Dimmick, I. (1947). SOIL BACTERIA SIMILAR IN MORPHOLOGY TO MYCOBACTERIUM AND CORYNEBACTERIUM. *J Bacteriol* **54**: 291-303.

Crowther, T.W., Maynard, D.S., Leff, J.W., Oldfield, E.E., McCulley, R.L., Fierer, N., et al. (2014). Predicting the responsiveness of soil biodiversity to deforestation: a cross-biome study. *Glob Chang Biol*

- Dale, V.H., Joyce, L.A., McNulty, S., Neilson, R.P., Ayres, M.P., Flannigan, M.D., et al.** (2001). Climate change and forest disturbances. *Bioscience* **51**: 723-734.
- Daniel, R.** (2005). The metagenomics of soil. *Nat Rev Microbiol* **3**(6): 470-478.
- de Kruif, P.** (1926). *Microbe hunters*. Orlando, FL: Harcourt, Inc.
- De Rijk, P., Neefs, J.M., Van der Peer, Y. and De Wachter, R.** (1992). Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res* **20**: 2075-2089.
- DeCáceres, M., Font, X., Oliva, F. and Vives, S.** (2007). GINKGO, A PROGRAM FOR NON-STANDARD MULTIVARIATE FUZZY ANALYSIS. *Adv in Fuzzy Sets & Systems* **2**(1): 41-56.
- Della-Marta, P.M., Haylock, M.R., Lutterbacher, J. and Wanner, H.** (2007). Doubled length of western European summer heat waves since 1880. *J Geophys Res* **112**: D15103.
- Delmont, T.O., Prestat, E., Keegan, K.P., Faubladiet, M., Robe, P., Clark, I.M., et al.** (2012). Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J* **6**(9): 1677-1687.
- den Bakker, H.C., Cummings, C.A., Ferreira, V., Vatta, P., Orsi, R.H., Degoricija, L., et al.** (2010). Comparative genomics of the bacterial genus *Listeria*: Genome evolution is characterized by limited gene acquisition and limited gene loss. *BMC Genomics* **11**: 688.
- Deng, Y., He, Z., Xu, M., Qin, Y., Van Nostrand, J.D., Wu, L., et al.** (2012). Elevated carbon dioxide alters the structure of soil microbial communities. *Appl Environ Microbiol* **78**(8): 2991-2995.
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., et al.** (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**(7): 5069-5072.
- Doerks, T., Copley, R.R., Schultz, J., Ponting, C.P. and Bork, P.** (2002). Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res* **12**(1): 47-56.
- Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. and Knight, R.** (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**(16): 2194-2200.
- Egamberdieva, D., Kamilova, F., Validov, S., Gafurova, L., Kucharova, Z. and Lugtenberg, B.** (2008). High incidence of plant growth-stimulating bacteria associated with the rhizosphere of wheat grown on salinated soil in Uzbekistan. *Environ Microbiol* **10**(1): 1-9.
- Eschbach, M., Möbitz, H., Rompf, A. and Jahn, D.** (2003). Members of the genus *Arthrobacter* grow anaerobically using nitrate ammonification and fermentative processes: anaerobic adaptation of aerobic bacteria abundant in soil. *FEMS MicrobiolLett* **223**(2): 227-230.
- Felicísimo, Á.M., Muñoz, J., Villalba, C.J. and Mateo, R.G.** (2011). Impactos, vulnerabilidad y adaptación al cambio climático de la biodiversidad española. 1. Flora y vegetación. Ministerio de Medio Ambiente y Medio Rural y Marino, Madrid.

- Fernández, H.R., Romero, F., Peralta, M. and Grosso, L.** (2001). La diversidad del zoobentos en ríos de montaña del noroeste de Argentina: comparación entre seis ríos. *Ecología Austral* **11**: 9-16.
- Fierer, N., Bradford, M.A. and Jackson, R.B.** (2007). Toward an ecological classification of soil bacteria. *Ecology* **88**(6): 1354-1364.
- Fierer, N., Lauber, C.L., Ramirez, K.S., Zaneveld, J., Bradford, M.A. and Knight, R.** (2012). Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J* **6**(5): 1007-1017.
- Fierer, N. and Lennon, J.T.** (2011). The generation and maintenance of diversity in microbial communities. *Am J Bot* **98**(3): 439-448.
- Fioretto, A., Papa, S. and Pellegrino, A.** (2005). Effects of fire on soil respiration, ATP content and enzyme activities in Mediterranean maquis. *Appl Veg Sci* **8**: 13–20.
- Fischer, A.** (1895). Untersuchungen über Bakterien. *Jahrb. wiss. Botan* **27**: 1-163.
- Fox, G.E., Wisotzkey, J.D. and Jurtshuk, P.** (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42**: 166–170.
- Gans, J., Woilinsky, M. and Dunbar, J.** (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309**: 1387–1390.
- Giorgi, F. and Lionello, P.** (2008). Climate change projections for the Mediterranean region. *Global Planet Change* **63**: 90-104.
- Gómez-Zotano, J., Moreno-Sánchez, J.J. and Rodríguez-Martínez, F.** (2005). El incendio de Sierra Nevada (22-24 de septiembre de 2005). Una catástrofe ecológica. *Cuadernos Geográficos* **037**: 205-214.
- Good, P., Moriondo, M., Giannakopoulos, C. and Bindi, M.** (2008). The meteorological conditions associated with extreme fire risk in Italy and Greece: relevance to climate models studies. *Int J Wildland Fire* **17**: 1-11.
- Goodfellow, M., Kämpfer, P., Busse, H.-J., Trujillo, M.E., Suzuki, K.-I., Ludwig, W., et al.** (2012). *Bergey's manual of systematic bacteriology*. (M. Goodfellow, P. Kämpfer, H.-J. Busse, M. E. Trujillo, K.-I. Suzuki, W. Ludwig, & W. B. Whitman, Eds.) (Second edi., pp. 578–624). Athens, GA.
- Griffiths, B.S. and Philippot, L.** (2013). Insights into the resistance and resilience of the soil microbial community. *FEMS Microbiol Rev* **37**(2): 112-129.
- Grimes, D.J., Atwell, R.W., Brayton, P.R., Palmer, L.M., Rollins, D.M., Roszak, D.B., et al.** (1986). The fate of enteric pathogenic bacteria in estuarine and marine environments. *Microbiol Sci* **3**: 324–329.
- Guénon, R., Vennetier, M., Dupuy, N., Ziarelli, F. and Gros, R.** (2011). Soil organic matter quality and microbial catabolic functions along a gradient of wildfire history in a Mediterranean ecosystem. *App Soil Ecol* **48**: 81-93.
- Hagedorn, C. and Holt, J.G.** (1975a). A nutritional and taxonomic survey of *Arthrobacter* soil isolates. *Can J Microbiol* **21**: 353-361.

- Hagedorn, C. and Holt, J.G.** (1975b). Ecology of soil Arthrobacters in Clarion-Webster toposequences of Iowa. *Appl Microbiol* **29**: 211-218.
- Halverson, L.J. and Stacey, G.** (1986). Signal Exchange in Plant-Microbe Interactions. *Microb Rev* **50**(2): 193-225.
- Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. and Goodman, R.M.** (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**: 245-249.
- Hansen, M.C., Tolker-Nielsen, T., Givskov, M. and Molin, S.** (1998). Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region. *FEMS Microbiol Ecol* **26**: 141-149.
- Harris, J.** (2009). Soil microbial communities and restoration ecology: facilitators or followers? *Science* **325**(5940): 573-574.
- Hartmann, M., Niklaus, P.A., Zimmermann, S., Schmutz, S., Kremer, J., Abarenkov, K., et al.** (2014). Resistance and resilience of the forest soil microbiome to logging-associated compaction. *ISME J* **8**(1): 226-244.
- Heck, K.L., van Belle, G. and Simberloff, D.** (1975). Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology* **56**: 1459-1461.
- Hill, G.T., Mitkowski, N.A., Aldrich-Wolfe, L., Emele, L.R., Jurkonie, D.D., Ficke, A., et al.** (2000). Methods for assessing the composition and diversity of soil microbial communities. *App Soil Ecol* **15**: 25-36.
- Hooper, L.V.** (2001). Commensal Host-Bacterial Relationships in the Gut. *Science* **292**(5519): 1115-1118.
- International Human Genome Sequencing Consortium.** (2004). Finishing the euchromatic sequence of the human genome. *Nature* **431**(7011): 931-45.
- IPCC.** (2007). Fourth Assessment Report, Climate Change 2007: Synthesis Report. Cambridge University Press, Cambridge, UK.
- Keddie, R.M. and Jones, D.** (1981). Saprophytic, aerobic coryneform bacteria pp. 1838–1878. Starr, M. P., Stolp, H., Trüper, H. G., Balows, A., and Schlegel, H. G. (ed.) *The prokaryotes: A handbook on habitats, isolation and identification of bacteria* Berlin Springer-Verlag.
- Kielak, A., Pijl, A.S., van Veen, J.A. and Kowalchuk, G.A.** (2008). 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* **63**(3): 372-382.
- Kuffner, M., Hai, B., Rattei, T., Melodelima, C., Schloter, M., Zechmeister-Boltenstern, S., et al.** (2012). Effects of season and experimental warming on the bacterial community in a temperate mountain forest soil assessed by 16S rRNA gene pyrosequencing. *FEMS Microbiol Ecol* **82**(3): 551-562.

- Kuiper, I., Legendijk, E.L., Bloemberg, G.V. and Lugtenberg, B.** (2004). Rhizoremediation: A Beneficial Plant-Microbe Interaction. *MPMI* **17**(1): 6-15.
- Lambais, M.R., Lucheta, A.R. and Crowley, D.E.** (2014). Bacterial Community Assemblages Associated with the Phyllosphere, Dermosphere, and Rhizosphere of Tree Species of the Atlantic Forest are Host Taxon Dependent. *Microb Ecol.*
- Lauber, C.L., Hamady, M., Knight, R. and Fierer, N.** (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* **75**(15): 5111-5120.
- Leal-García, J.** (2013). **Análisis del crecimiento vegetativo e inversión reproductiva de *Quercus pyrenaica* Willd. en tres localidades de Sierra Nevada.** Trabajo Fin de Máster. Máster Universitario en Conservación, Gestión y Restauración de la Biodiversidad. Universidad de Granada.
- Lederberg, J. and McCray, A.T.** (2001). 'Ome Sweet 'Omics-- A Genealogical Treasury of Words. *The Scientist* **15**(7): 8.
- Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., et al.** (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* **20**(2): 265-272.
- Legendre, P. and Legendre, L.** (1998). Numerical ecology. 2nd English edition. Elsevier Science B.V., Amsterdam.
- Lindemann, S.R., Moran, J.J., Stegen, J.C., Renslow, R.S., Hutchison, J.R., Cole, J.K., et al.** (2013). The epsomitic phototrophic microbial mat of Hot Lake, Washington: community structural responses to seasonal cycling. *Front Microbiol* **4**: 323.
- Lionello, P., Malanotte-Rizzoli, P., Boscolo, R., Alpert, P., Artale, V., Li, L., et al.** (2006). The Mediterranean climate: an overview of the main characteristics and issues. In: Lionello, P., Malanotte-Rizzoli, P., Boscolo, R. (Eds.), *Mediterranean Climate Variability*. Elsevier, Amsterdam, pp. 1-26.
- Lipson, D.A.** (2007). Relationships between temperature responses and bacterial community structure along seasonal and altitudinal gradients. *FEMS Microbiol Ecol* **59**(2): 418-427.
- Lombard, N., Prestat, E., van Elsas, J.D. and Simonet, P.** (2011). Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. *FEMS Microbiol Ecol* **78**(1): 31-49.
- Lozupone, C., Ladser, M.E., Knights, D., Stombaugh, J. and Knight, R.** (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J* **5**(2): 169-172.
- Lundberg, D.S., Lebeis, S.L., Paredes, S.H., Yourstone, S., Gehring, J., Malfatti, S., et al.** (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature* **488**(7409): 86-90.
- Lynch, M.D., Bartram, A.K. and Neufeld, J.D.** (2012). Targeted recovery of novel phylogenetic diversity from next-generation sequence data. *ISME J* **6**(11): 2067-2077.
- Maidak, B.L., Larsen, N., McCaughey, M.J., Overbeek, R., Olsen, G.J., Fogel, K., et al.** (1994). The Ribosomal Database project. *Nucleic Acids Res* **22**(17): 3485-3487.

- Mardis, E.R.** (2008). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* **9**: 387-402.
- Martínez-Parras, J.M. and Molero-Mesa, J.** (1982). Ecología y fitosociología de *Quercus pyrenaica* Willd. en la provincia Bética. Los melojares béticos y sus etapas de sustitución. *Lazaroa* **4**: 91-104.
- Martiny, J.B., Bohannon, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., et al.** (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**(2): 102-112.
- McMorrow, D.** (2010). The \$100 Genome: implications for the DoD. McLean, VA The MITRE Corporation: 1-54.
- Medini, D., Donati, C., Tettelin, H., Massignani, V. and Rappuoli, R.** (2005). The microbial pan-genome. *Curr Opin Genet Dev* **15**(6): 589-594.
- Megharaj, M., Avudainayagam, S. and Naidu, R.** (2003). Toxicity of hexavalent chromium and its reduction by bacteria isolated from soil contaminated with tannery waste. *Curr Microbiol* **47**:51-54.
- Mendes, R., Garbeva, P. and Raaijmakers, J.M.** (2013). The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol Rev* **37**: 634-663.
- Mendes, R., Kruijt, M., de Bruijn, I., Dekkers, E., van der Voort, M., Schneider, J.H., et al.** (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332**(6033): 1097-1100.
- Metzker, M.L.** (2010). Sequencing technologies - the next generation. *Nat Rev Genet* **11**(1): 31-46.
- Mizrahi-Man, O., Davenport, E.R. and Gilad, Y.** (2013). Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS One* **8**(1): e53608.
- Mongodin, E.F., Shapir, N., Daugherty, S.C., DeBoy, R.T., Emerson, J.B., Shvartzbeyn, A., et al.** (2006). Secrets of soil survival revealed by the genome sequence of *Arthrobacter aurescens* TC1. *PLoS Genet* **2**(12): e214.
- Monson, R.K., Lipson, D.L., Burns, S.P., Turnipseed, A.A., Delany, A.C., Williams, M.W., et al.** (2006). Winter forest soil respiration controlled by climate and microbial community composition. *Nature* **439**(7077): 711-714.
- Mosca, E., Montecchio, L., Sella, L. and Garbaye, J.** (2007). Short-term effect of removing tree competition on the ectomycorrhizal status of a declining pedunculate oak forest (*Quercus robur* L.). *Forest Ecol Manag* **244**: 129-140.
- Nakatsu, C.H., Barabote, R., Thompson, S., Bruce, D., Detter, C., Brettin, T., et al.** (2013). Complete genome sequence of *Arthrobacter* sp. strain FB24. *Stand Genomic Sci* **9**(1): 106-116.
- Nakatsu, C.H., Carmosini, N., Baldwin, B., Beasley, F., Kourtev, P. and Konopka, A.** (2005). Soil microbial community responses to additions of organic carbon substrates and heavy metals (Pb and Cr). *Appl Environ Microbiol* **71**(12): 7679-7689.

- Namiki, T., Hachiya, T., Tanaka, H. and Sakakibara, Y.** (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res* **40**(20): e155.
- Neefs, J.M., Van der Peer, Y., De Rijk, P., Chapelle, S. and De Wachter, R.** (1993). Compilation of small ribosomal subunit RNA structures. *Nucleic Acids Res* **21**(13): 3025-3049.
- Neefs, J.M., Van der Peer, Y., De Rijk, P., Goris, A. and De Wachter, R.** (1991). Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res* **19**: 1987-2015.
- Nihorimbere, V., Ongena, M., Smargiassi, M. and Thonart, P.** (2011). Beneficial effect of the rhizosphere microbial community for plant growth and health. *Biotechnol Agron Soc Environ* **15**: 327-337.
- Nordin, K., Unell, M. and Jansson, J.K.** (2005). Novel 4-chlorophenol degradation gene cluster and degradation route via hydroxyquinol in *Arthrobacter chlorophenolicus* A6. *Appl Environ Microbiol* **71**(11): 6538-44.
- Olivares, F.J., Barea-Azcón, J.M., Pérez-López, F.J., Tinaut, A. and Henares, I.** (2011). Sierra Nevada: una gran montaña, un pequeño continente. En: *Las mariposas diurnas de Sierra Nevada*. Consejería de Medio Ambiente, Junta de Andalucía (Ed.). pp. 75-117.
- Olsen, G.J., Larsen, N. and Woese, C.R.** (1991). The Ribosomal Database project. *Nucleic Acids Res* **19**: 2017-2021.
- Parks, D.H. and Beiko, R.G.** (2010). Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **26**(6): 715-721.
- Pausas, J.G.** (2006). Simulating Mediterranean landscape pattern and vegetation dynamics under different fire regimes. *Plant Ecol* **187**: 249-259.
- Pausas, J.G. and Fernández-Muñoz, S.** (2012). Fire regime in the Western Mediterranean Basin: from fuel-limited to drought-driven fire regime. *Climatic Change* **110**: 215-226.
- Pham, V.H. and Kim, J.** (2012). Cultivation of unculturable soil bacteria. *Trends Biotechnol* **30**(9): 475-484.
- Piganeau, G., Desdevises, Y., Derelle, E. and Moreau, H.** (2008). Picoeukaryotic sequences in the Sargasso sea metagenome. *Genome Biol* **9**(1): R5.
- Poretzky, R., Rodriguez, R.L., Luo, C., Tsementzi, D. and Konstantinidis, K.T.** (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One* **9**(4): e93827.
- Qiu, X., Wu, P., Zhang, H., Li, M. and Yan, Z.** (2009). Isolation and characterization of *Arthrobacter* sp. HY2 capable of degrading a high concentration of p-nitrophenol. *Bioresour Technol* **100**(21): 5243-5248.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al.** (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**(Database issue): D590-596.

- Quince, C., Curtis, T.P. and Sloan, W.T. (2008). The rational exploration of microbial diversity. *ISME J* **2**(10): 997-1006.
- Raaijmakers, J.M., Paulitz, T.C., Steinberg, C., Alabouvette, C. and Moëgne-Loccoz, Y. (2009). The rhizosphere: A playground and battlefield for soilborne pathogens and beneficial microorganisms. *Plant Soil* **321**: 341–361.
- Rincon-Florez, V., Carvalhais, L. and Schenk, P. (2013). Culture-Independent Molecular Tools for Soil and Rhizosphere Microbiology. *Diversity* **5**(3): 581-612.
- Riesner, D., Steger, G., Zimmat, R., Owens, R.A., Wagenhöfer, M., Hillen, W., *et al.* (1989). Temperature-gradient gel electrophoresis of nucleic acids: analysis of conformational transitions, sequence variations, and protein-nucleic acid interactions. *Electrophoresis* **10**(5-6): 377-89.
- Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., *et al.* (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat Protoc* **9**(5): 1038-1048.
- Rodriguez-Navarro, D.N., Dardanelli, M.S. and Ruiz-Sainz, J.E. (2007). Attachment of bacteria to the roots of higher plants. *FEMS Microbiol Lett* **272**(2): 127-136.
- Rodriguez-Valera, F., Martin-Cuadrado, A.B., Rodriguez-Brito, B., Pasic, L., Thingstad, T.F., Rohwer, F., *et al.* (2009). Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* **7**(11): 828-836.
- Roesch, L.F., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K., Kent, A.D., *et al.* (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* **1**(4): 283-290.
- Rothberg, J.M. and Leamon, J.H. (2008). The development and impact of 454 sequencing. *Nat Biotechnol* **26**: 1117–1124.
- Rosselló-Mora, R. and Amann, R.I. (2001). The species concept for prokaryotes. *FEMS Microbiol Rev* **25**: 39-67.
- Ruiz-Labourdette, D., Schmitz, M.F. and Pineda, F.D. (2013). Changes in tree species composition in Mediterranean mountains under climate change: Indicators for conservation planning. *Ecol Indic* **24**: 310-323.
- Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al.* (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**(3): e77.
- Sanger, F., Nicklen, S. and Coulson, R. (1977). DNA sequencing with chain-terminating inhibitors. *P Natl Acad Sci USA* **74**(12): 5463-5467.
- Santa-Cruz Calvo, L. (2014). **Rizorremediación de suelos contaminados por hidrocarburos mediante microorganismos tolerantes a la sequía**. Tesis Doctoral. Universidad de Granada, Departamento de Microbiología. Instituto del Agua.

- Sardans, J. and Peñuelas, J.** (2005). Drought decreases soil enzyme activity in a Mediterranean *Quercus ilex* L. forest. *Soil Biol Biochem* **37**: 455-461.
- Sardans, J., Peñuelas, J. and Estiarte, M.** (2008). Changes in soil enzymes related to C and N cycle and in soil C and N content under prolonged warming and drought in a Mediterranean shrubland. *App Soil Ecol* **39**: 223-235.
- Schadt, E.E., Turner, S. and Kasarskis, A.** (2010). A window into third-generation sequencing. *Hum Mol Genet* **19**(R2): R227-240.
- Schloss, P.D.** (2008). Evaluating different approaches that test whether microbial communities have the same structure. *ISME J* **2**(3): 265-275.
- Schloss, P.D. and Handelsman, J.** (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**(3): 1501-1506.
- Schmidt, M.W., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., et al.** (2011). Persistence of soil organic matter as an ecosystem property. *Nature* **478**(7367): 49-56.
- Schröter, D., Cramer, W., Leemans, R., Prentice, I.C., Araujo, M.B., Arnell, N.W., et al.** (2005). Ecosystem service supply and vulnerability to global change in Europe. *Science* **310**: 1333-1337.
- Shuldiner, A.R., Nirula, A. and Roth, J.** (1989) Hybrid DNA artefact from PCR of closely related target sequences. *Nucleic Acids Res.* **17**: 4409.
- Simon, C. and Daniel, R.** (2011). Metagenomic analyses: past and future trends. *Appl Environ Microbiol* **77**(4): 1153-1161.
- Singh, B.K., Bardgett, R.D., Smith, P. and Reay, D.S.** (2010). Microorganisms and climate change: terrestrial feedbacks and mitigation options. *Nat Rev Microbiol* **8**(11): 779-790.
- Singh, B.K., Millard, P., Whiteley, A.S. and Murrell, J.C.** (2004). Unravelling rhizosphere-microbial interactions: opportunities and limitations. *Trends Microbiol* **12**(8): 386-393.
- Singh, D., Shi, L. and Adams, J.M.** (2013). Bacterial diversity in the mountains of South-West China: climate dominates over soil parameters. *J Microbiol* **51**(4): 439-447.
- Skerman, V.B.D., McGowan, V. and Sneath, P.H.A.** (1980). Approved Lists of Bacterial Names. *Int J Syst Bacteriol* **30**: 225-420.
- Sleator, R.D., Shortall, C. and Hill, C.** (2008). Metagenomics. *Lett Appl Microbiol* **47**(5): 361-366.
- Smalla, K., Sessitsch, A. and Hartmann, A.** (2006). The rhizosphere: "Soil compartment influenced by the root". *FEMS Microbiol Ecol* **56**: 165.
- Smith, N.R., Kishchuk, B.E. and Mohn, W.W.** (2008). Effects of wildfire and harvest disturbances on forest soil bacterial communities. *Appl Environ Microbiol* **74**(1): 216-224.

- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R. et al.** (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *P Natl Acad Sci USA* **103**: 12115–12120.
- Solomon, K.V., Haitjema, C.H., Thompson, D.A. and O'Malley, M.A.** (2014). Extracting data from the muck: deriving biological insight from complex microbial communities and non-model organisms with next generation sequencing. *Curr Opin Biotechnol* **28C**: 103-110.
- Solomon, S.D., Qin, M., Manning, Z., Chen, Z., Marquis, M., Averyt, K.B., et al.** (Eds.). (2007). IPCC: Summary for Policymakers. *Climate Change 2007: the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge.
- Stackebrandt, E. and Woese, C.R.** (1981). The evolution of the prokaryotes. In: *Molecular and Cellular Aspects of Microbial Evolution. Symposium of the Society for General Microbiology 32* (Carlile, M.J., Collins, J.F. and Moseley, B.E.B., Eds.), pp. 1-3. Cambridge University Press, Cambridge.
- Stackebrandt, E. and Goebel, B.M.** (1994). Taxonomic Note: A Place for DNA-DNA Reassociation and 16 rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *Int J Syst Bacteriol* **44**: 846-849.
- Stephen, K.E., Homrighausen, D., DePalma, G., Nakatsu, C.H. and Irudayaraj, J.** (2012). Surface enhanced Raman spectroscopy (SERS) for the discrimination of *Arthrobacter* strains based on variations in cell surface composition. *Analyst* **137**(18): 4280-4286.
- Tamames, J., Abellan, J.J., Pignatelli, M., Camacho, A. and Moya, A.** (2010). Environmental distribution of prokaryotic taxa. *BMC Microbiol* **10**: 85.
- Thatcher D.R. and Hodson, B.** (1981). Denaturation of proteins and nucleic acids by thermal-gradient electrophoresis. *Biochem J* **197**(1): 105-9.
- Ter Braak, C.J.F. and Smilauer, P.** (2002). *Canoco for Windows version 4.5*. Biometris–Plant Research International, Wageningen.
- Tettelin, H., Massignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., et al.** (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *P Natl Acad Sci USA* **102**(39): 13950-13955.
- Tiselius, A.** (1937). Electrophoresis of serum globulin. *Biochem J* **31**(9): 1464-1477.
- Torres-Cortés, G., Millán, V., Fernández-González, A.J., Aguirre-Garrido, J.F., Ramírez-Saad, H.C., Fernández-López, M., et al.** (2012). Bacterial community in the rhizosphere of the cactus species *Mammillaria carnea* during dry and rainy seasons assessed by deep sequencing. *Plant Soil* **357**(1-2): 275-288.
- Torsvik, V., Goksoyr, J. and Daae, F.L.** (1990). High Diversity in DNA of Soil Bacteria. *Appl Environ Microbiol* **56**(3): 782-787.
- Torsvik, V. and Ovreas, L.** (2002). Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* **5**: 240-245.

Tringe, S.G. and Hugenholtz, P. (2008). A renaissance for the pioneering 16S rRNA gene. *Curr Opin Microbiol* **11**(5): 442-446.

Trivedi, P., Anderson, I.C. and Singh, B.K. (2013). Microbial modulators of soil carbon storage: integrating genomic and metabolic knowledge for global prediction. *Trends Microbiol* **21**(12): 641-651.

Uroz, S., Buee, M., Murat, C., Frey-Klett, P. and Martin, F. (2010). Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environ Microbiol Rep* **2**(2): 281-288.

Uroz, S., Calvaruso, C., Turpault, M.P. and Frey-Klett, P. (2009). Mineral weathering by bacteria: ecology, actors and mechanisms. *Trends Microbiol* **17**(8): 378-387.

Uroz, S., Ioannidis, P., Lengelle, J., Cebon, A., Morin, E., Buee, M., et al. (2013). Functional assays and metagenomic analyses reveals differences between the microbial communities inhabiting the soil horizons of a Norway spruce plantation. *PLoS One* **8**(2): e55929.

Uroz, S., Oger, P., Lepleux, C., Collignon, C., Frey-Klett, P. and Turpault, M.P. (2011). Bacterial weathering and its contribution to nutrient cycling in temperate forest ecosystems. *Res Microbiol* **162**(9): 820-31.

Valle, F. (2003). Mapa de series de vegetación de Andalucía. Ed. Rueda. **ISBN:** 84-7207-137-5.

Vessey, J.K. (2003). Plant growth promoting rhizobacteria as biofertilizers. *Plant Soil* **255**: 571–586.

Vos, M., Wolf, A.B., Jennings, S.J. and Kowalchuk, G.A. (2013). Micro-scale determinants of bacterial diversity in soil. *FEMS Microbiol Rev* **37**(6): 936-954.

Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**(16): 5261-5267.

Whelan, R.J. (1995). *The Ecology of Fire*. Cambridge (UK): Cambridge University Press.

Wintzingerode, F.V., Gobel, U.B. and Stackebrandt, E. (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**: 213-229.

Woese, C.R. (1987). Bacterial Evolution. *Microbiol Rev* **51**(2): 221-271.

Woese, C.R., Kandler, O. and Wheelis, M.L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *P Natl Acad Sci USA* **87**: 4576-4579.

Wooley, J.C. and Ye, Y. (2009). Metagenomics: facts and artifacts, and computational challenges. *J Comp Sci Technol* **25**: 71–81.

Xiang, X., Shi, Y., Yang, J., Kong, J., Lin, X., Zhang, H., et al. (2014). Rapid recovery of soil bacterial communities after wildfire in a Chinese boreal forest. *Sci Rep* **4**: 3829.

Xueju, L., Tfaily, M.M., Green, S.J., Steinweg, J.M., Chanton, P., Invittaya, A., et al. (2014). Microbial Metabolic Potential for Carbon Degradation and Nutrient (Nitrogen and Phosphorus) Acquisition in an Ombrotrophic Peatland. *Appl Environ Microbiol* **80**(11): 3531–3540.

Yuste, J.C., Barba, J., Fernández-González, A.J., Fernández-López, M., Mattana, S., Martínez-Vilalta, J., et al. (2012). Changes in soil bacterial community triggered by drought-induced gap succession preceded changes in soil C stocks and quality. *Ecol Evol* **2**(12): 3016-3031.

Yuste, J.C., Fernández-González, A.J., Fernández-López, M., Ogaya, R., Peñuelas, J., Sardans, J., et al. (2014). Strong functional stability of soil microbial communities under semiarid Mediterranean conditions and subjected to long-term shifts in baseline precipitation. *Soil Biol Biochem* **69**: 223-233.

Zerbino, D.R. and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**(5): 821-829.

Zhang, B., Liang, C., He, H. and Zhang, X. (2013). Variations in Soil Microbial Communities and Residues Along an Altitude Gradient on the Northern Slope of Changbai Mountain, China. *PLoS One* **8**(6): e66184.

